

## Personalized third-trimester fetal growth evaluation: comparisons of individualized growth assessment, percentile line and conditional probability methods

Russell L. Deter, Wesley Lee, Haleh Sangi-Haghpeykar, Adi L. Tarca, Jia Li, Lami Yeo & Roberto Romero

To cite this article: Russell L. Deter, Wesley Lee, Haleh Sangi-Haghpeykar, Adi L. Tarca, Jia Li, Lami Yeo & Roberto Romero (2015): Personalized third-trimester fetal growth evaluation: comparisons of individualized growth assessment, percentile line and conditional probability methods, The Journal of Maternal-Fetal & Neonatal Medicine, DOI: [10.3109/14767058.2014.995083](https://doi.org/10.3109/14767058.2014.995083)

To link to this article: <http://dx.doi.org/10.3109/14767058.2014.995083>

 [View supplementary material](#) 

 Published online: 25 Sep 2015.

 [Submit your article to this journal](#) 

 Article views: 1

 [View related articles](#) 

 [View Crossmark data](#) 

 Citing articles: 1 [View citing articles](#) 

ORIGINAL ARTICLE

## Personalized third-trimester fetal growth evaluation: comparisons of individualized growth assessment, percentile line and conditional probability methods

Russell L. Deter<sup>1</sup>, Wesley Lee<sup>1,2,3</sup>, Haleh Sangi-Haghpeykar<sup>1</sup>, Adi L. Tarca<sup>3</sup>, Jia Li<sup>4</sup>, Lami Yeo<sup>3,5</sup>, and Roberto Romero<sup>3,6,7</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX, USA, <sup>2</sup>Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Rochester, MI, USA, <sup>3</sup>Perinatology Research Branch, NICHD/NIH/DHHS, Detroit, MI, USA, <sup>4</sup>Department of Electrical and Computer Engineering, Oakland University, Rochester, MI, USA, <sup>5</sup>Department of Obstetrics and Gynecology, Wayne State University/Hutzel Hospital, Detroit, MI, USA, <sup>6</sup>Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI, USA, and <sup>7</sup>Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA

### Abstract

**Objective:** To compare third-trimester size trajectory prediction errors (average transformed percent deviations) for three individualized fetal growth assessment methods.

**Methods:** This study utilized longitudinal measurements of nine directly measured size parameters in 118 fetuses with normal neonatal growth outcomes. Expected value (EV) function coefficients and variance components were obtained using two-level random coefficient modeling. Growth models (IGA) or EV coefficients and variance components (PLM and CPM) were used to calculate predicted values at ~400 third-trimester time points. Percent deviations (%Dev) calculated at these time points using all three methods were expressed as percentages of IGA MA-specific reference ranges [transformed percent deviations (T%Dev)]. Third-trimester T%Dev values were averaged (aT%Dev) for each parameter. Mean ± standard deviation's for sets of aT%Dev values derived from each method (IGA, PLM and CPM) were calculated and compared.

**Results:** Mean aT%Dev values for nine parameters were: (i) IGA: -4.3 to 5.2% (9/9 not different from zero); (ii) PLM: -32.7 to 25.6% (4/9 not different from zero) and (iii) CPM: -20.4 to 17.4% (5/9 not different from zero). Seven of nine systematic deviations from zero were statistically significant when IGA values were compared to either PLM or CPM values. Variabilities were smaller for IGA when compared to those for PLM or CPM, with (i) 5/9 being statistically significant (IGA versus PLM), (ii) 2/9 being statistically significant (IGA versus CPM) and (iii) 5/9 being statistically significant (PLM versus CPM).

**Conclusions:** Significant differences in the agreement between predicted third-trimester size parameters and their measured values were found for the three methods tested. With most parameters, IGA gave smaller mean aT%Dev values and smaller variabilities. The CPM method was better than the PLM approach for most but not all parameters. These results suggest that third-trimester size trajectories are best characterized by IGA in fetuses with normal growth outcomes.

### Keywords

Fetal size trajectories, Rossavik size model, ultrasound

### History

Received 28 September 2014

Revised 23 November 2014

Accepted 2 December 2014

Published online 24 December 2014

### Introduction

Advances in human genomics have promoted efforts to diagnose and treat medical problems on a more individualized basis. This concept is frequently called “personalized medicine” and has many potential advantages [1]. Application of these principles to the evaluation of fetal growth reduces normal size parameter variability by using each fetus as its

own control [2]. An important clinical goal is to detect significant deviations of a size parameter (or set of parameters) from its predicted third-trimester trajectory, when most growth abnormalities occur [3]. Achieving this objective is significantly facilitated by the availability of individualized size standards.

Three methods are capable of generating individualized fetal size standards. The first is individualized growth assessment (IGA) which uses the slopes of second-trimester linear growth curves to specify Rossavik size models for one-, two- and three-dimensional parameters in fetuses with normal neonatal growth outcomes [2,4]. These models can be used to

Address for correspondence: Dr Russell L. Deter, MD, Department of Obstetrics and Gynecology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. Tel: +1-713-524-2877. E-mail: russellld@bcm.edu

predict third-trimester size trajectories that are in good agreement with actual measurements in normally growing fetuses [5].

The second method is the percentile line method (PLM), which has been used primarily to generate predicted birth weights [6–8] but is also the basis for age adjustments in calculating “customized weight percentiles” [9–12]. In this procedure, the percentile for a given size parameter measurement is derived from a population size curve, either at birth [8,9] or in the second or third trimesters [6,7]. The specified percentile is then used to identify a percentile line of the prenatal population size curve for the same population [7,8] or a different one [6]. This percentile line is assumed to be the expected trajectory for the size parameter in the fetus being studied. The PLM relies on the assumptions that individual fetuses grow along specific percentile lines of population size curves [8] and that individual size trajectories are parallel to each other. The current study represents an initial test of these assumptions.

