Published in final edited form as:

Stat Med. 2012 January 30; 31(2): . doi:10.1002/sim.4405.

A Linear Mixed Model for Predicting a Binary Event From Longitudinal Data Under Random Effects Misspecification

Paul S. Albert

Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics, and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA. albertp@mail.nih.gov

SUMMARY

The use of longitudinal data for predicting a subsequent binary event is often the focus of diagnostic studies. This is particularly important in obstetrics, where ultrasound measurements taken during fetal development may be useful for predicting various poor pregnancy outcomes. We propose a modeling framework for predicting a binary event from longitudinal measurements where a shared random effect links the two processes together. Under a Gaussian random effects assumption, the approach is simple to implement with standard statistical software. Using asymptotic and simulation results, we show that estimates of predictive accuracy under a Gaussian random effects distribution are robust to severe misspecification of this distribution. However, under some circumstances, estimates of individual risk may be sensitive to severe random effects misspecification. We illustrate the methodology with data from a longitudinal fetal growth study.

1 Introduction

The use of longitudinal data for predicting a subsequent binary event is often the focus of diagnostic studies. Interest is in developing a dynamic predictor which can be repeatedly applied to an individual's longitudinal profile to predict a subsequent binary event. The monitoring of fetal development with repeated ultrasound measurements is important for the clinical management of pregnant women, and this will serve as the motivating example for the methodological developments in this paper. Of importance in this application is using longitudinal ultrasound measurements to predict a subsequent endpoint at birth.

A natural way to formulate such a predictor is to jointly model the longitudinal measurements and a subsequent binary event with a shared random parameter model. In these models, the dependence between the two data types is induced through random effects which are shared between the two processes. The framework is introduced assuming that the random effects follow a very flexible random effects distribution. However, it is difficult to fit these models with standard statistical software. We therefore propose an approach which is much simpler to implement for the case where the random effects distribution is assumed to be Gaussian. In using the simpler approach, it is important to understand the robustness of the assumed Gaussian random effects distribution to model misspecification.

Examining the robustness of random effects models for longitudinal data to the Gaussian random effects assumption is an active areas of research. Various authors have examined the robustness of fixed effects estimation of linear and generalized linear mixed models to the assumed random effects distribution¹⁻⁴. Others have examined robustness of random effects estimation to the assumed random effects distribution⁵⁻⁶. The robustness of fixed effect estimation to the assumed random effects distribution has been studied for joint models between longitudinal and survival data⁷⁻⁸. The focus of this article is on the robustness of the shared random parameter model under a Gaussian random effects assumption for

estimation of the individual predictive probabilities and overall measures of diagnostic accuracy for predicting a binary outcome.

In Section 2, we introduce the joint model for a general case where the random effects distribution follows a mixture of normal distributions. We show that a much simpler two-stage estimation procedure is possible when it is assumed that the random effects follow a Gaussian distribution. In Section 3, we examine the asymptotic bias when the random effects distribution is misspecified as Gaussian when in fact it is truly a two-group mixture of Gaussian distributions. We examine this bias for estimating individual predictive probabilities and for assessing the overall performance of the predictor. In Section 4, we present simulations to examine the finite sample properties of the prediction approaches under random effects misspecification. We illustrate the proposed methodology using data from a fetal growth study in Section 5. A discussion follows in Section 6.

2 Modeling Framework

We assume that the longitudinal measurements follow a linear mixed model,

$$Y_{ij} = \mathbf{X}_{ij}\beta + \mathbf{Z}_{ij}\mathbf{b}_i + \varepsilon_{ij}, \quad (1)$$

where $\mathbf{X}_{ij}$ denotes a vector of fixed effect covariates, β denotes corresponding regression parameters, $\mathbf{Z}_{ij}$ denotes a vector of random effect covariates, and $\mathbf{b}_i$ denotes the corresponding random effects, where $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n_i$, and I denotes the number of subjects. Further, we assume that measurements are taken at repeated time points denoted as t_{ij} and that the residual errors ε_{ij} are distributed with independent normal distributions with mean 0 and variance σ_ε^2 . For fetal growth, the longitudinal profile for ultrasound anthropomorphic measurements can be characterized by a special case of (1):

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \mathbf{b}_{i0} + \mathbf{b}_{i1} t_{ij} + \mathbf{b}_{i2} t_{ij}^2 + \varepsilon_{ij} \quad (2)$$

where t_{ij} is the time of the j th ultrasound measurement on the i th fetus, and Y_{ij} is the associated log-transformed anatomical measurement.

The association between an adverse binary event, denoted by S_i , and the longitudinal process can be introduced by random effects which are shared between the two processes. We assume that S_i is the binary random variable of interest. The probability of this binary variable can be linked to the longitudinal process as

$$P(S_i=1|\mathbf{b}) = \Phi\{\mathbf{W}_i\eta + \alpha h(\mathbf{b}_i)\}, \quad (3)$$

where $h(\mathbf{b})$ is a linear function in the elements of $\mathbf{b}$, where $h(\mathbf{b}) = \mathbf{g}'\mathbf{b}$, $\mathbf{W}_i$ is a vector of subject-specific covariates, and α characterizes the strength of the association between the two processes. Assuming the quadratic growth model (2) for fetal growth, an appropriate model for the binary outcome (e.g., abnormal birth outcome such as macrosomia, defined as excessive weight at birth) is

$$P(S_i=1|\mathbf{b}_i) = \Phi\{\eta_0 + \alpha(b_{i0} + b_{i1}t_* + b_{i2}t_*^2)\}, \quad (4)$$

where t_* is a time point near the time of birth (e.g. at 39 weeks of gestation). In this case $\mathbf{g} = (1 t_* t_*^2)'$. The joint model (2) and (4) relates the longitudinal fetal growth pattern to the probability of an abnormal birth outcome through an individual's predicted measurement at time t_* .

