BIOMETRICS 55, 1022–1029
December 1999

Empirical Bayes Estimation of Random Effects Parameters in Mixed Effects Logistic Regression Models

Thomas R. Ten Have* and **A. Russell Localio**

Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Blockley Hall, 6th Floor, 423 Guardian Drive, Philadelphia, Pennsylvania 19104-6021, U.S.A.
* *email:* ttenhave@cceb.upenn.edu

SUMMARY. We extend an approach for estimating random effects parameters under a random intercept and slope logistic regression model to include standard errors, thereby including confidence intervals. The procedure entails numerical integration to yield posterior empirical Bayes (EB) estimates of random effects parameters and their corresponding posterior standard errors. We incorporate an adjustment of the standard error due to Kass and Steffey (KS; 1989, *Journal of the American Statistical Association* **84**, 717–726) to account for the variability in estimating the variance component of the random effects distribution. In assessing health care providers with respect to adult pneumonia mortality, comparisons are made with the penalized quasi-likelihood (PQL) approximation approach of Breslow and Clayton (1993, *Journal of the American Statistical Association* **88**, 9–25) and a Bayesian approach. To make comparisons with an EB method previously reported in the literature, we apply these approaches to crossover trials data previously analyzed with the estimating equations EB approach of Waclawiw and Liang (1994, *Statistics in Medicine* **13**, 541–551). We also perform simulations to compare the proposed KS and PQL approaches. These two approaches lead to EB estimates of random effects parameters with similar asymptotic bias. However, for many clusters with small cluster size, the proposed KS approach does better than the PQL procedures in terms of coverage of nominal 95% confidence intervals for random effects estimates. For large cluster sizes and a few clusters, the PQL approach performs better than the KS adjustment. These simulation results agree somewhat with those of the data analyses.

KEY WORDS: Bayesian estimation; Hierarchical models; Kass–Steffey numerical integration; Penalized quasi-likelihood; Random slope.

1. Introduction

We present two approaches to empirical Bayes (EB) inference for random effects parameters in mixed effects logistic regression models with both within- and between-cluster covariates: (1) an application of the Kass and Steffey (KS; 1989) adjustment to EB standard errors, and (2) an application of the penalized quasi-likelihood (PQL) approach of Breslow and Clayton (1993) based on a Taylor series expansion about EB estimates of the random effects parameters. We compare these two approaches through simulations. In addition, we compare them and a fully Bayesian approach in the analyses of two data sets: (1) an assessment of health care providers according to the risk of death from pneumonia, adjusted for patient-level covariates, and (2) a treatment–placebo, two-period crossover trial of a drug and placebo for palliative treatment of cerebrovascular deficiency as measured by a positive or negative electrocardiogram reading.

For the first example, we present random intercept model-based results to determine which hospitals serving a corporation's employees had significantly different rates of mortality for adult pneumonia patients, as reflected by random intercept estimates (Localio et al., 1995). The analysis of the pneumonia data was based on all computerized discharge data for adult pneumonia cases (DRGs 089-090) in 1989–1990 for 22 hospitals in central Pennsylvania. The cluster size ranged from 28 to 203, with a median of 104.5. We present results from modeling the death of an adult patient as a function of three 0–1 dummy variables corresponding to a four-level measure of patient severity (MedisGroups admission severity group, [ASG]), and two dummy variables corresponding to a three-level age variable (<40, 40–54, 55–64 years). This approach was used by Localio et al. (1997) in comparing hospitals with respect to coronary bypass surgical rates.

For the second example, we follow an existing EB approach in the analysis of the crossover data by Waclawiw and Liang (1994), who present EB estimates based on estimating equations under a random intercept logistic model with treatment and period as 0–1 binary covariates. With data from two centers, Waclawiw and Liang (and therefore we) analyzed only the 67 patients from the second center, of which 34 were randomly assigned to receive the active drug followed by the placebo, whereas the remaining patients were allocated to the reverse sequence (see Jones and Kenward, 1989, pp. 90–91, for more details). Waclawiw and Liang (1994) focused on EB estimation of the random intercepts for the purposes of estimating individual patients' risk for cerebrovascular deficiency.

EB estimation of random effects parameters under mixed effects models with both within- and between-cluster covariates has become an important issue in research and policy making. Localio et al. (1997) tested the PQL procedure in ranking health care providers according to the risk of adverse outcomes, e.g., death from coronary bypass surgery, while adjusting for patient- and hospital-specific risk factors. Using analytical, numerical, or Monte Carlo integration approaches, Legler and Ryan (1997) and Sammel, Ryan, and Legler (1997) relied on EB estimates of random effects parameters to measure the severity of an individual's risk of birth defects that underlies several different types of birth defect outcomes. However, they did not report standard errors. Alternatively, Waclawiw and Liang (1994) integrated the Stein estimation with a generalized estimating equations approach. Bayesian methods have also been used to obtain estimates of random effects (Christiansen and Morris, 1997). Although these Bayesian approaches may provide more accurate estimates of standard errors than EB approaches do (Kass and Steffey, 1989; Waclawiw and Liang, 1994), Bayesian approaches, which rely on Markov chain Monte Carlo techniques, have caveats (Natarajan and McCulloch, 1995; Hobert and Casella, 1996; Kass and Wasserman, 1996). Others have presented estimation approaches for cluster-level probabilities, given random intercept models with only between-cluster covariates (Thomas, Longford, and Rolph, 1994; Piegorsch and Casella, 1996; Booth and Hobert, 1998).

