Published in final edited form as:

J R Stat Soc Ser A Stat Soc. 2018 June ; 181(3): 825–842. doi:10.1111/rssa.12312.

Estimating onset time from longitudinal and cross-sectional data with an application to estimating gestational age from longitudinal maternal anthropometry during pregnancy and neonatal anthropometry at birth

Ana Maria Ortega-Villa,

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, 20850, United States Of America. 9609 Medical Center Dr, Room SG/7E132, Rockville, MD, 20850, United States Of America

Katherine L. Grantz, and

Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda Maryland, 20892, United States of America

Paul S. Albert*

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, 20850, United States Of America. 9609 Medical Center Dr, Room SG/7E146, Rockville, MD, 20850

Summary

Determining the date of conception is important for estimating gestational age and monitoring whether the fetus and mother are on track in their development and pregnancy. Various methods based on ultrasound have been proposed for dating a pregnancy in high resource countries. However, such techniques may not be available in under-resourced countries. We develop a shared random parameter model for estimating the date of conception using longitudinal assessment of multiple maternal anthropometry and cross-sectional neonatal anthropometry. The methodology is evaluated with a training-test set paradigm as well as with simulations to examine the robustness of the method to model misspecification. We illustrate this new methodology with data from the NICHD Fetal Growth Studies.

Keywords

Measurement error; Onset time estimation; Prediction; Regression - calibration

1. Introduction

An important problem in international health is identifying the gestational age of a newborn, so that proper intervention can be employed in case of preterm birth. The last menstrual

---

*Corresponding Author. albertp@mail.nih.gov, Tel: (240) 276 - 7593.

period (LMP) has been used as a method for assessing gestational age, where women are asked to recall the first day of their last menstrual period and a fixed number of days is added to approximate the time of conception. However, this is known to have sizable measurement error due to recollection problems and cycle irregularities among women (Johnson et al., 2011; Hall et al., 1985). Combinations of longitudinal maternal and cross-sectional neonatal anthropometry measurements to improve gestational age estimation are appealing given that they are low tech and low cost.

Several methods in obstetrics have been proposed for dating a pregnancy in high resource countries. Early ultrasound is often used as a gold standard in population based studies for dating gestational age. For example, measures such as crown-rump length (length from the top of the head to the bottom of the buttocks), biparietal diameter (transverse diameter of the head) and femur length are used in the early part of gestation to predict gestational age since there is very little variation across subjects (Taipale and Hiilesmaa, 2001). In low resource countries however, ultrasound-based measures of gestational age are not often available and LMP can only be assessed in some individuals. In these situations, others have proposed fundal height as a method for dating gestational age. White et al. (2012) proposed ad hoc statistical methodology for developing a predictor for gestational age based on longitudinal assessments of fundal height.

Various longitudinal maternal anthropometric measurements are collected as part of routine prenatal care across a woman's pregnancy in order to assess the status of the fetus. Some of these measurements are fundal height, a measurement from the pubic symphysis to the top of the gravid uterus, and weight. Neonatal anthropometric measurements typically include birthweight, length, and head circumference. Based on these measurements along with the additional study measurement of skin folds and neonatal abdominal circumference, we develop a methodological framework for predicting gestational age.

In this paper, we present a framework for predicting gestational age using univariate and multivariate longitudinal information from maternal anthropometry. Further, the approach can include cross-sectional information from neonatal anthropometry at birth. We develop a mixed modeling framework with the goal of creating a predictor for the onset time (conception) that can be used for estimating gestational age at birth. We evaluate the properties of the predictor with simulations and examine the predictive accuracy of longitudinal maternal and cross-sectional neonatal anthropometry for estimating gestational age with a unique data source. The data application comes from the NICHD Fetal Growth Studies, Louis et al. (2015), a prospective cohort study of 2334 non-obese pregnant women in the U.S. population.

In Section 2 we present the model formulation. In Section 3 we present estimations using the example data from the NICHD Fetal Growth Studies. In Section 4 we present a simulation study that examines the properties of the estimation procedures, as well as examines the robustness of the approach to model misspecification. We present concluding remarks in Section 5.

The programs that were used to analyse the data can be obtained from <http://wileyonlinelibrary.com/journal/rss-datasets/>

2. Model Formulation

2.1. Univariate Longitudinal Maternal Anthropometry

We consider the case where we have a gold standard measurement of gestation for individuals in the training set. Let Y_{ij} be the j th fundal height measurement on the i th participant of the maternal outcome. Denote G_i as the time from conception until the first maternal visit for the i th subject. We propose the model,

$$Y_{ij} = X_{ij}(G_i + t_{ij})'\beta + Z_{ij}(G_i + t_{ij})'\mathbf{b}_i + \varepsilon_{ij} \quad (1)$$

where, $i = 1, \dots, N$ represents the subject, $j = 1, \dots, n_i$ represents the measurement, $\mathbf{b}_i$ is assumed to have a multivariate normal distribution with mean $\mathbf{0}$ and variance $\Sigma_{\mathbf{b}}$, ε_{ij} is normal with mean 0 and variance σ_{ε}^2 , $X_{ij}(t_{ij}) = (1, t_{ij}^2, \dots, t_{ij}^{p_1})'$ and $Z_{ij}(t_{ij}) = (1, t_{ij}^2, \dots, t_{ij}^{p_2})'$ are polynomial representations of $G_i + t_{ij}$ with p_1 and p_2 being the orders of the polynomials for the fixed and random effects respectively. Recall, the random variable G_i can be interpreted as the time from conception to the first follow up visit for each subject i . We assume that G_i is either a normal or a truncated normal distribution with mean and variance parameters μ_G and σ_G^2 respectively, depending on whether a maternal anthropometry measure is only considered after a particular time in gestational age. In this scenario, we are considering the model presented by Laird and Ware (1982). Further, we assume that $t_{i1} = 0$ for all i , with the interpretation of t_{ij} being the time from the initial time of observation for each subject until the j -th measurement. Motivated by the data example, we set $p_1 = p_2 = 1$, and the model is

$$Y_{ij} = \beta_0 + \beta_1(G_i + t_{ij}) + b_{i0} + b_{i1}(G_i + t_{ij}) + \varepsilon_{ij}. \quad (2)$$

We also assume that μ_G does not depend on covariates. If μ_G does depend on covariates, it means that certain covariates could influence an individual to start her follow up visits earlier or later than others. Most fetal growth studies, however, have prescribed times for all subjects to enter the study.

