Published in final edited form as:

Stat Med. 2018 December 30; 37(30): 4743–4757. doi:10.1002/sim.7955.

Estimating onset time from longitudinal data in the presence of measurement error with application to estimating gestational age from maternal anthropometry during pregnancy

Ana M. Ortega-Villa and Paul S. Albert

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

Abstract

Accurate assessment of gestational age at birth is necessary for optimal pediatric care. In high resource countries, several methods using ultrasound have been proposed to assess gestational age at birth; however, these methods are not easily accessible for low-resource populations. We develop a shared random parameter model for estimating gestational age at birth from longitudinal maternal anthropometry that incorporates additional maternal information from the last menstrual period, a measure of gestational age collected with sizable measurement error. The proposed methodology is evaluated using simulation studies under a training-test set paradigm. In addition, we propose methodology to validate prediction when some measurements of the gold standard are collected with measurement error. We illustrate the proposed methodologies with data from the NICHD Fetal Growth Studies.

Keywords

measurement error; onset time estimation; prediction; validation under measurement error

1 | INTRODUCTION

Accurate assessment of gestational age at birth and identification of preterm birth is of critical importance in obstetrics. In high resource countries, gestational age is assessed via early ultrasound (gold standard test) by examining 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 due to the small variability across subjects.¹ However, in resource-limited countries the use of ultrasound has a large number of barriers that include the lack of training, equipment and maintenance capabilities.²

Correspondence: Paul S. Albert, 9609 Medical Center Drive, Room 7E146. Bethesda, MD 20892. albertp@mail.nih.gov.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

FINANCIAL DISCLOSURE

None reported.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Therefore, in under-resourced settings, alternative approaches are needed. Recalled last menstrual period (LMP) is routinely collected and used as a measure of gestational age. To obtain LMP, 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, the LMP is known to have substantial measurement error due to problems with recollection, cycle irregularities and potential language barriers between patients and obstetricians.^{3,4}

To estimate gestational age, researchers have focused on developing biomarkers or using anthropometric measures obtained from the mother during pregnancy or newborn at birth. White et al⁵ proposed a method that uses serial fundal height measurements to predict gestational age based on a combination of linear models. Ortega-Villa et al⁶ proposed a framework using longitudinal maternal and cross-sectional neonatal anthropometry measures where fundal height is the primary measure to predict gestational age. Our methodology expands this framework to incorporate last menstrual period information in order to improve the accuracy of prediction as well as allow us to develop a predictor when we do not have a gold standard test on every subject.

In this paper, we present a framework to predict the gestational age at birth of a newborn, using longitudinal measures of maternal fundal height collected throughout obstetric care, and cross-sectional last menstrual period information assessed at the first visit. We propose methodology to be implemented in two stages. In the first stage, a joint random effects model for gestational age and last menstrual period is fitted, and in the second stage predictions for newborns without obstetric gestational age assessment is performed. Additionally, we propose methodology to validate predictions when the gold standard is measured with error. We assess the properties of our proposed methodologies with a simulation study and illustrate their application using data from the NICHD Fetal Growth Studies,⁷ a prospective cohort study of 2334 non-obese pregnant women from the U.S. population.

The remainder of this paper is as follows: Section 2 contains the model formulation. Section 3 presents asymptotic variance estimations of the methodology. Section 4 presents an analysis based on data from the NICHD Fetal Growth Studies. Section 5 presents a set of simulation studies and Section 6 contains concluding remarks.

2 | MODEL FORMULATION

The objective of the proposed methodology is to predict gestational age at birth for a newborn without any previous obstetric gestational age assessment. The proposed predictors are to be implemented in two stages. The goal of stage one is to estimate the parameters of the model. Therefore, during stage one, available data where some subjects have obstetric gestational age assessment via ultrasound is used to fit a longitudinal joint random effects model in which gestational age is used as a covariate and fundal height as the response. In stage two, the goal is to predict the unknown gestational age at birth of a new woman/newborn. In this stage, the roles of gestational age and fundal height are reversed, and the fundal height observations of said woman are used to predict her unknown gestational age at birth. To emulate this situation, we adopt a training-test set paradigm where the first stage is

performed in the training set, and the accuracy of the predictor is assessed in the second stage with test-set data. We present the longitudinal model in Section 2.1, and the proposed predictors in Section 2.3.

2.1 | Univariate longitudinal mixed effects model

Let Y_{ij} represent the j -th measurement of fundal height for the i -th subject. Fundal height is the distance in centimeters from the fundus to the top of the symphysis pubis. Let G_i represent the time from conception as assessed by ultrasound, until the first maternal visit for the i -th subject, also known as the gold standard assessment of gestational age. Ortega-Villa et al⁶ proposed methodology for when G_i was observed for all the subjects in the training set by using the Laird and Ware⁸ model

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

where $i = 1, \dots, N$ represents the subject, $j = 1, \dots, n_i$ represents the measurement, $X_{ij}(t_{ij}) = (1, t_{ij}, t_{ij}^2, \dots, t_{ij}^{p_1})'$, and $Z_{ij}(t_{ij}) = (1, t_{ij}, t_{ij}^2, \dots, t_{ij}^{p_2})'$ 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, $\boldsymbol{\beta}$ represents the overall effect of $X_{ij}(t_{ij})$ on Y_{ij} , $\mathbf{b}_i = (b_{i0}, b_{i1}, \dots, b_{ip_2})^T$ represents individual level variability and is assumed to have a multivariate normal distribution with mean $\mathbf{0}$ and variance Σ_b , and ε_{ij} is normal with mean 0 and variance σ_ε^2 . In this situation, G_i is observed and follows a normal distribution with mean and variance parameters μ_G and σ_G^2 respectively. Motivated by the data example, the polynomials are set to be of order $p_1 = p_2 = 1$ with the model being simplified to

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

with the assumption 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.

We extend this methodology to allow for ultrasound-determined G_i to be observed in some of the subjects of the training set, and for others to have a value of G_i measured with error, denoted by L_i , which represents the LMP of subject i . We specifically consider the case where if a woman has observed G_i , she also has observed L_i . The most general case we study, uses a training set in which a fraction of women have both G_i and L_i , a fraction have only L_i and a fraction has neither.