In Sections 2 and 3, we present the KS-based numerical integration and the PQL approximation approaches to EB estimation of random effects for two models: (1) a model with fixed effects parameters and a random intercept (random intercept–only model), and (2) a random coefficient model with fixed effects parameters and a random intercept and slope (random intercept–slope model).

In Section 4, we make simulation-based comparisons between the KS and PQL approaches. We do not present simulation results based on the Waclawiw–Liang estimation technique because the public version of their program is under development (Waclawiw, personal communication). Simulations could not be undertaken for the Bayesian approach with BUGS software (Spiegelhalter et al., 1995) because of the extended time to complete a single run and the necessity to assess convergence after each run (to assure proper posterior distributions), an issue critical to the crossover trial data.

In Section 5, we analyze our two data sets under the random intercept logit model with the two approaches (KS and PQL) and the fully Bayesian approach undertaken with MCMC methods as implemented in BUGS. Section 6 summarizes the simulation and data analysis results.

2. Model

This section presents a mixed effects logistic model with a general random effects structure, thus accommodating both random intercept–only and random intercept–slope models. We assume I clusters, indexed by $i = 1, \ldots, I$, each of which consists of J_i observations, indexed by $j = 1, \ldots, J_i$. In the various contexts for which EB estimation of random effects is employed, a cluster may correspond to a hospital or a subject in the health care provider or crossover trial example, respectively.

Let Y_{ij} be the binary response for the jth observation of the ith cluster; $Y_{ij} = 1$ if observation j of cluster i exhibits the response of interest, and 0 otherwise, for all i and j. Also define $\mathbf{x}_{ij}$ to be the observed covariate vector, i.e., an intercept vector and within- and between-cluster covariates, corresponding to the fixed effects for the jth observation of the ith cluster.

Letting π_{ij} denote the probability of $Y_{ij} = 1$, given the ith cluster random effect $\boldsymbol{\tau}_i$, the conditional distribution for Y_{ij} is Bernoulli(π_{ij}), with the logistic model specified for π_{ij}

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{w}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\tau}_i + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}, \qquad (1)$$

where $\boldsymbol{\beta}$ is a vector of an intercept parameter and log odds ratios corresponding to unit changes in the elements of the covariate vector $\mathbf{x}_{ij}$ among observations within a cluster, $\mathbf{w}_{ij}$ is a vector of covariates for the random effects, $\boldsymbol{\tau}_i$ $(i = 1, \ldots, I)$ are identically distributed according to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance–covariance matrix $\mathbf{I}$, the identity matrix; and $\boldsymbol{\Sigma}$ is the lower-triangular matrix in the Cholesky decomposition of a symmetric positive definite variance–covariance matrix, for instance, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}$. See Hedeker and Gibbons (1994) and Ten Have et al. (1998) for similar parameterizations of mixed effects logistic models.

For applications with only a random intercept, e.g., example data sets, $\boldsymbol{\tau}_i$ consists of just an intercept τ_{0i}, which is a standard normal random variate, and $\mathbf{w}_{ij} = 1$. Accordingly, $\boldsymbol{\Sigma}$ consists of one nonnegative element σ_0. For applications with a random intercept and slope, the vector $\boldsymbol{\tau}_i$ is augmented with a random slope τ_{1i}, and $\mathbf{w}_{ij}$ is then a vector with the first component equal to 1 and the second component equal to a covariate value that changes within a subject, for instance, time. Here, the Cholesky decomposition lower-triangular matrix $\boldsymbol{\Sigma}$ is a 2×2 matrix with diagonal elements σ_0 and σ_1, lower off-diagonal element σ_{01}, and upper off-diagonal element 0. A property of the model in (1) is the conditional independence among the elements of $\mathbf{Y}_i^{\mathrm{T}} = (Y_{i1}, \ldots, Y_{iJ_i})$, given $\boldsymbol{\tau}_i$. Marginally, i.e., integrating $\boldsymbol{\tau}_i$ from the likelihood, the elements of $\mathbf{Y}_i^{\mathrm{T}} = (Y_{i1}, \ldots, Y_{iJ_i})$ are correlated.

The focus of this article is on estimating the unobserved random effect parameters $\boldsymbol{\eta}_i = \boldsymbol{\Sigma}\boldsymbol{\tau}_i$. Here $\boldsymbol{\Gamma}(= \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}})$ is the variance–covariance of $\boldsymbol{\eta}_i$. Under the random intercept–slope model, $\boldsymbol{\eta}^{\mathrm{T}} = (\eta_{0i}, \eta_{1i})$, where $\eta_{0i} = \sigma_0\tau_{0i}$ and $\eta_{1i} = \sigma_{01}\tau_{0i} + \sigma_1\tau_{1i}$, with variances equal to σ_0^2 and $\sigma_0^2 + \sigma_1^2$, respectively, and their covariance equal to $\sigma_0^2 + \sigma_0\sigma_1$.