In our data example, the primary maternal longitudinal anthropometry measurement is fundal height. Fundal height measurements are considered more biologically accurate after 20 weeks of gestation, therefore we only considered fundal height measurements taken after 20 weeks.

We can estimate the model parameters by maximizing the likelihood, $L = \log \prod_{i=1}^N L_i(Y_i, G_i)$, where $L_i(Y_i, G_i)$ is each individual's contribution to the likelihood, which is given by

$$L_i(\mathbf{Y}_i, G_i) = \left[\int_{\mathbf{b}_i} f(\mathbf{Y}_i | \mathbf{b}_i, G_i) h(\mathbf{b}_i) d\mathbf{b}_i \right] g(G_i) \quad (3)$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ij})$, $f(\mathbf{Y}_i | \mathbf{b}_i, G_i)$ is a normal distribution with mean $\boldsymbol{\mu}_{\mathbf{Y}_i}$ and variance $\sigma_{\epsilon}^2 I$, with $\boldsymbol{\mu}_{\mathbf{Y}_i} = X_{ij}(G_i + t_{ij})' \boldsymbol{\beta} + Z_{ij}(G_i + t_{ij})' \mathbf{b}_i$, G_i is normally distributed with mean μ_G and variance σ_G^2 and $\mathbf{b}_i$ is distributed normal with mean $\mathbf{0}$ and variance Σ_b .

Since the likelihood (3) factors as $L_i(\mathbf{Y}_i, G_i) = L_i(\mathbf{Y}_i | G_i) L_i(G_i)$, the parameters of the first component (1) can be estimated using standard statistical software, while the parameters of the second component can be estimated using standard optimization techniques. Specifically, we used the lme function (Pinheiro et al., 2014) in R for estimating the parameters of the first component. The parameters for the second component μ_G and σ_G^2 , can be estimated by maximizing a likelihood that explicitly accounts for the fact that fundal height is not considered before 20 weeks. This likelihood is given by

$$L_G = \prod_{i=1}^N \frac{\phi(G_i; \mu_G, \sigma_G^2)}{1 - \Phi(20; \mu_G, \sigma_G^2)} \quad (4)$$

where $\phi(G_i; \mu_G, \sigma_G^2)$ represents the normal density function and $\Phi(G_i; \mu_G, \sigma_G^2)$ represents the normal cumulative distribution.

In terms of gestational age prediction, we can compute $E[G_i | \mathbf{Y}_i, \boldsymbol{\theta}]$, where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \Sigma_b, \sigma_{\epsilon}^2, \mu_G, \sigma_G^2\}$.

$$\begin{aligned} E[G_i | \mathbf{Y}_i, \boldsymbol{\theta}] &= \frac{\int_G \int_{\mathbf{b}_i} G_i \left[\prod_j f(y_{ij} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i}{\int_G \int_{\mathbf{b}_i} \left[\prod_j f(y_{ij} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i} \quad (5) \\ &= \frac{\int_G G_i f(\mathbf{y}_i | G_i) g(G_i) dG_i}{\int_G f(\mathbf{y}_i | G_i) g(G_i) dG_i} \end{aligned}$$

2.2. Multiple Longitudinal Maternal Anthropometry Measures

In this section, we extend the univariate method, in order to include other maternal anthropometry measurements with the goal of accuracy improvement.

Consider the model

$$Y_{ijk} = X_{ijk}(G_i + t_{ij})' \beta_k + Z_{ijk}(G_i + t_{ij})' \mathbf{b}_{ik} + \varepsilon_{ijk} \quad (6)$$

where, $i = 1, \dots, N$ represents the subject, $j = 1, \dots, n_i$ represents the measurement, $k = 1, \dots, K$ represents the number of longitudinal maternal measures to be considered, $\mathbf{b}_{ik} \sim N(\mathbf{0}, \Sigma_b)$, $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ and $G_i \sim N(\mu_G, \sigma_G^2)$.

The maternal anthropometry measurements considered in our data example are fundal height, maternal weight change and subscapular skin fold, resulting in $K = 3$. Similarly, we assume that $t_{i1} = 0$ for all i , and that G_i is observed.

The estimation of (6) can be obtained by either estimating the multivariate linear mixed model directly when $K = 3$ or by fitting pairwise mixed models as proposed by Fieuws and Verbeke (2006). In the high-dimensional case, it is recommended to fit all possible bivariate models, and then obtain the needed coefficients by averaging over the estimates that are calculated more than once, and extracting those that have been uniquely estimated.

Gestational age G_i is estimated by evaluating

$$E[G_i | Y_{i1}, \dots, Y_{ik}, \theta], \text{ where } Y_{ik} = (Y_{i1k}, \dots, Y_{iKk}) \text{ and } \theta = \{\beta, \Sigma_b, \sigma_\varepsilon^2, \mu_G, \sigma_G^2\}.$$

$$\begin{aligned} E[G_i | Y_{i1}, \dots, Y_{ik}, \theta] &= \frac{\int_G \int_{\mathbf{b}_i} G_i \left[\prod_j f(y_{ij1}, \dots, y_{ijK} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i}{\int_G \int_{\mathbf{b}_i} \left[\prod_j f(y_{ij1}, \dots, y_{ijK} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i} \quad (7) \\ &= \frac{\int_G G_i f(y_{ij1}, \dots, y_{ijK} | G_i) g(G_i) dG_i}{\int_G f(y_{ij1}, \dots, y_{ijK} | G_i) g(G_i) dG_i} \end{aligned}$$

Similar to the univariate case from Section 2.1, we truncate fundal height at 20 weeks. However, since there are no restrictions on measurement times for anthropometric measurements other than fundal height, μ_G and σ_G^2 can be obtained as $\hat{\mu}_G = \frac{1}{N} \sum_i G_i$ and $\hat{\sigma}_G^2 = \frac{1}{N-1} \sum_i (G_i - \hat{\mu}_G)^2$, respectively.