A third approach, the conditional probability method (CPM), is a modification of the procedure described by Royston [13] for determining reference ranges based on conditional probability. The CPM has not been previously used to predict fetal size trajectories. In this method, a second-trimester value for an anatomical parameter is specified, together with its expected value and variance (derived from a set of longitudinal size curves). These data are used to calculate a *z*-score for the parameter value. Expected values are then determined for a set of third-trimester time points and the covariances between earlier and later expected values are calculated. Using these data and the function given by Royston [13], predicted values at third-trimester time points are calculated (both linear and quadratic functions can be fitted to longitudinal size curves; see Appendix for the variance and covariance calculations using these models).

Our study compares the utility of these three methods for predicting nine different fetal size parameters during the third trimester, using percent deviation (%Dev) measurements. The number of fetuses studied throughout pregnancy was relatively large and all had comprehensively defined normal, neonatal growth outcomes. Statistical parameters, needed to calculate predicted values and the %Dev comparisons themselves, were based on measurements obtained from the same sample. These characteristics significantly enhanced reliability of the results obtained.

## Methods

### Longitudinal study

This investigation was carried out using data obtained in a longitudinal growth study of 119 fetuses from the metropolitan Detroit area whose demographics have been described previously [14]. These fetuses had normal neonatal growth outcomes, as determined using multiple size parameters [the modified Neonatal Growth Assessment Score ( $m_3NGAS_{51}$ )] [15]. Briefly, 3D ultrasonography was used to scan fetuses at 3–4-week intervals from 17 to 40 weeks, menstrual age (MA) with measurement of the biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), femur diaphysis length (FDL), mid-thigh circumference (ThC),

humerus diaphysis length (HDL), mid-arm circumference (ArmC), fractional arm volume (AVol) and fractional thigh volume (TVol) [14]. Fetal age at the time of scan was determined primarily from crown-rump length measurements or last menstrual periods, the latter confirmed by second-trimester ultrasound examination (agreement within 7 days) [14]. Neonatal measurements of HC, AC, ThC, ArmC, crown-wheel length (CHL) and weight (WT) were made within 48 h of delivery [14,16]. The  $m_3NGAS_{51}$  values were calculated from growth potential realization index values [GPRI = (measured birth characteristics ÷ predicted birth characteristics) × 100] [16]. Only individuals with  $m_3NGAS_{51}$  values within the sample-specific normal range were included in this sample [14]. The Human Investigation Committee at William Beaumont Hospital and Institutional Review Board of the National Institute of Child Health and Human Development approved the research protocol.

### Data analysis

#### IGA

A detailed description of the IGA methods used in this investigation has been presented in a previous publication [14]. Briefly, linear functions were fit to 3–4 measurements before 28.2 weeks, MA, and their coefficient values estimated using the ordinary least squares (OLS) regression analysis. The slopes of these straight lines were used to specify Rossavik size models [ $P = c(t)^{k+s(t)}$ ] for each anatomical parameter in a given fetus [14]. These size models were used to generate predicted values for each third-trimester time point for which an actual measurement was available. Actual measurements were compared to predicted measurements and %Dev values calculated [2]:

$$\%Dev_{ijk} = \frac{\text{obs. measurement}_{ij} - \text{pred. measurement}_{ijk}}{\text{pred. measurement}_{ijk}} \times 100 \quad (1)$$

where subscripts *i*, *j* and *k* refer to the fetus, fetal age at time of measurement and method of obtaining the predicted value (*k*: 1 = IGA, 2 = PLM, 3 = CPM), respectively. This statistical parameter indicates how well actual measurements agree with predicted values and has an ideal value of 0%. If all other aspects are held constant except for the method of generating predicted values, %Dev differences provide the information needed to determine the optimal method for generating projected size trajectories for use in personalized growth evaluations.

A two-level statistical modeling procedure was then used to fit linear models to the %Dev values as a function of MA for each anatomical parameter [14]. Expected value function coefficients and variance component estimates were obtained as described by Royston [13] (Appendix, Reference 14). These data were used to calculate MA-specific 2 SD reference ranges for all time points in the third trimester at which a %Dev was calculated [14].

#### PLM

Specification of percentiles for anatomical measurements requires definition of expected values and their variances.

Such data can be obtained from longitudinal studies of fetal growth as described by Royston and Altman [17]. The first step in this procedure is to choose an appropriate mathematical model to describe individual size trajectories for each anatomical parameter. Based on scatter plots, linear and quadratic polynomials (dependent variable: anatomical measurement; independent variable: MA) were evaluated. Analysis of studentized residuals indicated that quadratic functions were optimal for BPD, HC, AC, FDL, ThC, HDL and ArmC. For AVol and TVol, natural logarithmic transformations of both the anatomical parameters and MA permitted the use of linear polynomials. For each anatomical parameter, a random coefficients model (a special case of hierarchical linear models) was used in a two-level modeling procedure [18] to obtain estimates of the expected value function coefficients and the variance components using SAS for Windows (Cary, NC), version 9.2 (exception: for FDL, model parameters obtained with MLwiN software (Bristol, UK) without intercept/slope covariance were used in order to assure positive total variances at all time points). Complete details of these mathematical procedures are summarized in the Appendix.