Parameter estimation under the proposed methodology can be obtained by maximizing the likelihood

$$\mathcal{L} = \prod_{i=1}^N [\mathcal{L}_i(\mathbf{Y}_i, G_i, L_i)]^{V_i} [\mathcal{L}_i(\mathbf{Y}_i, L_i)]^{1-V_i}, \quad (3)$$

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{id})$, V_i is a binary variable with the value of one when subject i has observed the ultrasound-consistent G_i and of zero when the subject only has L_i information, $\mathcal{L}_i(\mathbf{Y}_i, G_i, L_i)$ and $\mathcal{L}_i(\mathbf{Y}_i, L_i)$ are given by

$$\begin{aligned} \mathcal{L}_i(\mathbf{Y}_i, G_i, L_i) &= \left[\int_{\mathbf{b}_i} f(\mathbf{Y}_i | \mathbf{b}_i, G_i, L_i) f(\mathbf{b}_i) d\mathbf{b}_i \right] f(L_i | G_i) f(G_i) \quad (4) \\ \mathcal{L}_i(\mathbf{Y}_i, L_i) &= \int_{G_i} \mathcal{L}_i(\mathbf{Y}_i, G_i, L_i) dG_i \end{aligned}$$

where $f(\mathbf{Y}_i | \mathbf{b}_i, G_i, L_i)$ is a normal distribution with mean $\boldsymbol{\mu}_Y = X_{ij}(G_i + t_{ij})' \boldsymbol{\beta} + Z_{ij}(G_i + t_{ij})' \mathbf{b}_i$ and variance σ_e^2 , and $f(L_i | G_i)$ is a normal distribution with mean G_i and variance σ_M^2 . Note that $f(\mathbf{Y}_i | \mathbf{b}_i, G_i, L_i) = f(\mathbf{Y}_i | \mathbf{b}_i, G_i)$ since it is reasonable to assume that $\mathbf{Y}_i$ is independent of L_i when G_i is known, as all information of the relationship between fundal height and gestational age is contained in the measured without error gestational age at first visit G_i .

We estimate μ_G and σ_G^2 from the sample estimates of the observed G_i values in the training set, while all other parameters can be estimated by maximizing (3). Maximization can be done by using optimization functions in standard software such as `optim` in R. Initial values can be obtained by fitting the linear mixed model with verified data G_i using standard software such as the `lme` function in R.⁹

2.2 | Subjects without G_i or L_i

In this section, we consider the case in which a fraction of the subjects have information from an early ultrasound (G_i), a fraction recalls their LMP (L_i), and the remainder of the subjects has neither. In this situation, we consider the longitudinal linear mixed model from (1) and (2). Parameter estimations under this scenario can be obtained by maximizing the likelihood

$$\mathcal{L} = \prod_{i=1}^N [\mathcal{L}_i(\mathbf{Y}_i, G_i, L_i)]^{V_{i1}} [\mathcal{L}_i(\mathbf{Y}_i, L_i)]^{V_{i2}} [\mathcal{L}_i(\mathbf{Y}_i)]^{1-V_{i1}-V_{i2}}, \quad (5)$$

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{id})$, V_{i1} is a binary variable with the value of one when subject i has the ultrasound-consistent G_i and is zero otherwise, and V_{i2} is a binary variable with the value of one when subject i recalls her LMP L_i and is zero otherwise. $\mathcal{L}_i(\mathbf{Y}_i, G_i, L_i)$ and $\mathcal{L}_i(\mathbf{Y}_i, L_i)$ are given by (4) and $\mathcal{L}_i(\mathbf{Y}_i)$ is given by

$$\begin{aligned}\mathcal{L}_i(\mathbf{Y}_i) &= \int_{G_i} \mathcal{L}_i(\mathbf{Y}_i, G_i) dG_i \\ \mathcal{L}_i(\mathbf{Y}_i, G_i) &= \left[\int_{\mathbf{b}_i} f(\mathbf{Y}_i | \mathbf{b}_i, G_i) f(\mathbf{b}_i) d\mathbf{b}_i \right] f(G_i),\end{aligned}$$

where $f(\mathbf{Y}_i | \mathbf{b}_i, G_i)$ is a normal distribution with mean $\boldsymbol{\mu}_Y = X_{ij}(G_i + t_{ij})' \boldsymbol{\beta} + Z_{ij}(G_i + t_{ij})' \mathbf{b}_i$ and variance σ_ϵ^2 .

The maximization of the proposed likelihoods (3) and (5) was done using the optim function in R under the BFGS¹⁰ algorithm.

2.3 | Proposed predictors

We consider two predictors for gestational age. Predictor 1, corresponds to the conditional expectation of G_i given $\mathbf{Y}_i$ (fundal height measurements),⁶ and Predictor 2 is obtained from the conditional expectation of G_i given $\mathbf{Y}_i$ and L_i . Let $\theta = \{\boldsymbol{\beta}, \Sigma_b, \sigma_\epsilon^2, \mu_G, \sigma_G^2\}$ be the set of parameters obtained from the linear mixed model.

Predictor 1 $E[G_i | \mathbf{Y}_i, \theta]$

$$\begin{aligned}E[G_i | \mathbf{Y}_i, \theta] &= \frac{\int_G \int_{\mathbf{b}_i} G_i \left[\prod_j f(y_{ij} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i}{\int_G \int_{\mathbf{b}_i} \left[\prod_j f(y_{ij} | G_i, \mathbf{b}_i) \right] g(G_i) h(\mathbf{b}_i) dG_i d\mathbf{b}_i} \quad (6) \\ &= \frac{\int_G G_i f(\mathbf{y}_i | G_i) g(G_i) dG_i}{\int_G f(\mathbf{y}_i | G_i) g(G_i) dG_i}.\end{aligned}$$

Predictor 2 $E[G_i | \mathbf{Y}_i, L_i, \theta]$

$$\begin{aligned}E[G_i | \mathbf{Y}_i, L_i, \theta] &= \frac{\int_G \int_{\mathbf{b}_i} G_i \left[\prod_j f(y_{ij} | G_i, L_i, \mathbf{b}_i) \right] f(L_i | G_i) f(G_i) f(\mathbf{b}_i) dG_i d\mathbf{b}_i}{\left[\prod_j f(y_{ij} | G_i, L_i, \mathbf{b}_i) \right] f(L_i | G_i) f(G_i) f(\mathbf{b}_i) dG_i d\mathbf{b}_i} \quad (7) \\ &= \frac{\int_G G_i f(\mathbf{y}_i | G_i, L_i) f(L_i | G_i) f(G_i) dG_i}{\int_G f(\mathbf{y}_i | G_i, L_i) f(L_i | G_i) f(G_i) dG_i}.\end{aligned}$$

2.4 | Prediction interval for gestational age (G_i)

In this section, we provide the procedure to obtain a $(1-\alpha)\%$ prediction interval for $G_i | \mathbf{Y}_i, L_i, \hat{\theta}$. The procedure for $G_i | \mathbf{Y}_i, \theta$ is outlined in Ortega-Villa et al.⁶ The prediction interval for $G_i | \mathbf{Y}_i, L_i$ can be constructed as

$$E(G_i | \mathbf{Y}_i, L_i, \hat{\theta}) \pm z_{\alpha/2} \sqrt{\text{Var}(G_i | \mathbf{Y}_i, L_i, \hat{\theta})},$$

where

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

$\text{Var}[E(G_i | Y_i, L_i, \hat{\theta}) | \hat{\theta}]$ is calculated from the following bootstrap procedure:

Let N^* denote the number of subjects in the training set (60% of the total number). Repeat the following steps for a large number of bootstrap samples. We repeated the procedure 1000 times.