2.3. Multiple Longitudinal Maternal Anthropometry Measures and Neonatal Anthropometry at Birth

In this model we consider multiple longitudinal maternal anthropometry measurements, as presented in Section 2.2, along with several neonatal anthropometry measures measured once, at birth. Examples of the neonatal measurements include abdominal and head circumference, length and birthweight. The complete model for this scenario is given by the shared random effects model, where the shared random parameter provides a flexible and

simple way to introduce dependence between the longitudinal maternal and cross-sectional neonatal anthropometric measurements, see Albert and Follman (2008). The model is

$$\begin{aligned} Y_{ijk} &= X_{ijk}(G_i + t_{ij})' \beta_k + Z_{ijk}(G_i + t_{ij})' b_{ik} + \varepsilon_{ijk}, \quad k = 1, \dots, K, \\ W_{il} &= V_{il}(G_i + t_{i*})' \gamma_l + b_i' \psi_l + \varepsilon_{il}, \quad l = 1, \dots, L \end{aligned} \quad (8)$$

where, $i = 1, \dots, N$ represents the subject, $j = 1, \dots, n_i$ represents the measurement, $k = 1, \dots, K$ represents the maternal anthropometry measure and $l = 1, \dots, L$ represents the neonatal anthropometry measure, obtained at delivery, denoted as $(G_i + t_{i*})$, where t_{i*} is the time from the first measurement to delivery. In addition, $b_i = (b_{i1}, \dots, b_{iK})$, $b_{ik} \sim N(0, \Sigma_b)$,

$\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, $X_{ij}(t_{ij}) = (1, t_{ij}^2, \dots, t_{ij}^{p_1})'$ and $Z_{ij}(t_{ij}) = (1, t_{ij}^2, \dots, t_{ij}^{p_2})'$ are polynomial representations of G_i and t_{ij} with p_1 and p_2 being the orders of the polynomials for the fixed and random effects respectively. The parameter vector ψ_l characterizes the association between the longitudinal maternal and neonatal cross-sectional measurements by linking the random effects across the two processes. The mean structure for each neonatal measurement is governed through $V_{il}(t_i)$, where $V_{il}(t_i) = (1, t_i^2, \dots, t_i^{p_3})'$ is a polynomial representation of G_i and t_{i*} , where $t_i = G_i + t_{i*}$. The random variable G_i is either considered normally distributed with mean μ_G and variance σ_G^2 when there is no constraint on how early the first measurement is, or truncated normal when there is such a constraint.

The shared random parameter provides a flexible and simple way to introduce dependence between the longitudinal maternal and cross-sectional neonatal anthropometric measurements.

The likelihood of the full model is given by

$$L_i(Y_i, W_i, G_i) = \left[\int_{b_i} \prod_j f(Y_{i1}, \dots, Y_{ik}, W_{i1}, \dots, W_{iL} \mid b_i, G_i) h(b_i) db_i \right] g(G_i) \quad (9)$$

Where $Y_i = (Y_{i1}, \dots, Y_{iK})$ and $W_i = (W_{i1}, \dots, W_{iL})$. Parameters can be estimated by maximizing the full likelihood given by $L = \prod_i L_i(Y_i, W_i, G_i)$. However, to avoid such a computationally extensive task, we propose a two stage model: In stage 1, the longitudinal maternal anthropometric measurements are modeled as in (6), with estimations obtained by maximizing the likelihood of the joint model in the case of $K = 3$ or bivariate model estimations (Fieuw and Verbeke, 2006). In stage 2, the neonatal anthropometry model

$$W_{il} = V_{il}(G_i + t_{i*})' \gamma_l + \widehat{b_i}' \psi_{lk} + \varepsilon_{il}$$

is fit via linear regression, using $\widehat{b}_i$, the estimated random effects obtained from empirical Bayes estimation. Notice that this model can be simplified to include one maternal anthropometry measurement and several neonatal anthropometry measures by setting $K = 1$, or include multiple maternal measurements and one neonatal measure by setting $L = 1$. Estimates of μ_G and σ_G^2 can be obtained as described in Section 2.2.

Estimation of G_i is performed by evaluating $E[G_i | Y_i, W_i, \theta]$

$$E[G_i | Y_i, W_i, \theta] = \frac{\int_G \int_{b_i} G_i \left[\prod_j f(y_i, w_i | G_i, b_i) \right] g(G_i) h(b_i) dG_i db_i}{\int_G \int_{b_i} \left[\prod_j f(y_i, w_i | G_i, b_i) \right] g(G_i) h(b_i) dG_i db_i} \quad (10)$$

$$= \frac{\int_G G_i f(y_i, w_i | G_i) g(G_i) dG_i}{\int_G f(y_i, w_i | G_i) g(G_i) dG_i}.$$

where $f(y_i, w_i | G_i)$ is a multivariate normal distribution.

2.4. Prediction Interval for the Predicted Gestational Age

In this section, we provide a procedure to obtain a $(1 - \alpha)\%$ prediction interval for $G_i | Y_i, \theta$. The prediction interval is given by

$$E(G_i | Y_i, \hat{\theta}) \pm z_{\alpha/2} \sqrt{\widehat{Var}(G_i | Y_i, \hat{\theta})}$$

where

$$\begin{aligned} Var(G_i | Y_i, \hat{\theta}) &= E[Var(G_i | Y_i, \hat{\theta}) | \hat{\theta}] + Var[E(G_i | Y_i, \hat{\theta}) | \hat{\theta}] \\ &= Var(G_i | Y_i, \theta) + Var[E(G_i | Y_i, \hat{\theta}) | \hat{\theta}] \end{aligned}$$

