Published in final edited form as:

Stat Med. 2010 November 10; 29(25): 2656–2668. doi:10.1002/sim.4043.

Empirical constrained Bayes predictors accounting for non-detects among repeated measures

Reneé H. Moore^{a,*†}, Robert H. Lyles^b, and Amita K. Manatunga^b

^aDepartment of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, U.S.A.

^bDepartment of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA 30322, U.S.A.

Abstract

When the prediction of subject-specific random effects is of interest, constrained Bayes predictors (CB) have been shown to reduce the shrinkage of the widely accepted Bayes predictor while still maintaining desirable properties, such as optimizing mean-square error subsequent to matching the first two moments of the random effects of interest. However, occupational exposure and other epidemiologic (e.g. HIV) studies often present a further challenge because data may fall below the measuring instrument's limit of detection. Although methodology exists in the literature to compute Bayes estimates in the presence of non-detects (Bayes_{ND}), CB methodology has not been proposed in this setting. By combining methodologies for computing CBs and Bayes_{ND}, we introduce two novel CBs that accommodate an arbitrary number of observable and non-detectable measurements per subject. Based on application to real data sets (e.g. occupational exposure, HIV RNA) and simulation studies, these CB predictors are markedly superior to the Bayes predictor and to alternative predictors computed using *ad hoc* methods in terms of meeting the goal of matching the first two moments of the true random effects distribution.

Keywords

constrained Bayes; detection limits; left-censoring; mixed linear model; random effects; shrinkage

1. Introduction

When assessing occupational exposures on individual workers or characterizing RNA levels for individual HIV patients based on repeated measures, it is often advantageous to predict a true random mean exposure level for each worker or trajectory for each patient. A potential complication is the fact that exposure and laboratory assay data are often subject to a limit of detection (LOD), which should be accounted for when predicting subject-specific random effects. This article presents constrained Bayes predictors of subject-specific random effects that account for the fact that each subject in the study may have an arbitrary combination of observable and non-detectable exposure measurements.

In linear or non-linear mixed models, the posterior mean of the random effects given the data, sometimes termed the Bayes predictor, is well known to be optimal with respect to

unweighted or weighted squared-error loss [1–4]. For this reason, the posterior mean is generally the basis of the many approaches implemented for subject-specific prediction. For example, the estimated best linear unbiased predictor (BLUP), the most common method utilized to estimate the Bayes predictor in the mixed linear model context [3], is available in several statistical packages [5, 6]. However a well-known feature of the squared-error criterion underlying the Bayes predictor is its tendency to overshrink estimates toward a population mean. This overshrinkage makes the Bayes predictor subject to poor performance in specified regions (e.g. the tails) of the random effects distribution, and may compromise its performance as a classification tool [4, 7, 8].

Alternative predictors, commonly termed as constrained Bayes (CB) estimates, have been shown to reduce shrinkage in an appealing way. For example, by matching the sample mean and variance of the predictors to the posterior expected mean and variance of the random effects, the modified Bayes approach of Louis [2] and Ghosh [7] reduces the impact of overshrinkage in both tails of the random effects distribution. Shen and Louis [4] simultaneously focus on ranking, optimality of subject-specific estimates, and matching the histogram of the true parameter distribution. In the same spirit as Louis and Ghosh, Lyles and Xu [9] match a predictor's mean and variance with the assumed population mean and variance of the random effect based on normal-theory ANOVA and randomized regression models. An important feature of all of these CB methods is the fact that they optimize the traditional mean-squared error criterion, subject to the moment-matching constraints imposed.

Due to exposure values that fall below an LOD associated with the measuring instrument, there has been much statistical research into handling censored environmental and laboratory assay data. A common *ad hoc* approach is to substitute the LOD or a fraction of the LOD for all values below the limit [10]. Other authors have recommended implementing likelihood-based approaches via the expectation–maximization algorithm [11–13], Bayesian methods [14, 15], and Simplex and Marquardt algorithms [16]. Lyles *et al.* [17] maximize an integrated likelihood directly using quasi-Newton routines, to simultaneously handle problems of informative drop-out and left-censoring. Although these and other authors provide ample methodology to model censored longitudinal data, relatively less attention has been paid to the prediction of subject-specific random effects in settings where data are subject to a LOD.

Our objective is to provide a methodology to compute CB predictors when correlated continuous data are subject to non-detectable measurements (ND). Rather than seeking a full Bayesian analysis, we focus on the traditional mixed model context and aim to provide reduced-shrinkage alternatives to the Bayes predictor. In particular, the goal of the proposed CB predictors is for their distribution (i.e. mean and variance; parameter histogram) to match the assumed distribution of the true random effects. In Section 2, we introduce two different CB predictors (methods CB_{GND} and CB_{SM}) for log-scale mean exposure that take into account the fact that each individual may contribute an arbitrary number of ND. In Section 3, we apply this methodology to an occupational exposure data set and extend the CB_{GND} approach to the analysis of longitudinal HIV RNA data. In Section 4, we conduct simulation studies to evaluate the performance of the proposed CB predictors and compare them to *ad hoc* alternatives and to the estimated Bayes predictor (posterior mean). We close in Section 5 with a discussion that outlines related ongoing and future research.

2. Extension of constrained Bayes (CB) methodology to non-detectable data (ND) under the mixed model

2.1. Review of Bayes predictor with ND

The general linear mixed model, given below in equation (1), provides a powerful tool to estimate subject-specific random effects.

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\Psi} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \mathbf{b}_i \sim N(\mathbf{0}, \Delta), \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma), \quad (1)$$

where $\mathbf{Y}_i$ denotes the vector of repeated responses for the i th subject, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices for the fixed effects $\boldsymbol{\Psi}$ and random effects $\mathbf{b}_i$, respectively, and we assume independence of $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$. Owing to the fact that it minimizes the mean-square error [3], the Bayes predictor is the most commonly utilized basis for estimating subject-specific predictors of random effects. Based on (1), the Bayes predictor is: $\mathbf{b}_{i,B} = E(\mathbf{b}_i | \mathbf{Y}_i)$. Generally, each element of the random vector $\mathbf{b}_{i,B}$ is a weighted average of a population profile and a linear function of the observed data $\mathbf{y}_i$. Since in the Bayes context, the prior mean of the random effects is zero, substantial weight will often be allocated to the prior population profile. This phenomenon is commonly referred to as shrinkage [3, 18, 19].