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

Step 2: Fit (2) with the bootstrap data.

Step 3: Extract the parameters from Step 3, and calculate $E(G_i | Y_i, L_i, \hat{\theta})$ for the subjects in the unique test set.

Once the process is repeated, $\text{Var}[E(G_i | Y_i, L_i, \hat{\theta}) | \hat{\theta}]$ is the sample variance of the bootstrap estimates for each subject in the test set. The first component of $\text{Var}(G_i | Y_i, L_i, \hat{\theta})$ can be computed as

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

where $E(G_i | Y_i, L_i, \theta)$ is obtained from (7) and

$$E(G_i^2 | Y_i, L_i, \theta) = \frac{\int_G G_i^2 f(y_i | G_i, L_i) f(L_i | G_i) f(G_i) dG_i}{\int_G f(y_i | G_i, L_i) f(L_i | G_i) f(G_i) dG_i}.$$

Integration in (3)-(7) is done by implementing the trapezoidal rule; however, other numerical integration approximations can be used.

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:

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

Step 2: Only considering women in the training set with observed G_i and L_i , fit the following regression model, using their corresponding estimated coefficients $(\hat{\alpha}_{0i}, \hat{\alpha}_{1i})$ from Step 1.

$$G = A\phi + \phi_3 L + \varepsilon,$$

where L is the observed LMP, $G = \{G_1, \dots, G_{N^*}\}$, N^* represents the total number of subjects in the training set with observed G_i and L_i , $A = (\mathbf{a}_0, \mathbf{a}_1)$, $\mathbf{a}_0 = (a_{01}, \dots, a_{0N^*})^T$, $\mathbf{a}_1 = (a_{11}, \dots, a_{1N^*})^T$, and $\phi = (\phi_0, \phi_1, \phi_3)^T$.

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

$$\hat{G}_i = \hat{\phi}_0 + \hat{\phi}_1 \hat{\alpha}_{0i}^* + \hat{\phi}_2 \hat{\alpha}_{1i}^* + \hat{\phi}_3 L_i.$$

Note: To best prepare for situations in which women in the test set do not have L_i , an alternative model without L should be fit in Step 2 and implemented in Step 3 for best prediction accuracy.

2.6 | Validation under measurement error

In this section, we propose methodology to validate prediction in a situation where some of the subjects have the gold standard measurement G_i , and the remainder have only the imperfect gold standard L_i .

When all the subjects have the G_i measurements, the mean squared error of prediction is obtained in the test sample by

$$MSE_1 = \frac{1}{N} \sum_{i=2}^N (\hat{G}_i - G_i)^2, \quad (8)$$

where G_i represents the ultrasound verified gestational age at first visit and $\hat{G}_i$ represents the predicted value. We propose that when none of the subjects have an observed value of G_i , and instead just have L_i (last menstrual period information), which is measured with error the mean squared error of prediction in the test sample can be calculated from

$$MSE_2 = \frac{1}{N} \sum_{i=2}^N (\hat{G}_i - L_i)^2 - \hat{\sigma}_M^2, \quad (9)$$

where $\hat{\sigma}_M^2$ is the sample estimate of σ_M^2 , and L_i is defined as $L_i = G_i + \epsilon_i$.

We propose MSE_2 given by (9) from the derivation below:

$$\begin{aligned}
MSE_2 &= E[(\hat{G}_i - L_i)^2] \\
&= E(\hat{G}_i^2 - 2\hat{G}_i L_i + L_i^2) \\
&= E(\hat{G}_i^2) - 2E[\hat{G}_i(G_i + \epsilon_i)] + E[(G_i + \epsilon_i)^2] \\
&= E[(\hat{G}_i - G_i)^2] + E(\epsilon_i^2) \\
&= MSE_1 + \sigma_M^2.
\end{aligned}$$

Since σ_M^2 is unknown, we propose using a sample estimate which can be obtained by maximizing either likelihood (3) or (5) depending on the situation, or by obtaining a sample estimate either in the training or the test set from the squared differences of the observed G_i and L_i .

When a proportion p of the subjects has an observed G_i and the remaining subjects have L_i then we propose using all available information by creating a pooled estimated of the root mean squared of prediction, given by

$$MSE_3 = pMSE_1 + (1 - p)MSE_2. \quad (10)$$

3 | ASYMPTOTIC VARIANCE EVALUATION

We calculated the asymptotic variance estimates for the longitudinal model parameters obtained by maximizing the likelihood presented in (3), which we will refer to as likelihood V+NV, and the ones obtained by maximizing a likelihood which only considers information from subjects with ultrasound verified G_i labeled as V.

Let Σ_{V+NV} represent the asymptotic variance-covariance matrix for the likelihood V+NV and Σ_V represent the variance-covariance matrix from the verified subjects only likelihood. Then,

$$\begin{aligned}
\Sigma_{V+NV} &= - \left[pE \left(\frac{\partial^2 L_v}{\partial \theta^2} \right) + (1 - p)E \left(\frac{\partial^2 L_{nv}}{\partial \theta^2} \right) \right]^{-1} \quad (11) \\
\Sigma_V &= - \left[pE \left(\frac{\partial^2 L_v}{\partial \theta^2} \right) \right]^{-1},
\end{aligned}$$

where p is the proportion of verified subjects, $L_v = \mathcal{L}_i(Y_i, G_i, L_i)$ and $L_{nv} = \mathcal{L}_i(Y_i, L_i)$ as in Section 2.1. The expected values in (11) are difficult to compute in closed form; therefore, we estimated $E(X)$ via Monte Carlo sampling where $E(X) \approx \frac{1}{n} \sum X$, where X is generated by the distribution of X .