In a general form, the random effects distribution can be assumed to follow a mixture of normal distributions with different means. Slaughter et al.¹⁰ proposed this random effects structure for fetal growth, in which there are distinct groups of fetuses which grow at different inherent rates with heterogeneity within each group. The random effects

distribution can be expressed as $\mathbf{b}_i \sim \sum_{q=1}^Q p_q N(\boldsymbol{\mu}_q, V)$, where $\boldsymbol{\mu}_q$ and p_q are the mean and proportion, respectively, of observations in the q th normal component. Further, we assume that $\sum_{q=1}^Q \boldsymbol{\mu}_q p_q = 0$. Assuming the conditional independence of $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})'$ and S_i given $\mathbf{b}_i$, maximum-likelihood estimation can be conducted by maximizing the likelihood,

$$L = \prod_{i=1}^I \sum_{q=1}^Q \int \mathbf{b}_i \left\{ \prod_{j=1}^{n_i} f_q(y_{ij} | \mathbf{b}_i) \right\} P(S_i | \mathbf{b}_i) f_q(\mathbf{b}_i) p_q d\mathbf{b}_i, \quad (5)$$

where $f(y|\mathbf{b})$ is a univariate normal density and $f_q(\mathbf{b})$ is a multivariate normal density with mean $\boldsymbol{\mu}_q$ and variance V . Numeric integration over the random effects $\mathbf{b}_i$ can be performed using multivariate Gaussian quadrature¹¹. Further, the likelihood can be maximized using numerical optimization techniques such as a Newton-Raphson algorithm¹². Models can be fit for an increasing number of normal distributions in the mixture (Q) with the use of penalized likelihood techniques such as Akaike's information criteria (AIC) or the Bayesian information criteria (BIC) for model selection. Muthen and colleagues^{13–14} proposed a similar class of growth-curve models which incorporates a mixture of latent trajectories. A special case of the finite mixture of normals is when $V = 0$. In this case, the random effects distribution reduces to a pointwise finite mixture, which has been used to non-parametrically estimate the random effects distribution¹⁵.

Parameter estimation can be substantially simplified under the assumption that random effects are normally distributed. First note that for any random effects distribution, the

likelihood can be written in the form $L = L_1 \times L_2$, where $L_1 = \prod_{i=1}^I h(\mathbf{Y}_i)$ and

$L_2 = \prod_{i=1}^I P(S_i | \mathbf{Y}_i)$. Under a Gaussian random effects assumption, a two-stage pseudo-likelihood approach can be used for parameter estimation. In the first stage, we maximize L_1 with respect to the longitudinal model parameters by fitting the linear mixed model (1) with standard software. In the second stage, we maximize L_2 given the parameter estimates obtained from stage 1. By assuming a probit link function in (3) and by noting that $E[\Phi(X+W)] = \Phi(\mu+W)/\sqrt{1+\sigma^2}$ when $X \sim N(\mu, \sigma^2)$, we can express $P(S_i = 1 | \mathbf{Y}_i)$ as a probit function,

$$\begin{aligned}
P(S_i=1|Y_i) &= \int P(S_i=1|\mathbf{b})f(\mathbf{b}|Y_i)d\mathbf{b} \\
&= E_{\mathbf{b}|Y_i}[\Phi\{\mathbf{W}_i'\eta + \alpha h(\mathbf{b})\}] \\
&= \Phi\left[\frac{\mathbf{W}_i'\eta + \alpha h(\hat{\mathbf{b}}_i)}{\sqrt{1 + \alpha^2 \text{var}\{h(\hat{\mathbf{b}}_i - \mathbf{b}_i)\}}}\right] \\
&= \Phi\left[\frac{\mathbf{W}_i'\eta + \alpha \mathbf{g}'\hat{\mathbf{b}}_i}{\sqrt{1 + \alpha^2 \mathbf{g}'\text{var}(\hat{\mathbf{b}}_i - \mathbf{b}_i)\mathbf{g}}}\right]
\end{aligned} \tag{6}$$

where $\hat{\mathbf{b}}_i$ is the posterior mean of the random effects (empirical Bayes estimator) for the i th individual and $\text{var}(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \Sigma_{\mathbf{b}} - \text{var}(\mathbf{b}_i)$ where

$$\text{var}(\hat{\mathbf{b}}_i) = \Sigma_{\mathbf{b}} \mathbf{Z}_i' \left\{ \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \mathbf{X}_i' \left\{ \sum_{i=1}^I \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right\}^{-1} \mathbf{X}_i' \mathbf{V}_i^{-1} \right\} \mathbf{Z}_i \Sigma_{\mathbf{b}}, \text{ with}$$

$\mathbf{V}_i = \text{var}(Y_i) = \mathbf{Z}_i \Sigma_{\mathbf{b}} \mathbf{Z}_i' + I_{n_i} \sigma_{\varepsilon}^2$, and where I_{n_i} is an identity matrix with dimension n_i .

Equation (6) allows for the estimation of η and α accounting for the calibration error in using the plug-in estimator of $\hat{\mathbf{b}}_i$ for $\mathbf{b}_i$ ^{16,17} when the binary outcome is modeled with a probit link function.

The parameters η and α characterizing the probability of the binary outcome can be estimated in the second stage with a regression calibration approach by maximizing L_2 . For the case where the number of follow-up measurements and times are the same across all subjects, maximum-likelihood estimators of η and α can be obtained with simple probit regression. Specifically, we fit the probit regression $\Phi\{P(S_i=1)\} = \mathbf{W}_i \eta^* + \alpha^* \mathbf{g}_i' \hat{\mathbf{b}}_i$ and obtain MLEs $\hat{\eta}^*$ and $\hat{\alpha}^*$. Maximum-likelihood estimators of η and α can then be obtained by noting that $\alpha = \hat{\alpha}^* / \sqrt{(1 + \hat{\alpha}^{*2} \mathbf{g}' \text{var}(\hat{\mathbf{b}}_i - \mathbf{b}_i) \mathbf{g})}$ and $\eta = \hat{\eta}^* / \sqrt{1 - \hat{\alpha}^{*2} \mathbf{g}' \text{var}(\hat{\mathbf{b}}_i - \mathbf{b}_i) \mathbf{g}}$. More generally, L_2 can be maximized with a quasi-Newton Raphson procedure¹². Foulkes et al.¹⁸ proposed a two-stage model for prediction that does not explicitly account for the calibration error in using $\hat{\mathbf{b}}_i$ as a plug-in estimator for $\mathbf{b}_i$.

