Journal of Modern Applied Statistical Methods

Volume 8 | Issue 1 Article 7

5-1-2009

Covariate-Adjusted Constrained Bayes Predictions of Random Intercepts and Slopes. Sujit Ghosh is a

Robert H. Lyles
Emory University, rlyles@emory.edu

Reneé H. Moore
University of Pennsylvania, rhmoore@mail.med.upenn.edu

Amita K. Manatunga
Emory University, amanatu@emory.edu

Kirk A. Easley
Emory University, keasle2@emory.edu

Journal of Modern Applied Statistical Methods
May 2009, Vol. 8, No. 1, 81-94

1538 – 9472/09/$95.00

Covariate-Adjusted Constrained Bayes Predictions of Random Intercepts and Slopes

Robert H. Lyles
Emory University

Reneé H. Moore
University of Pennsylvania

Amita K. Manatunga
Emory University

Kirk A. Easley
Emory University

Constrained Bayes methodology represents an alternative to the posterior mean (empirical Bayes) method commonly used to produce random effect predictions under mixed linear models. The general constrained Bayes methodology of Ghosh (1992) is compared to a direct implementation of constraints, and it is suggested that the former approach could feasibly be incorporated into commercial mixed model software. Simulation studies and a real-data example illustrate the main points and support the conclusions.

Key words: Mixed linear model, prediction, random effects, shrinkage.

Introduction

The standard mixed linear model (e.g., Laird & Ware, 1982) remains a popular practical tool for analyzing longitudinal, repeated measures, or otherwise correlated continuous data. In such analyses, the prediction of linear combinations of fixed and random effects can be of great interest. The typical approach implemented in commercial software is to obtain empirical best linear unbiased predictors (EBLUPs), which estimate the posterior mean of the linear combination given the response data (Littell, et al., 2006). The general acceptance of these empirical Bayes-like predictions stems from their intuitive appeal and their theoretical underpinnings as minimal prediction mean squared error estimates (Searle, et al., 1992). They are also referred to as shrinkage estimators, given their characteristic of pulling subject-specific predictions toward a population mean.

Due to the shrinkage phenomenon, EBLUPs stemming from linear mixed models exhibit distributions that can be much narrower than those assumed to characterize the random variables being predicted. Several authors (e.g., Efron & Morris, 1971; Louis, 1984; Ghosh, 1992) have suggested potential drawbacks to this general feature and proposed methods that reduce shrinkage and/or more closely match the predictor and underlying true distributions.

One effect of overshrinkage in certain applications is that it can lead to a lack of sensitivity for identifying extreme experimental units relative to a fixed threshold (i.e., the probability that an EBLUP lies beyond a threshold given that the true random variable does can be quite small). To improve sensitivity in such a context, Lyles and Xu (1999) proposed constrained Bayes predictors of random intercepts and slopes aimed to minimize mean squared error of prediction (MSEP) given that the means and variances of the predictor distributions match those of the true random effects. Lyles, et al. (2007) introduced additional prediction criteria (e.g., regional bias and

Robert H. Lyles is an Associate Professor in the Department of Biostatistics at The Rollins School of Public Health. Email: rlyles@emory.edu. Reneé Moore is an Assistant Professor in the Department of Biostatistics and Epidemiology in the School of Medicine. Email: rhmoore@mail.med.upenn.edu. Amita Manatunga is a Professor in the Department of Biostatistics at The Rollins School of Public Health. Email: amanatu@emory.edu. Kirk Easley is a Senior Associate in the Department of Biostatistics at The Rollins School of Public Health. Email: keasle2@emory.edu.

MSEP) that are relevant when extreme subjects are of key interest and they suggested that the constrained Bayes approach can be an appealing alternative in such situations. Constrained Bayes prediction of random effects has not been widely advocated for use in the mixed linear model context.

The models considered by Lyles and Xu (1999) are extended here to use fixed and/or time-dependent covariates, and their direct constrained Bayes strategy is compared with the general paradigm advocated by Ghosh (1992). This comparison is relevant for two reasons.

First, while the criteria put forth by Lyles and Xu are specific to the mixed linear model, Ghosh's approach originates from a more general and decidedly Bayesian point of view. Ghosh provides a paradigm for minimizing a mean squared error criterion subject to matching the posterior expectation of the first two moments of a parameter distribution to corresponding moments of the histogram of the set of estimates. It is therefore useful to assess the performance of Ghosh's paradigm in the mixed model setting and to compare it against an approach that is directly rooted in that context.

Second, Ghosh's method is general, flexible, and implemented in a straightforward and consistent manner. Therefore its validation against an approach directly rooted in the mixed model setting could highlight, for practitioners and commercial mixed linear model software developers, the viability of an accessible alternative prediction method.

Methodology

Models and Posterior Mean Predictions

Two familiar normal-theory mixed linear models are used for illustration: the random intercept and random intercept/slope models, respectively.

The random intercept (or one-way random effects ANOVA) model is specified as follows (e.g., Searle, et al., 1992):

$$Y_{ij} = \mu + b_i + e_{ij} \qquad (1)$$

($i = 1,2,\ldots, k$; $j = 1,2,\ldots, n_i$), with i indexing the subject and j indexing the observation. Typical normality assumptions dictate that $b_i \sim N(0,\sigma_b^2)$ and $e_{ij} \sim N(0,\sigma_w^2)$, with independence across subjects and between the random terms b_i and e_{ij}.

Under model (1), a common objective is to predict the i^{th} subject's random subject-specific mean, i.e., $\mu_i = \mu + b_i$ ($i=1,\ldots,k$). The EBLUP, as provided by standard mixed model software, is an estimate of the posterior mean $E(\mu_i \mid \mathbf{Y}) = E(\mu_i \mid \mathbf{Y}_i)$, where $\mathbf{Y}$ and $\mathbf{Y}_i$ denote the complete and i^{th} subject-specific data vectors, respectively:

$$\tilde{\mu}_i = E(\mu_i \mid \mathbf{Y}_i = \mathbf{y}_i) = \nu_i \bar{y}_i + (1-\nu_i)\mu \qquad (2)$$

$$\text{where } \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij} \text{ , and}$$

$$\nu_i = \{1 + \sigma_w^2/(n_i\sigma_b^2)\}^{-1}.$$

The parameter ν_i governs the extent to which the predicted value shrinks toward the population mean μ, with more excessive shrinkage occurring when ν_i is small (i.e., when $\sigma_w^2/(n_i\sigma_b^2)$ is large). The BLUP is obtained by replacing μ in (2) by its best linear unbiased estimate (Searle, et al., 1992), whereas in practice the EBLUP also replaces the variance components in (2) by their estimates.

Next, consider the random intercept/ slope model, also known as a randomized regression or linear growth curve model (e.g., Diggle, et al., 1994):

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)t_{ij} + e_{ij} \qquad (3)$$

($i = 1,2,\ldots, k$; $j = 1,2,\ldots, n_i$), where t_{ij} denotes the time at which Y_{ij} is measured. Typically this model assumes independence across subjects and normally distributed random effects as follows:

$$\begin{pmatrix} a_i \\ b_i \\ e_{ij} \end{pmatrix} \sim N_3\left\{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right\},$$

with $\sigma_1^2, \sigma_2^2, \sigma_{12}$, and σ^2 denoting the variances of the subject-specific intercept and slope deviations, their covariance, and the random error variance, respectively.