Data was generated according to (2), with parameters $\beta_0 = 0.9706$, $\beta_1 = 0.9675$, $\mu_G = 24.2813$, $\sigma_G^2 = 7.2555$, $\sigma_e^2 = 1.5355$. L_i was generated from a normal distribution with mean G_i and variance $\sigma_M^2 = 2$. The random effects variance is given by

$$\Sigma_b = \begin{bmatrix} 3.0061 & -0.1023 \\ -0.1023 & 0.0043 \end{bmatrix}.$$

Results of the asymptotic variance estimates for one million Monte Carlo samples are presented in Table 1, where parameters obtained by maximizing likelihood (3) are labeled as V+NV, and parameters obtained by maximizing a likelihood that only considers verified subject information data are labeled as V. Note that results presented in Table 1, do not account for small-sample variability in cases where p is small, and hence, present optimistic estimates of the variability under the specified conditions. As it is expected, the variance of parameters both in V+NV and V increase as the proportion of verified subjects (p) decreases. Interestingly, the rate of increase is much more rapid for V than V+NV. Further, although there is some payoff for incorporating the non-verified cases into the analysis when p is large, there is a great advantage in doing so when p is small. Thus, incorporating subjects with only LMP gestational dating without ultrasound can result in a more accurate predictor of gestational age as compared to discarding them.

4 | PREDICTION OF GESTATIONAL AGE IN THE NICHD FETAL GROWTH STUDIES

We illustrate the application of the proposed predictors using data from the NICHD Fetal Growth Studies,⁷ a prospective US cohort of 2334 non-obese, low-risk, healthy pregnant women. Gestational age was based on self-reported last menstrual period consistent with an early gestation ultrasound measure. We consider longitudinal measurements of fundal height (cm) as our maternal anthropometry measure due to its performance as a predictor in previous studies.^{5,6} Since the NICHD Fetal Growth Study was designed to access fetal growth as measured by ultrasound, an eligibility requirement for the study was that G_i and L_i be within a narrow range to assure an accurate measurement of gestational age on each study participant. To illustrate the performance of the proposed methodology with this dataset, we simulate LMP with measurement error from a normal distribution with mean G_i (the individual gestational age at first visit) and measurement error variance $\sigma_M^2 = 2$.

Figure 1 presents profile plots of fundal height for fifty randomly selected subjects (A), along with a scatter plot of the generated LMP values vs. the observed gestational age (B). Panel (A) depicts the linear relationship between fundal height and gestational age, which supports the choice of linear mixed model (2). The departure from the 45 degree line in panel (B) reflects the simulated measurement error in the LMP.

To illustrate the performance of the proposed methodology, we considered two estimation cases. In Case 1, we show examples of the performance of the model presented in Section 2.1, when the training set is composed of a proportion p of subjects who have ultrasound

assessed G_i and observed last menstrual period L_i , and a proportion $(1 - p)$ without G_i but with observed L_i . In Case 2, we consider a training set composed of a proportion of subjects (p_1) with observed G_i , a proportion (p_2) with only observed L_i and a proportion $(1 - p_1 - p_2)$ with no gestational age assessment (ie, no G_i or L_i). We created 1000 distinct training-test set assignments. The results presented in this section correspond to the average obtained across all training-test set assignments. Validation is performed using ultrasound-confirmed G_i for all the subjects in the test set.

4.1 | Case 1: Subjects have either ultrasound-confirmed G_i or LMP L_i

In this section, we illustrate the application of the proposed methodology through the NICHD Fetal growth study. We consider four training set compositions in which the proportion of subjects with observed G_i (p) varies, in order to assess the performance of the proposed predictors. We consider three specifications for p , $p = 0.95$, $p = 0.5$ and $p = 0.05$. Specifically, when $p = 0.95$, 95% of the women in the training set have both ultrasound assessed G_i and LMP information L_i , and 5% only have L_i . Similarly, when $p = 0.5$, 50% of the women in the training set have G_i and L_i , and 50% only have LMP information and finally, when $p = 0.05$, 5% of women in the training set have observed G_i and L_i and 95% only have L_i .

Table 2 contains the root mean squared error of prediction for the estimations using the NICHD Fetal Growth Study Data. We present results for parameters obtained using only gold standard ultrasound dated G_i measurements in the training set as in Ortega-Villa et al,⁶ and using parameters obtained from likelihood (3), and Predictors 1 (6) and 2 (7). We found that estimates of σ_M^2 are sensitive to starting values, and therefore for starting values of σ_M^2 we used a sample estimate obtained from the squared difference between G_i and L_i in the training set. For all choices of p , Predictor 2 outperformed Predictor 1 by yielding a lower root mean squared error of prediction. As an example, consider the case where we estimate σ_M^2 , for $p = 0.05$ the root MSE for Predictor 2 is 0.968, while the value for Predictor 1 is 1.131.

Figure 2 presents an example of the estimation and calculation of 95% prediction intervals for the NICHD Fetal Growth Studies Data when $p = 0.95$ for one randomly selected training-test set assignment. Panels (A) and (B) present estimates versus observed values on test-set data for predictors 1 and 2 respectively. Points in panel (B) are closer to the 45 degree line, than those in panel (A), showing that Predictor 2 achieves more accurate estimations of gestational age at first visit. Panel (C) presents estimates with 95% prediction intervals for a random sample of 25 subjects. The size of these intervals varies between subjects showing variability at the individual level. Estimates from Predictor 2 in panel (C) appear to be closer to the 45 degree line. The 95% prediction interval included the observed gestational age 96% of the time for both predictors, demonstrating nominal coverage rates for both intervals. The results presented in Table 2 and Figure 2 show that when all the values of G_i cannot be observed in the training set, incorporating LMP (L_i) information into the predictor improves the accuracy of prediction.

4.2 | Case 2: Subjects have either ultrasound-confirmed G_i or LMP L_i or neither

Motivated by the application of this methodology in a setting where access to ultrasound is extremely difficult, we evaluate the case in which the training set is composed of 50% of women who do not have ultrasounds assessed G_i , but have L_i information (ie, $p_1 = 0$, $p_2 = 0.5$), and 50% do not have any gestational age assessment whatsoever prior to delivery. The root MSE of prediction for Predictor 1 in this situation is 1.151 with a standard deviation of 0.12, and for Predictor 2 1.146 with a standard deviation of 0.12. Due to the difference in root MSE, Predictor 2 outperforms Predictor 1. However, since this methodology is sensitive to starting values, we suggest that at the very least, initial values be explored through a thorough grid search. We believe that designs with no gestational age assessment prior to delivery (ie, no G_i or L_i) are very sensitive to modeling assumptions, they should not be used.