Recognizing that many biomedical and epidemiological data are subject to ND, several authors (e.g. Lyles *et al.* [17]) have adapted the Bayes predictor to account for ND via:

$$\text{Bayes}_{\text{ND}} = E(\mathbf{b}_i | \mathbf{Y}_i^*) = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{b}_i f^*(\mathbf{Y}_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{f^*(\mathbf{Y}_i; \theta)}, \quad (2)$$

where $\mathbf{Y}_i^*$ is the n_i -vector of responses containing n_{i1} detectable values and $n_i - n_{i1}$ ND ($n_{i1} = 0, \dots, n_i$) based on the ND-adjusted likelihood for (1) given by

$$L(\theta; \mathbf{Y}^*) = \prod_{i=1}^k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^*(\mathbf{Y}_i | \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i, \quad (3)$$

where

$$f^*(\mathbf{Y}_i | \mathbf{b}_i) = \left\{ \prod_{j=1}^{n_{i1}} \mathbf{f}(\mathbf{y}_{ij} | \mathbf{b}_i) \right\} \left[\prod_{j=n_{i1}+1}^{n_i} F(\text{LOD}_{ij} | \mathbf{b}_i) \right]$$

and the asterisk (*) denotes that each subject may or may not have ND, θ represents the vector of parameters ($\boldsymbol{\Psi}$ and variance components), and LOD_{ij} is the detection limit corresponding to the (i, j) th non-detected measurement. In each of the motivating studies presented in this paper, the LOD is the same for all (i, j) and henceforth, we proceed with LOD in place of LOD_{ij} . Expression (3) can often be maximized numerically using built-in optimization routines that support user-specified likelihood functions in numerous statistical software packages. This expression can also be modified to accommodate right and/or interval censoring in addition to left censoring as presented here.

2.2. Method I: extending Ghosh's CB estimation to ND (CB_{GND})

The shrinkage property of the Bayes predictor motivates our work to develop CB predictors that adjust for ND. Louis [2] recognized the overshrinkage inherent in the Bayes estimate and modified it by producing estimates whose first two moments match the corresponding moments from the posterior histogram of normal means. Subsequently, Ghosh [7] provided a 'recipe' to generalize Louis' modified Bayes predictor for use with any distribution. Therefore, as long as one can compute the underlying components while properly accounting for censoring, the non-linearity imposed by the presence of ND is not a major deterrent when employing Ghosh's approach. We appeal to Ghosh's general methodology and introduce CB predictors that account for left censored data. We will denote these CB predictors as CB_{GND} (Ghosh's CB modified for ND).

There are three general steps required to compute CB_{GND}. Step one requires maximizing the ND-adjusted likelihood (3) to obtain the maximum likelihood estimates (MLEs) for the components of θ . In step two, we substitute the MLEs into the analytic expression for the Bayes predictor (2) to provide empirical Bayes estimates accounting for ND (Bayes_{ND}). Once we have computed the Bayes estimate (2), in step three, we compute CB_{GND} according to the general paradigm of Ghosh [7]. In particular, for a univariate random effect b_i ,

$$CB_{GND} = \omega E(b_i|Y_i^*) + (1 - \omega) \bar{E}(b_i|Y_i^*), \quad (4)$$

where

$$\begin{aligned} \bar{E}(b_i|Y_i^*) &= k^{-1} \sum_{i=1}^k E(b_i|Y_i^*), \\ \omega &\equiv \omega(\mathbf{y}) = [1 + \{H_1(\mathbf{y})/H_2(\mathbf{y})\}]^{1/2}, \\ H_1(\mathbf{y}) &= (1 - k^{-1}) \sum_{i=1}^k \text{Var}(b_i|Y_i^*) = (1 - k^{-1}) \sum_{i=1}^k \{E(b_i^2|Y_i^*) - [E(b_i|Y_i^*)]^2\} \end{aligned}$$

and

$$H_2(\mathbf{y}) = \sum_{i=1}^k \{E(b_i|Y_i^*) - \bar{E}(b_i|Y_i^*)\}^2.$$

The term $E(b_i^2|Y_i^*)$ is obtained by numerically evaluating an integral analogous to (2). In practice, we use the corresponding 'empirical' CB predictor obtained by replacing unknown parameters by their MLEs based on (3).

2.3. Method II: CB estimation via subject-specific maximization of a censored-data likelihood (CB_{SM})

Lyles and Xu [9] introduced a CB predictor based upon the idea of Louis' modified Bayes estimate. Instead of matching the posterior mean and variance, they provided a reduced-shrinkage predictor by matching its first two moments to the prior mean and variance, which correspond in the mixed linear model context to the true assumed moments of the random effects. To illustrate the methodology of Lyles and Xu, we utilize the random intercept only model as described by (1) but for which the only fixed effect is the overall mean (i.e. $\mathbf{X}_i = \mathbf{1}_{n_i}$ and $\Psi = \mu$, so that $E(\mathbf{Y}_i) = \mu \mathbf{1}_{n_i}$), i.e.

$$Y_{ij} = \mu + b_i + e_{ij}, i = 1, \dots, k, j = 1, \dots, n_i, \quad (5)$$

where the response Y_{ij} is the j th exposure measurement for the i th individual, and μ is the population mean. We make the typical assumptions that $b_i \stackrel{\text{iid}}{\sim} N(0, \sigma_b^2)$ and $e_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_\omega^2)$, with $\text{Cov}(b_i, e_{ij}) = 0$ for all (i, j) . Under (5), Lyles and Xu [9] compared predictors of the form

$$\tilde{\mu}_i = c_i \bar{Y}_i + d_i, \quad (6)$$

where $\bar{Y}_i = (n_i)^{-1} \sum_{j=1}^{n_i} Y_{ij}$.

For the Bayes predictor,

$$c_i = \gamma_i \quad \text{and} \quad d_i = (1 - \gamma_i)\mu, \quad (7)$$

where $\gamma_i = n_i \sigma_b^2 / (n_i \sigma_b^2 + \sigma_\omega^2)$ (e.g. Searle [3]).

In contrast, the CB predictor takes

$$c_i = \sqrt{\gamma_i} \quad \text{and} \quad d_i = (1 - \sqrt{\gamma_i})\mu. \quad (8)$$

This CB predictor is defined so that $E(\tilde{\mu}_i) = E(\mu_i) = \mu$ and $\text{Var}(\tilde{\mu}_i) = \text{Var}(\mu_i) = \sigma_b^2$, where $\mu_i = \mu + b_i$ is the i th subject's random mean. In practice, the parameters μ , σ_b^2 , and σ_ω^2 are replaced by their estimates to provide the corresponding empirical CB predictor. Note that Lyles and Xu [9] did not accommodate data subject to ND when defining this reduced-shrinkage predictor.