The joint model can be used to develop a predictor of the binary outcome from longitudinally collected measurements. For example, in the fetal growth application, we are interested in predicting an adverse pregnancy outcome from a series of ultrasound measurements taken at various gestational ages. Denote $\mathbf{Y}^P = (y_{s1}, y_{s2}, \dots, y_{sL})'$ as a vector of longitudinal measurements taken at time points $s_1, s_2, \dots, s_L$, where L is the number of repeated measurements in the predictor. The superscript P is used to distinguish measurements used in prediction from measurements used to fit the joint model. Denote S^P as the binary response we wish to predict. In general, the predictor based on longitudinal data can be written as

$$P(S^P=1|\mathbf{Y}^P) = \frac{\int \mathbf{b} f(\mathbf{y}^P|\mathbf{b}) P(S^P=1|\mathbf{b}) f(\mathbf{b}) d\mathbf{b}}{\int \mathbf{b} f(\mathbf{y}^P|\mathbf{b}) f(\mathbf{b}) d\mathbf{b}} \tag{7}$$

where $f(\mathbf{y}|\mathbf{b})$ is the product of univariate normal densities and $f(\mathbf{b})$ is the multivariate random effects density. For a general multivariate random effects distribution, multivariate

integration is required to evaluate the predictor. Prediction is substantially simplified under a Gaussian random effects modeling assumption. Specifically,

$$P(S^P=1|Y^P)=\Phi\left[\frac{\{W'\eta+\alpha g'\hat{b}^P\}}{\sqrt{1+\alpha^2 g'var(\hat{b}^P-b)g}}\right], \quad (8)$$

where $\hat{b}^P = \Sigma_b Z^P V^{P-1} (y^P - X^P \beta)$, V^P is the variance of Y^P , and $var(\hat{b}^P - b)$ is calculated with an expression similar to the one for $var(\hat{b}_i - b)$ in (6).

We need to assess the predictive ability of the longitudinal classifiers. One such measure is the receiver operator curve (ROC) or the area under this curve (AUC). Predictors can also be assessed by absolute measures of risk¹⁹ such as the mean-squared error (MSE) of prediction. For any chosen measure, there are various ways to correctly validate a predictor, which can include cross-validation as well as fitting the model on one set of data (training set) and estimating the predictive ability of the classifier on another set of data (test set). Without loss of generality, we will assume that model parameters are estimated using the joint model estimated from the training set (using data from I_{tr} individuals) and that the resulting estimated predictors are validated using a test data set. Denote S_i^P and Y_i^P as values of the binary outcome and longitudinal measurements for the i th subject in the test set, where $i = 1, 2, \dots, I_{te}$, and I_{te} is the number of subjects in the t

The quality of the predictor can be evaluated using the ROC curves. The ROC curve is a plot of 1-specificity versus sensitivity for multiple cut-off values of the predictor.

Specifically, we plot $1 - \widehat{Spec}(k)$ versus $\widehat{Sens}(k)$ where

$$\widehat{Sens}(k) = \frac{1}{n_{S1}} \sum_{i=1}^{I_{te}} \{I(\hat{P}(S_i^P=1|Y_i^P)) \geq C_k\} I(S_i^P=1), \quad (9)$$

$$\widehat{Spec}(k) = \frac{1}{n_{S0}} \sum_{i=1}^{I_{te}} \{I(\hat{P}(S_i^P=0|Y_i^P)) < C_k\} I(S_i^P=0), \quad (10)$$

where $I(x)$ is an indicator function which is equal to 1 if x is true and equal to 0, otherwise. Further,

Further, $\hat{P}(S_i^P=1|Y_i^P)$ are obtained by fitting the joint model to the training data and by plugging the maximum-likelihood estimators into (9) or (10). Further, we denote

$n_{S1} = \sum_{i=1}^{I_{te}} I(S_i^P=1)$, $n_{S0} = I_{te} - n_{S1}$, and $C_k, k = 1, 2, \dots, K$ are the K unique predictive values among the I_{te} subjects in the test data set. The area under the ROC curve (AUC) evaluated on the test set data can be estimated as

$AUC = \frac{1}{2} \sum_{k=2}^K \{\widehat{Sens}(k) + \widehat{Sens}(k-1)\} \{\widehat{Spec}(k) - \widehat{Spec}(k-1)\}$. The ROC curve for assessing predictive accuracy has been criticized for lack of clinical relevance and for not accounting for the prevalence of disease. Alternative measures for assessing absolute risk have been proposed in the literature¹⁹. For example, the MSE for prediction can be

calculated as $\widehat{MSE} = \frac{1}{I_{te}} \sum_{i=1}^{I_{te}} [\hat{P}(S_i^P=1|Y_i^P) - S_i^P]^2$. Obtaining individual estimates of risk and assessing the overall performance of a predictor are more computationally difficult when the random effects distribution is a mixture of normal distributions as compared with a normal distribution.

3 Asymptotic Bias for Misspecified Random Effects

Interest is on examining the asymptotic bias for estimating predictive probabilities as well as in assessing the overall accuracy of the predictor for the two-stage approach which assumes Gaussian random effects. Specifically, we examine the bias that exists when we assume that the random effects distribution is a normal distribution (model M) when the true random effects distribution is a mixture of two normal distributions (T). Asymptotic bias for the estimated predictive probabilities is under the assumption that the number of individuals in the training set gets large ($I_{tr} \rightarrow \infty$). Asymptotic bias for estimating the ROC, AUC, and MSE is under the additional assumption that the number of individuals in the test set is large ($I_{te} \rightarrow \infty$). Further, we assume that the longitudinal measurements in the training set are taken at time points $t_1, t_2, \dots, t_J$, while the longitudinal measurements in the test data set are taken at time points $s_1, s_2, \dots, s_L$.