Percentile determination was made for each parameter ( $p$ ) in each fetus at 22 weeks, MA, the approximate mid-point of most second-trimester size curves. The estimated  $p$  value at 22 weeks ( $p_{22}$ ) was calculated using a linear function fit to the data before 28 weeks (see ‘‘IGA’’ section). This procedure minimized measurement error, provided parameter values at the same time point in all fetuses and gave a reference based on several second-trimester measurements (similar to the second-trimester slope values used in the IGA method). With the data described in the previous paragraph, the expected value at 22 weeks ( $EV_{22}$ ) and its variance ( $Var_{22}$ ) were calculated as given in the Appendix. The standard deviation at 22 weeks ( $SD_{22}$ ) was the square root of  $Var_{22}$ . The parameter value at a specified percentile for any given age ( $PV_i$ ) is related to the number ( $\lambda_i$ ) of SD units above or below the expected value. The following function was used to calculate the  $PV_i$  above the  $EV_i$  for a given  $\lambda_i$ :

$$PV_i = EV_i + (\lambda_i)(SD_i) \quad (2)$$

This function can be re-arranged to give the value of  $\lambda_i$  at 22 weeks for any  $PV_i$ :

$$\lambda_i = (PV_i - EV_i)/SD_i = z\text{-score}_i(z_i) \quad (3)$$

$$z_{22} = (PV_{22} - EV_{22})/SD_{22}$$

The value of  $\lambda$  calculated at 22 weeks, MA ( $z_{22}$ ), for each measurement defined the percentile line that the PLM assumes to be the individual size curve of an anatomical parameter being studied in a given fetus. At subsequent time points in the third trimester,  $z_{22}$  and the MA-specific  $EV_i$  and  $SD_i$  values were used to calculate third-trimester predicted parameter values ( $pPV_i$ ):

$$pPV_i = EV_i + (z_{22})(SD_i) \quad (4)$$

Actual measurements and predicted values were then used to calculate %Dev values [2].

### CPM

Predicted values using CPM method require calculation of  $PV_{22}$ ,  $EV_{22}$  and  $Var_{22}$ , as well as third-trimester MA-specific

$EV_i$  and  $var_i$  values. They also require the covariance ( $cov_{22,i}$ ) between  $EV_{22}$  and the third-trimester  $EV_i$  values at measurement time points. In the ‘‘Appendix’’ section, the equations for calculating  $cov_{22,i}$  for linear and quadratic functions are given. With these data, the predicted value at any third-trimester time point ( $pPV_i$ ) can be calculated using the following function [13]:

$$pPV_i = EV_i + (PV_{22} - EV_{22})(cov_{22,i}/Var_{22}) \quad (5)$$

Actual measurements and predicted values were again used to calculate percent deviations [2].

### Comparisons of IGA, PLM and CPM percent deviations

#### Data transformations.

##### 1. Reference range normalization

The meaning of differences in %Dev values cannot be ascertained unless they are compared to the reference range specified for the parameter studied at the time of measurement. For example, a 1% difference in %Dev values is 20% of a  $\pm 5\%$  2SD reference range (large difference) but only 4% of  $\pm 25\%$  2SD reference range (small difference). Reference ranges can vary substantially with the anatomical parameter (e.g. HC) being studied and the fetal age ( $t$ ) at which the %Dev measurement is made [14]. Parameter- and MA-specific reference ranges were calculated for IGA %Dev values using the method of Royston (Appendix, Reference 14). Since the expected value of any %Dev value is zero when growth is normal, these reference ranges were defined as extending from  $-2$  SD below zero to  $+2$  SD above zero. For subsequent analyses, each MA-specific %Dev value, obtained using any of the three methods studied [IGA ( $k=1$ ), PLM ( $k=2$ ) and CPM ( $k=3$ )], was transformed (T%Dev) by dividing it by the appropriate, MA-specific IGA  $+2SD$  range (to prevent a change in sign) [14]. Although the IGA value was used, any MA-specific  $+2SD$  range (i.e. from either PLM or CPM) could have been utilized. However, only IGA reference ranges are currently available. An example of these calculations, using HC, is presented below:

$$\text{IGA T\%Dev}_{i\text{HC } t1} = \frac{\%Dev_{i\text{HC } t1}}{\text{IGA } 2SD_{\text{HC } t}} \times 100$$

$$\text{PLM T\%Dev}_{i\text{HC } t2} = \frac{\%Dev_{i\text{HC } t2}}{\text{IGA } 2SD_{\text{HC } t}} \times 100$$

$$\text{CPM T\%Dev}_{i\text{HC } t3} = \frac{\%Dev_{i\text{HC } t3}}{\text{IGA } 2SD_{\text{HC } t}} \times 100$$

This procedure expressed all %Dev values in a common form as a percentage of the appropriate reference range, maintained the original relationships between %Dev values obtained with different methods for generating predicted values, and prevented changes in sign after transformation.

##### 2. Third-trimester T%Dev averaging

For most fetuses in this sample, there were 3–4 (range: 2–6) transformed %Dev values in the third trimester. The number and ages of measurements varied between fetuses but were identical for the three predicted value methods studied (IGA, PLM and CPM). To obtain a more representative indicator of transformed third-trimester %Dev measurements and to

simplify subsequent analysis, the available T%Dev values for each fetus were averaged (aT%Dev) as illustrated below for HC:

$$\begin{aligned} \text{IGA aT\%Dev}_{i\text{HC}1} &= \frac{\sum_{i=1}^{n_i\text{HC}} \text{T\%Dev}_{i\text{HC}1}}{n_i\text{HC}} \\ \text{PLM aT\%Dev}_{i\text{HC}2} &= \frac{\sum_{i=1}^{n_i\text{HC}} \text{T\%Dev}_{i\text{HC}2}}{n_i\text{HC}} \\ \text{CPM aT\%Dev}_{i\text{HC}3} &= \frac{\sum_{i=1}^{n_i\text{HC}} \text{T\%Dev}_{i\text{HC}3}}{n_i\text{HC}} \end{aligned}$$

$n_i\text{HC}$  = number of T%Dev<sub>*i* HC</sub> values available for Fetus *i*

### Statistical analysis

For each fetus, the aforementioned procedure resulted in three aT%Dev<sub>*i*</sub> values for each anatomical parameter, one for each method used to generate predicted values (a total of 27 for the 9 anatomical parameters). In evaluating the three methods for any given anatomical parameter, three criteria were assessed:

- (1) Which method gave the mean of the set of aT%Dev values closest to the ideal value of 0%?
- (2) Which method had the smallest mean of the set of aT%Dev values?
- (3) Which method had the smallest aT%Dev variance?