The example data analyses assume random intercept–only models, i.e., $\boldsymbol{\eta}_i = \eta_{0i} = \sigma_0\tau_{0i}$. For example, η_{0i} would represent the departure in the hospital-level odds of mortality from the average for all hospitals in the sample, whereas in the crossover trial example, η_{0i} represents the underlying subject predisposition to the cerebrovascular outcome after controlling for treatment and period differences within a subject. Finally, the specification in (1) and the numerical estimation procedure described later for $\boldsymbol{\eta}_i$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$ may be extended in a straightforward way to other members of the class of generalized linear mixed effects models.

3. Estimation

Under both the random intercept–only and random intercept–slope models, we present a two-stage procedure for EB estimation of $\boldsymbol{\eta}_i$. The first step is the maximum likelihood estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, which is necessary for the second step of empirical Bayes estimation of $\boldsymbol{\eta}_i$. We also present an estimation of the variance of the EB estimator of $\boldsymbol{\eta}_i$.

3.1 *Maximum Likelihood Estimation*

The marginal likelihood, which is maximized to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, is obtained by replacing the integral with a weighted sum, where the weights are binomial mass functions (Mauritsen, 1984; Ten Have et al., 1998).

Letting K be the dimension of $\boldsymbol{\tau}_i$, we approximate the following integral for cluster i:

$$\int_{-\infty}^{\infty} \prod_{j=1}^{J_i} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}} (2\pi)^{-K/2} \times \exp\left(\sum_{k=0}^{K-1} \tau_{ki}^2/2\right) d\tau_{0i} \cdots d\tau_{K-1,i}$$
$$\approx \sum_{\nu_0=0}^{V} \cdots \sum_{\nu_{K-1}=0}^{V} \prod_{j=1}^{J_i} \pi_{ij\nu}^{y_{ij}} (1-\pi_{ij\nu})^{1-y_{ij}} \times \prod_{k=1}^{K} \left[\binom{V}{\nu_{ki}} (1/2)^{\nu_{ik}} (1-1/2)^{V-\nu_k}\right], \tag{2}$$

where

$$\pi_{ij\nu} = [1+\exp(-\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij} - (\boldsymbol{\Sigma}\mathbf{w}_{ij})^{\mathrm{T}}\boldsymbol{\tau}_{i\nu})]^{-1},$$

the kth element of $\boldsymbol{\tau}_{i\nu}$, $\tau_{ik\nu}$, is defined as a standardized binomial random variable $\tau_{ik\nu} = [\nu_{ik} - V(1/2)]/[V(1/2)(1-1/2)]^{1/2}$, $\nu_{ik} \sim$ binomial$(V, 1/2)$, and $k = 1$ or 2 for the random intercept or slope, respectively. Increasing V improves the accuracy of the binomial approximation (Mauritsen, 1984). For the simulations and data analyses, $V = 20$. The approximation in (2) can be used for other generalized linear mixed models by replacing $\Pi_{j=1}^{J_i} \pi_{ij\nu}^{y_{ij}}(1-\pi_{ij\nu})^{1-y_{ij}}$ with another appropriate conditional distribution, e.g., Poisson distribution.

We maximize (2) with respect to the parameters in (1), using a quasi-Newton approach with first derivatives based on (2) and initial values for the fixed effects parameters obtained from logistic regression assuming independence. The inverse of the Hessian of the final marginal likelihood yields standard errors for the corresponding estimates of the fixed effects parameters.

3.2 *EB Estimates*

Given the resulting maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$, we obtain EB estimates of the elements of $\boldsymbol{\eta}_i$ by following Sammel et al. (1997). We first present the estimation of η_{0i} under the random intercept–only model, i.e., $\boldsymbol{\eta}_i = \eta_{0i}$ and $\boldsymbol{\Sigma} = \sigma_0$

$$\hat{\eta}_{0i} = E\left(\hat{\sigma}_0\tau_{0i} \mid \mathbf{y}_i, \mathbf{x}_i, \hat{\sigma}_0, \hat{\boldsymbol{\beta}}\right) = \int_{-\infty}^{\infty} \hat{\sigma}_0\tau_{0i}\mathrm{pr}(\tau_{0i} \mid \mathbf{y}_i, \mathbf{x}_i, \hat{\sigma}_0, \hat{\boldsymbol{\beta}}) d\tau_{0i}$$
$$\approx \sum_{\nu=0}^{V} \hat{\sigma}_0\tau_{0i\nu}M_{i\nu} \Big/ \sum_{\nu=0}^{V} M_{i\nu}, \tag{3}$$

where

$$M_{i\nu} = \prod_{j=1}^{J_i} \hat{\pi}_{ij\nu}^{y_{ij}} (1-\hat{\pi}_{ij\nu})^{1-y_{ij}} \left[\binom{V}{\nu} (1/2)^{\nu} (1-1/2)^{V-\nu}\right],$$

$\mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iJ_i})$, and $\mathbf{x}_i^{\mathrm{T}} = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iJ_i})$.