We use the idea behind this alternative CB proposal [(6) and (8)] to formulate a second CB prediction method that accounts for the presence of ND. Note that c_i in (8) can be re-written

as $c_i = \sigma_b / \sqrt{\text{Var}(\bar{Y}_i)}$, where $\text{Var}(\bar{Y}_i) = \sigma_b^2 + (\sigma_\omega^2 / n_i)$ in the absence of ND. Also note that $\text{Var}(\bar{Y}_i) = \sigma_b^2 + \text{Var}(\bar{Y}_i | b_i)$, and that without ND, $\bar{Y}_i$ is the MLE for μ_i given a fixed b_i .

When ND are present, we seek the analogue to $\bar{Y}_i$, i.e. the best estimator for μ_i given b_i fixed. Hence for the construction of CB_{SM} (the subscript denoting 'subject-specific maximization'), we replace $\bar{Y}_i$ by the MLE for the i th subject's mean, denoted as $\hat{\mu}_{i,\text{ML}}$. That is, we maximize the following likelihood separately for each subject ($i = 1, \dots, k$) based on n_i measurements:

$$L(\mu_i; \mathbf{Y}_i) = \left[\prod_{j=1}^{n_{i1}} f(y_{ij}; \mu_i, \sigma_\omega^2) \right] \{ \Pr(Y_{ij} < \text{LOD}; \mu_i, \sigma_\omega^2) \}^{n_{i2}}, \quad (9)$$

where $f(y_{ij}; \mu_i, \sigma_\omega^2)$ is the $N(\mu_i, \sigma_\omega^2)$ density with corresponding CDF such that $\Pr(Y_{ij} < \text{LOD}) = \Phi((\text{LOD} - \mu_i) / \sigma_\omega)$, the first n_{i1} measurements on subject i are detectable, and $n_{i2} = n_i - n_{i1}$ measurements are below the LOD. We maximize this left-censored normal data likelihood

numerically with respect to only μ_i , with σ_w^2 replaced by its MLE based on (3). Note that if a subject has no measurements below the LOD, then maximizing (9) results in $\hat{\mu}_{i,ML} = \bar{Y}_i$ and $\widehat{\text{Var}}(\hat{\mu}_{i,ML}|b_i) = \hat{\sigma}_w^2/n_i = \widehat{\text{Var}}(\bar{Y}_i|b_i)$. When ND are present, we compute $\widehat{\text{Var}}(\hat{\mu}_{i,ML}|b_i)$ by inverting the observed information matrix associated with (9). Thus with ND, we propose the following second approach to construct a CB predictor:

$$\text{CB}_{SM} = \hat{c}_i \hat{\mu}_{i,ML} + \hat{d}_i, \quad (10)$$

where $\hat{c}_i = \hat{\sigma}_b / \sqrt{\widehat{\text{Var}}(\hat{\mu}_{i,ML})}$ and $\hat{d}_i = \hat{\mu}(1 - \hat{c}_i)$, with $\widehat{\text{Var}}(\hat{\mu}_{i,ML}) = \hat{\sigma}_b^2 + \widehat{\text{Var}}(\hat{\mu}_{i,ML}|b_i)$. As with σ_w^2 , the parameters μ and σ_b^2 are replaced by their estimates based on maximizing (3).

3. Application to motivational studies

3.1. Application of CB_{GND} and CB_{SM} to occupational dust exposure data

A study was conducted in the Netherlands to investigate the effects of occupational dust exposure on respiratory and lung function among people working in animal feed production [20]. Obtaining subject-specific estimates of workers' mean occupational dust exposure levels is of interest to explore possible links with respiratory and lung function problems. Dust exposures of workers in eight different job categories were studied. Here we utilize 53 dust exposure measurements from a group of 12 facility operator workers. Each worker provided from 2–6 measurements, with the average number of measurements (n_i) per worker equal to 4.4. Due to limitations of the instrument used to measure dust levels, 8 of the 12 workers had at least one dust measurement below the LOD (0.2mg/m³) and 21 per cent of the 53 dust exposure levels were below the LOD.

To obtain an estimate of each worker's dust exposure level, we utilize the random intercept only model described by (5), a special case of (1). In addition to computing ND-adjusted subject-specific predictors for this occupational exposure data set, we also compute empirical CB predictors using *ad hoc* methods. As previously mentioned, a common expedient approach to handling ND is to substitute in the LOD or a fraction thereof in their place (e.g. Hornung and Reed, [10]). Using this *ad hoc* approach, one assumes the substituted ND measurements to be observed and therefore uses the following basic likelihood:

$$L(\mu, \sigma_b^2, \sigma_w^2; Y) = \prod_{i=1}^k \left[\int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(Y_{ij}|\mu_i) f(\mu_i) d\mu_i \right]. \quad (11)$$

Note that in the standard mixed model integrated likelihood (11), both the detectable measurements and ND contribute in the same way through $f(Y_{ij}|\mu_i)$. Only if each subject has all detectable values will the adjusted likelihood (3) reduce to (11). The 'naïve' MLEs corresponding to (11) are available in a closed form [3].

Given the right-skewness in the data, we log transformed the dust exposure measurements in the example data set into better satisfy normality conditions. Maximizing (11) for the occupational dust exposure data set yielded the following MLEs: $\hat{\mu}(\text{std error}) = -0.194$ (0.21), $\hat{\sigma}_b^2 = 0.336$, $\hat{\sigma}_w^2 = 0.803$, which are used in the computation of the '*ad hoc*' predictors of log-scale mean occupational dust exposure levels. In contrast, maximization of the ND-adjusted likelihood (3) yields $\hat{\mu} = -0.335$ (0.23), $\hat{\sigma}_b^2 = 0.326$, and $\hat{\sigma}_w^2 = 1.243$ (used in the

calculation of the empirical Bayes_{ND}, CB_{GND}, and CB_{SM} predictors). In comparison to the ND-adjusted likelihood, the naïve approach yields larger estimates for μ and σ_b^2 and a smaller estimated σ_ω^2 , reflecting the inconsistency induced by failure to properly account for ND.