For the purposes of evaluating asymptotic bias, we will assume a model similar to the one described by (2) and (4). Specifically,

$$Y_{ij} | \mathbf{b}_i = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + b_{i0} + b_{i1} t_j + b_{i2} t_j^2 + \varepsilon_{ij} \quad (11)$$

$$P(S_i = 1 | \mathbf{b}_i) = \Phi\{\eta_0 + \alpha(b_{i0} + b_{i1} t_j + b_{i2} t_j^2)\} \quad i=1, 2, \dots, I, j=1, 2, \dots, J, \quad (12)$$

where $\mathbf{b}_i = (b_{i0}, b_{i1}, b_{i2})'$. The working model M assumes that $\mathbf{b}_i \sim N(0, \Sigma_b)$, while the true model T is a mixture of two normals where $\mathbf{b}_i \sim p_1 N(\boldsymbol{\mu}_1, V) + (1 - p_1) N(\boldsymbol{\mu}_2, V)$. Denote $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ as fixed effect parameters, Σ_b as the variance of the random effects, and σ_ε^2 as the residual variance estimator under model (T). Further, denote $\boldsymbol{\beta}^*$ as the estimated fixed effect parameters, $\hat{\Sigma}_b^*$ as the estimated variance of the random effects, and $\hat{\sigma}_\varepsilon^{*2}$ as the estimated residual variance estimator under model (M) when the true model is model (T). Following from Verbeke and Lesaffre¹, $\boldsymbol{\beta}^*$ and $\hat{\sigma}_\varepsilon^{*2}$ are consistent estimators of $\boldsymbol{\beta}$ and σ_ε^2 , respectively. Additionally, $\hat{\Sigma}_b^*$ is a consistent estimators of Σ_b^* , where

$$\Sigma_b^* = \text{var}(\mathbf{b}_i) = \Sigma_b + p_1(1 - p_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^2.$$

Asymptotic bias for model parameter estimators of the second stage model, η_0 and α under model misspecification can be obtained by maximizing the expected log-likelihood of model M where the expectation is taken with respect to model T ^{4,20}. Denote $f_M(S|Y)$ as the probability of an abnormal outcome S given the longitudinal process Y under the misspecified model M . Under model misspecification, the estimated parameters of the predictive model $\hat{\eta}_0^*$ and $\hat{\alpha}^*$ converge to η_0^* and α^* , where

$$\begin{aligned} & (\eta_0^*, \alpha^*) \\ &= \underset{(\eta_0, \alpha)}{\text{argmax}} E_T \left[\log f(S | Y, \beta, \Sigma_b^*, \sigma_\varepsilon^2) \right] = \underset{(\eta_0, \alpha)}{\text{argmax}} E_T \left[S \log \Phi \left(\frac{\eta_0 + \alpha \mathbf{g}' \mathbf{b}(Y)}{\sqrt{1 + \alpha^2 \mathbf{g}' \text{Var}\{\mathbf{b}(Y)\} \mathbf{g}}} \right) \right. \\ & \quad \left. + (1 - S) \log \left(1 - \Phi \left(\frac{\eta_0 + \alpha \mathbf{g}' \mathbf{b}(Y)}{\sqrt{1 + \alpha^2 \mathbf{g}' \text{Var}\{\mathbf{b}(Y)\} \mathbf{g}}} \right) \right) \right] \end{aligned} \quad (13)$$

where $\mathbf{g}=(1, t_j, t_j^2)'$, $\mathbf{b}(\mathbf{Y})=\Sigma_{\mathbf{b}}^* \mathbf{Z}' V^{-1}(\mathbf{Y}-\mathbf{Z}\beta)$, and where

$$\mathbf{Z}=\begin{pmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_j \\ t_1^2 & t_2^2 & \dots & t_j^2 \end{pmatrix}', V=\mathbf{Z}'\Sigma_{\mathbf{b}}^*\mathbf{Z}+I_j\sigma_{\varepsilon}^2. \text{ Further, } Var\{\mathbf{b}(\mathbf{Y})\}=\Sigma_{\mathbf{b}}^*-\Sigma_{\mathbf{b}}^*\mathbf{Z}V^{-1}\mathbf{Z}\Sigma_{\mathbf{b}}^*.$$

Finally, under misspecification of the random effects distribution, the estimated predicted value based on a vector of longitudinal measurements $\mathbf{Y}^P$, $P(\hat{S}=1|\mathbf{Y}^P)$, will converge to $P^*(S=1|\mathbf{Y}^P)$, where

$$P^*(S=1|\mathbf{Y}^P)=\Phi\left[\frac{\eta_0^*+\alpha^*\mathbf{g}'\mathbf{b}(\mathbf{Y}^P)}{\sqrt{1+\alpha^{*2}\mathbf{g}'Var\{\mathbf{b}(\mathbf{Y}^P)\}\mathbf{g}}}\right], \quad (14)$$

$\mathbf{Z}=\begin{pmatrix} 1 & 1 & \dots & 1 \\ s_1 & s_2 & \dots & s_L \\ s_1^2 & s_2^2 & \dots & s_L^2 \end{pmatrix}'$, and $\mathbf{g}$, $\mathbf{b}(\mathbf{Y}^P)$, and $Var\{\mathbf{b}(\mathbf{Y}^P)\}$ are calculated similarly to what was presented previously. We note that (14) depends on the two-group mixture only through the form for $\Sigma_{\mathbf{b}}$.

Expression (7) is evaluated using multivariate Gaussian quadrature with 50 quadrature points²¹ for each of the three dimensions (corresponding to each of the three random effects). Evaluating (13) is intractable in closed form due to the difficulty in evaluating the expectation with respect to model T . Instead, $E_T[\log f(S|\mathbf{Y}, \beta, \Sigma_{\mathbf{b}}^*, \sigma_{\varepsilon}^2)]$ is approximated by

$\frac{1}{K}\sum_{k=1}^K \log f(S_k|\mathbf{Y}_k^P, \beta, \Sigma_{\mathbf{b}}^*, \sigma_{\varepsilon}^2)$, where $(S_k, \mathbf{Y}_k^P)$ is simulated under model T with $K=20,000$. The maximization required for evaluating (13) was conducted using a quasi-Newton-Raphson algorithm¹¹.