5 | SIMULATION

We evaluated the performance of the proposed predictors and the validation under measurement error approach through four simulation studies. We designed Simulation 1 to evaluate accuracy of prediction of the two proposed predictors by varying the composition of the training set. Simulation 2 was created to assess the robustness of the proposed methodology to misspecification in the distribution of LMP (L_i). In Simulation 3, we evaluate the performance of a simple linear predictor, and in Simulation 4, we evaluated the performance of the proposed mean squared error of prediction when the gold standard is measured with error on a fraction of the women in the test set.

5.1 | Simulation 1: Accuracy of prediction evaluation

In this simulation, we explored accuracy of parameter estimation and prediction of gestational age at first visit (G_i) under the two proposed estimation approaches and the two proposed predictors. We considered eight different training set compositions to evaluate our predictions. We consider four assignments to evaluate predictions associated with maximizing likelihood (3), where the training set contains women who have a gestational age assessment either through early ultrasound (G_i) or LMP (L_i). In addition, we consider four assignments to evaluate predictions linked to maximizing likelihood (5), where the training set contains women who have G_i or L_i or neither (ie, no gestational age assessment).

Let p_1 represent the proportion of women with observed G_i , p_2 the proportion of women with observed L_i but not G_i , and finally $p_3 = 1 - p_1 - p_2$ represent the proportion of women with no gestational age assessment. We refer to these training test assignments as scenarios 1–8, where coefficient estimations for scenarios 1–4 are obtained through maximizing likelihood (3) and for scenarios 5–8 through maximizing likelihood (5). In each scenario we generated 1000 datasets of 2000 subjects each, where in each dataset 60% of the subjects were randomly selected and assigned to the training set and the remaining 40% are assigned to the test set. Data was generated according to (2), with parameters $\beta_0 = 0.9706$, $\beta_1 = 0.9675$, $\mu_G = 24.2813$, $\sigma_G^2 = 7.2555$, $\sigma_e^2 = 1.5355$, and the random effects variance is given by

$\Sigma_b = \begin{bmatrix} 3.0061 & -0.1023 \\ -0.1023 & 0.0043 \end{bmatrix}$. Additionally, L_i was generated from a normal distribution with mean G_i and variance $\sigma_M^2 = 2$.

Details of each of the training test assignment scenarios presented in Table 3. Table 4 presents the results of the simulation study, by providing the average root MSE of prediction and its standard deviation for each of the proposed predictors along with the average difference between them across 1000 simulated datasets. We found that in all simulation scenarios Predictor 2 has a smaller root MSE of prediction, and positive values when looking at the difference, providing evidence that incorporating LMP into the prediction improves accuracy. In addition, we found that predictions for Scenarios 4 and 8, in which no available information of G_i was considered were sensitive to starting values. Prediction under this situation has to be done with care, and a thorough exploration of starting values should be done, where at the very least an extensive grid search should be conducted.

5.2 | Simulation 2: Misspecification of the distribution of L_i

In this simulation, we studied the performance of prediction when the distribution of L_i is misspecified. Specifically, we consider the case in which the distribution of L_i is no longer normal with mean G_i and variance σ_M^2 but instead comes from a mixture of normal distributions (12). The motivation behind this simulation is to evaluate the impact of the violation of this normality assumption on the accuracy of prediction. This simulation allows us to determine how sensitive the proposed model and predictors are, to the case in which some women tend to underestimate their LMP, some women tend to recall their actual LMP and some tend to overestimate their LMP.

$$L_i = p_1 N(-4, \sigma_m^2) + p_2 N(0, \sigma_m^2) + p_3 N(4, \sigma_m^2) + G_i, \quad (12)$$

where G_i is normally distributed with mean $\mu_G = 24.28$ and variance $\sigma_G^2 = 7.25$, and σ_m^2 was calculated to be consistent with σ_M^2 . Details of the derivation of σ_m^2 can be found in the supplementary material.

Similarly to Simulation 2, we generated 1000 datasets of 2000 subjects each, where in each dataset 60% of the subjects were randomly selected and assigned to the training set and the remaining 40% are assigned to the test set. We considered three scenarios for p , the proportion of subjects in the data that have verified G_i , $p = 0.95$, $p = 0.5$, and $p = 0.05$. Data was generated according to (2), with parameters $\beta_0 = 0.9706$, $\beta_1 = 0.9675$, $\mu_G = 24.2813$, $\sigma_G^2 = 7.2555$, $\sigma_e^2 = 1.5355$. The random effects variance is given by

$$\Sigma_b = \begin{bmatrix} 3.0061 & -0.1023 \\ -0.1023 & 0.0043 \end{bmatrix}.$$

Table 5 presents the root mean squared error of prediction for Simulation 4. We performed comparisons between the predictions obtained from maximizing likelihood (3) using the two

predictors presented in Section 2.3. As in the previous simulations, we assumed G_i was observed for all women in the test set for validation purposes. We found that in situations where the percentage of verified subjects is low, results from Predictor 2 are very close to those of Predictor 1. This means that, in order for the incorporation of L_i to yield valuable accuracy improvements, the distribution for LMP needs to be correctly specified.

5.3 | Simulation 3: Simple linear predictor

In this simulation study, we evaluated the accuracy of prediction of the proposed simple predictor in comparison with that of predictors 1 and 2. Section 5.3.1 illustrates a situation where the proportion (p) of subjects with observed G_i in the training set varies. In section 5.3.2, we present an evaluation of prediction accuracy considering the case where women have either $G_i(p_1)$, $L_i(p_1)$ or neither (p_3).

We generated 1000 datasets of 2000 subjects each, where in each dataset 60% of the subjects were randomly selected and assigned to the training set and the remaining 40% are assigned to the test set. Data was generated according to (2), with parameters $\beta_0 = 0.9706$, $\beta_1 =$

0.9675 , $\mu_G = 24.2813$, $\sigma_G^2 = 7.2555$, $\sigma_\varepsilon^2 = 1.5355$, and the random effects variance is given by

$\Sigma_b = \begin{bmatrix} 3.0061 & -0.1023 \\ -0.1023 & 0.0043 \end{bmatrix}$. Additionally, L_i was generated from a normal distribution with mean G_i and variance σ_M^2 .