The mean of a set of aT%Dev values (mean aT%Dev) is a measure of the systematic deviates from zero and the aT%Dev variance can be considered a measure of its random variability. Both these characteristics would be minimized by a method that had the smallest mean aT%Dev value (not significantly different from zero) and the smallest aT%Dev variance. For each anatomical parameter, the mean aT%Dev value obtained with IGA, PLM and CPM were compared to zero using the *t*-test ( $p < 0.05$ ). Mean aT%Dev values for the three different methods were compared to each other using paired *t*-tests with a Bonferroni adjustment for non-independence (significance levels: IGA versus PLM and IGA versus CPM:  $p < 0.02$ ; PLM versus CPM:  $p < 0.01$ ). The three aT%Dev variances were compared using the Pitman test for correlated variances [19] using the same Bonferroni adjustment employed with the paired *t*-tests.

A secondary analysis of these data was carried out using a two-level mixed modeling procedure. Although the analysis described above is considered adequate, its power for

detecting true differences between methods is reduced. It uses only one value (the average) per fetus to represent the third-trimester T%Dev values obtained with each method instead of all values. For this reason, we repeated the analyses of systematic deviates and their variances using a two-level mixed-effects modeling procedure. This procedure and the results obtained are described in Supplementary File S1.

## Results

### Comparisons of systematic deviates to zero

Table 1 presents the mean aT%Dev values, our measures of systematic deviates from zero, for the three methods studied in nine anatomical parameters. Their *p* values for the comparisons to zero are also given. For IGA, 9/9 values were not significantly different from zero. With PLM the mean values not significantly different from zero decreased to 4/9, these four (AC, ThC, AVol and TVol) being soft tissue measures. For CPM, 5/9 values were not significantly different from zero (AC, ThC, AVol, TVol and FDL).

### Comparisons of systematic deviates between methods

Table 2 compares systematic deviates from zero for nine anatomical parameters obtained using IGA, PLM and CPM. For 8/9 parameters (exception: AVol), IGA mean aT%Dev values were smaller than those for PLM and CPM (Table 1). Seven of these differences were statistically significant when compared to the corresponding means obtained with either PLM or CPM (exceptions: AVol and TVol). Comparisons of PLM means to those for CPM indicated that seven were smaller for CPM and two for PLM (Table 1). Five of these differences were statistically significant (exceptions: AC, FDL, ThC and HDL).

### Comparisons of deviate variabilities between methods

Table 3 presents 2 SD ranges for aT%Dev values obtained using IGA, PLM and CPM for the nine anatomical parameters studied. Corresponding IGA ranges were smaller for 9/9 parameters compared to those for PLM while 6/9 were smaller and 3/9 essentially the same when compared to the CPM ranges. Five statistically significant differences were

Table 1. Comparisons of mean average third-trimester transformed percent deviations to zero.

Growth parameter	<i>n</i>	IGA aT%Dev mean value	<i>p</i> value <sup>a</sup>	PLM aT%Dev mean value	<i>p</i> value <sup>a</sup>	CPM aT%Dev mean value	<i>p</i> value <sup>a</sup>
BPD (cm)	117	5.2	0.19	25.6	0.0001*	17.4	0.0001*
HC (cm)	117	1.5	0.68	19.7	0.0001*	12.2	0.001*
AC (cm)	118	-4.3	0.26	4.8	0.24	2.3	0.55
FDL (cm)	117	1.1	0.77	9.6	0.03*	5.6	0.18
ThC (cm)	113	-0.3	0.94	-5.9	0.20	-5.3	0.21
HDL (cm)	118	2.1	0.57	16.7	0.0001*	14.1	0.001*
ArmC (cm)	118	0.2	0.96	-32.7	0.0001*	-20.4	0.0001*
AVol (mL)	118	2.5	0.54	0.0	0.99	-2.4	0.55
TVol (mL)	118	-0.4	0.92	-5.4	0.28	-7.2	0.13

aT%Dev = average third trimester transformed percent deviation; BPD = biparietal diameter; HC = head circumference; AC = abdominal circumference; FDL = femur diaphysis length; ThC = mid-thigh circumference; HDL = humerus diaphysis length; ArmC = mid-arm circumference; AVol = fractional arm volume; TVol = fractional thigh volume.

<sup>a</sup>One sample *t*-test; statistical significance at  $p < 0.05$  and designated by asterisk (\*).

observed in the IGA/PLM variance comparisons (BPD, FDL, ThC, HDL and TVol) and two (ThC and TVol) in the IGA/CPM comparisons. Ranges for PLM were larger for 9/9 parameters when compared to those for CPM. Variance differences were statistically significant for five parameters (BPD, FDL, HDL AVol and TVol). The IGA 2 SD ranges were more consistent among anatomical parameters than are those obtained with either PLM or CPM.

### Comparisons using two-level mixed-effects modeling procedure

There was a high degree of agreement between the results given in Tables 2 and 3 and those obtained using two-level mixed-effects modeling (Table A; Supplementary File S1). For the systematic deviates from zero, 20/27 of the comparisons were in agreement and two more (IGA versus PLM; ThC; IGA versus CPM: ThC) were almost in agreement. For variances, 21/27 values were in agreement despite the difference in variance parameters. Of the six that were not in agreement, only two involved IGA (IGA versus PLM; ThC; IGA versus CPM: ThC).

## Discussion

### Elimination of confounding variables

Essential to any assessment of differences between predictive size trajectory methods is the elimination of confounding variables. Only a procedure that minimizes such variables can give reliable information on how similar or different two or more methods perform. In our investigation, every effort was made to exclude variables that could affect comparisons between methods. Hence, the three different methods in our investigation were carried out under near ideal conditions. All %Dev were obtained in exactly the same way using the same measurements except that predicted values were determined with different procedures: IGA, PLM or CPM. Studies without adequate controls are likely to give less reliable results.

