Supplemental material for:
A Class of Generalized Linear Mixed Models Adjusted for Marginal Interpretability

Jeffrey J. Gory, Peter F. Craigmile, Steven N. MacEachern

Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

Last updated October 20, 2020

S1 Proofs

The proofs of Propositions 1 and 2 are immediate, and are omitted.

Proof of Proposition 3: For each $i = 1, \ldots, N$, dividing both sides of (5) by $\exp(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{d}_i^T \mathbf{a_i})$ and then taking the natural logarithm of both sides yields

$$-\mathbf{d}_i^T \mathbf{a_i} = \log\left\{ \int \exp(\mathbf{d}_i^T \mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} \right\}.$$

The integral in this equation is equal to $M_{\mathbf{U}}(\mathbf{d}_i)$. Thus, multiplying both sides by -1, we obtain $\mathbf{d}_i^T \mathbf{a_i} = -\log\{M_{\mathbf{U}}(\mathbf{d}_i)\}$, as required.

Proof of Corollary 1: This result follows immediately from Proposition 3. For each $i = 1, \ldots, N$, if $\mathbf{d}_i^T \mathbf{a_i}$ exists, then $\mathbf{d}_i^T \mathbf{a_i} = -\log\{M_{\mathbf{U}}(\mathbf{d}_i)\}$, and $M_{\mathbf{U}}(\mathbf{d}_i)$ must exist. Conversely, if $M_{\mathbf{U}}(\mathbf{d}_i)$ exists, then $\mathbf{d}_i^T \mathbf{a_i} = -\log\{M_{\mathbf{U}}(\mathbf{d}_i)\}$ also exists because $M_{\mathbf{U}}(\mathbf{d}_i)$ is strictly positive and thus lies in the domain of $\log(\cdot)$.

Proof of Corollary 2: For this model, $d_i = 1$ for all $i = 1, \ldots, N$ and $M_U(t) = \exp(\sigma^2 t^2/2)$. Thus, $a_i = -\log\{M_U(d_i)\} = -\log\{\exp(\sigma^2/2)\} = -\sigma^2/2$ for all $i = 1, \ldots, N$, as required.

Proof of Theorem 1: To simplify notation, we suppress the subscript i. For any choice of $f_{\mathbf{U}}$,

$$-\infty < \ell = \ell \int f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} \leq \int h(\mathbf{x}^T\boldsymbol{\beta} + \mathbf{d}^T\mathbf{u} + \mathbf{d}^T\mathbf{a})f_{\mathbf{U}}(\mathbf{u})d\mathbf{u}$$
$$\leq u \int f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = u < \infty.$$

Consequently, the integral $\int h(\mathbf{x}^T\boldsymbol{\beta}+\mathbf{d}^T\mathbf{u}+\mathbf{d}^T\mathbf{a})f_{\mathbf{U}}(\mathbf{u})d\mathbf{u}$ exists. Further, $\int h(\mathbf{x}^T\boldsymbol{\beta}+\mathbf{d}^T\mathbf{u}+\mathbf{d}^T\mathbf{a})f_{\mathbf{U}}(\mathbf{u})d\mathbf{u}$ is a continuous function of $\mathbf{d}^T\mathbf{a}$ and, provided $\mathbf{d} \neq \mathbf{0}$, the following two limits hold:

$$\lim_{\mathbf{d}^T\mathbf{a}\to-\infty} \int h(\mathbf{x}^T\boldsymbol{\beta} + \mathbf{d}^T\mathbf{u} + \mathbf{d}^T\mathbf{a})f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = \ell,$$
$$\lim_{\mathbf{d}^T\mathbf{a}\to\infty} \int h(\mathbf{x}^T\boldsymbol{\beta} + \mathbf{d}\mathbf{g}^T\mathbf{u} + \mathbf{d}^T\mathbf{a})f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = u.$$

Since $\ell \leq h(\mathbf{x}^T\boldsymbol{\beta}) \leq u$, continuity implies that for any value of $h(\mathbf{x}^T\boldsymbol{\beta})$ there exists an adjustment $\mathbf{d}^T\mathbf{a}$ such that (4) holds. When $\mathbf{d} = \mathbf{0}$, (4) trivially holds because $h(\mathbf{x}^T\boldsymbol{\beta} + \mathbf{d}^T\mathbf{u} + \mathbf{d}^T\mathbf{a}) = h(\mathbf{x}^T\boldsymbol{\beta})$.