5.3.1 | Observed G_i on a proportion of women, L_i observed on all subjects—

In this section, we present a comparison of prediction accuracy obtained by implementing predictors 1 and 2 from Section 2.3 and the simple linear predictor (SLP) from Section 2.5. We considered the specific case in which the proportion of women with observed G_i and L_i (p) in the training set is either $p = 0.95$, 0.5 , or 0.05 , for for three values of the measurement error variance associated with LMP ($\sigma_M^2 = 2, 4, 6$).

Table 6 presents the average root MSE of prediction for the considered p and σ_M^2 assignments. We found that in situations where there is at least a proportion of the subjects with observed G_i , even in situations where the measurement error variance is large (ie, $\sigma_M^2 = 6$), the simple linear predictor outperforms predictors 1 and 2.

5.3.2 | G_i and L_i observed on a proportion of women—In this section, we consider three situations in which a proportion p_1 of women have observed G_i and L_i , a proportion p_2 do not have G_i but have L_i and a proportion p_3 has neither. In this situation, we can only consider women in the training set with observed G_i to perform step 2 and information from a proportion $(1-p_1)$ of women will be lost. Here, L_i was generated from a normal distribution with mean G_i and variance $\sigma_M^2 = 2$.

Specifically we consider three scenarios:

Scenario 1: $p_1 = 0.90$; $p_2 = 0.05$, $p_3 = 0.05$

Scenario 2: $p_1 = 0.1$; $p_2 = 0.40$; $p_3 = 0.5$

Scenario 3: $p_1 = 0.05$; $p_2 = 0.05$; $p_3 = 0.90$.

Table 7 presents the root MSE of prediction for the simulation study. As expected, as the proportion of women with observed $G_i(p_1)$ decreases, the accuracy of prediction of the three predictors also decreases. In this situation, Predictor 2 outperforms both Predictor 1 and SLP, which means that L_i is the driving force behind the accuracy in prediction of SLP, and when it's absent the simple model does not perform as well as Predictor 2.

5.4 | Simulation 4: Mean squared error of prediction evaluation

In this simulation we assessed the performance of our proposed estimator for the root mean squared error of prediction, when some subjects have a gold standard measured with error in both the training and the test set. All results presented for this simulation consider estimations obtained by implementing Predictor 1, since the objective is to assess the performance of the root mean squared error estimation and not the predictor. For this simulation we consider L_i the measure of LMP to be the error infused gestational age variable, and G_i as the true gold standard for the accuracy assessment.

We generated 1000 datasets of 1000 subjects each, where in each dataset 60% of the subjects were randomly selected and assigned to the training set and the remaining 40% are assigned to the test set. We considered four scenarios for p , the proportion of subjects in the entire data that have verified G_i (training and test-set), $p = 0.95, 0.5, 0.10$, and 0.05 , respectively. Data was generated according to (2), with parameters $\beta_0 = 0.9706$, $\beta_1 = 0.9675$, $\mu_G = 24.2813$, $\sigma_G^2 = 7.2555$, $\sigma_\varepsilon^2 = 1.5355$. The random effects variance is given by

$$\Sigma_b = \begin{bmatrix} 3.0061 & -0.1023 \\ -0.1023 & 0.0043 \end{bmatrix}.$$

L_i was generated from a normal distribution with mean G_i and variance $\sigma_M^2 = 2$. Predictions of G_i were calculated by following the procedure from Section 2.1 maximizing the likelihood (3) and using Predictor 1 (6). True estimates of the root MSE were obtained by using the predicted values of G_i and the generated values of G_i , as in (8). Additionally we calculated four estimates of MSE_3 (10), using the true value of σ_M^2 , using the sample estimate of σ_M^2 from the verified training set data, using σ_M^2 , and using the sample estimate of σ_M^2 from the verified data in the test set.

Table 8 shows the average MSE for the different methods when only a proportion p of women have the gold standard in both the training and the test set when $\sigma_M^2 = 2$ as specified in the previously mentioned simulations. The proposed approaches resulted in biased assessment of MSE when p is small, due primarily to biased estimates of σ_M^2 . This bias is reduced for smaller amounts of measurement error (results for $\sigma_M^2 = 1$ and 0.1 in Table 8) or when the simulation is done with a sample size of 5000 rather than 1000 (data not shown). $\sigma_M^2 = 2$, only using verified data is more efficient than the proposed approach even for

relatively small p . However, there is a clear efficiency gain for our proposed approach when σ_M^2 is reduced and p is small. In summary, using the proposed estimator of MSE that uses L_i rather than G_i only pays off when the measurement error variance is small and the proportion of verified subjects is small, otherwise, calculating the root mean squared error of prediction with only the verified subjects results in minimal loss of efficiency.

6 | DISCUSSION

In this article, we have proposed an alternative to predict gestational age when we have longitudinal measurements of maternal fundal height during pregnancy and a cross-sectional measurement of recalled last menstrual period, a measure of gestational age subject to measurement error. This methodology is geared towards the prediction of gestational age in under developed countries, in which ultrasound based measures in the test-set are not easily accessible. In addition, we propose an estimator of the mean squared error of prediction, that can be applied in the test set when the gold standard is measured with error.

We assessed the asymptotic properties of the methodologies and concluded that in the presence of measurement error (ie, LMP measurements), the inclusion of said measurements by maximizing (3) reduces the variance of parameter estimates relative to discarding this information. In addition, we studied the properties of the methodology in four simulation studies. We found that the methodology is sensitive to starting values when the proportion of verified subjects is zero, and that even a small proportion of verified subjects (ie, 5%) greatly improves this issue. For the specific case that the training set is composed of subjects who only have LMP measurements, we suggest an extensive grid search in order to obtain the starting values for the optimization. For prediction, we found that using L_i when it is available (Predictor 2) results in improved performance as compared to not using it (Predictor 1). However, this improvement can be substantially diminished when the distribution of L_i is misspecified.

We examined the performance of incorporating individuals with this error-prone gestational age measurement (L_i) in the test-set for evaluating predictive performance when the gold standard measurement (G_i) is not observed. Our results show that major gains can be achieved by incorporating those subjects with only L_i into the MSE estimation in the test-set when both the measurement error and the fraction of test-set women with G_i observed is small. Of course, as with the prediction in the training set, it is important to correctly specify the measurement error distribution of L_i .