The first potential confounding variable is the nature of the sample used to provide comparative data. Comparison of fetal growth assessment methods requires use of samples without growth pathology. Normal fetal growth is well controlled as can be seen in the constant [20] or regularly changing [21] relationships between anatomical measurements during

Table 2. Comparisons of mean aT%Dev between groups.

Growth parameter	<i>n</i>	IGA versus PLM aT%Dev <i>p</i> value <sup>a</sup>	IGA versus CPM aT%Dev <i>p</i> value <sup>a</sup>	PLM versus CPM aT%Dev <i>p</i> value <sup>a</sup>
BPD (cm)	117	<0.02*	<0.02*	<0.01*
HC (cm)	117	<0.02*	<0.02*	<0.01*
AC (cm)	118	<0.02*	<0.02*	0.23
FDL (cm)	117	<0.02*	<0.02*	0.02
ThC (cm)	113	<0.02*	<0.02*	0.75
HDL (cm)	118	<0.02*	<0.02*	<0.05
ArmC (cm)	118	<0.02*	<0.02*	<0.01*
AVol (mL)	118	0.43	0.06	<0.01*
TVol (mL)	118	0.19	0.06	<0.01*

aT%Dev = average third trimester transformed percent deviation; BPD = biparietal diameter; HC = head circumference; AC = abdominal circumference; FDL = femur diaphysis length; ThC = mid-thigh circumference; HDL = humerus diaphysis length; ArmC = mid-arm circumference; AVol = fractional arm volume; TVol = fractional thigh volume.

<sup>a</sup>Paired t-test; Bonferroni adjustment: IGA versus PLM ( $p = 0.02$ ), IGA versus CPM ( $p = 0.02$ ), PLM versus CPM ( $p = 0.01$ ); statistical significance designated by asterisk (\*).

Table 3. Comparisons of variances for average third trimester transformed percent deviations between groups.

Growth parameter	<i>n</i>	IGA aT%Dev 2 SD range	PLM aT%Dev 2 SD range	CPM aT%Dev 2 SD range	IGA versus PLM variances <i>p</i> value <sup>a</sup>	IGA versus CPM variances <i>p</i> value <sup>a</sup>	PLM versus CPM variances <i>p</i> value <sup>a</sup>
BPD (cm)	117	85.4	100.2	88.2	<0.02*	>0.10	<0.01*
HC (cm)	117	78.4	87.4	79.2	<0.05	>0.10	<0.02
AC (cm)	118	83.0	88.2	82.8	>0.10	>0.10	>0.10
FDL (cm)	117	82.0	95.4	88.9	<0.02*	>0.10	<0.01*
ThC (cm)	113	82.8	96.2	90.0	<0.02*	<0.02*	>0.10
HDL (cm)	118	80.2	97.4	85.4	<0.02*	>0.10	<0.01*
ArmC (cm)	118	84.4	89.6	85.4	>0.10	>0.10	>0.10
AVol (mL)	118	87.4	95.6	87.0	>0.10	>0.10	<0.01*
TVol (mL)	118	85.2	106.8	101.4	<0.02*	<0.02*	<0.01*

IGA = Individualized Growth Assessment; PLM = Percentile Line Method; CPM = Conditional Probability Method; aT%Dev = average third trimester transformed percent deviation; BPD = biparietal diameter; HC = head circumference; AC = abdominal circumference; FDL = femur diaphysis length; ThC = mid-thigh circumference; HDL = humerus diaphysis length; ArmC = mid-arm circumference; AVol = fractional arm volume; TVol = fractional thigh volume.

<sup>a</sup>Pitman correlated variance test; Bonferroni adjustment: IGA vs PLM, IGA vs CPM ( $p = 0.02$ ), PLM vs CPM ( $p = 0.01$ ); statistical significance designated by asterisk (\*).

pregnancy. Growth pathology is a result of interference with these regulatory systems [22,23]. Consequently, different interactions between disordered growth and different growth assessment methods may occur. Our longitudinal study design provided information on the neonatal growth status of each fetus that is not available in cross-sectional studies. The use of a GPRI-based, weighted, five parameter  $m_3NGAS_{51}$  [15] for determining neonatal growth outcome accounts for the major variables confounding neonatal growth assessment. For that reason, it provides a more comprehensive basis for deciding which neonates had normal growth outcomes. All fetuses in our sample had normal neonatal growth outcomes (as indicated by the  $m_3NGAS_{51}$  and a sample-specific normal range), providing strong evidence that this study was carried out in normally growing fetuses.

A second confounding variable is the use of a different sample to determine percentile lines from the one used to compare predicted values to actual measurements [6–9,24,25]. Percentile lines based on one sample may not be the same as those of a second sample even if both samples are from the same population. This problem was eliminated in our longitudinal study through the use of two-level statistical modeling. This statistical procedure provided the means for obtaining estimates of the expected value function coefficients and the variance components needed to determine age-specific SD from longitudinal data sets (Appendix) [13]. Expected values and SD are the statistics used in determining percentile line predictions. In our investigation, the same sample was used to determine all specific percentile lines and to calculate the predicted values that were compared with actual measurements in the third trimester.