Under the random intercept–slope model, with $\boldsymbol{\eta}_i^{\mathrm{T}} = (\eta_{0i}, \eta_{1i})$ and $\boldsymbol{\Sigma}$ consisting of nonzero elements σ_0, σ_{01}, and σ_1, the EB estimator of the random intercept takes the form

$$\hat{\eta}_{0i} = E(\hat{\sigma}_0\tau_{0i} \mid \mathbf{y}_i, \mathbf{x}_i, \widehat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\beta}}) \approx \frac{\sum_{\nu_0,\nu_1=0}^{V} \hat{\sigma}_0\tau_{0i\nu}M_{i\nu_0\nu_1}}{\sum_{\nu_0,\nu_1=0}^{V} M_{i\nu_0\nu_1}},$$

where

$$M_{i\nu_0\nu_1} = \prod_{j=1}^{J_i} \hat{\pi}_{ij\nu}^{y_{ij}} (1-\hat{\pi}_{ij\nu})^{1-y_{ij}} \times \left[\binom{V}{\mu_0} (1/2)^{\nu_0} (1-1/2)^{V-\nu_0}\right] \times \left[\binom{V}{\mu_1} (1/2)^{\nu_1} (1-1/2)^{V-\nu_1}\right], \tag{4}$$

and $\Sigma_{\nu_0,\nu_1=0}^{V}$ denotes the double sum $\Sigma_{\nu_0=0}^{V}\Sigma_{\nu_1=0}^{V}$. The EB estimator of the random slope is

$$\hat{\eta}_{1i} = E(\hat{\sigma}_{01}\tau_{0i} + \hat{\sigma}_1\tau_{1i} \mid \mathbf{y}_i, \mathbf{x}_i, \widehat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\beta}})$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\hat{\sigma}_{01}\tau_{0i} + \hat{\sigma}_1\tau_{1i}) \times \mathrm{pr}(\tau_{0i}, \tau_{1i} \mid \mathbf{y}_i, \mathbf{x}_i, \widehat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\beta}}) d\tau_{0i}\, d\tau_{1i}$$
$$\approx \frac{\sum_{\nu_0,\nu_1=0}^{V} (\hat{\sigma}_{01}\tau_{0i\nu_0} + \hat{\sigma}_1\tau_{1i\nu_1})M_{i\nu 0\nu_1}}{\sum_{\nu_0,\nu_1=0}^{V} M_{i\nu_0\nu_1}}.$$

3.3 *Variances of EB estimates*

We first extend the approach of Piegorsch and Casella (1996), who focus on obtaining the EB estimates of binomial probabilities that preclude within-cluster covariates, to computation of the estimate of the variance of $\hat{\eta}_{0i}$ under the random intercept–only model

$$\widehat{\mathrm{var}}(\hat{\eta}_{0i}) = E\left(\hat{\sigma}_0^2\tau_{0i}^2 \mid \mathbf{y}_i, \mathbf{x}_i, \hat{\sigma}_0, \hat{\boldsymbol{\beta}}\right) - E\left(\hat{\sigma}_0\tau_{0i} \mid \mathbf{y}_i, \mathbf{x}_i, \hat{\sigma}_0, \hat{\boldsymbol{\beta}}\right)^2$$
$$\approx \frac{\sum_{\nu=0}^{V} \hat{\sigma}_0^2\tau_{0i\nu}^2 M_{i\nu}}{\sum_{\nu=0}^{V} M_{i\nu}} - \hat{\eta}_{0i}^2. \tag{5}$$

To adjust $\widehat{\text{var}}(\hat{\eta}_{0i})$ for the variability in the estimate of $\hat{\sigma}_0$, we consider adding to $\widehat{\text{var}}(\hat{\eta}_{0i})$ a special case of an adjustment due to Kass and Steffey (1989) for the general class of parametric EB models. This adjustment factor, denoted as $A_{\text{KS}i}$, approximates posterior variances that are marginal with respect to all other estimated parameters such as $\boldsymbol{\beta}$ and σ_0. We refer to this adjustment as the KS approach. Letting $\boldsymbol{\theta}^{\text{T}} = (\boldsymbol{\beta}^{\text{T}}, \sigma_0)$, $A_{\text{KS}i}$ is defined for cluster i as

$$A_{\text{KS}0i} = \widehat{\text{cov}}(\hat{\boldsymbol{\theta}}) \left(\left. \frac{\partial \hat{\eta}^*_{0i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right) \left(\left. \frac{\partial \hat{\eta}^*_{0i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right)^{\text{T}}, \quad (6)$$

where $\hat{\eta}^*_{0i}$ equals (3) but with $\boldsymbol{\theta}$ replacing $\hat{\boldsymbol{\theta}}$, $\widehat{\text{cov}}(\hat{\boldsymbol{\theta}})$ is minus the inverse Hessian of the likelihood based on (2), and $(\partial \hat{\eta}^*_{0i}/\partial \boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ is the vector of partial derivatives of $\hat{\eta}^*_{0i}$ with respect to the elements of $\boldsymbol{\beta}$ and σ_0, with these parameters replaced with their estimates. Specifically, the element of $(\partial \hat{\eta}^*_{0i}/\partial \boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ corresponding to σ_0 is

$$\frac{\sum_{\nu=0}^{V} \tau_{0i\nu} M_{i\nu}}{\sum_{\nu=0}^{V} M_{i\nu}} + \sigma_0 \left[\frac{\sum_{\nu=0}^{V} \tau_{0i\nu} \frac{\partial M^*_{i\nu}}{\partial \sigma_0}}{\sum_{\nu_0=0}^{V} M_{i\nu}} - \frac{\left(\sum_{\nu=0}^{V} \tau_{0i\nu} M_{i\nu} \right) \sum_{\nu=0}^{V} \frac{\partial M^*_{i\nu}}{\partial \sigma_0}}{\left(\sum_{\nu=0}^{V} M_{i\nu} \right)^2} \right] \Bigg|_{\sigma_0=\hat{\sigma}_0},$$