Four different subject-specific estimates (*ad hoc* and ND-adjusted) for each facility operator worker's mean dust exposure level are displayed in Figure 1. In Figure 1, we note as expected that the LOD-adjusted CBs (CB_{GND}: sample variance = 0.356, CB_{SM}: sample variance = 0.383) impose less shrinkage toward the prior mean of -0.335 than the ND-adjusted Bayes_{ND} (sample variance = 0.201). CBs computed using the *ad hoc* approach (and naïve likelihood) are larger than the ND-adjusted CBs because they are based on the larger (inconsistent) estimate $\hat{\mu} = -0.194$, versus the ND-adjusted $\hat{\mu} = -0.335$ used in the computation of CB_{GND} and CB_{SM}.

3.2. Application of CB_{GND} to HIV RNA data

The HIV Epidemiology Research Study (HERS) was a large-scale epidemiologic investigation that included a prospective, multisite cohort study of HIV infection. Two primary objectives were to describe the natural history of HIV progression among women and to identify factors that are associated with prognosis [21]. We utilize HERS data collected between 1993 and 1995 that include 3–5 longitudinal measurements of HIV RNA levels from 528 women. 40 per cent of the 1864 HIV RNA levels fall below the assay's LOD (below 500 copies per ml). Lyles *et al.* [17] used this data set to compute the 528 Bayes predictions of the random intercepts and slopes for these women while accounting for ND. However, they did not consider shrinkage-resistant alternatives. We adapt the CB_{GND} approach to this more complex mixed model setting, taking advantage of its ready generalizability (see Discussion).

We present CB_{GND} predictions of the random intercept-slope model of (1), with fixed effects components defined by $\mathbf{X}_i = (\mathbf{1}_{n_i}, \mathbf{t}_{n_i})$ and $\Psi = (\alpha, \beta)'$ so that $E(\mathbf{Y}_i) = \alpha \mathbf{1}_{n_i} + \beta \mathbf{t}_{n_i}$, where $\mathbf{t}_i$ is a vector containing observation times for all subjects, i.e.

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)t_{ij} + \varepsilon_{ij}, i = 1, 2, \dots, k, j = 1, \dots, n_i. \quad (12)$$

Here, Y_{ij} is the logged HIV RNA level measured at j th time point t_{ij} for i th woman. We assume $b_i \stackrel{\text{iid}}{\sim} N(0, \sigma_2^2)$ are independent of the random errors $\varepsilon_{ij} \sim N(0, \sigma^2)$, with $\text{cov}(a_i, b_i) = \sigma_{12}$. To compute LOD-adjusted CB predictors for the random intercepts ($\alpha_i = \alpha + a_i$) and slopes ($\beta_i = \beta + b_i$) in model (12), we begin by maximizing the following likelihood [18] to obtain the MLE of $\theta = (\alpha, \beta, \sigma_1^2, \sigma_2^2, \sigma_{12}, \sigma^2)^T$, where d denotes the LOD:

$$L(\theta; \mathbf{Y}) = \prod_{i=1}^k f^*(\mathbf{Y}_i; \theta) = \prod_{i=1}^k \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n_{i1}} f(Y_{ij} | \alpha_i, \beta_i) \right\} \left\{ \prod_{j=n_{i1}+1}^{n_i} F_Y(d | \alpha_i, \beta_i) \right\} f(\alpha_i | \beta_i) f(\beta_i) d\alpha_i d\beta_i \right]. \quad (13)$$

As before, n_{i1} is the number of observed measurements for subject i , and the asterisk(*) denotes that each subject may or may not contribute ND measurements.

The next step in constructing the LOD-adjusted CB estimates is calculating the Bayes predictor. Utilizing the MLEs obtained via maximizing (13) and the following analytic expression, empirical Bayes estimates of intercepts or slopes can be obtained:

$$e_i^B(y) = E(\phi_i | Y_i) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_i f^*(Y_i; \theta) d\alpha_i d\beta_i}{f^*(Y_i; \theta)}. \quad (14)$$

Note that for our model, $\phi_i = (\alpha + a_i)$ or $\phi_i = (\beta + b_i)$ depending on whether we are interested in predicting random intercepts or random slopes, respectively and $f^*(Y_i; \theta)$ is as expressed in (13). After inserting MLEs to obtain the empirical Bayes estimates, Ghosh's recipe as outlined in equation (4) can be followed to obtain empirical CB_{GND} for the intercepts and the slopes.

Maximizing (13), we find the adjusted MLEs to be $\alpha = 2.887$, $\beta = 0.058$, $\sigma_1^2 = 0.719$, $\sigma_2^2 = 0.0372$, $\sigma_{12} = -0.061$, $\sigma_2 = 0.383$. In contrast, naïve MLEs when replacing ND by the 1/2 LOD are: $\alpha = 3.214$, $\beta = 0.039$, $\hat{\sigma}_1^2 = 0.236$, $\hat{\sigma}_2^2 = 0.0149$, $\sigma_{12} = -0.012$, and $\sigma_2 = 0.201$. Note that relative to the analysis that accounts explicitly in the likelihood for ND estimates, the *ad hoc* approach gives larger estimates of the α and σ_{12} parameters, and smaller estimates of β and the other variance components. Summary statistics for the empirical Bayes and CB estimates of the HERS intercepts and slopes based on *ad hoc* methods (naïve likelihood) and ND-adjusted likelihood (13) are displayed in Table I. Note the successful matches between the sample means and variances of the empirical CB_{GND} predictors, and the MLES $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_b^2, \hat{\sigma}_\omega^2)$ based on the ND-adjusted likelihood.

Based on the methodology of Lange and Ryan [22], we present a quantile–quantile (q–q) plot to assess the normality assumption corresponding to the random slopes in the HERS data example. We conclude that there are no obvious violations of normality (Figure 2). Verbeke and Molenberghs [18] provide further discussions about normality diagnostic procedures in the context of mixed models.

4. Simulation studies

We conducted several simulation studies to assure the translation of the CB properties from theory to practice and to examine the impact of the necessity to replace unknown parameters by their MLEs when calculating the CB predictions. The flexibility of simulations also allows us to examine how well the ND-adjusted methods perform under different conditions (e.g. varying parameter assumptions or varying proportions of ND).

4.1. Simulation studies for the random intercept only model

In the first set of simulation studies (Table II), we simulated 1000 data sets with $k = 100$ subjects, $n_i = 5$ measurements per subject, and under conditions similar to those of the occupational dust exposure data set [20]: $\mu = -0.335$, $\sigma_b^2 = 0.326$, $\sigma_\omega^2 = 1.243$, and with two varying proportions of ND. The imposed LOD in the first simulation study (Table II, middle panel) was equal to $-1.69 [\ln(0.2)]$, the same as for example data set, and 15 per cent (77 135/500 000) of the simulated data values fell below the LOD. In the second simulation (Table II, right panel), the LOD imposed was equal to -0.65 and 40 per cent of the 500 000 measurements fell below the LOD.