Under model (3), it is common to seek predictions of the i^{th} subject's random intercept ($\alpha_i = \alpha + a_i$) and slope ($\beta_i = \beta + b_i$). As with model (1) and most feasible mixed linear models, standard software provides EBLUPs for these quantities. In this case, they are estimates of the posterior means $E(\alpha_i \mid \mathbf{Y}) = E(\alpha_i \mid \mathbf{Y}_i)$ and $E(\beta_i \mid \mathbf{Y}) = E(\beta_i \mid \mathbf{Y}_i)$. The normality assumptions accompanying model (3) yield

$$\begin{aligned} \tilde{\beta}_i &= E(\beta_i \mid \mathbf{Y}_i = \mathbf{y}_i) \\ &= \beta + (\sigma_{12}\mathbf{1}'_{n_i} + \sigma_2^2 \mathbf{t}'_i)\Sigma_i^{-1}(\mathbf{y}_i - \alpha \mathbf{1}_{n_i} - \beta \mathbf{t}_i) \end{aligned} \quad (4)$$

where $\mathbf{\Sigma}_i = \mathrm{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{\Delta} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}$, $\mathbf{Z}_i$ is the design matrix for the simple linear regression of $\mathbf{Y}_i$ on time ($\mathbf{t}_i$) for subject i, and $\mathbf{\Delta} = \mathrm{Var}\,(a_i, b_i)'$. Assuming $n_i \geq 2$, Lyles and Xu (1999) showed that $E(\beta_i \mid \mathbf{Y}_i)$ takes an appealing form:

$$\begin{aligned} \tilde{\beta}_i &= E(\beta_i \mid \mathbf{Y}_i) \\ &= \gamma_{i1} + \gamma_{i2}\hat{\alpha}_{i,ols} + \gamma_{i3}\hat{\beta}_{i,ols} \end{aligned} \quad (5)$$

where $\hat{\alpha}_{i,ols}$ and $\hat{\beta}_{i,ols}$ represent the ordinary least squares (OLS) intercept and slope from regressing $\mathbf{Y}_i$ on $\mathbf{t}_i$. The coefficients in (5) are given by:

$$\gamma_{i2} = (\sigma_{12} v_{\beta i} - \sigma_2^2 c_{\alpha\beta i})/\delta_i,$$

$$\gamma_{i3} = (\sigma_2^2 v_{\alpha i} - \sigma_{12} c_{\alpha\beta i})/\delta_i,$$

and

$$\gamma_{i1} = \beta(1 - \gamma_{i3}) - \alpha\gamma_{i2},$$

with

$$\delta_i = (v_{\alpha i} v_{\beta i} - c_{\alpha\beta i}^2),$$

$$v_{\alpha i} = \mathrm{Var}(\hat{\alpha}_{i,ols}) = \sigma_1^2 + \sigma^2[1/n_i + \bar{t}_i^2/\{(n_i - 1)s_{ti}^2\}],$$

$$v_{\beta i} = \mathrm{Var}(\hat{\beta}_{i,ols}) = \sigma_2^2 + \sigma^2/\{(n_i - 1)s_{ti}^2\},$$

$$c_{\alpha\beta i} = \mathrm{Cov}(\hat{\alpha}_{i,ols}, \hat{\beta}_{i,ols}) = \sigma_{12} - \bar{t}_i\sigma^2/\{(n_i - 1)s_{ti}^2\},$$

and where $\bar{t}_i$ and s_{ti}^2 denote the sample mean and variance of the observation times $\mathbf{t}_i = (t_{i1}, ..., t_{in_i})'$. Similarly, it can be shown that

$$\tilde{\alpha}_i = E(\alpha_i \mid \mathbf{Y}_i) = \tau_{i1} + \tau_{i2}\hat{\alpha}_{i,ols} + \tau_{i3}\hat{\beta}_{i,ols} \quad (6)$$

with

$$\tau_{i2} = (\sigma_1^2 v_{\beta i} - \sigma_{12} c_{\alpha\beta i})/\delta_i,$$

$$\tau_{i3} = (\sigma_{12} v_{\alpha i} - \sigma_1^2 c_{\alpha\beta i})/\delta_i,$$

and

$$\tau_{i1} = \alpha(1 - \tau_{i2}) - \beta\tau_{i3}.$$

Consider the problem of predicting the unknown response under model (3) for subject i at some clinically or otherwise significant point in time (t_i^*). In other words, seeking to predict the value of

$$Y_{it}^* = E(Y_{ij} \mid \alpha_i, \beta_i, t_{ij} = t_i^*) = \alpha_i + \beta_i t_i^*.$$

The posterior mean of Y_{it}^* is

$$\tilde{Y}_{it}^* = E(Y_{it}^* \mid \mathbf{Y}_i) = \tilde{\alpha}_i + \tilde{\beta}_i t_i^* \quad (7)$$

where $\tilde{\beta}_i$ and $\tilde{\alpha}_i$ are as defined in (5) and (6), for $n_i \geq 2$. EBLUPs for $\tilde{\beta}_i$ and $\tilde{\alpha}_i$ are obtained by inserting parameter estimates into the general expressions for $E(\beta_i \mid \mathbf{Y}_i)$ and $E(\alpha_i \mid \mathbf{Y}_i)$, where $n_i = 1$ is permissible. The EBLUP for Y_{it}^* inserts the EBLUPs for $\tilde{\beta}_i$ and $\tilde{\alpha}_i$ into (7).

Constrained Bayes Predictions

The constrained Bayes (CB) approach (Louis, 1984) was extended by Ghosh (1992)

into a flexible paradigm. Lyles and Xu (1999) suggested that this general idea provides a natural alternative to the EBLUP in the mixed linear models context when overshrinkage could detract from the desired application of predicted values. They applied a slight adaptation of the CB concept under models (1) and (3) by minimizing prediction mean squared error (MSEP) among unbiased candidates whose variances match that of the assumed random effects distribution. While this necessarily results in some sacrifice in overall MSEP relative to the posterior mean, it provides a set of predictions that more faithfully reproduce the underlying distribution of interest and are less likely to under-represent the extremeness of experimental units in the tails.