where $M^*_{i\nu} = M_{i\nu}$, but with σ_0 replacing $\hat{\sigma}_0$, and

$$(\partial M^*_{i\nu}/\partial \sigma_0)|_{\sigma_0=\hat{\sigma}_0} = M_{i\nu} \tau_{0i\nu} \Sigma_{j=1}^{J_i} (y_{ij} - \hat{\pi}_{ij\nu}).$$

The element of $(\partial \hat{\eta}^*_{0i}/\partial \boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ corresponding to β_l, the lth element of $\boldsymbol{\beta}$, is

$$\sum_{\nu=0}^{V} \hat{\sigma}_0 \left[\frac{\sum_{\nu=0}^{V} \tau_{0i\nu} \frac{\partial M^*_{i\nu}}{\partial \beta_l}}{\sum_{\nu=0}^{V} M_{i\nu}} - \frac{\left(\sum_{\nu=0}^{V} \tau_{0i\nu} M_{i\nu} \right) \sum_{\nu=0}^{V} \frac{\partial M^*_{i\nu}}{\partial \beta_l}}{\left(\sum_{\nu=0}^{V} M_{i\nu} \right)^2} \right] \Bigg|_{\beta_l=\hat{\beta}_l}.$$

For the random intercept–slope model, the variance estimate for $\hat{\eta}_{0i}$ is the same as (5) but with $M_{i\nu_0\nu_1}$, defined in (4), replacing $M_{i\nu}$. The variance estimate for $\hat{\eta}_{1i}$ is obtained as

$$\begin{aligned}
\widehat{\text{var}}(\hat{\eta}_{1i}) &= \hat{\sigma}^2_{01} \widehat{\text{var}}(\hat{\tau}_{0i}) + \hat{\sigma}^2_1 \widehat{\text{var}}(\hat{\tau}_{1i}) + 2\hat{\sigma}_{01}\hat{\sigma}_1 \widehat{\text{cov}}(\hat{\tau}_{0i}, \hat{\tau}_{1i}) \\
&\approx \frac{\sum_{\nu_0,\nu_1=0}^{V} \hat{\sigma}^2_{01} \tau^2_{0i\nu_0} M_{i\nu_0\nu_1}}{\sum_{\nu_0,\nu_1=0}^{V} M_{i\nu_0\nu_1}} - \hat{\sigma}^2_{01} \hat{\tau}^2_{0i} \\
&\quad + \frac{\sum_{\nu_0,\nu_1=0}^{V} \hat{\sigma}^2_1 \tau^2_{1i\nu_1} M_{i\nu_0\nu_1}}{\sum_{\nu_0,\nu_1=0}^{V} M_{i\nu_0\nu_1}} - \hat{\sigma}^2_1 \hat{\tau}^2_{1i} \\
&\quad + 2\frac{\sum_{\nu_0,\nu_1=0}^{V} \hat{\sigma}_{01}\hat{\sigma}_1 \tau_{0i\nu_0} \tau_{1i\nu_1} M_{i\nu_0\nu_1}}{\sum_{\nu_0,\nu_1=0}^{V} M_{i\nu_0\nu_1}} \\
&\quad - \hat{\sigma}_{01}\hat{\sigma}_1 \hat{\tau}_{0i} \hat{\tau}_{1i}.
\end{aligned}$$

Under the random intercept–slope model, the variance of $\hat{\eta}_{0i}$ is augmented with $A_{\text{KS}0i}$, as defined in (6) for $\hat{\eta}_{0i}$, but with minor changes not presented here because of space limitations. Details are presented in a technical report that may be obtained from the authors. Letting $\boldsymbol{\theta}^{\text{T}} = (\boldsymbol{\beta}, \sigma_0, \sigma_{01}, \sigma_1)$, the variance of $\hat{\eta}_{1i}$ is augmented with

$$A_{\text{KS}1i} = \widehat{\text{cov}}(\hat{\boldsymbol{\theta}}) \left(\left. \frac{\partial \sigma_{01} \hat{\tau}^*_{0i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \left. \frac{\partial \sigma_1 \hat{\tau}^*_{1i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right) \times \left(\left. \frac{\partial \sigma_{01} \hat{\tau}^*_{0i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \left. \frac{\partial \sigma_1 \hat{\tau}^*_{1i}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right)^{\text{T}}.$$

We implemented these procedures in FORTRAN90.

3.4 *PQL Approach*

We performed the PQL estimation with the GLIMMIX SAS macro, an iteratively reweighted estimation routine based on Proc Mixed. Much has been written on the PQL approach. Therefore, we limit its consideration to a cursory description of the EB estimation. Briefly, estimation of the random effects occurs as a by-product of the estimation of the fixed effects $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, as the Taylor series approximation behind the PQL approach involves expanding around EB estimates of the η_i parameters, which are obtained using a linear approximation. The standard errors of the PQL-based EB estimates are obtained at the last iteration of Proc Mixed. Thus, the details for this may be found in the technical report for Proc Mixed (SAS Institute, 1996) in conjunction with the linear expansions discussed in Wolfinger and O'Connell (1993). For the data analysis and simulations, we did not adjust these standard errors as we did for the integration-based approaches, because we wanted to assess the quality of a standard package such as the SAS GLIMMIX macro (Localio et al., 1997).