The asymptotic bias is calculated for different two-group normal mixture random effects distributions with a common variance and with probability of 0.5 of being in each of the two mixture groups. The different two-group mixture models (A) to (D) correspond to a mixture of normals where the separation of the two normals in the mixture is 20%, 40%, 200%, and 400% of the standard deviation, respectively. Figure 1 shows histograms of the random effects distribution under models (A–D). In evaluating asymptotic biases in both individual prediction as well as overall quality of the predictor, we chose values of the fixed effects that result in patterns which are consistent to what would be expected in fetal growth.

Specifically, we chose a quadratic mean structure and variance structure parameters which would be consistent with longitudinal biomarkers and imaging data in fetal growth studies. Further, we chose $t_j=j$ and $J=4$ in (11) and (12) which is consistent with (i) the four longitudinal measurements taken in the example fetal growth study analyzed in Section 5, (2) linking the longitudinal process with the binary event through an individual's predicted longitudinal measurement near the time of the binary event (e.g., the final measurement which in the fetal growth study is close to the time of measuring the birth outcome). Figure 2 shows simulated individual realizations for random effects distributions (A) to (D) with the mean growth curve and whether the binary response is positive or negative (denoted as 0 or 1 at time 4) superimposed on each graph. The asymptotic bias calculations are also done assuming that there are three measurements in the predictor ($L=3$) taken at follow-up times s_1, s_2 , and s_3 , and the time point for the binary prediction is at $t_j=4$. The measurement times for prediction may either be different, reflecting longitudinally collected measurements, or replicate values taken at a single time point. Taking replicate measurements at a single time point may be advantageous when σ_{ε}^2 is large and when only a single time point is available

for prediction. Table 1 shows the asymptotic bias of predictive values ($P^*(S = 1|Y^P) - P(S = 1|Y^P)$) for different longitudinal measurements values and times under the different random effects misspecifications. Values of Y_p were obtained by simulating from (11). Further, measurement times (s_1, s_2, s_3) were chosen as either early in the observation period (three replicate measurements taken at time 1), late in the observation period (three replicate measurements taken at time 4), or three time points taken at times 1, 2, and 3. The results suggest that there is very little bias for individual prediction under small misspecifications of the random effects distribution (A and B). The bias can be much more substantial under moderate or severe misspecifications (C and D). The degree of bias is substantially increased when predicting from earlier measurements that are taken far from the final time point that links the two processes. For example, when the three measurements are taken as replicates at time point 4 there is only mild bias even under the most severe random effects misspecification (D). However, when the three time points are replicate measurements at time point 1, far from the time of the binary outcome we are trying to predict, bias is very substantial even for random effects misspecification (C).

Focus is often on estimating the quality of a predictor based on the receiver operator curve (ROC). When the predictor is developed under model (M) when the truth is model (T), the estimated ROC curve as computed by (9) and (10) converges to the curve characterized by plotting $1 - \widehat{Spec}^*(c)$ versus $\widehat{Sens}^*(c)$ for all continuous values of c between 0 and 1, where $\widehat{Sens}^*(c) = E_T[P^*(S = 1|Y^P) - c|S = 1]$ and $\widehat{Spec}^*(c) = E_T[P^*(S = 1|Y^P) < c|S = 0]$, and where $P^*(S = 1|Y^P)$ is computed under the Gaussian random effects assumption as (14), and the expectation is taken under the true (T) joint model (Y^P, S). Under the correct model, the estimators $\widehat{Sens}^*(c)$ and $\widehat{Spec}^*(c)$ will converge to $Sens(c)$ and $Spec(c)$, which can be expressed as $Sens(c) = E_T[P(S = 1|Y^P) > c|S = 1]$, $Spec(c) = E_T[P(S = 1|Y^P) < c|S = 0]$, where $P(S = 1|Y^P)$ is evaluated as in (7). The expectation with respect to the true model required for calculating $\widehat{Sens}^*(c)$, $\widehat{Spec}^*(c)$, $Sens(c)$, and $Spec(c)$ can be approximated by simulating (Y^P, S) from the true model (we simulated 20,000 realizations) and averaging values. The asymptotic bias in AUC can be evaluated by computing the area of these two ROC curves.

The expected mean-squared error is a measure of the absolute risk of the predictor. In large samples, the MSE converges to $MSE = E_T[(P(S = 1|Y^P) - S)^2]$ under the correct model (T) and $MSE^* = E_T[(P(S = 1|Y^P) - S)^2]$ under the misspecified Gaussian model (M). Table 2 shows asymptotic bias of the model accuracy for increasing departures from normality (A–D) when $s_l = l, l = 1, 2, 3$ and $L = 3$. The results indicate that the AUC is asymptotically unbiased even under severe random effects misspecification. Table 2 also shows that the asymptotic bias for the MSE is very small even for large random effects misspecification.

4 Simulation

We examine the finite sample properties of the predictors under the correctly specified and misspecified models. We simulate under model (11) and (12) with b_i following a two-group pointwise mixture with mean 0. Specifically, we simulate b_i to be (b_0, b_1, b_2) with probability

p_1 or $-\frac{p_1}{1-p_1}(b_0b_1b_2)$ with probability $1-p_1$ and estimate the joint model under the correctly specified finite mixture model and the misspecified Gaussian random effects model. The two-group pointwise mixture is a special case of the two-group mixture of normals when $V=0$. We examined a two-group pointwise mixture model as compared to the two-group mixture of normal distributions examined in the asymptotic calculations, since estimating the parameters of model (11) and (12) with a two-group mixture model is substantially more computationally intensive than estimating the parameters with a two-group pointwise mixture (i.e., no multivariate numerical integration is required for the pointwise mixture). Further, the pointwise mixture is more of a departure from a Gaussian

random effects distribution than the more general two-group mixture of normals, thereby providing a more extreme comparison for demonstrating robustness. The parameters for the two-group pointwise mixture is the same as those specified in (C) and (D) in Figure 2 with no between-subject variation within each of the mixture groups. The simulations were conducted based on a training set sample size of $I_{tr} = 1000$ and a test set sample size of $I_{te} = 1000$. Further, we assume that longitudinal observations for the test set are $t_j = j$, $j = 1, 2, 3$, and 4 with $t_J = 4$. The simulations are conducted for a predictor whose longitudinal measurements are taken at three time points, $s_l = 1$, $l = 1, 2$, and 3.