Under model (1), the CB predictor for μ_i recommended by Lyles and Xu is obtained directly by forcing the first two moments of the $\tilde{\mu}_i$ and μ_i distributions to match:

$$\tilde{\mu}_{i,LX} = \sqrt{v_i}\,\bar{y}_i + (1-\sqrt{v_i})\,\mu \qquad (8)$$

The square root is indicative of the reduction in shrinkage relative to the posterior mean in (2). Under model (3), use of a Lagrangian multiplier to enforce equality of the second moments while minimizing MSEP yields a constrained Bayes alternative to the posterior mean in (5):

$$\tilde{\beta}_{i,LX} = \gamma_{i1} + \gamma_{i2}\hat{\alpha}_{i,ols} + \gamma_{i3}\hat{\beta}_{i,ols} \qquad (9)$$

The coefficients in (9) are defined as

$$\gamma_{i1} = \beta(1-\gamma_{i3}) - \alpha\gamma_{i2}\,,$$

$$\gamma_{i2} = \pm\eta_i[\sigma_2^2/\{v_{\beta i} + \eta_i(2c_{\alpha\beta i} + \eta_i v_{\alpha i})\}]^{1/2},$$

and

$$\gamma_{i3} = \gamma_{i2}/\eta_i\,,$$

where

$$\eta_i = (v_{\beta i}\sigma_{12} - \sigma_2^2 c_{\alpha\beta i})(v_{\alpha i}\sigma_2^2 - \sigma_{12}c_{\alpha\beta i})^{-1}.$$

The ± sign in front of γ_{i2} is needed because there are two roots, although the positive root is usually correct. The positive or negative root is taken for γ_{i2} depending on which yields the lower value of the MSEP criterion:

$$\begin{aligned} MSEP &= E(\tilde{\beta}_i - \beta_i)^2 \\ &= (\gamma_{i2}^2 v_{\alpha i} + \gamma_{i3}^2 v_{\beta i} + 2\gamma_{i2}\gamma_{i3}c_{\alpha\beta i}) - 2(\gamma_{i2}\sigma_{12} + \gamma_{i3}\sigma_2^2) + \sigma_2^2 \end{aligned} \qquad (10)$$

The definitions of η_i and γ_{i2} serve to correct a subtle error in the result originally put forth by Lyles and Xu (1999). The Appendix provides analogous constrained Bayes predictors for α_i and Y_{it}^*, which are both new to the literature. Empirical constrained Bayes (ECB) predictions are obtained for practical use by replacing unknown parameters by their estimates in equations (8), (9), (A1), and (A3), and when calculating the MSEP criterion in (10).

In contrast to the preceding direct model-specific CB predictors, consider the general CB paradigm provided by Ghosh (1992). Using β_i under model (3) to illustrate, $\tilde{\beta}_{i,B}$ is first taken to indicate the posterior mean (or Bayes) predictor for subject i. An algebraic expression for $\tilde{\beta}_{i,B}$ was given in (5). Ghosh's approach defines the CB estimate ($\tilde{\beta}_{i,G}$) as follows:

$$\tilde{\beta}_{i,G} = w\tilde{\beta}_{i,B} + (1-w)\,\bar{\tilde{\beta}}_B \qquad (11)$$

where

$$\bar{\tilde{\beta}}_B = k^{-1}\sum_{h=1}^{k}\tilde{\beta}_{h,B}\,,\ w = (1 + H_1/H_2)^{1/2},$$

$$H_2 = \sum_{h=1}^{k}(\tilde{\beta}_{h,B} - \bar{\tilde{\beta}}_B)^2\,,$$

and

$$H_1 = tr\{Var(\boldsymbol{\beta} - \bar{\beta}\mathbf{1}_k \mid \mathbf{Y})\} = (1-k^{-1})\sum_{h=1}^{k} Var(\beta_h \mid \mathbf{Y}_h) \qquad (12)$$

with $\boldsymbol{\beta}$ representing the k-vector $(\beta_1, \beta_2, ..., \beta_k)'$.

The latter equality is supplied in (12) as a result of assumed independence across experimental units for the class of mixed models under consideration here. Note that in addition to the posterior means, this paradigm requires only the corresponding posterior variances. Using the previous notation (see equation (4) and Appendix), results in:

$$\mathrm{Var}(\beta_i \mid \mathbf{Y}_i) = \sigma_2^2 - \{\sigma_{12}\mathbf{1}'_{n_i} + \sigma_2^2\mathbf{t}'_i\}\boldsymbol{\Sigma}_i^{-1}\{\sigma_{12}\mathbf{1}'_{n_i} + \sigma_2^2\mathbf{t}'_i\}', \quad (13)$$

$$\mathrm{Var}(\alpha_i \mid \mathbf{Y}_i) = \sigma_1^2 - \{\sigma_1^2\mathbf{1}'_{n_i} + \sigma_{12}\mathbf{t}'_i\}\boldsymbol{\Sigma}_i^{-1}\{\sigma_1^2\mathbf{1}'_{n_i} + \sigma_{12}\mathbf{t}'_i\}', \quad (14)$$

and

$$\mathrm{Var}(Y^*_{it} \mid \mathbf{Y}_i) = \mathrm{Var}(Y^*_{it}) - \{\psi_{i1}\mathbf{1}'_{n_i} + \psi_{i2}\mathbf{t}'_i\}\boldsymbol{\Sigma}_i^{-1}\{\psi_{i1}\mathbf{1}'_{n_i} + \psi_{i2}\mathbf{t}'_i\}'. \quad (15)$$

ECB predictions for practical use can be obtained by replacing unknown parameters by their estimates when computing the posterior means and variances, and the building blocks for these calculations are already built into standard software for mixed linear models.

Incorporating Fixed or Time-Dependent Covariates

Consider the following extensions of models (1) and (3) to include a set of T covariates, some of which may be time-dependent:

$$Y_{ij} = \mu + b_i + \sum_{t=1}^{T}\theta_t c_{ijt} + e_{ij} \quad (16)$$

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)\,t_{ij} + \sum_{t=1}^{T}\theta_t c_{ijt} + e_{ij} \quad (17)$$

where c_{ijt} represents the observed value of the t^{th} covariate for subject i at time point j (t=1,..,T; i=1,..,k; j=1,..,n_i). Let $\mathbf{c}'_{ij} = (c_{ij1}, c_{ij2}, ..., c_{ijT})$ and form the $n_i \times T$ matrix $\mathbf{C}_i$ by stacking the row vectors $\mathbf{c}'_{ij}$ in order. Next, define the transformed observed data vector $\mathbf{y}^{\bullet}_i = \mathbf{y}_i - \mathbf{C}_i\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_T)'$. The extension to the posterior mean formula in (2) is

$$\tilde{\mu}_i = \mathrm{E}(\mu_i \mid \mathbf{Y}_i, \mathbf{C}_i) = \nu_i\bar{y}^{\bullet}_i + (1-\nu_i)\mu \quad (18)$$

with μ_i and ν_i defined exactly as before and $\bar{y}^{\bullet}_i = n_i^{-1}\sum_{j=1}^{n_i} y^{\bullet}_{ij}$. In practice, predicting $\tilde{Y}_{ij} = \mathrm{E}(Y_{ij} \mid b_i, \mathbf{C}_i) = \mu_i + \mathbf{c}'_{ij}\boldsymbol{\theta}$ may be more likely. Standard mixed linear model software typically provides the EBLUP for b_i, from which EBLUPs for μ_i and $\tilde{Y}_{ij}$ are easily obtained.