We note that the illustration of the proposed prediction methodology was made with data from the NICHD Fetal Growth Study of Singleton Gestations using a training-test set paradigm. However, in practice, the prediction will be made for out of sample neonates. To examine how the prediction performed when the out of sample prediction was different than the study population, Ortega-Villa et al⁶ performed a simulation study in which σ_G^2 was larger in the test set than in the training set, for Predictor 1 in a setting where all subjects had ultrasound verified G_i . Ortega-Villa et al⁶ found that for twofold increases in the variance the methodology provides nearly unbiased estimations of the true root-mean-squared error of

prediction. These results provide some evidence to show that these predictors can be trained and applied in separate cohorts, with acceptable prediction results.

The proposed predictors rely on the assumption that $G_i \sim N(\mu_G, \sigma_G^2)$. When G_i is observed in every woman in the training set, this assumption can be relaxed, and the observations can be used to calculate the empirical distribution function for G_i . We considered this scenario and found that the accuracy gain from using the empirical distribution function in prediction is minimal. Details about the estimation procedure can be found in the supplementary material. However, when G_i is not observed on everyone in the training set relaxing the distribution assumption of G_i is more difficult. One approach would be to extend the proposed approach by replacing the Gaussian assumption by a point-wise mixture distribution, an approach often used for the non-parametric estimation of a random effect distribution. Since G_i is approximately normal in the NICHD Fetal Growth Study, we leave this to future research.

In addition to the proposed predictors, we consider a three-step simple linear predictor (SLP), which uses the observed G_i and L_i in the training set to fit linear regression models that can later be used to predict G_i in women who require gestational age assessment. SLP can only be applied in situations where at least a small proportion of women in the training set have observed both G_i and L_i . When the training set is exclusively composed of women who have had a gestational age assessment (ie, G_i on some and L_i observed on everyone), SLP outperforms Predictor 1 and 2. However, when the training set also includes women without any gestational age assessment (ie, G_i or L_i missing on some women), Predictor 2 outperforms Predictor 1 and SLP. These results suggest that the selection of which predictor to use will be dictated by the available information in the training set.

We include the R function `Pred.G.LMP` in the Supplementary Materials to obtain estimates of gestational age at birth for individual women who have at least two fundal height measurements and a single observation of recalled last menstrual period (LMP) collected during the first visit (Predictor 2). Example code for women with no LMP information (Predictor 1) can be found in the supplementary material of Ortega-Villa et al.⁶

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the reviewers and the associate editor for their valuable input. We also 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 thank the NICHD fetal growth studies as well. This research was supported by the National Cancer Institute, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development intramural funding and included American Recovery and Reinvested Act funding via contract numbers HHSN275200800013C, HHSN27500800002I, HHSN275000006, 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.

Funding information

National Cancer Institute; Eunice Kennedy Shriver National Institute of Child Health and Human Development; American Recovery and Reinvested Act, Grant/Award Number: HHSN275200800013C, HHSN275008000021, HHSN27500006, HHSN27502008000031C, HHSN275200800014C, HHSN275200800012C, HHSN275200800028C, and HHSN275201000009C; Bill and Melinda Gates Foundation, Grant/Award Number: Grand Challenge Exploration Grant

REFERENCES

1. Taipale P, Hiilesmaa V. Predicting delivery date by ultrasound and last menstrual period in early gestation. *Obstet Gynecol.* 2001;97(2):189–194. [PubMed: 11165580]
2. Shah S, Bellows BA, Adedipe AA, Totten JE, Backlund BH, Sajed D. Perceived barriers in the use of ultrasound in developing countries. *Crit Ultrasound J.* 2015;7(1):1–5. [PubMed: 25852842]
3. Alexander GR, Tompkins ME, Petersen DJ, Hulsey TC, Mor J. Discordance between LMP-based and clinically estimated gestational age: implications for research, programs, and policy. *Public Health Rep.* 1995;110(4):395–402. [PubMed: 7638326]
4. Mongelli M, Wilcox M, Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. *Am J Obstet Gynecol.* 1996;174(1):278–281. [PubMed: 8572021]
5. White LJ, Lee SJ, Stepniewska K, et al. Estimation of gestational age from fundal height: a solution for resource-poor settings. *J Royal Soc Interface.* 2012;9(68):503–510.
6. Ortega-Villa AM, Grantz KL, Albert PS. 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. *J Royal Stat Soc Ser A.* 2018;181(3):825–842.
7. Louis GMB, Grewal J, Albert PS, et al. Racial/ethnic standards for fetal growth: the NICHD Fetal Growth Studies. *Am J Obstet Gynecol.* 2015;213(4):449–e1. [PubMed: 26410205]
8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics.* 1982;38(4):963–974. [PubMed: 7168798]
9. Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: linear and nonlinear mixed effects models. R package version 3.1–118. 2014 <http://CRAN.R-project.org/package=nlme>. Accessed June 2016.
10. Fletcher R, Reeves CM. Function minimization by conjugate gradients. *Comput J.* 1964;7(2):149–154.

FIGURE 1.

A, Profile plots of fundal height for fifty randomly selected subjects; B, Example of generated last menstrual period vs observed gestational age at first visit for data from the NICHD Fetal Growth Studies

FIGURE 2.

Example of the estimation and 95% Prediction intervals for the NICHD Fetal Growth Studies Data under $p = 95\%$ verified subjects scenario. Estimates versus truly observed gestational age for (A) predictor 1 and (B) predictor 2. (C) Estimates with prediction interval for a random sample of 25 subjects

TABLE 1

Asymptotic variance calculations for longitudinal model parameters under likelihood (3) (V+NV) and the verified only (V) likelihood, when the proportion of verified subjects is p

		$\sigma_{b_0}^2$	$\sigma_{b_1}^2$	σ_{b_0, b_1}	β_0	β_1	σ_ε^2
$p = 0.95$	V+NV	6.19	0.02	0.92	0.30	10.42	0.01
	V	6.46	0.02	0.93	0.32	10.93	0.01
$p = 0.50$	V+NV	6.75	0.02	1.52	0.30	10.91	0.01
	V	12.28	0.04	1.77	0.61	20.76	0.02
$p = 0.05$	V+NV	8.03	0.02	4.46	0.31	11.95	0.01
	V	122.83	0.39	17.69	6.06	207.62	0.23

TABLE 2

Root mean squared error of prediction for the NICHD Fetal Growth Study data

p	Only Gold Standard G_i	Max. Likelihood With Known σ_M^2		Max. Likelihood With Estimated σ_M^2	
	Pred. 1	Pred. 1	Pred. 2	Pred. 1	Pred. 2
0.95	1.066 (0.06)*	1.067 (0.06)	0.935 (0.03)	1.072 (0.07)	0.937 (0.04)
0.50	1.070 (0.07)	1.077 (0.07)	0.937 (0.03)	1.078 (0.07)	0.937 (0.03)
0.05	1.094 (0.10)	1.131 (0.11)	0.968 (0.07)	1.131 (0.11)	0.968 (0.07)

* Average $\sqrt{MSE}$ (std dev.)