4. Simulations

To assess the accuracy of the proposed EB estimates and confidence intervals under KS and PQL, we performed simulations (500 iterations per simulation) for bias, mean-squared error (MSE), and coverage of nominal 95% confidence intervals, first under the random intercept model and then under the random intercept–slope model. We specified a logit model with an intercept, a log odds ratio for a 0–1 binary covariate, and a normally distributed random intercept having mean 0 and variance σ_0^2. Technical details on the simulation are given in our technical report.

There were a number of goals in varying these conditions. First, the intent was to assess the behavior of the different approaches under two different contexts: (1) many clusters with small cluster sizes, as in the crossover trial and studies of birth defects, and (2) few clusters with large cluster sizes, typically the case for the comparison of health care providers. Second, varying $\mathbf{\Sigma}$ allowed the assessment under different magnitudes of within-cluster correlation.

The simulation results reveal that KS and PQL perform differently under different conditions. Under the random intercept–only model, for many clusters ($I = 100$) and a small cluster size ($J = 2$), KS, in general, is superior to PQL in terms of confidence interval coverage (see Table 1). For a few clusters ($I = 20$) and a large cluster size ($J = 100$), the converse is true (see Table 1). This observation holds for a range of random effects variances and covariances and baseline risks of outcome. We note that increasing the number of clusters (I) to at least 200 (and up to 400) while holding $J = 2$ improves KS confidence interval coverage to at least 94% for both the random intercept–only and random intercept–slope models. The only exception to these general results, including the KS versus PQL relationship, occurs for small random effects variances and covariances under the random intercept–slope model for $J = 2$.

Unlike confidence interval coverage, the minimum and maximum biases and MSEs of $\hat{\eta}_{0i}$ under the random intercept-only model increase with increasing σ_0 regardless of the estimation method, number of clusters, cluster size, and the size of β_0. These KS and PQL statistics are of similar magnitudes, which contrasts with the reversal of superiority between the KS and PQL coverage levels.

Some of the differences between KS and PQL observed for EB confidence interval coverage under the random intercept–only model carry over to the random intercept–slope model. Again, there is a reversal of superiority between the two approaches with respect to coverage when I and J are varied. However, unlike under the random intercept model, the KS coverage levels for $I = 100$ and $J = 2$ start poorly and then increase to acceptable levels under the random intercept–slope model. With many clusters ($I \geq 200$) and with $J = 2$, the KS coverage levels under the random intercept–slope model exceed 94.0% for all attempted random effects variance and baseline risk specifications, except for small random effects variation. The remaining KS coverage levels with $I = 20$ and $J = 100$ and for PQL in general behave similarly to their counterpart coverage levels under the random intercept–only model. Moreover, as with the random intercept model, PQL and KS yield very comparable MSE and bias statistics. Our technical report provides detailed results.

5. Data Analysis

In this section, we present the results obtained by applying the two approaches, KS and PQL, to analyzing the two data sets: (1) the treatment–placebo, two-period crossover trial of a treatment of cerebrovascular deficiency; and (2) the assessment of health care providers according to the risk of death from pneumonia. For both examples, we present results based on random effects estimates and their respective standard errors under three approaches: (1) the numerical integration method, which corresponds to the KS approach; (2) the PQL approach; and (3) the Bayesian approach using BUGS based on 20,000 iterations after a 20,000 iteration burn-in period. The Bayesian procedure was implemented assuming the following priors: (1) a vague inverse gamma prior for σ_0 with mean 1 and variance 1000, (2) vague independent univariate normal priors for fixed effects parameters, each with mean 0 and variance 10^5, and (3) a normal prior for the random effects parameters with mean 0 and variance σ_0^2.

Assessment of convergence was based on two chains, using the Gelman–Rubin shrinkage factor for each fixed and random effects parameter, obtained in CODA (Cowles and Carlin, 1996). Evidence of departures from convergence, i.e., 97.5% of the shrinkage factors less than 5%, diminished within 20,000 iterations for the hospital data, but not for the cerebral data. At the suggestion of a referee, for the latter data, we increased the magnitude of the parameters of the normal (variance = 1000) and gamma hyperpriors (variance = 10) for β and $1/\sigma_0$, respectively. The convergence criteria appeared to be satisfied under these hyperpriors, although the resulting fixed estimates are 10–15% smaller than the original estimates.

Table 1

Simulated coverage of nominal 95% confidence intervals for η_{0i}, $i = 1, \ldots, I$, under KS and PQL approaches, using different combinations of the random effects standard deviation (σ_0) and log baseline risk (β_0). Simulation results are for $I = 100$ and $J_i = 2$ and for $I = 20$ and $J_i = 100$.