Similarly, extensions to (4) and (5) under the randomized regression model (17) are

$$\tilde{\beta}_i = \mathrm{E}(\beta_i \mid \mathbf{Y}_i, \mathbf{C}_i) = \beta + (\sigma_{12}\mathbf{1}'_{n_i} + \sigma_2^2\mathbf{t}'_i)\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}^{\bullet}_i - \alpha\mathbf{1}_{n_i} - \beta\mathbf{t}_i)$$

and

$$\tilde{\beta}_i = \mathrm{E}(\beta_i \mid \mathbf{Y}_i, \mathbf{C}_i) = \gamma_{i1} + \gamma_{i2}\hat{\alpha}_{i,ols} + \gamma_{i3}\hat{\beta}_{i,ols} \quad (19)$$

where β_i, γ_{i1}, γ_{i2}, and γ_{i3} are defined as before, but with $\hat{\alpha}_{i,ols}$ and $\hat{\beta}_{i,ols}$ now representing the OLS intercept and slope from regressing $\mathbf{y}^{\bullet}_i$ on $\mathbf{t}_i$. The algebraic expression in (19) requires $n_i \geq 2$. Standard software typically provides EBLUPs for a_i and b_i, from which EBLUPs for α_i and β_i follow directly. In turn, the analogue to equation (7) becomes

$$\tilde{Y}^*_{it} = \mathrm{E}(Y^*_{it} \mid \mathbf{Y}_i, \mathbf{C}_i) = \tilde{\alpha}_i + \tilde{\beta}_i t^*_i + \mathbf{c}_{i,t^*}\boldsymbol{\theta} \quad (20)$$

which can arguably be defined only for non-time-dependent covariates unless the values of

any time dependent ones are known at time t_i^* (as indicated by the notation $\mathbf{c}_{i,t^*}$).

Extensions of the CB predictors $\tilde{\mu}_{i,LX}, \tilde{\beta}_{i,LX}$, and $\tilde{\alpha}_{i,LX}$ in equations (8), (9), and (A1) with covariate adjustment according to models (16) and (17) require no changes to the coefficients already given, once the transformation $\mathbf{y}_i^{\bullet} = \mathbf{y}_i - \mathbf{C}_i\boldsymbol{\theta}$ is made. The same is true for $\tilde{Y}^*_{it,LX}$ in equation (A3), except the term $\mathbf{c}_{i,t^*}\boldsymbol{\theta}$ is added as in (20). ECB predictions for practical use follow, once estimates of the mixed linear model parameters are inserted.

In adapting the paradigm of Ghosh (1992) as in (11) and (12), ECB predictions appear straightforward for a broad class of general linear mixed models because (i) EBLUPs accounting for covariates come directly out of standard software, and (ii) the required conditional variances [e.g., (13)-(15)] are unchanged by the addition of covariates. In the case of $\tilde{Y}^*_{it}$, Ghosh's paradigm requires a separate application of posterior mean and variance calculations analogous to those in (11) and (12) for each unique value of t_i^* (Moore, 2006).

Example

Consider longitudinal data on CD4 cell counts collected for the Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted (P^2C^2) HIV Infection Study (The P^2C^2 Study Group, 1996). This National Heart, Lung, and Blood Institute-funded study enrolled infants born to HIV-positive women during the years 1990-1993, and followed them prospectively during the first few years of life. Specifically, data was analyzed on 59 vertically infected infants who contributed a total of 539 CD4 counts over time, with the number of measurements per child ranging from 3 to 19. Initial CD4 counts were typically observed at or within a few weeks of birth. The length of follow-up on children ranged from 1 to 6 years, with a median of 3.5 years. Also recorded for each child was the age at which he or she was determined to have reached Class A (mildly symptomatic) HIV status (Centers for Disease Control and Prevention, 1994). Across the 59 subjects, this age ranged from 0.4 to 16 months.

A mixed linear model was fit to these data, with age as the longitudinal metameter. While there was some indication of right skewness in the CD4 counts, standard transformations tended to overcorrect this and for the sake of clarity the untransformed CD4 counts were analyzed. For an illustration with covariate adjustment, the child's gender (1 for male, 0 for female) and the concurrent CD8 cell count were accounted for via the following model:

$$CD4_{ij} = (\alpha + a_i) + (\beta + b_i)AGE_{ij} + \theta_1 GENDER_i + \theta_2 CD8_{ij} + e_{ij} \quad (21)$$

The primary objective was to compare EBLUP and ECB predictions of the random intercepts ($\alpha_i = \alpha + a_i$) and random slopes ($\beta_i = \beta + b_i$). For this purpose, both the direct ECB approach patterned after Lyles and Xu (1999; 'LX ECB') and the general ECB method following Ghosh (1992) were investigated.

Next, EBLUP and Ghosh ECB predictions of Y^*_{it} were compared, where $Y^*_{it} = \alpha_i + \beta_i t_i^* + \theta_1 GENDER_i + \theta_2 CD8_i$ represents the unknown model-based CD4 count at time t_i^*. For this latter purpose, t_i^* was defined as the age at which the child was diagnosed with Class A HIV disease, and model (21) was re-fit with the initial CD8 count ($CD8_i$) in place of the time-dependent version in light of the fact that CD8 was unrecorded at the times t_i^*. Table 1 provides the coefficient and variance component estimates from fitting both versions of model (21) by maximum likelihood via SAS PROC MIXED (SAS Institute, Inc., 2004a). The table indicates a highly significant average decline of approximately 400 CD4 cells per year, little effect of gender, and a significant positive association with the CD8 count, regardless of

whether the latter was measured only initially or treated as time-dependent.

In Figure 1A, EBLUPs are plotted for the random intercepts α_i against the corresponding Ghosh ECB predictions, based on the model treating CD8 as time-dependent. The EBLUPs were obtained directly from the mixed linear model software, and the Ghosh ECBs were computed readily using the EBLUPs and posterior variance calculations with variance components replaced by their MLEs (see e.g., eqns. 11-15). The reduction in shrinkage afforded by the CB method is evidenced by the characteristic tilting in the pattern of plotted points.

Figure 1B plots the LX ECB predictions of α_i versus the Ghosh ECBs. To obtain the LX ECBs, the MLEs for variance components were inserted into the formulae provided herein, with covariate adjustment as described in Section 2.3. With a few exceptions, the two approaches produce essentially identical results. The sample means of the 59 EBLUP, Ghosh ECB, and LX ECB predicted values were 1675.5, 1675.5, and 1675.3, respectively. The corresponding sample variances were 365470, 475026, and 473752. Comparing these to $\hat{\alpha} = 1675.5$ and $\hat{\sigma}_1^2 = 468832$ (Table 1) highlights the moment matching characteristics of the CB approaches, as well as the overshrinkage of the EBLUP.

Figure 2 is the counterpart to Figure 1, for the predicted random slopes (β_i). The tilting remains prominent in Figure 2A, while Figure 2B reveals somewhat more pronounced discrepancies between the Ghosh and LX ECB point predictions than in the case of the intercepts. The sample means of the EBLUP, Ghosh ECB, and LX ECB predicted values were −388.2, −388.2, and −395.3, respectively, with sample variances of 27904, 48316, and 49401. Comparing these to $\hat{\beta} = -388.2$ and $\hat{\sigma}_2^2 = 47843$ (Table 1) again highlights the ECB moment-matching properties in action.