TABLE 3

Specifications for simulation scenarios in Simulation 1

Scenario	p_1	p_2	p_3
1	0.95	0.05	0
2	0.50	0.50	0
3	0.05	0.95	0
4	0	1.0	0
5	0.90	0.05	0.05
6	0.10	0.40	0.50
7	0.05	0.05	0.90
8	0	0.50	0.50

TABLE 4

Prediction results for Simulation 1, considering training where a proportion of subjects have G_i , a proportion L_i only and a proportion has neither

Scenario	Mean $\sqrt{MSE}$	Mean $\sqrt{MSE}$	Mean difference
	Pred. 1	Pred. 2	Pred. 1 - Pred. 2
1	1.036 (0.03) *	0.936 (0.04)	0.099 (0.049)
2	1.031 (0.03)	0.933 (0.03)	0.098 (0.026)
3	1.043 (0.04)	0.947 (0.05)	0.096 (0.062)
4	1.032 (0.03)	0.933 (0.03)	0.099 (0.035)
5	1.032 (0.03)	0.931 (0.03)	0.101 (0.022)
6	1.040 (0.03)	0.938 (0.04)	0.102 (0.05)
7	1.055 (0.05)	0.956 (0.06)	0.099 (0.08)
8	1.032 (0.03)	0.931 (0.03)	0.101 (0.02)

Abbreviation: SLP, simple linear predictor.

* Average $\sqrt{MSE}$ (std dev.)

TABLE 5

Average root mean squared error of prediction in Simulation 2: prediction evaluation under model misspecification

	95% Verified		50% Verified		5% Verified	
	Pred. 1	Pred. 2	Pred. 1	Pred. 2	Pred. 1	Pred. 2
σ_M^2 Known	1.163 (0.06)	1.136 (0.06)	1.164 (0.06)	1.143 (0.06)	1.201 (0.08)	1.180 (0.08)
Estimated σ_M^2	1.177 (0.07)	1.157 (0.07)	1.168 (0.06)	1.149 (0.06)	1.217 (0.09)	1.198(0.10)

Average $\sqrt{MSE}$ (std dev.)

TABLE 6

Root mean squared error of prediction to compare Predictor 1, Predictor 2, and the simple linear predictor (SLP), when G_i is observed in a portion p of women

p	$\sigma_M^2 = 2$			$\sigma_M^2 = 4$			$\sigma_M^2 = 6$		
	Pred 1	Pred 2	SLP	Pred 1	Pred 2	SLP	Pred 1	Pred 2	SLP
0.95	0.943	0.854	0.740	0.963	0.911	0.811	0.983	0.954	0.845
0.50	1.031	0.933	0.794	1.031	0.974	0.859	1.041	1.008	0.931
0.05	1.035	0.935	0.838	1.043	0.974	0.860	1.046	1.020	0.931

TABLE 7

Root mean squared error of prediction to compare Predictor 1, Predictor 2, and the simple linear predictor (SLP), when G_i is not observed in a portion $p_2 + p_3$ of women

	Predictor 1	Predictor 2	SLP
Scenario 1	1.032	0.931	0.950
Scenario 2	1.041	0.938	0.964
Scenario 3	1.062	0.956	0.974

TABLE 8

Root mean squared error of prediction in Simulation 4: mean squared error of prediction evaluation

$\sigma_M^2 = 2$	$p=0.05$	$p=0.10$	$p=0.50$	$p=0.95$
True value	1.046 (0.04) *	1.037 (0.03)	1.031 (0.03)	1.036 (0.03)
True value of σ_M^2	1.046 (0.08)	1.035 (0.08)	1.031 (0.06)	1.037 (0.04)
Sample estimate of σ_M^2 train set	1.013 (0.20)	1.021 (0.14)	1.031 (0.06)	1.037 (0.04)
Likelihood obtained σ_M^2	1.020 (0.20)	1.023 (0.14)	1.033 (0.06)	1.038 (0.04)
Sample estimate of σ_M^2 test set	1.009 (0.22)	1.007 (0.15)	1.030 (0.06)	1.037 (0.04)
Only Verified Subjects	1.041 (0.14)	1.034 (0.09)	1.030 (0.04)	1.038 (0.04)
$\sigma_M^2 = 1$	$p=0.05$	$p=0.10$	$p=0.50$	$p=0.95$
True value	1.043 (0.04)	1.037 (0.03)	1.030 (0.03)	1.030 (0.03)
True value of σ_M^2	1.044 (0.06)	1.036 (0.05)	1.031 (0.04)	1.031 (0.03)
Sample estimate of σ_M^2 train set	1.030 (0.11)	1.030 (0.08)	1.031 (0.04)	1.031 (0.03)
Likelihood obtained σ_M^2	1.030 (0.11)	1.030 (0.08)	1.031 (0.04)	1.031 (0.03)
Sample estimate of σ_M^2 test set	1.027 (0.12)	1.024 (0.08)	1.030 (0.04)	1.031 (0.03)
Only Verified Subjects	1.038 (0.14)	1.034 (0.09)	1.030 (0.04)	1.030 (0.03)
$\sigma_M^2 = 0.1$	$p=0.05$	$p=0.10$	$p=0.50$	$p=0.95$
True value	1.043 (0.04)	1.042 (0.05)	1.097 (0.26)	1.052 (0.11)
True value of σ_M^2	1.043 (0.04)	1.042 (0.06)	1.097 (0.26)	1.052 (0.11)
Sample estimate of σ_M^2 train set	1.042 (0.04)	1.041 (0.06)	1.097 (0.26)	1.052 (0.11)
Likelihood obtained σ_M^2	1.035 (0.05)	1.034 (0.06)	1.042 (0.26)	1.018 (0.12)
Sample estimate of σ_M^2 test set	1.042 (0.04)	1.041 (0.06)	1.097 (0.26)	1.052 (0.11)
Only Verified Subjects	1.037 (0.14)	1.038 (0.10)	1.096 (0.26)	1.052 (0.11)

* Average $\sqrt{MSE}$ (std dev.)