We calculate $Var(G_i | Y_i, \theta)$ as follows:

$$Var(G_i | Y_i, \theta) = E(G_i^2 | Y_i, \theta) - [E(G_i | Y_i, \theta)]^2$$

where $E(G_i | Y_i, \theta)$ is calculated as in (5), and $E(G_i^2 | Y_i, \theta)$ is given by

$$\begin{aligned}
 E[G_i^2 | Y_i, \theta] &= \frac{\int_G \int_{b_i} G_i^2 \left[\prod_j f(y_{ij} | G_i, b_i) \right] g(G_i) h(b_i) dG_i db_i}{\int_G \int_{b_i} \left[\prod_j f(y_{ij} | G_i, b_i) \right] g(G_i) h(b_i) dG_i db_i} \\
 &= \frac{\int_G G_i^2 f(y_i | G_i) g(G_i) dG_i}{\int_G f(y_i | G_i) g(G_i) dG_i}
 \end{aligned}$$

We calculate $\text{Var}[E(G_i | Y_i, \hat{\theta}) | \hat{\theta}]$ using a bootstrap approach outlined in the following steps which should be repeated a large number of times (1000). Denote by N^* , the number of subjects present in the training set, which contains 60% of the subjects in the data.

Step 1: Sample N^* subjects with replacement from the training set.

Step 2: Using the created bootstrap training set, estimate the linear mixed effects model, and when applicable the shared random effects model.

Step 3: Using the parameters from Step 2, calculate the predictor, $E(G_i | Y_i, \hat{\theta})$, in the test set.

Once the process is repeated the prescribed number of times, obtain $\text{Var}[E(G_i | Y_i, \hat{\theta}) | \hat{\theta}]$ as the variance of the bootstrap estimates for each subject in the test set.

2.5. Simple Linear Predictor

In this section we present a three-step procedure that can be used to calculate a simple predictor for G_i using longitudinal assessments of fundal height.

Step 1: For each subject i in the training set, fit the following simple linear regression model:

$$Y_{ij} = \alpha_{0i} + \alpha_{1i} t_{ij} + \varepsilon_{ij}$$

Step 2: Since G_i is known in the training set, fit the following regression model, using all the estimated coefficients $(\hat{\alpha}_{0i}, \hat{\alpha}_{1i})$

$$G = A\phi + \varepsilon$$

where $G = \{G_1, \dots, G_{N^*}\}$, N^* represents the total number of subjects in the training set, $A = (\mathbf{a}_0, \mathbf{a}_1)$, $\mathbf{a}_0 = (\alpha_{01}, \dots, \alpha_{0N^*})^T$, $\mathbf{a}_1 = (\alpha_{11}, \dots, \alpha_{1N^*})^T$ and $\phi = (\phi_0, \phi_1, \phi_2)^T$.

Step 3: For each subject in the test set repeat Step 1, and extract the estimated coefficients $(\hat{\alpha}_{0i}, \hat{\alpha}_{1i})$, now labeled $(\hat{\alpha}_{0i}^*, \hat{\alpha}_{1i}^*)$. Once this process is completed, G_i is predicted as

$$\hat{G}_i = \hat{\phi}_0 + \hat{\phi}_1 \hat{\alpha}_{0i}^* + \hat{\phi}_2 \hat{\alpha}_{1i}^*$$

3. Prediction of Gestational Age in the NICHD Fetal Growth Studies

We develop a predictor of gestational using data from the NICHD Fetal Growth Studies, (Louis et al., 2015). This study was a prospective cohort study of 2334 non-obese pregnant women in the U.S. population.

We studied three maternal anthropometry measurements: fundal height (cm), weight change (lb) and subscapular skin fold cm). Figure 1 presents the profile plots of twenty randomly selected subjects for the chosen maternal anthropometric measurements in parts (a), (b) and (c) respectively. We also considered four neonatal birth anthropometry measures, abdominal and head circumference, length and birthweight. Figure 2 presents the neonatal data of the participants in parts (a), (b), (c) and (d) respectively.

Accuracy of prediction is assessed via root mean squared error. We estimated eleven models, of different combinations of three longitudinal maternal anthropometry measurements (FH, W and SF) and four neonatal anthropometry measures (AC, HC, L and BW). Fundal height measurements were only considered after 20 weeks of gestation. The parameter estimates of all eleven methods using a training set of 60% of randomly chosen subjects are presented on Tables 1 and 2. The estimated random effect variances are presented in the on-line supplement.

Figure 3 presents an example of the fit for the basic fundal height model for 25 randomly selected subjects. The raw fundal height measurements for each subject are presented in part (a) in the form of profile plots, with each line corresponding to a different subject. Part (b) presents the fitted curves for the selected participants. In addition, Figure 4 presents an example of the predicted vs observed gestational age for one of the evaluated test sets (a), along with the respective 95% prediction intervals for a random sample consisting of 5% of the subjects (b). Note that the size of these intervals varies between subjects, representing individual level variability. Also note, that empirically, in 95% of the subjects the prediction intervals cross the 45 degree line, which means they contain the true value of the parameter, advocating for the accuracy of the methodology. This calculation shows that empirically, this method provides the theoretical 95% claim.

Recall that the main goal of the methodology is to obtain accurate predictions of gestational age at first visit. In order to assess this accuracy, we calculated root mean squared error of prediction in the test set composed by the remaining 40% of the subjects. In this case, it is possible to calculate this measure, because in this study, there was a gold standard measurement of gestational age at first visit for all subjects. Table 3 presents prediction $\sqrt{\text{MSE}}$ values for a subset of the estimated methods. The presented results are based on the repeated formation of training and test sets (1000 repetitions), where the reported $\sqrt{\text{MSE}}$ values are calculated as an average across the replication.

From the results presented in Table 3 we can conclude that the proposed method of longitudinal maternal fundal height can produce results within $2.06(1.96\sqrt{\text{MSE}})$ weeks of the true value 95% of the time. These results can be improved by including other longitudinal measurements, for example, weight change (within 1.99 weeks with 95% probability) or combinations of maternal and neonatal anthropometry (within 1.91 weeks on average). The

decreases in root mean squared error obtained by including additional maternal and neonatal anthropometry, represent at most one day, which clinically is not relevant. Therefore, in clinical practice we recommend the use of the most parsimonious model, i.e. the fundal height alone model.