Four other sources of variation between individuals can make comparisons of growth assessment methods unreliable: (i) anatomical parameter estimates rather than direct measurements, (ii) differences in measurement errors in specifying the growth potential parameter, (iii) differences in the number and age at different ultrasound examinations and (iv) differences in normal variability at different ages for different anatomical parameters. The present study was limited to parameters that could be directly measured using ultrasonography. Estimated WT was not evaluated because it cannot be measured directly [26]. However, the parameters used in WT estimation procedures were included so it is likely that similar results would be obtained with EWT. A linear function was fit to the data before 28.2 weeks, MA and either the slope or reference measurement (at the same fixed age for all fetuses) was defined. This procedure minimized potential errors associated with early measurement of anatomical parameters and gave reference measurements with similar characteristics. It also provided a uniform way to define “growth potential” in all fetuses for the three growth assessment methods evaluated. Applying all three methods to exactly the same large sample of fetuses and third-trimester time points eliminated sampling error between methods. Expressing the comparison measurements (%Dev) obtained with all three methods as percentages of the same reference ranges (T%Dev) resulted in a statistical parameter normalized for differences in variability over time and between parameters. The average of transformed %Dev values is a more representative indicator of T%Dev values in the third trimester therefore is appropriate for determining

group characteristics and comparisons between anatomical parameters.

### Principal findings of the study

All nine anatomical parameters evaluated with IGA gave systematic deviations that were not significantly different from zero (100%) and most were smaller than those of PLM and CPM. Comparable values for PLM and CPM were 44% and 56%. Fifty six percent of PLM deviation variances and twenty percent of PLM variances were larger than those for IGA. Approximately half of the systematic deviations for PLM and CPM were different from each other, with most being smaller for CPM. Of the nine deviation variances, 56% differed between PLM and CPM with all CPM variances being smaller. These results indicate that IGA was the optimal method for determining individualized third-trimester size standards, with CPM being better than PLM for most parameters.

### Previous studies

#### *Specification of third-trimester size trajectories*

Direct comparisons of methods capable of prospectively specifying third-trimester growth trajectories are very limited. No published comparisons of IGA or PLM with CPM have been reported and only one comparison of IGA with PLM. Shields et al. [7] evaluated the relationships between predicted and actual third-trimester measurements for BPD, HC, AC and FDL in a low risk Hispanic population. They concluded that PLM gave more accurate predictions than IGA, in direct contradiction to the results presented here. However, assessment of their results requires consideration of several issues related to the validity of their conclusion: (i) IGA prediction errors were presented as absolute percentage values (combining systematic and random prediction errors) so comparisons with other IGA studies that use signed percentage values are not possible; (ii) the nature of their reported BPD measurement errors and incorrectly cited data for other parameters make direct comparison of their results to other studies difficult; (iii) percentile lines in the Shield’s study were obtained from a cross-sectional sample from another population and many essential characteristics of this sample (e.g. number of patients, scans per patient included, fetal age determination method used, normalcy of growth) were not specified and (iv) direct comparisons of systematic and random differences were not carried out as done in the current study.

#### *Previous IGA, PLM and CPM investigations*

In earlier publications, IGA was used to evaluate third-trimester growth and predict birth characteristics, to detect growth and size abnormalities and to determine perinatal complication risks associated with different growth/size categories [2,27,28]. The IGA procedures in the current investigation are essentially the same as those used previously except that they were derived from a much larger sample [14].

The PLM has been used to evaluate third-trimester WT estimates and predict birth weight, detect size abnormalities and determine perinatal complication risks associated with different size categories [6–9,24,25,29,30]. Most growth analyses that have been based on individual percentile lines

were carried out at birth [8,9,29] or in the mid-to-late third trimester [6,24,25] using either single measurements [6,8,9,25,29] or the average of several measurements [7,24] taken at different time points in different fetuses/neonates. The data required for percentile line specification were obtained from the same population but a different sample [7–9,24] or from a different population [6,25,29]. Only three investigations [6,7,31] used second-trimester measurements (single or multiple varying time points in different fetuses) for percentile line specification and in only one [31] was specification data obtained from the same sample. The results of this latter study, using FDL, indicated that only 13% of fetuses remained in the quartile specified by their second-trimester measurements during the third trimester [31].

The CPM has not been previously used to specify third-trimester growth trajectories. However, it has been used to evaluate third-trimester size, predict birth weights and abdominal profile areas, detect size abnormalities and determine perinatal complication risks associated with different size categories [32–34]. Single measurements, obtained serially at 4-week intervals beginning at 20 weeks [32] or at different early [33,34] or mid [33] third-trimester time points have been used to specify subsequent conditioned reference ranges at later time points. Statistical parameters used in these calculations either came from the same sample [33,34] or from a different population [34]. The prior applications of conditional probability (e.g. generation of conditional reference ranges for use in cross-sectional evaluations) differ significantly from the objective in this investigation (e.g. generation of individual predicted size trajectories).

### Novel aspects of this study

This investigation provides the most reliable comparisons between IGA and PLM in generating third-trimester size trajectory predictions. We introduce a new method (CPM) for generating such trajectories and directly compared all three approaches using two different statistical methods. Results are provided for nine fetal anatomical size parameters.

### Strengths and limitations

No previous studies using either PLM or CPM have had the following characteristics of the current investigation:

- (i) Use of data from a prospective, longitudinal study of fetal growth that included only fetuses with rigorously defined, normal neonatal growth outcomes.
- (ii) Growth potential specification at the same time point in the second trimester, derived from multiple measurements, for all fetuses.
- (iii) Calculation of predicted values using statistical parameters derived from the same sample used for the evaluation of predicted values.

These evaluations were only possible using longitudinal data sets and two-level statistical modeling. Our approach corrects for most, if not all, postulated causes for the failure of CPM to improve risk assessment based on birth weight categories [34]. These causes will most likely affect results obtained with PLM also.

An important study limitation is the possible loss of power in detecting differences resulting from the use of

third-trimester aT%Dev values, instead of individual T%Dev values. However, the same analyses using two-level mixed-effects modeling based on individual T%Dev values, yielded similar results (Supplementary File S1).