Table 3 shows the results of simulations which vary the random effects distribution (b_0 b_1 b_2), the mixing proportion p_1 , and the strength of the relationship between the random effects and the probability of the binary endpoint(α). Over a wide range of parameter values, the average estimated AUC under the true mixture model ($avg\widehat{AUC}_T$) was nearly the same as the average estimated under the misspecified Gaussian random effects model ($avg\widehat{AUC}_M$). In addition, the standard errors were very close to each other. The average estimated prediction MSE under the misspecified Gaussian random effects distribution was also very close to the MSE evaluated under the correctly specified two-group pointwise mixture distribution.

5 Example

An important issue in obstetrics is how to deliver the baby. If the fetus is too large, a cesarean delivery may be necessary. Thus, the development of simple and accurate methods for predicting a large baby using longitudinal ultrasound measurements will be useful. We illustrate the two-stage approach using longitudinal and birth outcome data from the National Institute of Child Health and Human Development (NICHD) study of successive small-for-gestational-age births in Scandinavia²². The study was designed to collect longitudinal ultrasound measurements at 17, 25, 33, and 37 weeks of gestational age.

To illustrate the methodology, we focus on the use of longitudinal ultrasound abdominal diameter in predicting macrosomia (defined by a birthweight larger than 4000 grams). We focus on 1203 subjects who had all four longitudinal ultrasound measurements along with birth outcome data. The median abdomen diameter (3rd quartile-4th quartile) at 17, 25, 33, and 37 weeks of gestational age are 39cm (36–41), 65cm (62–67), 92cm (89–95), and 105cm (101–108), respectively. The proportion of macrosomia was estimated as 0.175.

We split the data into a training set and test set of approximately equal lengths (600 in the training set and 603 in the test set) and (i) fit the two-stage modeling approach to the training set data with log-transformed abdomen diameter, and (ii) validated the resulting predictor using the test set data. All four longitudinal measurements were used in estimating the predictor. We then estimated the ROC curve, AUC, and MSE of prediction with only the test set data. Table 4 shows parameter estimates for the two-stage approach using models (2) and (4) fit to the training set data. Parameters for the slope and quadratic terms characterizing fetal growth are all highly significant, suggesting that there is statistically significant curvature to the mean structure on the log scale. The parameter α , which links the two processes, is positive and highly statistically significant, suggesting that fetal growth is positively related to the probability of macrosomia. Further, the between-subject variation (particularly for the intercept) is relatively high, and there is a sizable negative correlation between the random intercept and slope.

Q-Q plots for the random effects distribution²³ showed little departure from normality (data not shown). However, caution should be used in interpreting such plots since they may fail to detect non-normality when the error distribution is large and can falsely detect normality

for the random effects when the error distribution is non-normal⁵. A scatter plot of predicted individual means versus residuals showed no systematic pattern suggesting that the quadratic mean structure and random effects provide a reasonable representation of the mean structure (data not shown).

We estimated overall assessments of diagnostic accuracy for prediction using different measurement times from the test set data. Specifically, we estimated AUC (MSE) using all four longitudinal measurements, the last three of these measurements, the last two measurements, and the final measurement as 0.76 (0.12), 0.79 (0.11), 0.79 (0.11), and 0.80 (0.11), respectively. These results suggest that, in terms of overall predictive accuracy, there is little advantage in taking additional observations above the last available measurement. Further, a single measurement at 25 weeks and 33 weeks resulted in an AUC (MSE) of 0.71 (0.12) and 0.64 (0.13), respectively. Thus, taking the single measurement as close to birth as possible is advantageous for overall prediction.

6 Discussion

In this article, we have proposed a shared random parameter framework for predicting a binary event from longitudinal measurements. Specifically, we linked the two processes through a set of shared random effects. First, we consider a general model where the random effects distribution is assumed to follow a mixture of normal distributions. Second, a simple two-stage model which can be implemented with standard statistical software was considered. However, this approach can be implemented only under a Gaussian random effects assumption. A probit model is specified for the binary response in order to obtain a closed-form expression to account for calibration error (expression (6)). Alternatively, an approximate version of expression (6) can be developed for the logistic model using a cumulative Gaussian approximation to the logistic function²⁴.

Through asymptotic bias calculations and simulations, I demonstrate that measures of overall diagnostic accuracy such as the AUC of the ROC curve and MSE are robust to even severe misspecification of the random effects distribution. This result is consistent with McCulloch and Neuhaus⁶ who show that the MSE for random effects estimation is robust to the random effect misspecification. However, individual risk predictions may be sensitive to random effects misspecification. These results have important implications for practitioners. First, even when the random effects distribution is clearly non-normal, the simple two-stage approach will be a useful tool for assessing the overall quality of the predictor based on MSE or summaries of the ROC curve. Second, for predictors that have overall high predictive accuracy, a careful assessment of the random effects distribution is necessary for using the predictor to assess individual risk. This would involve fitting models with richer random effects distributions such as the mixture of normals discussed in Section 3.

The focus of this paper was on examining robustness to the Gaussian random effects assumption when predicting a binary event from longitudinally collected data. Of course, prediction may also be sensitive to departures from the mean structure. An advantage of the two-stage approach is that standard model diagnostic methods for checking the mean and variance structure in linear mixed modeling methodology can be applied. For example, a scatter plot of standardized residuals versus predicted values can easily be constructed to examine mean structure misspecification.