Gestational age is often approximated by cross-sectional fundal height by assigning the number of cm of the symphysis pubis fundal height to the gestational age at that visit (Corton et al., 2009). Andersson and Bergström (1995) conducted a study in rural Africa in which the first recorded fundal height was used as a proxy for gestational age and used it to predict delivery, and obtained a root mean squared error of prediction (at delivery) of 3.45 weeks. Following these guidelines, we applied this approach to the NICHD Fetal Growth Studies Data, where we repeated the training and test set procedure to have comparable results to those from the proposed methodologies. In this case, we estimated the gestational age at first visit for the subjects of the test set by using the cross-sectional value of fundal height obtaining a root mean squared error of prediction for gestational age of 1.57. Additionally, we used the last measurement of fundal height to predict gestational age, obtaining a root mean squared error of 2.08. Clearly, there is substantial efficiency gain in incorporating repeated measures of fundal height, where the root mean squared error of prediction is 1.05. These results show that the proposed methodologies are substantially more accurate than the cross-sectional method, used in current practice.

In Section 2.5 we provided details for a simple two-stage linear predictor for estimating gestational age. We conducted one thousand iterations of the training and test set scenario for the NICHD Fetal Growth Studies data, and obtained a root mean squared error of prediction of 1.11, which presents a 5.8% increase in root mean squared error relative to our proposed approach. However, when the number of follow up visits is small, the increase in mean squared error can be more substantial. For example when we only include a maximum of three measurements per participant, the root mean squared error of prediction for the proposed longitudinal fundal height approach is 1.27 as compared with the simple predictor where it is 1.80, a 41% increase in error.

4. Simulation

We conducted five simulation studies to evaluate our proposed methodologies. In simulation 1, we evaluated the properties of the proposed methodologies in terms of accuracy of parameter estimation and the prediction of gestational age at first visit. The predictor was developed under a model with parametric distribution assumptions and the predictive accuracy will be largest (lowest MSE) when these assumptions are satisfied. A set of robustness simulations are considered (2–5) to examine how much predictive accuracy is lost when the model is misspecified. In simulation 2 we considered violations in the normality assumption of G_i . In this case, G_i is no longer normally distributed, but instead, comes from a mixture of normal distributions. In simulation 3, we evaluated violations of the normality assumption of the random effects, in which we considered a mixture for the intercept. In Simulation 4, we considered cases in which the distribution of G_i is different in the training and test sets and finally, in simulation 5 we considered an informative relationship between the number of visits and the value of G_i .

4.1. Simulation 1: Accuracy of Prediction Evaluation

In this simulation, we explored accuracy of parameter estimation and prediction accuracy. We considered the same eleven combinations of the proposed methodologies as the ones estimated in the NICHD Fetal Growth data. In each case we generated 1000 datasets, of 2000 subjects each. For each dataset 60% of the subjects were randomly assigned to the training set, and 40% to the test set. In all cases G_i , gestational age corresponding to the first visit, was generated from a normal distribution with mean 24.28 weeks, and standard deviation of 2.69 weeks, corresponding to the truncated gestational age (at first visit) parameters in the NICHD Fetal Growth data.

Results show that parameter estimates and variance estimation for the parameter estimates are all nearly unbiased. In terms of prediction, the proposed method results in nearly unbiased estimation of the predictive accuracy. Further details on this simulation can be obtained in the on-line supplement.

4.2. Simulation 2: Misspecification of the distribution of G_i

We studied the performance of the model presented in (1) when G_i is no longer $N(\mu_G, \sigma_G^2)$ but instead, comes from a mixture of normal distributions. The motivation behind this simulation is to evaluate how the accuracy of the estimation is impacted by the violation of the normality assumption. This scenario is of importance, since an assumption of the proposed method is that μ_G in the model in (1) does not depend on covariates. This simulation allows us to determine how sensitive the proposed model is when μ_G depends on unobserved covariates, and hence G_i comes from a mixture of normals as in (11). In practical terms, we are assuming that there are no hidden mechanisms behind when a woman comes for a medical visit for the first time, this simulation aims to determine how well the method performs when a group of women comes in for their first visit, earlier than the other.

In this case we let

$$G_i \sim \pi N(\mu_{G_1}, \sigma_G^2) + (1 - \pi) N(\mu_{G_2}, \sigma_G^2). \quad (11)$$

We performed comparisons between the model in (1) and the model obtained by maximizing the likelihood $L = \log \prod_{i=1}^N L_i$ presented in (3) with the correct G_i distribution from equation (11) which we will refer to as the true model in this section. Note that we used the true values of the mixture parameters in the calculation of the root mean squared error for the correctly specified model.

We considered three simulation settings with 1000 data sets, with $N = 2000$ subjects, of which 60% were selected to be in the training set and 40% were chosen to be in the test set. The random variable G_i was generated according to equation (11), with different settings presented in the on-line supplement, corresponding on average to a mean value of gestational age equal to the one used in the simulations in Section 4.1. Figure 5 presents an

example of the generated G_i values for the three simulation settings. Further simulation setting details are presented in the on-line supplement.

We present simulation results of the parameter estimation for the three models for longitudinal maternal fundal height in the on-line supplement. Additionally, we present results of root mean squared error of prediction for the three models, under the true model (11) and under the proposed model (2), in Table 4. The information presented in Table 4, shows that the proposed methodology is robust to misspecification of the Gaussian assumption for G_i .

4.3. Simulation 3: Normality of b_i Violation

In this Simulation we studied the performance of the model presented in (1) when b_i is no longer distributed as $N(0, \Sigma_b)$, but a mixture of normal distributions instead.

$$b_i \sim \pi N\left(\begin{bmatrix} \mu_{b1} \\ 0 \end{bmatrix}; \Sigma_{b*}\right) + (1 - \pi)N\left(\begin{bmatrix} \mu_{b2} \\ 0 \end{bmatrix}; \Sigma_{b*}\right) \quad (12)$$

We computed the true prediction $\sqrt{MSE}$ by calculating $E(G_i^* Y_i, \theta)$ under the mixture distribution for the random effect. The estimated $\sqrt{MSE}$ was obtained by fitting model (1) and computing $E(G_i^* Y_i, \theta)$ under the Gaussian assumption. results show that the method provides unbiased estimations of the true $\sqrt{MSE}$ of prediction for moderate departure from normality. However, large departures from normality in the random effects distribution can result in more substantial bias in the $\sqrt{MSE}$. Further details on this simulation can be obtained in the on-line supplement.