Proof of Proposition 4: Let $\mathbf{U}_i \sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$ and $\epsilon \sim \mathrm{N}(0, 1)$, and define $W = \epsilon - \mathbf{d}_i^T\mathbf{U}_i$ so that $W \sim \mathrm{N}(0, 1 + \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i)$. Then,

$$\int \Phi(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{u} + \mathbf{d}_i^T\mathbf{a}_i)f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = \int \mathrm{P}(\epsilon \leq \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{u} + \mathbf{d}_i^T\mathbf{a}_i)f_{\mathbf{U}}(\mathbf{u})d\mathbf{u}$$
$$= \mathrm{P}(\epsilon \leq \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{U}_i + \mathbf{d}_i^T\mathbf{a}_i) = \mathrm{P}(\epsilon - \mathbf{d}_i^T\mathbf{U}_i \leq \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{a}_i)$$
$$= \mathrm{P}(W \leq \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{a}_i) = \Phi\left\{\frac{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{a}_i}{(1 + \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i)^{1/2}}\right\}.$$

Consequently,

$$\Phi(\mathbf{x}_i^T\boldsymbol{\beta}) = \int \Phi(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{u} + \mathbf{d}_i^T\mathbf{a}_i)f_{\mathbf{U}}(\mathbf{u})d\mathbf{u} = \Phi\left\{\frac{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{a}_i}{(1 + \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i)^{1/2}}\right\}.$$

Applying $\Phi^{-1}(\cdot)$ to both sides yields $\mathbf{x}_i^T\boldsymbol{\beta} = (\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{d}_i^T\mathbf{a}_i)(1 + \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i)^{-1/2}$. Solving for $\mathbf{d}_i^T\mathbf{a}_i$ we obtain $\mathbf{d}_i^T\mathbf{a}_i = \big((1 + \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i)^{1/2} - 1\big)\mathbf{x}_i^T\boldsymbol{\beta}$, as required.

Proof of Proposition 5: We show the limit for $\mathbf{x}_i^T\boldsymbol{\beta} \to -\infty$; the limit for $\mathbf{x}_i^T\boldsymbol{\beta} \to \infty$ follows from symmetry. Let $\kappa = \mathbf{x}_i^T\boldsymbol{\beta}$, $a = \mathbf{d}_i^T\mathbf{a}_i$, and $\tau^2 = \mathbf{d}_i^T\boldsymbol{\Sigma}\mathbf{d}_i$. Then from Proposition 7 it follows that,

$$\frac{e^{\kappa}}{1+e^{\kappa}} = \int_{\mathbb{R}^q} \frac{e^{\kappa+\mathbf{u}+a}}{1+e^{\kappa+\mathbf{u}+a}} f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} = \int_{\mathbb{R}} \frac{e^{\kappa+v+a}}{1+e^{\kappa+v+a}} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(- \frac{1}{2\tau^2} v^2 \right) dv.$$

Dividing by $\exp(\kappa)$ on both sides and taking the limit as $\kappa \to -\infty$ we obtain

$$1 = e^a \int_{\mathbb{R}} e^v \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(- \frac{1}{2\tau^2} v^2 \right) dv.$$

Recognizing the integral on the right-hand side of this equation as the moment generating function of a $\mathrm{N}(0, \tau^2)$ random variable evaluated with argument $t = 1$, we obtain $\exp(-a) = \exp(\tau^2/2)$, which implies that $a = -\tau^2/2$. Thus, $\mathbf{d}_i^T \mathbf{a}_i = -\frac{1}{2} \mathbf{d}_i^T \boldsymbol{\Sigma} \mathbf{d}_i$ for each $i = 1, \ldots, N$, as required.

Proof of Proposition 6: Expanding the square in (8) and then rearranging terms yields

$$(\mathbf{d}_i^T \mathbf{a}_i)^2 + 2(\mathbf{x}_i^T \boldsymbol{\beta})(\mathbf{d}_i^T \mathbf{a}_i) + \mathrm{Var}(\mathbf{d}_i^T \mathbf{U}_i) = 0.$$

Application of the quadratic formula then leads to

$$\begin{aligned} \mathbf{d}_i^T \mathbf{a}_i &= \frac{1}{2}[-2\mathbf{x}_i^T \boldsymbol{\beta} \pm \{(2\mathbf{x}_i^T \boldsymbol{\beta})^2 - 4\mathrm{Var}(\mathbf{d}_i^T \mathbf{U}_i)\}^{1/2}] \\ &= -\mathbf{x}_i^T \boldsymbol{\beta} \pm \{(\mathbf{x}_i^T \boldsymbol{\beta})^2 - \mathrm{Var}(\mathbf{d}_i^T \mathbf{U}_i)\}^{1/2}. \end{aligned}$$

Thus, subject to the constraint $\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{d}_i^T \mathbf{a}_i \geq 0$, we have

$$\mathbf{d}_i^T \mathbf{a}_i = -\mathbf{x}_i^T \boldsymbol{\beta} + \{(\mathbf{x}_i^T \boldsymbol{\beta})^2 - \mathrm{Var}(\mathbf{d}_i^T \mathbf{U}_i)\}^{1/2}.$$