We use these simulation studies to compare the ND-adjusted CB_{GND} and CB_{SM} predictors not only to each other but also to ND-adjusted $Bayes_{ND}$, to expedient *ad hoc* approaches (substituting LOD and 1/2 LOD for ND) and to the true simulated random effects (μ_i). In computing Empirical Bayes and CB predictors, we maximized likelihood (3) for each of the 1000 simulated data sets and utilized the respective 1000 sets of MLEs of μ , σ_b^2 , σ_ω^2 in the

calculations. For the ‘Theoretical’ predictors, we use the true parameter values $\mu = -0.335$, $\sigma_b^2 = 0.326$, $\sigma_\omega^2 = 1.243$ in the calculation of Bayes_{ND}, CB_{GND}, and CB_{SM} for all 1000 simulated data sets. This allows us to examine the effect of estimating the unknown parameters as one must do in practice.

For both simulation studies presented in Table II, the corresponding Empirical and Theoretical ND-adjusted predictors have very similar sample means and variances. For example in the simulation with 40 per cent ND (right panel), the sample mean and variance of the Empirical ND-adjusted CB_{SM} predictor are -0.347 and 0.326 , respectively, which closely match those of the Theoretical CB_{SM} (mean -0.345 , variance 0.321). This implies that the effect of parameter estimation upon the predictor histogram is minimal under conditions similar to those in our example. As expected, the ND-adjusted CBs are far superior in terms of matching the sample mean and variance of the true μ_i ’s when compared with the two *ad hoc* CB predictors (Table II). For instance, in the simulation with 40 per cent ND, the sample variances of the Empirical ND-adjusted CB_{GND} (0.332) and CB_{SM} (0.326) much more closely match the sample variance of the true μ_i ’s (0.329) than do the sample variances of the *ad hoc* CBs that substitute the LOD (0.125) or $1/2$ the LOD (0.225) for all ND.

In Figure 3, one can see a graphical display of the histograms of the predictors resulting from the simulation with 15 per cent ND (summary statistics Table III, middle panel). The distributions of CB_{SM} (Figure 3(c)) and CB_{GND} (not pictured) closely match the distribution of the true μ_i (Figure 3(d)), unlike the *ad hoc* CBs (Figure 3(b)) and the ND-adjusted Bayes_{ND} (Figure 3(a)) predictors. In particular, note the shrinkage of the Bayes_{ND} toward the overall population mean (-0.335), and notice the poor performance of both the Bayes_{ND} and *ad hoc* CB in matching the distribution of the true μ_i , especially in the left tail of the distribution where the ND have greater impact (Figure 3).

We also conducted several simulation studies to more comprehensively evaluate the Theoretical CB_{SM} and CB_{GND} distributions by varying the frequency of left censoring, the number of observations per subject (n_i), and the values of the variance components. Seven of these many simulation studies are summarized in Table III. The results suggest that both CB_{SM} and CB_{GND} are generally effective at extending the moment-matching features of the CB paradigm when ND are present. However, these additional simulations also reveal some useful points regarding the implementation and performance of the two methods. Regarding implementation, we noted in the case of CB_{SM} that if all of the subject’s measurements fall below the LOD, then we fail to obtain a true CB estimate because of identifiability problems associated with likelihood (9). CB_{GND}, on the other hand, remains a viable CB predictor in that instance. Therefore, if in practice one’s method of choice is CB_{SM}, we propose substituting CB_{GND} in place of CB_{SM} on a subject-specific basis in the generally rare event that all of the subject’s measurements are ND. This is the approach taken in the simulation studies presented in Table II with 40 per cent ND, where 4 per cent (3945/100 000) of simulated subjects had all measurements below the LOD.

The performances of the predictors are dependent on the number of measurements falling below the LOD and the relative magnitude of the variance components (e.g. intra-class correlation coefficient, $\rho = \sigma_b^2 / (\sigma_b^2 + \sigma_\omega^2)$). In terms of performance, we note that the numerical integration-based approaches (e.g. CB_{GND} utilizing Bayes_{ND}) generally matched the first two moments (μ and σ_b^2) extremely well. Only cases of a small ρ and large σ_ω^2 (e.g. VI) or small to moderate ρ and a large percentage of ND (e.g. IV), were mild exceptions to this rule. Simulations I–III demonstrate that as the percentage of measurements below the LOD decreases, the proposed estimators match the first two moments of the true distribution more

closely. The CB_{SM} approach tended to match the first two moments well across the board. However, we note that the calculation of CB_{SM} requires more computer time and memory due to maximizing a non-standard likelihood (9) for each subject. In general, we find both CB_{GND} and CB_{SM} to be quite effective at matching the predictor distribution to that of the true random effects when ND are present. This is the case even with small sample size. For example in a simulation study with $k = 12$ subjects and with 40 per cent simulated measurements below LOD (mirroring the motivating example), the sample means (variances) of CB_{GND} and CB_{SM} were $-0.36(0.33)$ and $-0.35(0.32)$, respectively that closely match the mean (var) of the true $\mu_i = -0.34(0.33)$ [Moore 2006, unpublished dissertation].

As an indication of tangible effects and benefits of the empirical CB approach for random effects prediction, we present Figure 4. The left panel (Figure 4(a)) displays results for a random sample of 100 of 1000 simulated subject-specific means based on the random intercept model with true parameters similar to those estimated for the Heederik data. Overall, the empirical Bayes_{ND} predictions (open circles) achieve a tighter spread about the 45° line than do the empirical CB_{GND} predictions (closed circles), as expected. While both prediction methods are approximately unbiased overall, this point about precision is also reflected in the corresponding overall observed prediction mean-squared errors in the simulation (0.144 for Bayes versus 0.168 for CB). In contrast, Figure 4(b) displays results for the most extreme 100 of 1000 simulated subjects (i.e. for the simulated subjects with the highest 50 (5 per cent) and lowest 50 per cent true μ_i values). In this case, it is apparent that the closed circles tend to be closer to the 45° line as a result of reduced shrinkage. Consequences of this include a reduction in conditional bias associated with the CB versus the Bayes method, and the fact that the observed prediction mean squared error is noticeably smaller for these subjects based on the CB as opposed to Bayes approach (0.229 versus 0.339). Such benefits of the CB approach are often valuable, because predicting random effects for ‘extreme’ subjects can of particular interest in medical and epidemiological studies [8].