4.4. Simulation 4: Different σ_G^2 in developing vs. applying the predictor

The predictor is often developed in a controlled fetal growth study with a certain distribution of G_i . The predictor is then applied, in a much less controlled setting where the variation of G_i is higher. In this simulation, we examine the situation where the distribution of G_i in the training and test set are both normally distributed with the same mean, but the variance being much larger in the test set. We considered three simulation scenarios, $\sigma_{G, \text{Test}}^2 = 2\sigma_{G, \text{Train}}^2$, $\sigma_{G, \text{Test}}^2 = 4\sigma_{G, \text{Train}}^2$ and $\sigma_{G, \text{Test}}^2 = 10\sigma_{G, \text{Train}}^2$, with all other parameters remaining equal to those of Simulation 1.

We performed comparisons between the predictions obtained using the correct variance (true model), and the misspecified variance (proposed model), to determine robustness of our method to this misspecification. To obtain these comparisons we considered 1000 data sets of $N = 2000$ subjects, divided in a training (60%) and a test (40%) set.

The results presented in Table 5 show that the method provides nearly unbiased estimation of the true root mean squared error of prediction for a two fold increase in the variance, with a slight inflation for larger departures from the true variance.

4.5. Simulation 5: Variations in informativeness of follow-up visits

The developed methodologies ignore relationships between the number of follow up visits and the value of G_i , when in practice it could happen that women who come earlier in gestation, have more visits than women who come later. This simulation considers the informative scenario in which the number of visits follows a Poisson distribution and is related to G_i through a shared random effect.

We generated 1000 data sets, with $N=2000$ subjects of which 60% were selected to be in the training set and 40% were assigned to the test set. The number of visits (C_i) was

generated from a Poisson ($\lambda_i/40$), where $\lambda_i = e^{b_i^*}/10$ and G_i is generated from a normal distribution with mean $\mu_G = \mu + \theta b_i^*$ and variance σ_G . Details of parameter selection can be found in the on-line supplement.

Estimations were performed using the univariate fundal height model and a root means squared error or prediction of 1.07 was obtained. Comparing this value to the one obtained by using correctly specified informative set up showed nearly identical results (root MSE of 1.08). Thus, ignoring a strong informative measurement number mechanism does not lose much predictive accuracy relative to correctly modeling it. Further details of the simulation can be found in the on-line supplement.

5. Discussion

In this article, we have proposed mixed and shared random parameter models to predict gestational age when we have longitudinal measurements of maternal anthropometry and cross-sectional data of neonatal anthropometry. The methodology was applied to data from the NICHD Fetal Growth Studies by Louis et al. (2015), a unique prospective cohort study with six targeted study visits that collected maternal anthropometry and detailed neonatal anthropometry conducted within the United States. The methodology in conjunction with this important study data allowed us to develop a predictor of gestational age that may be useful in under-resourced countries. The analysis suggested that longitudinal assessments of fundal height are most useful for the prediction, while other longitudinal measurements such as maternal weight gain and subscapular skin fold play little role in predicting gestational age. Further, we found that including neonatal cross-sectional measures improved prediction slightly. However, from a clinical perspective there is little payoff for including these cross-sectional measurements. A careful analysis showed that repeated fundal height measurements provides an accuracy of within 2 weeks in terms of gestational age prediction ($+/- 1.96\sqrt{MSE}$).

We studied the properties of the methodology with simulation studies that examined robustness to model misspecification. Linear mixed model estimation is robust to modeling assumptions such as normality of the error or random effects distribution. For example, it is well known that the estimates of the fixed effects are robust to the normality assumption of the random effects distribution in the standard linear mixed model. This fact, along with the fact that random effect estimates are not robust to the random effects distributions have been reported in Verbeke and Lessafre (1997). Similar to the situation where empirical Bayes

estimates of the random effects distribution are not robust to normality assumptions in the standard linear mixed model, we were concerned that estimates of G_i may lack robustness as well. Our simulations were designed to answer this question. Not surprisingly, the method performs well under the correct model misspecification. However, we do show that the estimation of prediction MSE is robust to moderate to severe model misspecification. This includes misspecification of the distribution of G_i and the distribution of b_i . We have taken a likelihood-based approach which is easily implemented. We note that a Bayesian alternative may provide some advantages, such as a flexible framework for updating the prediction with increasing information. However, we are satisfied with the ease of implementation and accuracy that the proposed methodologies provide.

A comparison of the proposed methodology to the current clinical practice of assessing fundal height at a single time point, and using its value as an estimate of gestational age, showed marked efficiency gains for our proposed approach. These results suggest that there is real potential for our approach in clinical practice. A potential barrier in the application of the proposed methodologies by health workers, is that it requires some programming for its implementation. We include an R program in the supplement that can be used to implement this methodology, where parameters estimates are obtained from the complete NICHD Fetal Growth Study data. Another potential barrier in the application of the proposed methodologies, is the need for repeated fundal height measurements. However, the World Health Organization guidelines for antenatal care recommend at least a four visit focused antenatal care model, in which fundal height can be obtained. We therefore anticipate that our methodology will become increasingly useful as these recommendations are followed.

As an alternative to the proposed methodology for predicting gestational age from longitudinal fundal height we propose a simple two-stage linear model for estimation. Although simpler, the two-stage approach is less efficient, particularly when the number of repeated fundal height measurements is small.