Figure 3 illustrates the reduction in shrinkage of the Ghosh ECB predictions (open circles) of CD4 cell counts at the time of Class A disease (Y_{it}^*), relative to the EBLUPs (closed circles). Separate plots are presented for females and males, with overlays of the population average regression lines calculated at the overall mean of the 59 initial CD8 counts (1294.7 cells). The lines provide a relevant visual reference based on the fit of model (21) (Table 1), although the plotted points were not expected to directly follow these linear trends given that subjects with less rapidly declining CD4 counts theoretically reach Class A disease at later ages.

Results

While the close agreement of the sample means and variances of the ECB predictions to the corresponding estimated moments ($\hat{\alpha}$ and $\hat{\sigma}_1^2$, $\hat{\beta}$ and $\hat{\sigma}_2^2$) in the real-data example is indicative, simulation studies are required to further assess the quality of the variance match and to compare the performances of the Ghosh and LX ECB methods in practical settings. Several combinations of covariates and true parameter values were examined and qualitatively similar results were found. In the interest of brevity and relevance to the application presented in the previous section, simulations designed to mimic the conditions observed in the example are summarized. Simulations were carried out using matrix manipulations and standard random number generating functions available in the SAS IML package (SAS Institute, Inc., 2004b).

Performance comparison: LX vs. Ghosh CB predictors

Data was generated according to model (21) for 20,000 hypothetical subjects, with true parameter values equal to the estimates listed in the top half of Table 1. The fabricated CD4 data were unbalanced with n_i ranging randomly between 2 and 10, and measurements were unequally timed over approximate 2 month intervals. Simulated subjects were male or female with probability 0.5. For simplicity, time-varying CD8 counts were generated at each visit from a normal distribution mimicking the sample mean and variance of the initial CD8 counts in the actual example. To illustrate results for predicting Y_{it}^*, the same simulation exercise was repeated except with a time independent

Table 1: Summary of mixed linear models fit to CD4 cell count data *

Model †	Coefficient	Estimate (standard error)	Variance Component	Estimate
CD8 as time-dependent	α	1675.50 (138.27)	σ_1^2	468832
	β	−388.17 (38.06)	σ_2^2	47843
	θ_1	−163.41 (146.61)	σ_{12}	−103226
	θ_2	0.26 (0.03)	σ^2	477810
CD8 as time-independent (initial value)	α	1735.88 (188.60)	σ_1^2	429957
	β	−417.51 (40.57)	σ_2^2	55206
	θ_1	−105.28 (146.61)	σ_{12}	−102537
	θ_2	0.27 (0.10)	σ^2	529062

* Data from P^2C^2 HIV Infection Study (The P^2C^2 Study Group, 1996)

† $CD4_{ij} = (\alpha + a_i) + (\beta + b_i)\,AGE_{ij} + \theta_1 GENDER_i + \theta_2 CD8 + e_{ij}$

Figure 1: EBLUP (panel A) and LX ECB (panel B) vs. Ghosh ECB predictions for random intercepts (α_i) based on the fit of model (21) with CD8 count as time-dependent

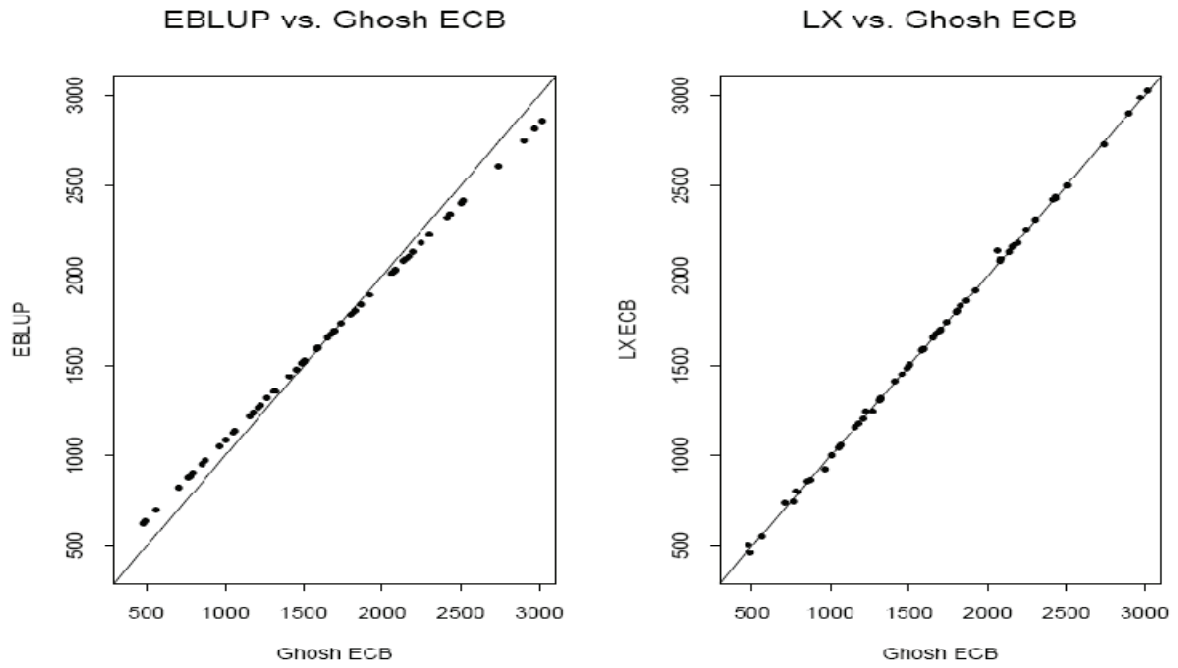


Figure 2: EBLUP (panel A) and LX ECB (panel B) vs. Ghosh ECB Predictions for Random Slopes (β_i) Based on the Fit of Model (21) with CD8 Count as Time-Dependent