### Implications for research and clinical care

Our study compares three methods for generating third-trimester size trajectories in normally growing fetuses. When compared to the IGA approach, population-based percentile lines do not appear to optimally characterize the growth trajectories of individual fetuses. Such discrepancies may contradict a fundamental assumption of EFW evaluation procedures that are commonly used in clinical practice (i.e. that population percentile lines are the actual size trajectories of individual fetuses).

### Conclusions

Comparisons between actual and predicted third-trimester measurements for nine fetal growth parameters indicate significant differences between IGA, PLM and CPM. Fetal size trajectories, generated using the IGA method, provide the smallest errors in individualized evaluation of normal fetal growth during the third trimester. This personalized approach may improve the detection of pregnancies that require closer surveillance or may benefit from therapeutic intervention. However, the precise relationships between deviations from normal size trajectories and adverse perinatal outcomes require further investigation. Freely available computer software called individualized Growth Assessment Program (iGAP) now permits personalized size assessment in both fetuses and neonates (<http://igap.research.bcm.edu>).

### Acknowledgements

The authors wish to acknowledge the technical assistance of Melissa Powell, RDMS and Beverley McNie, BS, CCRP. We also wish to thank M. Balasubramaniam for providing the data given in the Appendix.

### Declaration of interest

This research was supported (in-part) by the Perinatology Research Branch, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, DHHS. R. Romero contributed to this work as part of his official duties as an employee of the United States Federal Government. None of the other authors have disclosed a conflict of interest

### References

1. Hamburg M, Collins FS. The path to personalized medicine. *N Engl J Med* 2010;363:301–4.
2. Deter RL. Individualized growth assessment: evaluation of growth using each fetus as its own control. *Sem Perinatol* 2004;28:23–32.
3. Rodeck CH, Whittle MJ. *Fetal Medicine: basic science and clinical practice*. London: Elsevier Health Sciences; 2008.
4. Deter RL, Rossavik, IK, Harrist RB, Hadlock FP. Mathematical modeling of fetal growth: development of individual growth curve standards. *Obstet Gynecol* 1986;68:156–61.
5. Hata T, Deter RL, Hill RM. Individual growth curve standards in triplets: prediction of third-trimester growth and birth characteristics. *Obstet Gynecol* 1991;78:379–84.

6. Santonja-Lucas JJ, Armero C, Martinez-Gonzalez L. Long-term prediction of birth weight. *J Ultrasound Med* 1993;12:431–6.
7. Shields LE, Huff RW, Jackson GM, et al. Fetal growth: a comparison of growth curves with mathematical modeling. *J Ultrasound Med* 1993;5:271–4.
8. Mongelli M, Gardosi J. Gestation-adjusted projection of estimated fetal weight. *Acta Obstet Gynecol Scand* 1996;75:28–31.
9. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995;6:168–74.
10. McCowan LM, Harding JE, Stewart AW. Customized birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG* 2005;112:1026–33.
11. Hutcheon JA, Walker M, Platt RW. Assessing the value of customized birth weight percentiles. *Am J Epidemiol* 2011;173:459–67.
12. Figueras F, Figueras J, Meler E, et al. Customised birthweight standards accurately predict perinatal morbidity. *Arch Dis Child Fetal Neonatal Ed* 2007;92:F277–80.
13. Royston P. Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Stat Med* 1995;14:1417–36.
14. Deter RL, Lee W, Sangi-Haghpeykar H, et al. Individualized fetal growth assessment: critical evaluation of key concepts in the specification of third trimester growth trajectories. *J Matern Fetal Neonatal Med* 2014;27:537–42.
15. Deter RL, Spence LR. Identification of macrosomic, normal and intrauterine growth retarded neonates using the modified Neonatal Growth Assessment Score. *Fetal Diagn Ther* 2004;19:58–67.
16. Deter RL, Hill RM, Tennyson LM. Predicting the birth characteristics of normal fetuses 14 weeks before delivery. *J Clin Ultrasound* 1989;17:89–93.
17. Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995;6:307–12.
18. Goldstein H. Multilevel statistical models. 2nd ed. London: Arnold; 1995.
19. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames, Iowa: The Iowa State University Press; 1980.
20. Callen, PW. Ultrasonography in obstetrics and gynecology. 4th ed. Philadelphia: W.B. Saunders Company; 2000.
21. Deter RL, Hadlock FP, Harrist RB. Evaluation of normal fetal growth and the detection of intrauterine growth retardation. In: Callen PW, ed. Ultrasonography in obstetrics and gynecology. 1st ed. Philadelphia: W.B. Saunders Company; 1983:113–40.
22. Baschat AA. Pathophysiology of fetal growth restriction: implications for diagnosis and surveillance. *Obstet Gynecol Surv* 2004;59:617–27.
23. Miller J, Turan S, Baschat AA. Fetal growth restriction. *Semin Perinatol* 2008;32:274–80.
24. Hedriana HL, Moore TR. A comparison of single versus multiple growth ultrasono-graphic examinations in predicting birth weight. *Am J Obstet Gynecol* 1994;170:1600–6.
25. Pressman EK, Bienstock JL, Blakemore KJ, et al. Prediction of birth weight by ultrasound in the third trimester. *Obstet Gynecol* 2000;95:502–6.
26. Lee W, Balasubramaniam M, Deter RL, et al. New fetal weight estimation models using fractional limb volume. *Ultrasound Obstet Gynecol* 2009;34:556–65.
27. Simon, NV, Deter RL, Kofinas AD, et al. Small-for-menstrual-age infants: different subgroups detected with individualized fetal growth assessment. *J Clin Ultrasound* 1994;22:3–10.
28. Hata T, Kuno A, Akiyama M, et al. Detection of small-for-gestational-age infants with poor perinatal outcomes using individualized growth assessment. *Gynecol Obstet Invest* 1999;47:162–5.
29. Hutcheon JA, Zhang X, Crattingius S, et al. Customized birth weight percentiles: does adjusting for maternal characteristic matter? *Brit J Obstet Gynecol* 2008;115:1397–404.
30. Gardosi J. Customized fetal growth standards: rationale and clinical application. *Sem Perinatol* 2004;28:33–40.
31. Bjornerem A, Johnsen SL, Nguyen TV, et al. The shifting trajectory of growth in femur length during gestation. *J Bone Min Res* 2010;25:1029–33.
32. Owen P, Ogston S. Conditional centiles for the quantification of fetal growth. *Ultrasound Obstet Gynecol* 1998;11:110–7.
33. Owen P, Burton K, Ogston S, et al. Using unconditional and conditional standard deviation scores of fetal abdominal area measurements in the prediction of intrauterine growth restriction. *Ultrasound Obstet Gynecol* 2000;16:439–44.
34. Hutcheon JA, Egeland GM, Morin L, et al. The predictive ability of conditional fetal growth percentiles. *Paediatr Perinatol Epidem* 2010;24:131–9.