We present results considering fundal height after 20 weeks of gestation in both the training and the test set of the NICHD Fetal Growth example data, obtaining a prediction root mean squared error of 1.05 weeks. In practice, determining whether the 20 weeks condition is met is difficult and therefore we computed the unrestricted root mean squared error of prediction for the fundal height alone model, and obtained a value of 1.07 weeks. We also calculated the root mean squared of prediction using fundal height larger than or equal to 20 cm, which is what could be considered approximately 20 weeks of gestation in clinical practice, and obtained a value of 1.12 weeks. Additionally, we conducted the comparison between models containing different anthropometric variables obtaining similar results among models with and without the restriction. Clinically, there is no practical difference between these results as the largest difference is of one day.

Although the methodology only used maternal and neonatal anthropometry data for prediction, other biomarkers measured repeatedly in the mother and in the neonate can be used to increase predictive accuracy. A major initiative of the Bill and Melinda Gates Foundation is on using neonatal biomarkers to predict gestational age. This methodology can be used to predict gestational age using these neonatal biomarkers in conjunction with

longitudinal anthropometry data or biomarkers collected serially in mothers. We note that the interest of these methodologies is to predict gestational age at birth, and therefore the analysis is conditional on having a live birth.

The methodology assumes that we have gold standard assessments of gestational age based on ultrasound. However, particularly in low resource countries, such data may be hard to find. We are extending the methodology in order to develop predictors in situations for which we have measures of G_i that are subject to measurement error (for example, using last menstrual period, rather than early ultrasound).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Center for Information Technology, National Institutes of Health, for providing access to the Biowulf cluster computer system, and its high-performance computational capabilities. We also thank the NICHD fetal growth studies. This research was supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development intramural funding and included American Recovery and Reinvestment Act funding via contract numbers HHSN275200800013C, HHSN27500800002I, HHSN27500006, HHSN2750200800003IC, HHSN275200800014C, HHSN275200800012C, HHSN275200800028C, and HHSN275201000009C. This research was also funded by a Grand Challenge Exploration Grant (GCE) from the Bill and Melinda Gates Foundation.

References

- Albert, PS., Follman, D. Shared-parameter models. In: Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., editors. Longitudinal Data Analysis. Boca Raton, FL: CRC Press; 2008. p. 433–452.
- Andersson R, Bergström S. Use of fundal height as a proxy for length of gestation in rural Africa. The journal of tropical medicine and hygiene. 1995; 98(3):169–172. [PubMed: 7783274]
- Corton, MM., Leveno, K., Bloom, S., Hauth, J., Rouse, D., Spong, C. Williams Obstetrics. 23. New York, NY: McGraw-Hill Education;
- Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics. 2006; 62(2):424–431. [PubMed: 16918906]
- Hall MH, Carr-Hill RA, Fraser C, Campbell D, Samphier ML. The extent and antecedents of uncertain gestation. BJOG: An International Journal of Obstetrics and Gynaecology. 1985; 92(5):445–451.
- Johnson S, Shaw R, Parkinson P, Ellis J, Buchanan P, Zinaman M. Home pregnancy test compared to standard-of-care ultrasound dating in the assessment of pregnancy duration. Current Medical Research and Opinion. 2011; 27(2):393–401. [PubMed: 21175374]
- Laird N, Ware L. Random-Effects Models for Longitudinal Data. Biometrics. 1982; 38(4):963–974. [PubMed: 7168798]
- Louis GM, Grewal J, Albert PS, Sciscione A, Wing DA, Grobman WA, Newman RB, Wapner R, D'Alton ME, Skupski D, Nageotte MP, Ranzini AC, Owen J, Chien EK, Craigo S, Hediger ML, Kim S, Zhang C, Grantz KL. Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. American journal of obstetrics and gynecology. 2015; 213(4):449–e1. [PubMed: 26410205]
- Pinhoiro, J., Bates, D., Saikat, D., Sarkar, D. R Core Team. R package version 3.1–118. Adaptive Designs Johnson and Johnson Pharmaceutical Research and Development; Raritan, USA: 2014. nlme: linear and nonlinear mixed effects models. Available at: <http://CRAN.Rproject.org/package=nlme> [Accessed June 2016]
- Taipale P, Hiilesmaa V. Predicting delivery date by ultrasound and last menstrual period in early gestation. Obstetrics and Gynecology. 2001; 97(2):189, 194. [PubMed: 11165580]

- Verbeke G, Lesaffre E. The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data. *Computational Statistics and Data Analysis*. 1997; 23:541–556.
- White LJ, Lee SJ, Stepniewska K, Simpson JA, Dwell SLM, Arunjerdja R, Singhasivanon P, White NJ, Nosten F, McGready R. Estimation of gestational age from fundal height: a solution for resource-poor settings. *Journal of The Royal Society Interface*. 2012; 9(68):503–510.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Profile plots of twenty randomly selected subjects of three longitudinal maternal anthropometric measures, (a) fundal height in cm, (b) weight change in lb and (c) subscapular skin fold in cm. The solid line represents a linear regression of the maternal measure vs $G_i + t_{ij}$.

Fig. 2. Data plots of four neonatal anthropometry measures (a) abdominal circumference, (b) head circumference (c) birthweight and (d) length at birth. The solid line represents a linear regression of the neonatal measure vs $GA_i + t_i^*$.

Fig. 3. Example of the fundal height model for the NICHD Fetal Growth Studies data for 25 randomly selected subjects. The fundal height measurements for each subject are presented in (a), and fitted curves for the selected participants (b).

Fig. 4. Example of the estimation and 95% Prediction intervals for the fundal height model for the NICHD Fetal Growth Studies data. (a) Raw estimates (b) Estimates with prediction interval for a random sample of 5% of the subjects.

Fig. 5. Histogram of the generated gestational age at first visit for the robustness simulation study, (a) $\mu_{G_1} = 22.28$ and $\mu_{G_2} = 26.28$, (b) $\mu_{G_1} = 20.28$ and $\mu_{G_2} = 28.28$ and (c) $\mu_{G_1} = 18.28$ and $\mu_{G_2} = 30.28$.

Table 1

Parameter estimation results in the NICHD Fetal Growth Study example data, for combination of longitudinal maternal anthropometry measurements.