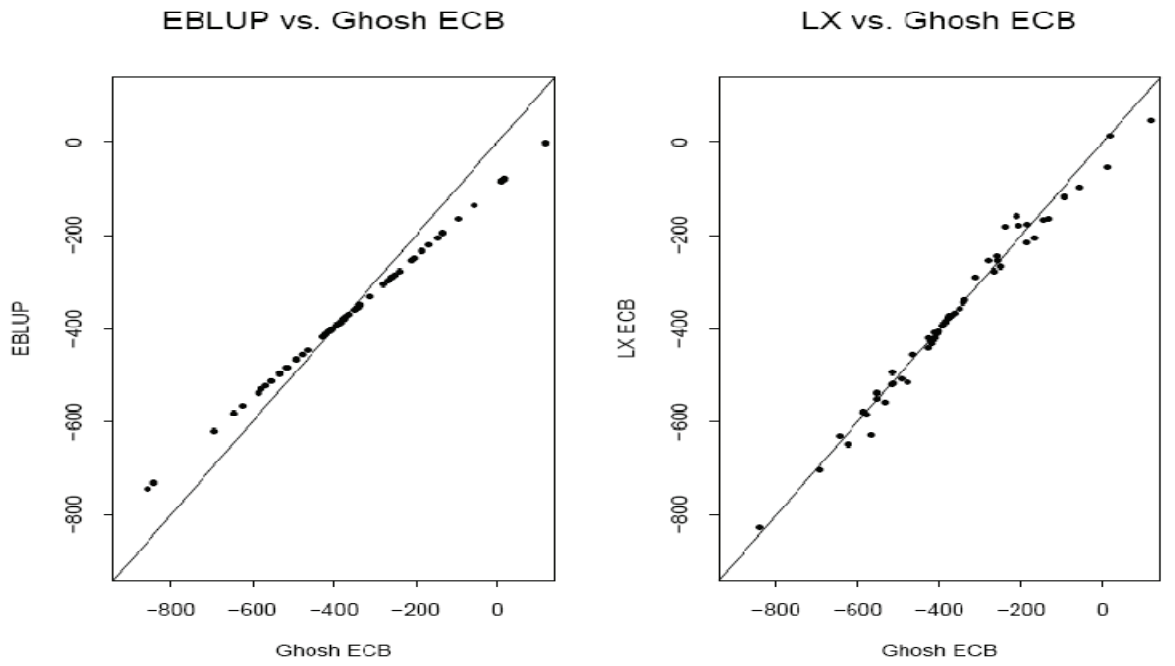
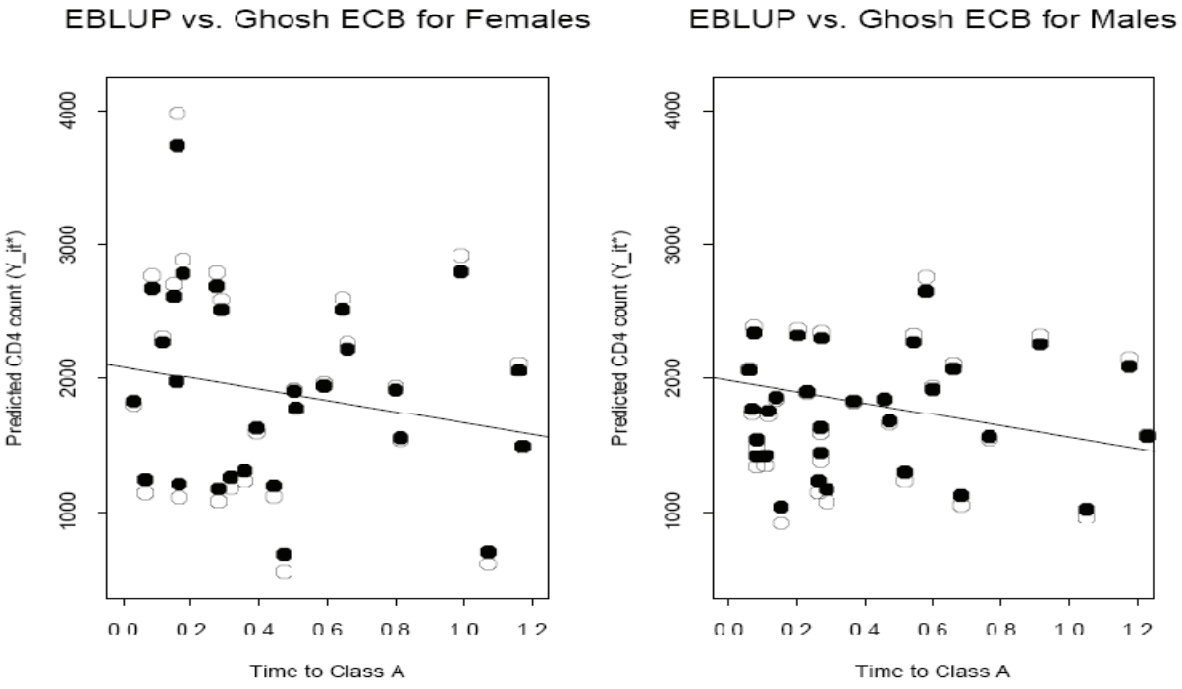


Figure 3: EBLUP (dark circle) vs. Ghosh ECB (open circle) Predictions of $Y_{it}^{*} = \alpha_i + \beta_i t_i^{*} + \theta_1 \mathrm{GENDER}_i + \theta_2 \mathrm{CD8}_i$ for Females (panel A) and Males (panel B), with Initial CD8 Count as a Time-Independent Covariate



initial CD8 count in place of the time-varying version. The time point of interest (t_i^*) was taken to occur at 2 years for each simulated subject.

Table 2 summarizes the simulation results for predicting the α_i's and β_i's, and Table 3 summarizes the results for predicting Y_{it}^*. In each case, the sample means of the BLUPs and the two CB predictors closely match the true mean of the random variable being predicted.

The sample variances over 20,000 simulated subjects for both the LX and Ghosh CB methods are very close to the corresponding true variances in each case, while the overshrinkage of the BLUPs is evident by their notably tighter sampling distributions. As a final note, the empirical prediction MSEs of the LX and Ghosh methods are similar, though predictably somewhat larger than those for the corresponding BLUPs. In each case, the Ghosh method achieved a small MSE advantage relative to the LX approach.

Table 2: Simulation Results for Random Intercept and Slope Predictions*†

	True α_i's	$\widetilde{\alpha}_{i,BLUP}$	$\widetilde{\alpha}_{i,LX}$	$\widetilde{\alpha}_{i,G}$
Mean	1675.5	1680.8	1681.2	1680.8
Variance	468832	376252	475834	474736
Prediction MSE	--	98600	105400	104469
	True β_i's	$\widetilde{\beta}_{i,BLUP}$	$\widetilde{\beta}_{i,LX}$	$\widetilde{\beta}_{i,G}$
Mean	–388.2	–389.4	–386.2	–389.4
Variance	47843	16115	48693	48134
Prediction MSE	--	31593	40375	40125

*Data simulated to mimic model (21) with parameters equal to estimates in Table 1 (top)
†Predictions computed assuming parameter values that generated the data

Table 3: Simulation Results for Y_{it}^* Predictions*†

	True Y_{it}^*'s	$\widetilde{Y}_{it,BLUP}^*$	$\widetilde{Y}_{it,LX}^*$	$\widetilde{Y}_{it,G}^*$
Mean	1156.4	1158.4	1158.3	1158.4
Variance	289054	177184	289249	288880
Prediction MSE	--	110636	128884	124112

*Data simulated to mimic model (21) with parameters equal to estimates in Table 1 (bottom)
†Predictions computed assuming parameter values that generated the data

Flexibility of Ghosh's Approach under More General Covariance Structures

The LX approach, while presentable in closed form for the models considered thus far, relies upon a strict form for candidate predictors and may be cumbersome or infeasible to extend to arbitrary mixed linear models. For example, consider an extension of model (17) to incorporate serially correlated random errors, e.g., via an AR(1) structure. Rather than $\sigma^2 \mathbf{I}_{n_i}$, the covariance matrix of the i^{th} vector of random errors $(\mathbf{e}_i)$ now takes the form

$$\mathrm{Var}(\mathbf{e}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ & 1 & \rho & \dots & \rho^{n_i-2} \\ & & 1 & \rho & . \\ & & & & . \\ & & & & . \\ & & & & 1 \end{pmatrix} = \sigma^2 \mathbf{P}_{AR(1)}$$

The structured error covariance makes it less reasonable to restrict to the class of predictors that are linear combinations of $\hat{\alpha}_{i,ols}$ and $\hat{\beta}_{i,ols}$ [see eqn. (5)] in order to develop a CB predictor via the LX approach. Further, the MSEP becomes a much more difficult objective function to work with analytically.