4.2. Simulation studies for random intercept-slope model

To evaluate the performance of CB_{GND} estimates of intercepts and slopes relative to the ND-adjusted Bayes estimates and to *ad hoc* methods, we present two simulation studies loosely based on the HERS data. For both simulation studies, we simulated 500 data sets, each containing 200 subjects with four measurements each, based on the following true parameters: $\alpha = 3$, $\beta = 0.05$, $\sigma_1^2 = 0.75$, $\sigma_2^2 = 0.25$, $\sigma_{12} = -0.05$, $\sigma_2 = 0.40$. In the first simulation study, the LOD was equal to 2.0 ($\log_{10}$ 100) copies per ml and this resulted in 18 per cent of the 400 000 simulated values falling below the LOD and 38 per cent of simulated subjects with at least one value below the LOD, with 4 per cent having non-detects for all four values. The sample means and variances in Table IV (middle panel: 18 per cent measurements < LOD) reveal that, for both intercepts and slopes, ND-adjusted CB_{GND} is superior to Bayes_{ND} and to the *ad hoc* approaches at matching the true random effect distribution. This is further verified by examining Figure 5, where the distribution of ND-adjusted CB_{GND} predictor (Figure 5(c)) provides an excellent match to the distribution of the true random intercepts (5d), especially in comparison with *ad hoc* methods (Figure 5(b)) and Bayes_{ND} (Figure 5(a)). The results were similar for the random slopes (not shown).

The summary statistics for the second simulation study in which the LOD was equal to 2.0 are presented in Table IV (right panel). Similar to the actual HERS data set, 40 per cent of the measurements fell below this LOD with 65 per cent of the simulated women having at least one value below the LOD, and for 15 per cent of the women all four of their simulated measurements fell below the LOD. It is clear that the ND-adjusted CBs of the intercepts and

the slopes more closely match the sample means and sample variances of the true intercepts and slopes as compared with the *ad hoc* CBs and the ND-adjusted Bayes (Table IV, right panel). The histograms (not shown) also display this superior matching. We note that these histograms reflect one tall bin on the left side due to the fact that 15 per cent of the simulated women have all four measurements below the LOD, and thus receive the same Bayes and CB estimate [Moore 2005, Proceedings of the American Statistical Association].

5. Discussion

We have developed methodology to compute CB predictors for repeated measures or longitudinal data subject to a detection limit. ND are quite frequent in many applications (e.g. environmental and HIV RNA epidemiologic studies), and CB methodology can be an attractive alternative when predicting subject-specific random effects [2, 16]. The CB predictor alleviates the overshrinkage of the Bayes predictor, especially in the tails of the distribution, thus offering potentially improved performance as a classification tool and/or enhanced precision for ‘extreme’ subjects [8, 9]. The approach is also attractive in that it maintains optimization of the traditional mean-squared error criterion, subject to the moment-matching constraints. Thus, while some sacrifice in squared error loss must be accepted when compared with the widely used Bayes predictor, CB methodology remains attentive to that criterion and outperforms the Bayes predictor with respect to the goal of matching the histogram of the true parameter distribution.

Prior to our research, the CB predictor had not been adapted to account for ND. The ND-adjusted CB predictors (CB_{GND} and CB_{SM}) presented here each achieved effective reduction in shrinkage, as suggested in the occupational exposure and HIV RNA data sets (Figure 1, Table I, respectively) and verified in the simulation studies (Figures 3–5, Tables II–IV). Furthermore, as designed, the ND-adjusted CB predictors much more closely match the true distribution of the random effects than *ad hoc* approaches or ND-adjusted Bayes predictors (Figures 3–5; Tables II–IV). Further simulation studies demonstrated that the properties of the ND-adjusted CB predictors remained reliable for small k and for larger proportions (40 per cent) below the LOD.

We acknowledge that a fully Bayesian approach could also be an appealing option, and that there are other viable routes for investigating prediction in the presence of non-detectable measurements (e.g. one could focus on posterior modes, rather than means). However, our current objective is to work in the traditional mixed model framework and propose an alternative to the empirical Bayes predictor (EBLUP) in a setting that would accomplish the CB goals, while accounting for left censoring.

For the random intercept only model, the development and comparison of the two CB methods to handle ND in the case of Model (1) is appealing in terms of approaching the CB objective from slightly different perspectives, as represented in the literature [7, 9]. In comparing the methods presented here, we note an implementation advantage for the CB_{GND} predictor in that it remains directly applicable when a subject has all ND measurements, and a possible performance advantage for CB_{SM} in that it offered slightly better moment-matching for larger values of the within-subject variance in our simulation studies (e.g. Table III: VI). In practice, we are comfortable recommending either approach under the random intercept model, with the caveat that CB_{GND} be applied for any subject with all ND. The R and SAS IML code utilized to obtain estimates of CB_{GND} and CB_{SM} are available upon request from the authors.

It is important to emphasize, however, that the CB_{GND} method enjoys a distinct advantage in terms of flexibility; the Ghosh [7] paradigm outlined in equations (4) continues to apply

with relatively little additional conceptual or computational effort when the underlying mixed linear model involves covariates and/or a more complicated random effects covariance structure. This is the reason why we focused attention on the CB_{GND} approach when extending to the random intercept-slope model in Section 3.2. The flexibility (e.g. with regard to inclusion of covariates and/or multiple random effects) of the CB paradigm for mixed linear and other models in the absence of ND has been discussed by several groups [23–25]. The CB_{GND} approach also has the benefit of straightforward adaptation to provide reduced-shrinkage predictions under more general random effects models (e.g. the mixture model proposed by Chu *et al.* [26]). While it is beyond our current scope to provide a detailed study of sensitivity to the assumed random effects distribution, this generalizability facilitated by Ghosh's paradigm means that one could make such a study in practice by comparing predictions based on alternative models.