	σ_0	0.5		1.0		2.0		5.0	
	β_0	−0.5	−3.0	−0.5	−3.0	−0.5	−3.0	−0.5	−3.0
$I = 100$	KS	94.6	96.5	91.8[a]	91.8[a]	93.5[a]	92.6[a]	94.5	94.1[a]
$J_i = 2$	PQL	93.7	84.6	89.1	87.1	86.5	88.8	82.1	80.9
$I = 20$	KS	92.3	91.0	81.6	75.7	56.5	34.5	46.6	34.1
$J_i = 100$	PQL	93.5	91.8	94.0	92.8	93.5	92.4	91.5	90.4

[a] Denotes that with increasing number of clusters (200–400), the coverage is within 0.5% of the nominal level of 95%.

Table 2

Random effects results obtained by analyzing the pneumonia and cerebral data sets with KS and PQL. Minimum and maximum statistics are presented for the random effect estimate $\hat{\eta}_i$ and its standard error $SE(\hat{\eta}_i)$, under each approach. In addition, we present the proportion of subjects for which $|\hat{\eta}_i/SE(\hat{\eta}_i)|$ exceeds 1.96 (*i.e., proportion of subjects for which $\hat{\eta}_i$ differs significantly from* 0), *which we denote by $\eta_i \neq 0$.*

			η_{0i}		SE($\hat{\eta}_{0i}$)	
Example	Approach	$\eta_{0i} \neq 0$	Minimum	Maximum	Minimum	Maximum
Pneumonia	KS	0.00%	−0.46	0.60	0.36	0.47
Data set	PQL	9.09%	−0.92	0.85	0.29	0.58
	Bayes	0.00%	−0.43	0.50	0.30	0.50
Cerebral	KS	0.00%	−6.39	3.12	2.08	3.60
Data set	PQL	31.34%	−6.96	3.43	0.86	2.41
	Bayes	22.39%	−11.70	6.02	3.07	8.05

The BUGS results for the cerebral data are based on these hyperpriors that resulted in apparent convergence. Because of space limitations, we present only the results for the EB estimates and standard errors based on these data sets. Results based on estimates of $\boldsymbol{\beta}$ appear in our technical report.

There is some correspondence between the random effects estimates from the example data sets and the simulation results, as is shown below. For the hospital results in Table 2, with a few clusters ($I = 22$) and large cluster sizes ($28 \leq J_i \leq 203$), the only procedure that indicates significant differences between individual random effects and 0 is PQL (2 out of 22 hospitals). None of the other approaches rejected any hospitals as being different. Note that in Table 2, the largest PQL EB standard error is larger than the largest Bayes posterior standard error, which contrasts with previous reports in the literature that Bayesian standard errors and confidence intervals are routinely larger and wider, respectively, than are EB standard errors and confidence intervals. The simulation results suggest that the PQL EB estimates and standard errors are credible, given the large cluster sizes. With large clusters, there might be sufficient binary information to account for the variability of $\hat{\sigma}_0$ in standard error estimation; therefore, Bayesian methodology need not be relied upon to account for this variability. This might not be the case with small clusters, as we next see for the cerebrovascular data, for which there are two observations per cluster.

For the cerebrovascular (crossover) random effects estimates in Table 2, there is more heterogeneity among the three approaches than for the hospital data in terms of indicating which subjects are significantly different from 0. The Bayesian method picked out 15 of 67 subjects with underlying risks of cerebrovascular problems that differ from the population mean. PQL identified an additional six patients, whereas KS did not pick out any. The Bayesian posterior estimates of η_{0i} cover a wider range than do the EB estimates. In addition, the Bayes posterior standard errors for η_{0i} are the largest among the standard errors of the different approaches. The PQL estimates and standard errors are the smallest, as one would expect from the literature on the conditions that are necessary for appropriate asymptotic behavior of the PQL estimates (Lin and Breslow, 1996). Recall that the Bayesian MCMC approach encountered convergence problems for this data set.

6. Discussion

We have shown through simulation and data analyses that the EB estimation performances of two mixed effects logistic regression procedures, the numerical integration-based KS and PQL approaches, depend on a number of factors: cluster size relative to the number of clusters, the size of the variances and covariances of the random effects parameters, the level of baseline risk, and the presence or absence of random slopes.

The simulation results show that cluster size relative to the number of clusters is the primary factor in distinguishing the EB performance of KS and PQL. In terms of confidence interval coverage, KS does better than PQL does for small cluster sizes and many clusters, whereas the converse is true for large cluster sizes and a few clusters. The two approaches perform very similarly in terms of bias and MSE for all attempted combinations of factors, with bias and MSE increasing with random effects variation.

The presence or absence of a random slope appears to influence the coverage properties of KS. With random slopes, KS does well for intermediate standard deviations, not as well for large standard deviations, and even worse for small standard deviations. Again, with a large enough number of clusters, the KS coverage levels are within 1.0% of nominal coverage for all levels of attempted random effects variances and covariances, except for the very weak levels of random effects variation. For large cluster sizes and a few clusters under the random intercept model, the performance of KS deteriorates rapidly as the random effects standard deviations increase, regardless of the presence of a random slope. PQL performed much more adequately for a few clusters and large cluster sizes, although actual coverage did not quite achieve nominal levels.