The proposed methodology focuses on predicting a binary outcome from a single longitudinally collected outcome. Although the shared random parameter model with the more general random effects model could be extended to the multivariate case, it is much

simpler to extend the two-stage approach to this multivariate setting. This topic is the subject of future research.

Acknowledgments

I thank the Associate Editor and two reviewers for their constructive comments on this article. I thank Dr. Jun Zhang for helpful comments on this manuscript. I thank the Center for Information Technology, National Institutes of Health, for providing access to the high-performance computational capabilities of the Biowulf cluster computer system. The research was supported by the Intramural Research Program of the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development.

REFERENCES

1. Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*. 1997; 23:541–556.
2. Neuhaus JM, Hauck WW, Kalbfleisch JD. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*. 1992; 79:755–762.
3. Litieri S, Alonso A, Molenberghs G. Type I and Type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*. 2007; 63:1038–1044. [PubMed: 17425642]
4. Heagerty PJ, Kerland BF. Misspecified maximum-likelihood estimates and generalized linear mixed models. *Biometrika*. 2001; 88:973–985.
5. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random effects population. *The Journal of the American Statistical Association*. 1996; 91:217–221.
6. McCulloch CE, Neuhaus JM. Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics*. 2010 In press at.
7. Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*. 2006; 62:1037–1043. [PubMed: 17156277]
8. Song XA, Davidian M, Tsiatis AA. A semiparametric likelihood approach to jointly modeling of longitudinal data and time-to-event data. *Biometrics*. 2002; 58:742–753. [PubMed: 12495128]
9. Deter RL. Individualized growth assessments: evaluation of growth using each fetus as its own control. *Seminars in Perinatology*. 2004; 28:23–32. [PubMed: 15058899]
10. Slaughter JC, Herring AH, Thorp JM. A Bayesian latent variable mixture model for longitudinal fetal growth. *Biometrics*. 2009; 65:1233–1242. [PubMed: 19432784]
11. Evans, M.; Swartz, T. *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press; 200.
12. Thisted, RA. *Elements of Statistical Computing: Numerical Computation*. New York: Chapman and Hall; 1988.
13. Muthen B, Brown C, Masyn K, Jo B, Khoo S, Yang C, Wang C, Kellam S, Carlin J, Liaoj J. General growth mixture modeling for randomized prevention intervention. *Biostatistics*. 2002; 3:459–475. [PubMed: 12933592]
14. Wang CP, Brown CH, Bandeen-Roche K. Residual diagnostics for growth mixture models: examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*. 2006; 100:1054–1076.
15. Laird N. Nonparametric maximum-likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*. 1978; 73:805–811.
16. Carroll RJ, Spiegelman CH, Lan KKG, Abbott RD. On errors-in-variables for binary regression models. *Biometrika*. 1984; 71:19–25.
17. Wang CY, Wang N, Wang S. Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics*. 2000; 56:487–495. [PubMed: 10877308]
18. Foulkes AS, Azzoni L, Li X, Johnson M, Smith C, Mounzer K, Montaner LJ. Prediction-based classification for longitudinal biomarkers. *Annals of Applied Statistics*. 2010; 4:1476–1497. [PubMed: 21274424]

19. Gail M, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Bio-statistics*. 2005; 6:227–239.
20. White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982; 50:1–26.
21. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*. New York: Dover; 1974.
22. Bakketeig LS, Jacobsen G, Hoffman HJ, Lindmark G, Bergsjö P, Moline K, Rodsten J. Pre-pregnancy risk factors of small-for-gestational age births among parous women in Scandinavia. *Acta Obstetrica et Gynecologica Scandinavica*. 1993; 72:273–279. [PubMed: 8389514]
23. Lang N, Ryan L. Assessing normality in random effects models. *The Annals of Statistics*. 1989; 17:624–642.
24. Johnson, NL.; Kotz, S. *Distributions in Statistics, Continuous Univariate Distributions*. Vol. Vol 2. Boston: Houghton-Mifflin; 1970. p. 6

Figure 1.

Histograms for the random effect intercept corresponding to two-group mixture models A, B, C, and D. The differing misspecified two-group mixture models A–D correspond to a separation between the two normals of 20%, 40%, 200%, and 400% respectively in the standard deviation of the normal distribution. The random effects distribution for misspecified model D is characterized by $\mu_1 = 2(\sqrt{V_{11}}, \sqrt{V_{22}}, \sqrt{V_{33}})'$ and $\mu_2 = -2(\sqrt{V_{11}}, \sqrt{V_{22}}, \sqrt{V_{33}})'$, where V_{11} , V_{22} , and V_{33} are the variances of the random intercept, slope, and quadratic effect, respectively. The figures show the distribution of the

random intercepts for $V_{11} = 0.04$. Histograms for the linear and quadratic random effects are similar to those for the random intercept and are not shown.

Figure 2.

Simulated longitudinal profiles under two group mixture models (A–D). We assume (16) and (17) with $\beta_0 = 0$, $\beta_1 = 3$, $\beta_2 = -0.4$, $\eta_0 = -1$, and $\sigma_\epsilon^2 = 0.5$. The true random effects distribution is a two-group mixture of normals with $p_1 = 0.5$, and $V = \text{diag}(0.04, 0.01, 0.0025)$. The data from twenty individuals is simulated for each panel. Solid lines reflect the mean growth profile in each panel, while dashed lines denote individual curves. The ones and zeros at time 4 reflect simulated binary outcomes that are linked to the underlying longitudinal process at time 4.

Asymptotic bias for individual predictive values under a Gaussian random effects model when the true model is a two-group mixture model. We assume (11) and (12) with $\beta_0 = 0$, $\beta_1 = 3$, and $\beta_2 = -0.4$, $\eta_0 = -1$, $\alpha = 2$, and $\sigma_\varepsilon^2 = 0.5$. The true random effects distribution is a two-group mixture of normals with $p_1 = 0.5$, and $V = \text{diag}(0.04, 0.01, 0.0025)$. We present $P(S = 1|Y)$, $P^*(S = 1|Y^P)$, and the asymptotic bias ($P^*(S_i = 1|Y^P) - P(S = 1|Y^P)$) for different Y^P , measurement times (s_1, s_2, s_3) and random effects misspecifications. Values of Y^P are simulated from (11).