As is common in occupational exposure and HIV RNA studies, our example data exhibit right-skewness. To better justify normality conditions required for the models presented, we have followed the common procedure of working with the natural logged exposures. However, Rappaport [27, 28] argued that for chronic diseases, the biologically relevant exposure of interest is often the true mean exposure on the original scale. Among other predictors, prior authors [29] presented a CB predictor of mean exposure for lognormal-scale data, based on model (1) for natural logged exposure measurements Y_{ij} . Our ongoing work extends this CB predictor on the original (lognormal) scale to adjust for data subject to a LOD. In addition to lognormal-scale prediction with left censoring, we are also developing ND-adjusted CB predictors to account for interval-censored exposure data encountered in environmental epidemiology. This ongoing work suggests that the methods presented here can be readily manipulated to handle various censoring mechanisms as well as multiple LODs, that may be encountered when exposure assays are refined or altered over the course of an epidemiologic study.

Acknowledgments

R. H. L. and A. K. M. were supported in part by an R01 from NIEHS (ES012458). We thank Dr Dick Heederik and Remko Houba for the use of the occupational exposure data and Dr Lytt Gardner and the HERS Study team for the HIV RNA data. The author are grateful to Dr Andrew N. Hill for engaging in many discerning conversations throughout this research endeavor.

References

1. Efron B, Morris C. Limiting the risk of Bayes and empirical Bayes estimators—part I: the Bayes case. *Journal of the American Statistical Association*. 1971; 66:807–815.
2. Louis TA. Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*. 1984; 79:393–398.
3. Searle, SR.; Casella, G.; McCulloch, CE. *Variance Components*. New York: Wiley; 1992.
4. Shen W, Louis TA. Triple-goal estimates in two-stage hierarchical models. *Journal of the Royal Statistical Society, Series B*. 1998; 60:455–471.
5. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: 2010. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>
6. Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. *SAS[®] for Mixed Models*. 2nd edn. Cary, NC: SAS Institute, Inc.; 2006.
7. Ghosh M. Constrained Bayes estimation with applications. *Journal of the American Statistical Association*. 1992; 87:533–540.
8. Lyles RH, Manatunga AK, Moore RH, Bowman FD, Cook CB. Improving point predictions of random effects for subjects at high risk: a case study. *Statistics in Medicine*. 2007; 26:1285–1300. [PubMed: 16810716]

9. Lyles RH, Xu J. Classifying individuals based on predictors of random effects. *Statistics in Medicine*. 1999; 18:35–52. [PubMed: 9990691]
10. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Applied Occupational Environmental Hygiene*. 1990; 5:46–51.
11. Pettitt AN. Censored observations, repeated measures and mixed effects models: an approach using the EM algorithm and normal errors. *Biometrika*. 1986; 73:635–643.
12. Hughes JP. Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*. 1999; 55:625–629. [PubMed: 11318225]
13. Wu L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to aids studies. *Journal of the American Statistical Association*. 2002; 97:955–964.
14. Carriquiry AL, Gianola D, Fernando RL. Mixed-model analysis of a censored normal distribution with reference to animal breeding. *Biometrics*. 1987; 43:929–939. [PubMed: 3509964]
15. Putter H, Heisterkamp SH, Lange JM, de Wolf F. A Bayesian approach to parameter estimation in HIV dynamical models. *Statistics in Medicine*. 2002; 21:2199–2214. [PubMed: 12210633]
16. Jacqmin-Gadda H, Théibaut R, Chêne G, Commenges D. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*. 2000; 1:355–368. [PubMed: 12933561]
17. Lyles RH, Lyles CM, Taylor DJ. Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Applied Statistics*. 2000; 49:485–497.
18. Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. New York: Springer; 2000.
19. Carlin, BP.; Louis, TA. *Bayesian Methods for Data Analysis*. 3rd edn. London, Boca Raton, FL: Chapman & Hall/CRC Press; 2008.
20. Heederik D, Boleij JSM, Kromhout H, Smid T. Use and analysis of exposure monitoring data in occupational epidemiology: an example of an epidemiological study in the Dutch animal food industry. *Applied Occupational Environmental Hygiene*. 1991; 6:458–464.
21. Smith DK, Warren DL, Vlahov D, Schuman P, Stein MD, Greenberg BL, Holmberg SD. Design and baseline participant characteristics of the Human Immunodeficiency Virus Epidemiology Research (HER) study: a prospective cohort of human immunodeficiency virus infection in US women. *American Journal of Epidemiology*. 1997; 146:459–469. [PubMed: 9290506]
22. Lange N, Ryan L. Assessing normality in random effects models. *The Annals of Statistics*. 1989; 17:624–642.
23. Frey J, Cressie N. Some results on constrained Bayes estimators. *Statistics and Probability Letters*. 2003; 65(4):389–399.
24. Lyles RH, Moore RH, Manatunga AK, Easley KA. Covariate-adjusted constrained Bayes predictions of random intercepts and slopes. *Journal of Modern Applied Statistical Methods*. 2010 in press.
25. Ghosh M, Kim MJ, Kim DH. Constrained Bayes and empirical Bayes estimation under random effects normal ANOVA model with balanced loss function. *Journal of Statistical Planning and Inference*. 2008; 138(7):2017–2028.
26. Chu H, Nie L, Kensler TW. A Bayesian approach estimating treatment effects on biomarkers containing zeros with detection limits. *Statistics in Medicine*. 2008; 27:2497–2508. [PubMed: 18167630]
27. Rappaport SM. Assessment of long-term exposures to toxic substances in air. *Annals of Occupational Hygiene*. 1991; 35:61–121. [PubMed: 2035954]
28. Rappaport SM. Biological considerations in assessing exposures to genotoxic and carcinogenic agents. *International Archives of Occupational and Environmental Health*. 1993; 65:S29–S35. [PubMed: 8406935]
29. Lyles RH, Kupper LL, Rappaport SM. On prediction of lognormal-scale mean exposure levels in epidemiologic studies. *Journal of Agricultural, Biological, and Environmental Statistics*. 1997; 2:417–439.

Figure 1. Four candidate predictors of mean log-scale occupational dust exposure for 12 facility operator workers (data from Heederik *et al.*, 1991). The *ad hoc* CB (sample mean = -0.194 and sample variance = 0.336) substitutes in the LOD for all ND and uses the naïve likelihood. The Empirical Bayes_{ND}, CB_{GND} and CB_{SM} (sample means = -0.365, -0.365, -0.313 and sample variances = 0.201, 0.355, 0.383, respectively) utilize ND-adjusted likelihoods and methodology. The vertical line drawn is $\hat{\mu}$ from the ND-adjusted likelihood.

Figure 2.
Quantile–Quantile plot of slopes of HIV RNA levels from women in the HERS study (1997).

Figure 3.