The simulation-based coverage results are somewhat supported by previous work on the asymptotic behavior of the numerical integration and PQL approaches. First, the small number of clusters precludes the good coverage properties of the numerical integration approach (as shown by Ten Have et al., 1998) from taking hold of large within-cluster correlations. Second, the relatively good coverage and low bias of PQL estimators under large cluster size agree with the work of Vonesh and Chinchilli (1997, p. 353), who show that asymptotic bias under the class of methods including PQL diminishes with increasing cluster size. Second-order corrections of the PQL approach (Lin and Breslow, 1996) may improve the coverage of confidence intervals. This is an area of future research.

The data analysis-based comparisons are somewhat congruent with the simulation results. By not rejecting any of the hospitals as different, "PQL" yields a unique inference among the three procedures for the hospital data with a few clusters and large cluster size. The convergence problems encountered with the Bayesian approach for the cerebral data notwithstanding, the results for this example data set are compatible with the simulation results, which show that KS does well for many clusters and small cluster sizes.

Résumé

Nous généralisons une approche pour estimer les paramètres d'effets aléatoires sous un modèle de régression logistique à valeur à l'origine et pente aléatoires en incluant les erreurs standard et également les intervalles de confiance. La procédure nécessite des intégrations numériques pour fournir les estimations de Bayes empiriques a posteriori (EB) des paramètres des effets aléatoires et des erreurs standard correspondantes. Nous introduisant un ajustement de l'erreur standard dû à Kass et Steffey (KS; 1989) afin de prendre en *compte la variabilité* dans l'estimation des composantes de variance des distributions des effets aléatoires. Dans le cadre d'une étude visant à évaluer la structure de soins vis-à-vis de la mortalité par pneumonie adulte des comparaisons ont été faites entre l'approximation par la quasi-vraisemblance pénalisée (PQL) proposée par Breslow et Clayton (1993) et l'approche bayésienne. Nous appliquons ces démarches sur un essai en cross-over analysé précédemment par la méthode des estimations d'équations EB de Waclawiw et Liang (1994) afin de les comparer avec une méthode EB proposée dans la bibliographie. Nous avons également réalisé des simulations afin de comparer les approches KS et PQL. Ces deux démarches conduisent à des estimations EB des paramètres à effets aléatoires avec des biais asymptotiques équivalents. Cependant, en présence d'un grand nombre de classes de petites tailles, l'approche KS est supérieure aux procédures PQL relativement à l'amplitude des intervalles de confiance au niveau nominal 95% pour les estimations des effet aléatoires. L'approche PQL est meilleure par contre en présence d'un faible nombre de classes de grandes tailles. Les résultats des simulations concordent quelque peu avec ceux des analyses réalisées.

References

Booth, J. G. and Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association* **93,** 262–272.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association* **88,** 9–25.

Christiansen, C. L. and Morris, C. N. (1997). Hierarchical Poisson regression modeling. *Journal of the American Statistical Association* **92,** 618–632.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91,** 883–904.

Hedeker, D. and Gibbons, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics* **50,** 933–944.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* **91,** 1461–1473.

Jones, B. and Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials.* New York: Chapman and Hall.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes methods). *Journal of the American Statistical Association* **84,** 717–726.

Kass, R. E. and Wasserman, S. (1996). Formal rules for selecting prior distributions: A review and annotated bibliography. *Journal of the American Statistical Association* **91,** 1343–1370.

Legler, J. M. and Ryan, L. M. (1997). Latent variables for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association* **92,** 13–20.

Lin, X. and Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91,** 1007–1016.

Localio, A. R., Hamory, B. H., Fisher, A. C., and Ten Have, T. R. (1997). The public release of hospital and physician mortality data in Pennsylvania: A case study. *Medical Care* **35,** 272–286.

Localio, A. R., Hamory, B. H., Sharp, T. J., Weaver, S. L., Ten Have, T. R., and Landis, J. R. (1995). Comparing hospital mortality in adult patients with pneumonia. *Annals of Internal Medicine* **122,** 125–132.

Mauritsen, R. (1984). Logistic regression with random effects. Ph.D. thesis, University of Washington, Seattle.

Natarajan, R. and McCulloch, C. E. (1995). A note on the existence of the posterior distribution for a class of mixed effects models for binomial responses. *Biometrika* **82,** 639–643.

Piegorsch, W. W. and Casella, G. (1996). Empirical Bayes estimation for logistic regression and extended parametric regression models. *Journal of Agricultural, Biological, and Environmental Statistics* **1,** 231–247.

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B* **59,** 667–678.

SAS Institute (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.12.* Cary, North Carolina: SAS Institute.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1995). *BUGS Bayesian Inference Using Gibbs Sampling,* Version 0.50. MRC Biostatistics Unit, Cambridge.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., and Landis, J. R. (1998). Mixed effects logistic models for longitudinal binary data with informative drop-out. *Biometrics* **54,** 367–383.

Thomas, N., Longford, N. T., and Rolph, J. E. (1994). Empirical Bayes methods for estimating hospital-specific mortality rates. *Statistics in Medicine* **13,** 889–903.

Waclawiw, M. A. and Liang, K.-Y. (1994). Empirical Bayes estimation and inference for the random effects model with binary response. *Statistics in Medicine* **13,** 541–551.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Communications in Statistics, Simulation and Computation* **22,** 1079–1106.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* New York: Marcel Dekker.

Received February 1998. *Revised December* 1998.
Accepted December 1998.