Table 1

Rand. Eff. Missp.	(s_1, s_2, s_3)			Y^P	$P(S = 1 Y^P)$	$P^*(S = 1 Y^P)$	Asymp Bias
A	1	2	3	2.1	5.3	5.8	0.543
	4	4	4	5.2	5.9	5.1	0.116
	1	1	1	2.4	2.6	3.1	0.341
B	1	2	3	2.1	5.3	5.8	0.550
	4	4	4	5.2	5.9	5.1	0.116
	1	1	1	2.4	2.6	3.1	0.346
C	1	2	3	2.6	5.6	6.0	0.846
	4	4	4	6.0	5.7	6.3	0.454
	1	1	1	2.6	3.5	2.5	0.654
D	1	2	3	2.6	4.9	5.4	0.656
	4	4	4	5.8	6.2	6.2	0.603
	1	1	1	3.0	2.4	2.5	0.562

Asymptotic bias for the AUC of the ROC curve and MSE of prediction under a Gaussian random effects model when the true model is a two-group mixture model. We assume (11) and (12) with $\beta_0 = 0$, $\beta_1 = 3$, $\beta_2 = -0.4$, $\eta_0 = -1$, and $\sigma_\epsilon^2 = 0.5$. The true random effects distribution is a two-group mixture of normals with $p_1 = 0.5$, and $V = \text{diag}(0.04, 0.01, 0.0025)$. We assume $t_j = j$ with $j = 1, 2, \dots, 4$ and that the predictor is based on $s_l = 1$ with $l = 1, 2$, and 3. The asymptotic bias for the AUC and the MSE is $AUC^* - AUC$ and $MSE^* - MSE$, respectively.

Table 2

	Rand. Eff. Missp.	α	AUC	AUC*	AUC* - AUC	MSE	MSE*	MSE* - AUC
A	1	1	0.754	0.754	<0.001	0.153	0.153	<0.001
	2	2	0.828	0.828	<0.001	0.152	0.152	<0.001
B	1	1	0.766	0.766	<0.001	0.1510	0.1510	<0.001
	2	2	0.833	0.833	<0.001	0.1522	0.1523	<0.001
C	1	1	0.907	0.907	<0.001	0.143	0.143	<0.001
	2	2	0.962	0.962	<0.001	0.0740	0.0757	0.002
D	1	1	0.983	0.983	<0.001	0.035	0.038	0.003
	2	2	0.999	0.999	<0.001	0.0060	0.0074	0.0014

Table 3

Simulation results for estimating the AUC and the MSE of a predictor under the correct (T) and misspecified (M) random effects distribution. with $\beta_0 = 0$, $\beta_1 = 3$, $\beta_2 = -0.4$, $\eta_0 = -1$, and $\sigma_\epsilon^2 = 0.5$. The true random effects distribution $\mathbf{b}_i$ is given by a two-group point mixture with mean $\mathbf{0}$ where $\mathbf{b}_i = (b_0, b_1, b_2)$ with probability p_1 or $\frac{p_1}{1-p_1}(b_0b_1b_2)$ with probability $1-p_1$. We assume $t_j = j$ with $j = 1, 2, \dots, 4$ and that the predictor is based on $s_l = l$ with $l = 1, 2$, and 3. We assume $I_{tr} = I_{te} = 1000$. The columns $avg\widehat{AUC}_T$ and $avg\widehat{MSE}_T$ denote the average AUC and MSE using the predictor with the correct random effects distribution, while the columns $avg\widehat{AUC}_M$ and $avg\widehat{MSE}_M$ denote the average AUC and MSE using the misspecified random effects distribution. Monte-Carlo standard deviations are presented in parentheses. Simulations are based on 1000 simulated datasets.

b_0	b_1	b_2	p_1	α	$avg\widehat{AUC}_T$	$avg\widehat{AUC}_M$	$avg\widehat{MSE}_T$	$avg\widehat{MSE}_M$
0.2	0.1	0.05	0.5	1	0.864 (0.010)	0.865 (0.011)	0.121 (0.0053)	0.137 (0.0056)
0.4	0.2	0.1	0.5	1	0.983 (0.0038)	0.983 (0.0043)	0.017 (0.0037)	0.021 (0.0035)
0.4	0.2	0.1	0.2	1	0.918 (0.014)	0.918 (0.014)	0.043 (0.0074)	0.054 (0.0053)
0.2	0.1	0.05	0.5	0.50	0.749 (0.016)	0.751 (0.016)	0.141 (0.0059)	0.145 (0.0063)
0.2	0.1	0.05	0.2	0.50	0.634 (0.027)	0.641 (0.025)	0.137 (0.0085)	0.139 (0.0075)
0.4	0.2	0.1	0.5	0.50	0.865 (0.0094)	0.865 (0.012)	0.118 (0.0056)	0.123 (0.0055)
0.4	0.2	0.1	0.2	0.50	0.781 (0.021)	0.782 (0.021)	0.113 (0.011)	0.117 (0.0066)

Table 4

Parameter estimates and standard errors from fitting the two-stage model (2) and (4). The longitudinal ultrasound data were log-transformed abdomen diameters, while the binary outcome was macrosomia. Standard errors were estimated using the bootstrap with 800 re-samples.

Parameter	Estimate	SE
β_0	1.99	0.022
β_1	0.119	0.0016
β_2	-0.0013	2.89×10^{-5}
σ_{b0}	0.140	0.0040
σ_{b1}	0.0042	1.25×10^{-4}
σ_{b2}	7.41×10^{-9}	6.04×10^{-10}
ρ_{01}	-0.97	0.0026
ρ_{02}	0.00	0
ρ_{12}	0.00	0
σ_ε	0.0478	0.0011
η_0	-1.58	0.142
α	35.20	3.51