Histograms for comparing the distribution of CB_{SM} (c) to the Distributions of the *ad hoc* CB (b), $Bayes_{ND}$ (a) and the true random effects (d), based on 1000 simulated data sets each with 100 subjects contributing four measurements. The overlaid normal curve corresponds to the distribution of the true random effects, where $\mu = -0.335$, $\sigma_b^2 = 0.326$, and $\sigma_\omega^2 = 1.243$. The LOD imposed was equal to -1.69 .

Figure 4.

Subject-specific plots from simulation study of 1000 subjects with distribution of true random effects similar to the Heederik data set $\mu = -0.335$, $\sigma_b^2 = 0.326$, and $\sigma_\omega^2 = 1.243$. The left panel (a) plots the Bayes_{ND} and CB_{GND} estimates for 100 randomly selected simulated subjects. The right panel (b) plots the Bayes_{ND} and CB_{GND} estimates for the 100 most extreme (based on true μ_i) simulated subjects in the group of 1000.

Figure 5.

Histograms for comparing the distributions of *ad hoc* CB (b), ND-adjusted CB (c), and ND-adjusted Bayes (a) Predictors to the distribution of the true random intercepts (d), based on 500 simulated data sets each with 200 subjects contributing 4 measurements. The overlaid normal curve corresponds to the distribution of the true random intercepts, where $\alpha = 3$ and $\sigma_1^2 = 0.75$.

Table I

LOD-adjusted CB predictors* of individual intercept and slope HIV RNA levels for the HERS data.

	Intercepts		Slopes	
	Sample mean	Sample variance	Sample mean	Sample variance
<i>Ad hoc</i> methods				
Substitute LOD				
Bayes	3.214	0.163	0.0393	0.0050
CB	3.214	0.236	0.0393	0.0149
Substitute 1/2 LOD				
Bayes	3.089	0.241	0.0425	0.0074
CB	3.089	0.342	0.0425	0.0214
ND-Adjusted [†]				
Bayes _{ND}	2.887	0.503	0.0583	0.0113
CB _{GND}	2.887	0.721	0.0583	0.0372

* On the log₁₀ scale.

[†] Note: $\hat{\alpha} = 2.887$, $\hat{\sigma}_1^2 = 0.719$, $\hat{\beta} = 0.058$, $\hat{\sigma}_2^2 = 0.0372$.

Table II

Simulation summary statistics for comparing the distributions of *ad hoc* CBs to ND-adjusted CBs under the random intercept model*.

	15 per cent of simulated measurements < LOD = -1.69		40 per cent of simulated measurements < LOD = -0.65	
	Sample mean	Sample variance	Sample mean	Sample variance
<i>Ad hoc methods</i>				
Substitute LOD				
Bayes	-0.233	0.137	0.022	0.068
CB	-0.233	0.239	0.022	0.125
Substitute 1/2 LOD				
Bayes	-0.291	0.198	-0.257	0.125
CB	-0.291	0.236	-0.257	0.225
<i>Empirical ND-adjusted</i>				
Bayes _{ND}	-0.356	0.203	-0.367	0.188
CB _{GND}	-0.356	0.339	-0.367	0.332
CB _{SM}	-0.334	0.324	-0.347	0.326
<i>Theoretical ND-adjusted</i>				
Bayes _{ND}	-0.357	0.198	-0.365	0.179
CB _{GND}	-0.357	0.338	-0.365	0.327
CB _{SM}	-0.334	0.323	-0.345	0.321
True μ_i	-0.335	0.326	-0.338	0.329

* Results summarized over 1000 simulated data sets where $\mu = -0.335$, $\sigma_b^2 = 0.326$, $\sigma_\omega^2 = 1.242$, $k = 100$ subjects, and $n_i = 5$ measurements per subject.

Summary statistics from seven simulation studies used to further evaluate the theoretical distributions of CB_{GND} and $CB_{SM}(\text{random intercept only model})$.

Table III

Simulation study	I	II	III	IV	V	VI	VII
Per cent measurements < LOD	37.00	16.00	8.00	36.00	16.00	16.00	8.00
Per cent persons with all measurements < LOD	21.00	7.00	2.60	14.00	3.50	1.00	0.30
True σ^2_ω	1.00	1.00	1.00	1.00	1.00	3.00	1.50
True $\rho = \sigma^2_b / (\sigma^2_b + \sigma^2_\omega)$	0.75	0.75	0.75	0.50	0.50	0.25	0.25
Sample mean: CB_{GND}	0.01	0.01	0.01	-0.10	0.00	-0.18	-0.02
Sample mean: CB_{SM}	0.00	0.01	0.01	-0.03	0.00	-0.01	0.00
True μ	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sample variance: CB_{GND}	3.02	2.97	3.00	1.13	0.99	1.23	0.53
Sample variance: CB_{SM}	2.89	2.92	2.99	1.07	0.98	1.03	0.50
True σ^2_b	3.00	3.00	3.00	1.00	1.00	1.00	0.50

Based on one simulated data set with $k = 25000$ subjects, $n_i = 3$ observations per subject for all seven simulation studies.

Table IV
Simulation summary statistics for comparing the distributions of ND-adjusted Bayes and CB to *ad hoc* CBs of intercepts and slopes.

18 per cent simulated measurements < LOD				40 per cent simulated measurements < LOD			
Intercepts (α_i)		Slopes (β_i)		Intercepts (α_i)		Slopes (β_i)	
Sample mean	Sample variance	Sample mean	Sample variance	Sample mean	Sample variance	Sample mean	Sample variance
<i>Ad hoc</i> methods							
Substitute LOD							
Bayes	3.09	0.43	0.07	0.08	0.25	0.07	0.05
CB	3.09	0.54	0.07	0.18	0.32	0.07	0.12
Substitute 1/2 LOD							
Bayes	3.04	0.50	0.08	0.10	0.33	0.07	0.06
CB	3.04	0.54	0.08	0.22	0.42	0.07	0.14
Empirical ND-adjusted							
Bayes	2.99	0.60	0.05	0.12	0.57	0.04	0.10
CB _{GND}	2.99	0.75	0.05	0.25	0.75	0.04	0.24
True	2.99	0.75	0.05	0.25	0.75	0.05	0.25

Based on 500 simulated data sets, each containing $k = 200$ subjects, $n_i = 4$ observations per subject, and parameters $\alpha = 3$, $\beta = 0.05$, $\sigma_1^2 = 0.75$, $\sigma_2^2 = 0.25$, $\sigma_{12} = -0.05$, $\text{LOD} = 2.0$ (log10 100) copies per ml.