Fortunately, the general paradigm of Ghosh (1992) encounters no difficulty with such an extension. In particular, the EBLUP remains available via common mixed linear model software, and the MVN theory-based posterior variance remains straightforward, with the only adjustment necessary to equations (13) and (14) being that the matrix $\mathbf{\Sigma}_i = \mathrm{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{\Delta} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}$ becomes $\mathbf{\Sigma}_i = \mathbf{Z}_i \mathbf{\Delta} \mathbf{Z}_i' + \sigma^2 \mathbf{P}_{AR(1)}$.

Table 4 displays the results of an additional simulation under the AR(1) error model. Data were generated under model (21) using the same true parameter values as for the simulation summarized in the top half of Table I, except with an AR(1) error structure for the covariance matrix of the random errors. The value ρ=0.30 was assumed. There were 5,000 simulated subjects, each with n_i=8 observations. The model was fit via SAS PROC MIXED and the ECB versions of $\tilde{\alpha}_{i,G}$ and $\tilde{\beta}_{i,G}$ were computed as in (11) and (12), by incorporating the EBLUPs produced by the software together with the estimated posterior variances as in (13) and (14).

As Table 4 shows, excellent matches were achieved between the sample means and variances of the ECB predictions, and the corresponding estimated population moments $(\alpha, \beta, \sigma_1^2, \sigma_2^2)$. Figure 4 displays histograms of the ECBs, which almost perfectly match the overlaid estimated theoretical normal distributions. In contrast, histograms of the EBLUPs (not shown) are characterized by markedly narrow spread as expected, thus dramatically failing to match the underlying theoretical distribution. Potential drawbacks of this overshrinkage in certain applications have been discussed at length in the literature (e.g., Louis, 1984; Ghosh, 1992; Shen & Louis, 1998; Stern & Cressie, 1999). The current example further highlights the flexibility of the Ghosh paradigm as a general approach to ECB prediction under the mixed linear model.

Conclusion

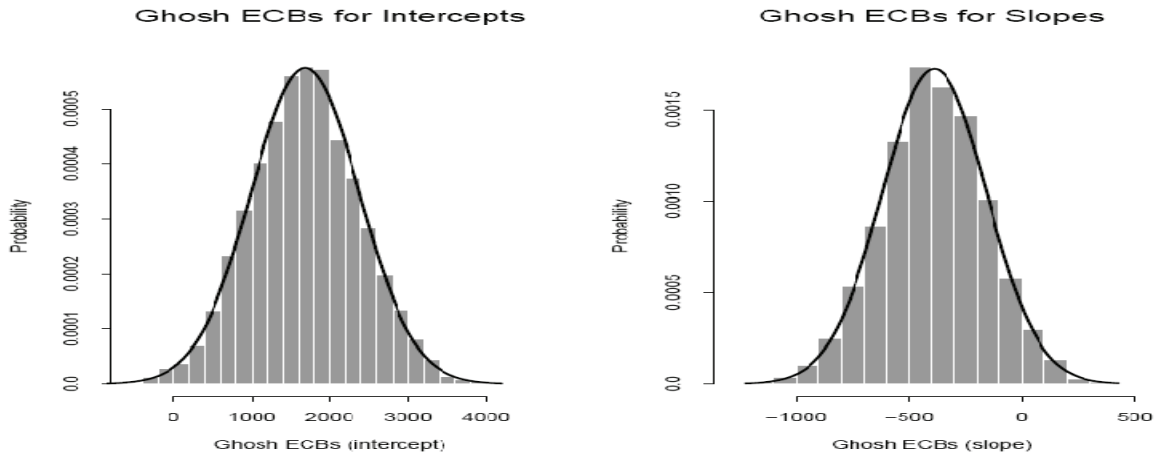
Louis (1984) and Ghosh (1992) discussed the motivation and potential benefits of constrained Bayes estimation, which seeks to optimize a traditional MSE criterion subject to matching the posterior expectation of the first two moments of a parameter distribution to the corresponding true moments. In particular, the known overall MSE advantage of the traditional posterior mean approach (which underlies the BLUP in the mixed linear model setting) is sometimes worth sacrificing to obtain a set of predictions with a histogram more closely matching a true distribution of random effects. For specific discussions of contexts in which constrained Bayes and related approaches offer tangible appeal, see Shen and Louis (1998), Lyles and Xu (1999), Stern and Cressie (1999), and Lyles, et al. (2007).

Table 4: Simulation Results for Random Intercept and Slope Predictions Under AR(1) Error Model*†

Parameter Estimates‡	ECB Sample Moments	$\tilde{\alpha}_{i,G}$	$\tilde{\beta}_{i,G}$
$\hat{\alpha} = 1683.04$ $\hat{\sigma}_1^2 = 481899$	Mean	1683.04	–389.21
$\hat{\beta} = -389.21$ $\hat{\sigma}_2^2 = 53551$	Variance	481961	53556

*Data simulated to mimic model (21) with k=5000, n_i=8 ($\forall$ i), true parameters set equal to estimates in Table 1 (top), and ρ=0.30
†Ghosh ECB predictions computed by inserting MLEs of parameters
‡MLEs; Other parameter estimates: $\hat{\theta}_1 = -172.30, \hat{\theta}_2 = 0.23, \hat{\sigma}_{12} = -112073,$ $\hat{\sigma}^2 = 510618, \hat{\rho} = 0.29$

Figure 4: ECB Histograms Using Simulated Data from AR(1) Model (Table IV)



The purpose of this article has been to outline and compare in detail the application of a direct (LX) CB approach considered by Lyles and Xu (1999) for certain mixed linear models, as opposed to the general method of Ghosh (1992). Both approaches were explored in the presence of covariates (possibly time-dependent), and it was concluded based on simulations and a real-data example that both may be effectively applied to achieve the moment-matching goals of the CB paradigm.

The LX approach, while presentable in closed form for the models considered herein, relies upon a strict form for candidate predictors and may not be straightforward to extend to arbitrary mixed linear models. However, as highlighted previously, the general method of Ghosh (1992) appears remarkably flexible and consistent in its application. In practice, it requires only EBLUPs and estimates of the posterior variances of the random effects being predicted, with the latter readily obtainable under normal-theory mixed models. It thus seems natural to compare the performance of the Ghosh method versus the LX approach in mixed model settings where the latter is available. The simulation studies summarized (and others, unreported) consistently show the Ghosh approach to be as effective as the direct LX method at matching moments, and also suggest slight prediction MSE gains via its use for unbalanced data.