FH	FH+W	FH+SF	SF+W	FH+W+SF
$\beta_0 = 0.97$	$\beta_{01} = 0.96$	$\beta_{01} = 0.97$	$\beta_{01} = 18.77$	$\beta_{01} = 0.87$
	$\beta_{02} = -5.76$	$\beta_{02} = 18.65$	$\beta_{02} = -5.76$	$\beta_{02} = -5.81$
				$\beta_{03} = 18.93$
$\beta_1 = 0.96$	$\beta_{11} = 0.96$	$\beta_{11} = 0.96$	$\beta_{11} = 0.12$	$\beta_{11} = 0.96$
	$\beta_{12} = 0.49$	$\beta_{12} = 0.13$	$\beta_{02} = 0.49$	$\beta_{12} = 0.96$
				$\beta_{13} = 0.12$
$\sigma_{\varepsilon}^2 = 1.53$	$\sigma_{\varepsilon 1}^2 = 1.52$	$\sigma_{\varepsilon 1}^2 = 1.62$	$\sigma_{\varepsilon 1}^2 = 13.04$	$\sigma_{\varepsilon 1}^2 = 1.54$
	$\sigma_{\varepsilon 2}^2 = 1.51$	$\sigma_{\varepsilon 2}^2 = 4.57$	$\sigma_{\varepsilon 1}^2 = 4.39$	$\sigma_{\varepsilon 2}^2 = 1.51$
				$\sigma_{\varepsilon 3}^2 = 4.47$

Table 2

Parameter estimates in the NICHD Fetal Growth example data for combination of longitudinal maternal an neonatal anthropometry measures.

FH+AC	FH+HC	FH+L	FH+BW	FH+All	FH+W+HC
$\beta_0 = 0.97$	$\beta_0 = 0.97$	$\beta_0 = 0.97$	$\beta_0 = 0.97$	$\beta_0 = 0.97$	$\beta_{01} = 0.98$ $\beta_{02} = -5.78$
$\beta_1 = 0.96$	$\beta_1 = 0.96$	$\beta_1 = 0.96$	$\beta_1 = 0.96$	$\beta_1 = 0.96$	$\beta_{01} = 0.96$ $\beta_{12} = 0.50$
$\gamma_0 = 9.67$	$\gamma_0 = 15.99$	$\gamma_0 = 16.58$	$\gamma_0 = -853.43$	$\gamma_{10} = 9.67$ $\gamma_{20} = 15.99$ $\gamma_{30} = 16.58$ $\gamma_{40} = -3853.44$	$\gamma_0 = 16.62$
$\gamma_1 = 0.59$	$\gamma_1 = 0.46$	$\gamma_1 = 0.85$	$\gamma_1 = 182.82$	$\gamma_{11} = 0.59$ $\gamma_{21} = 0.46$ $\gamma_{31} = 0.85$ $\gamma_{41} = 182.82$	$\gamma_1 = 0.44$
$\psi_1 = 0.40$	$\psi_1 = 0.31$	$\psi_1 = 0.72$	$\psi_1 = 111.50$	$\psi_{111} = 0.40$ $\psi_{211} = 0.31$ $\psi_{311} = 0.72$ $\psi_{411} = 111.50$	$\psi_{111} = 0.32$ $\psi_{121} = -0.036$
$\psi_2 = 21.31$	$\psi_2 = 14.60$	$\psi_2 = 28.64$	$\psi_2 = 6496.58$	$\psi_{112} = 24.31$ $\psi_{212} = 14.60$ $\psi_{312} = 28.64$ $\psi_{412} = 6496.58$	$\psi_{112} = 13.05$ $\psi_{122} = -0.42$
$\sigma_\varepsilon^2 = 1.53$	$\sigma_\varepsilon^2 = 1.53$	$\sigma_\varepsilon^2 = 1.53$	$\sigma_\varepsilon^2 = 1.53$	$\sigma_\varepsilon^2 = 1.53$	$\sigma_{\varepsilon 1}^2 = 1.54$ $\sigma_{\varepsilon 2}^2 = 1.47$
$\sigma_\varepsilon^2 = 3.78$	$\sigma_\varepsilon^2 = 1.66$	$\sigma_\varepsilon^2 = 5.02$	$\sigma_\varepsilon^2 = 124797.4$	$\sigma_{\varepsilon 1}^2 = 3.78$	$\sigma_\varepsilon^2 = 1.67$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FH+AC	FH+HC	FH+L	FH+BW	FH+All	FH+W+HC
$\sigma^2_{\varepsilon 2} = 1.66$					
$\sigma^2_{\varepsilon 3} = 5.02$					
$\sigma^2_{\varepsilon 4} = 124797.4$					

Table 3

Root Mean Squared Error of prediction for gestational age in the test set of eleven models of maternal longitudinal measures and neonatal birth anthropometry for the NICHD Fetal Growth Studies data.

Model	$\sqrt{\text{MSE}}$
FH > 20	
FH	1.05
FH+W	1.01
FH+S	1.02
W+SF	3.52
FH+W+SF	1.03
FH+AC	1.04
FH+HC	1.04
FH+L	1.03
FH+BW	1.03
FH+AC+HC+L+BW	1.02
FH+W+HC	0.97

Table 4

Root Mean Squared Error of prediction for gestational age in the test set of three models of maternal fundal height (FH) where G_j is distributed according to three normal mixtures.

Model	True $\sqrt{\text{MSE}}$	Estimated $\sqrt{\text{MSE}}$
Simulation 1	1.05	1.04
Simulation 2	1.05	1.05
Simulation 3	1.05	1.05

Table 5

Root Mean Squared Error of prediction for gestational age in the test set of three models of maternal fundal height (FH) where the variance of G_i in the test set is larger than in the training set.

Model	True $\sqrt{\text{MSE}}$	Estimated $\sqrt{\text{MSE}}$
$\sigma_{G, \text{Test}}^2 = 2\sigma_{G, \text{Train}}^2$	1.05	1.07
$\sigma_{G, \text{Test}}^2 = 4\sigma_{G, \text{Train}}^2$	1.05	1.12
$\sigma_{G, \text{Test}}^2 = 10\sigma_{G, \text{Train}}^2$	1.05	1.23