## Appendix

### Calculation of statistical parameters required for the percentile line and CPMs

Royston [13] has provided a method for calculating age-specific reference ranges for fetal anatomical parameters based on longitudinal data. This method utilizes the random coefficients model and two-level statistical modeling [18]. Coefficients for the model used to calculate expected values ( $EV_i$ ) at various time points ( $Age_i$ ) are generated, as well as estimates of coefficient variances (var), their covariances (cov) and the random error (varE). Royston [13] also gives equations for calculating the variance (varT) associated with any expected value and the covariance (covT) between expected values at different time points. These functions are specific for the linear model but can be extended to the quadratic model [32,33].

Expected value functions

$$\text{Linear model: } EV_i = B_0 + B_1(Age_i)$$

$$\text{Quadratic model: } EV_i = B_0 + B_1(Age_i) + B_2(Age_i)^2$$

Expected value variances

$$\text{Linear model: } \text{Var}T_i = \text{var}_{B_0} + \text{var}_{B_1}(Age_i)^2 + 2\text{cov}_{B_0,B_1}(Age_i) + \text{varE}$$

$$\text{Quadratic model: } \text{Var}T_i = \text{var}_{B_0} + \text{var}_{B_1} + \text{var}_{B_2} + 2\text{cov}_{B_0,B_1}(Age_i) + 2\text{cov}_{B_0,B_2}(Age_i)^2 + 2\text{cov}_{B_1,B_2}(Age_i) + \text{varE}$$

Table A1. Expected value function coefficients and variance components for growth measurements.

Parameter	$B_0$	$B_1$	$B_2$	$\text{var}_{B_0}$	$\text{var}_{B_1}$	$\text{var}_{B_2}$	$\text{COV}_{B_0,B_1}$	$\text{COV}_{B_0,B_2}$	$\text{COV}_{B_1,B_2}$	varE
BPD	-4.617	0.579	-0.00548	0.3998	0.00356	0.000001141	-0.0365	0.00058	-0.00006	0.0333
HC	-16.235	2.113	-0.02064	8.0881	0.05947	0.00002200	-0.6787	0.01264	-0.00112	0.2738
AC	-9.472	1.308	-0.00357	13.6154	0.10100	0.00004000	-1.1604	0.02262	-0.00198	0.3574
FDL	-4.046	0.441	-0.00372	0.0245	0.00000	0.0000000615	0.00000	0.00000	0.000000	0.0258
ThC	-3.118	0.384	0.00422	5.1081	0.03361	0.00001200	-0.4082	0.00729	-0.00061	0.1557
HDL	-3.309	0.400	-0.00382	1.1847	0.00748	0.000002380	-0.0930	0.00163	-0.00013	0.0144
ArmC	-1.433	0.238	0.00324	3.9379	0.02388	0.000008643	-0.3010	0.00542	-0.00044	0.0859
$\text{Log}_e$ AVol	-11.340	4.080	-	0.5038	0.04221	-	-0.1436	-	-	0.0103
$\text{Log}_e$ TVol	-11.794	4.4501	-	0.2875	0.02487	-	-0.0823	-	-	0.0120

$\text{Log}_e$  MA used as independent variable in  $\text{Log}_e$  AVol and  $\text{Log}_e$  TVol functions.

Estimates obtained using REML regression analyses (SAS) [exception: FDL and RIGLS (MLwiN)].

BPD = biparietal diameter; HC = head circumference; AC = abdominal circumference; FDL = femur diaphysis length; ThC = mid-thigh circumference; HDL = humerus diaphysis length; ArmC = mid-arm circumference; AVol = fractional arm volume; TVol = fractional thigh volume.

Covariance between expected values at Age<sub>1</sub> and Age<sub>i</sub>

Linear model:  $\text{Cov}\Gamma_{1,i} = \text{var}_{B_0} + \text{var}_1(\text{Age}_1)(\text{Age}_i) + \text{cov}_{B_0,B_1}(\text{Age}_1 + \text{Age}_i)$

Quadratic model:  $\text{Cov}\Gamma_{1,i} = \text{var}_{B_0} + \text{var}_{B_1}(\text{Age}_1)(\text{Age}_i) + \text{var}_{B_2}(\text{Age}_1)^2 (\text{Age}_i)^2 + \text{cov}_{B_0,B_1}(\text{Age}_1 + \text{Age}_i) + \text{cov}_{B_0,B_2}(\text{Age}_1^2 + \text{Age}_i^2) + \text{cov}_{B_1,B_2}[(\text{Age}_1 \text{Age}_i^2) + (\text{Age}_i \text{Age}_1^2)]$

The square root of the total variance at a given age is the SD and 2 SD includes 95% of the measurements at that age. The reference range is determined by adding and subtracting 2 SD from the expected value. Statistical parameters used in this investigation are given in the table above: they were obtained using complete data sets (17–40 weeks, MA) without prior knowledge of the IGA results obtained with these data sets.

**Supplementary material available online**

Supplementary File S1