Because the primary aim was to serve as proponents of the ECB approach under the mixed linear model, the results of the current study are encouraging. The CB paradigm of Ghosh (1992) relies on building blocks that are available in commercial software for mixed linear models (e.g., SAS PROC MIXED and similar procedures in other packages such as Splus, R, SPSS, STATA or BMDP). It was shown that it performs well relative to a direct, but far less flexible, CB approach developed expressly for mixed linear models. Although further assessments will be necessary, it is hoped that these results will encourage software developers to consider the possible inclusion of options to produce the Ghosh ECB predictions in future releases. This software advance would be welcome, for the purpose of allowing practitioners the freedom to select a validated alternative to the traditional EBLUP when overshrinkage could run counter to the objective at hand.

Acknowledgements

R. H. L. and A. K. M. were supported in part by an R01 from the National Institute of Environmental Health Sciences (ES012458). We thank the P^2C^2HIV study investigators and the National Heart, Lung and Blood Institute for use of their database, and appreciate support provided by the Biostatistics Core of the Emory Center for AIDS Research (P30 AI050409).

References

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. NY: Oxford University Press.

Efron, B., & Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators, part I: the Bayes case. *Journal of the American Statistical Association*, *66*, 807-815.

Centers for Disease Control and Prevention. (1994). 1994 revised classification for human immunodeficiency virus infection in children less than 13 years of age. *Morbidity and Mortality Weekly Report*, *43*, 1-10.

Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, *87*, 533-539.

Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963-974.

Littell, R. C., et al. (2006). *SAS for mixed models, 2nd Edition*. Cary, NC: SAS Institute, Inc.

Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, *79*, 393-398.

Lyles, R. H. & Xu, J. (1999). Classifying individuals based on predictors of random effects. *Statistics in Medicine*, *18*, 35-52.

Lyles, R. H., Manatunga, A. K., Moore, R. H., Bowman, F. D., & Cook, C. (2007). Improving point predictions of random effects for subjects at high risk. *Statistics in Medicine*, *26*, 1285-1300.

Moore, R.H., 2006. Prediction of random effects when data are subject to a detection limit. Unpublished Ph.D. dissertation, Department of Biostatistics, Emory University.

P^2C^2 HIV Study Group. (1996). The pediatric pulmonary and cardiovascular complications of vertically transmitted human immunodeficiency virus (P^2C^2 HIV) infection study: Design and methods. *Journal of Clinical Epidemiology*, *49*, 1285-1294.

SAS Institute, Inc. (2004a). *SAS/STAT 9.1 user's guide*. Cary, NC: SAS Institute, Inc.

SAS Institute, Inc. (2004b). *SAS/IML 9.1 user's guide*. Cary, NC: SAS Institute, Inc.

Searle, S.R., Casella, G., & McCullogh, C.E. (1992). Variance components. NY: Wiley.

Shen, W., & Louis, T.A. (1998). Triple-goal estimates in two-stage hierarchical models. *Journal of the Royal Statistical Society Series B*, *60*, 455-471.

Stern, H.S., & Cressie, N. (1999). Inference for extremes in disease mapping. In: A. Lawson, et al., (*Eds*). *Disease mapping and risk assessment for public health*, 63-84. Chichester: Wiley.

Appendix

A constrained Bayes predictor for the i^{th} subject's random intercept (α_i) may be obtained via calculations similar to those leading to $\tilde{\beta}_{i,LX}$ in equation (8), as follows:

$$\tilde{\alpha}_{i,LX} = \tau_{i1} + \tau_{i2}\hat{\alpha}_{i,ols} + \tau_{i3}\hat{\beta}_{i,ols}, \quad \text{(A1)}$$

where

$$\tau_{i1} = \alpha(1-\tau_{i2}) - \beta\tau_{i3},$$

$$\tau_{i2} = \pm[\sigma_1^2/\{v_{\alpha i} + \kappa_i(2c_{\alpha\beta i} + \kappa_i v_{\beta i})\}]^{1/2},$$

and

$$\tau_{i3} = \kappa_i\tau_{i2},$$

with

$$\kappa_i = (v_{\alpha i}\sigma_{12} - \sigma_1^2 c_{\alpha\beta i})(v_{\beta i}\sigma_1^2 - \sigma_{12}c_{\alpha\beta i})^{-1}.$$

Specifically, $\tilde{\alpha}_i$ defined in this way minimizes MSEP among predictors of the form (A1) subject to the constraints that $E(\tilde{\alpha}_i) = E(\alpha_i) = \alpha$ and $Var(\tilde{\alpha}_i) = Var(\alpha_i) = \sigma_1^2$, where the MSEP criterion is

$$E(\tilde{\alpha}_i - \alpha_i)^2 = (\tau_{i2}^2 v_{\alpha i} + \tau_{i3}^2 v_{\beta i} + 2\tau_{i2}\tau_{i3}c_{\alpha\beta i}) - 2(\tau_{i2}\sigma_1^2 + \tau_{i3}\sigma_{12}) + \sigma_1^2 \quad \text{(A2)}$$

In an analogous manner, constrained Bayes predictor for Y_{it}^* is defined as

$$\tilde{Y}_{it,LX}^* = \varphi_{i1} + \varphi_{i2}\hat{\alpha}_{i,ols} + \varphi_{i3}\hat{\beta}_{i,ols}, \quad \text{(A3)}$$

where

$$\varphi_{i1} = \alpha(1-\varphi_{i2}) - \beta(\varphi_{i3} - t_i^*),$$

$$\varphi_{i2} = \pm[\psi_{i3}/\{v_{\alpha i} + \omega_i(2c_{\alpha\beta i} + \omega_i v_{\beta i})\}]^{1/2},$$

and

$$\varphi_{i3} = \omega_i\varphi_{i2},$$

with

$$\omega_i = (v_{\alpha i}\psi_{i2} - \psi_{i1}c_{\alpha\beta i})(v_{\beta i}\psi_{i1} - \psi_{i2}c_{\alpha\beta i})^{-1},$$

$$\psi_{i1} = \sigma_1^2 + t_i^*\sigma_{12}, \quad \psi_{i2} = \sigma_{12} + t_i^*\sigma_2^2,$$

and

$$\psi_{i3} = \sigma_1^2 + t_i^{*2}\sigma_2^2 + 2t_i^*\sigma_{12}.$$

This minimizes MSEP for predictors of the form (A3), subject to the constraints

$$E(\tilde{Y}_{it}^*) = E(Y_{it}^*) = \alpha + \beta t_i^*$$

and

$$Var(\tilde{Y}_{it}^*) = Var(Y_{it}^*) = \varphi_{i2}^2 v_{\alpha i} + \varphi_{i3}^2 v_{\beta i} + 2\varphi_{i2}\varphi_{i3}c_{\alpha\beta i}$$

As with γ_{i2} in equation (9), technically the choice of the positive or negative root to define τ_{i2} and φ_{i2} should be based on which minimizes the corresponding MSEP criterion. However, it has been observed that the negative roots have never applied except in the case of γ_{i2}.