Proof of Proposition 7: The marginal likelihood for each Y_i, $i = 1, \ldots, N$, is

$$\begin{aligned} f_Y(Y_i) &= \int f_{Y|U}(Y_i | U_i = u) f_U(u) du \\ &= \int \mathrm{E}[Y_i | U_i = u]^{Y_i} (1 - \mathrm{E}[Y_i | U_i = u])^{1-Y_i} f_U(u) du \\ &= \int h(\mathbf{x}_i^T \boldsymbol{\beta} + u + a_i)^{Y_i} \big(1 - h(\mathbf{x}_i^T \boldsymbol{\beta} + u + a_i)\big)^{1-Y_i} f_U(u) du. \end{aligned}$$

Because the model is marginally interpretable, if $Y_i = 1$ we have

$$f_Y(Y_i) = \int h(\mathbf{x}_i^T \boldsymbol{\beta} + u + a_i) f_U(u) du = h(\mathbf{x}_i^T \boldsymbol{\beta}),$$

whereas if $Y_i = 0$ we have

$$f_Y(Y_i) = \int \big(1 - h(\mathbf{x}_i^T\boldsymbol{\beta} + u + a_i)\big) f_U(u)du = 1 - h(\mathbf{x}_i^T\boldsymbol{\beta}).$$

Since $Y_i \in \{0, 1\}$, we can therefore write

$$f_Y(Y_i) = h(\mathbf{x}_i^T\boldsymbol{\beta})^{Y_i}\big(1 - h(\mathbf{x}_i^T\boldsymbol{\beta})\big)^{1-Y_i},$$

and f_Y is completely independent of σ^2, as required.

Proof of Proposition 8: Suppose for each $i = 1, \dots, N$ we have $\mathbf{U}_i \sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$ and write $\mathbf{U}_i = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}_i$, where $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is a square root matrix for $\boldsymbol{\Sigma}$ and $\mathbf{Z}_i \sim \mathrm{N}_q(\mathbf{0}, \mathbf{I}_q)$. We can define a random variable $V = \mathbf{d}^T\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}$ such that $V \sim \mathrm{N}(0, \tau^2)$, where $\tau^2 = \mathbf{d}^T\boldsymbol{\Sigma}\mathbf{d}$. It is also possible to define $q - 1$ additional random variables $\mathbf{W} = (W_1, \dots, W_{q-1})^T$ that span the orthogonal complement of V relative to $\mathbb{R}^q$ such that $\mathbf{W}$ follows a $(q-1)$-dimensional Normal distribution with density $f_{\mathbf{W}}(\cdot)$. Given such a V and $\mathbf{W}$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^q} h(\kappa + \mathbf{d}^T\mathbf{u} + a)\Big(\frac{1}{2\pi}\Big)^{\frac{q}{2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\Big(-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}\Big)d\mathbf{u} \\
&= \int_{\mathbb{R}^q} h(\kappa + \mathbf{d}^T\boldsymbol{\Sigma}^{1/2}\mathbf{z} + a)\Big(\frac{1}{2\pi}\Big)^{\frac{q}{2}} \exp\Big(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\Big)d\mathbf{z}. \\
&= \int_{\mathbb{R}^q} h(\kappa + v + a)\frac{1}{\sqrt{2\pi\tau^2}} \exp\Big(-\frac{1}{2\tau^2}v^2\Big) f_{\mathbf{W}}(\mathbf{w})d\mathbf{w}dv \\
&= \int_{\mathbb{R}} h(\kappa + v + a)\frac{1}{\sqrt{2\pi\tau^2}} \exp\Big(-\frac{1}{2\tau^2}v^2\Big)dv \int_{\mathbb{R}^{q-1}} f_{\mathbf{W}}(\mathbf{w})d\mathbf{w} \\
&= \int_{\mathbb{R}} h(\kappa + v + a)\frac{1}{\sqrt{2\pi\tau^2}} \exp\Big(-\frac{1}{2\tau^2}v^2\Big)dv.
\end{aligned}$$

S2 Efficient and accurate evaluation of logistic-normal integrals

Monahan and Stefanski (1992) approximate the inverse logit function $h(z)$ with a weighted mixture of normals

$$h_k^*(z) = \sum_{i=1}^{k} p_{k,i}\Phi(zs_{k,i}),$$

where the weights $p_{k,i}$ and $s_{k,i}$ are chosen to minimize the maximum approximation error over all values of z. This leads to to the integral approximation

$$\int h(z)\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)dz \approx \int h_k^*(z)\frac{1}{\sigma}\phi\left(\tfrac{z-\mu}{\sigma}\right)dz = \sum_{i=1}^{k} p_{k,i}\Phi\left\{\frac{\mu s_{k,i}}{(1+\sigma^2 s_{k,i}^2)^{1/2}}\right\}, \tag{S1}$$

which is within 2.1×10^{-9} of the true value of the integral for all values of κ and σ with $k = 8$ mixture weights. Decreasing k can improve computational efficiency, but the increase in speed from using fewer weights is small relative to the corresponding loss of accuracy. We therefore recommend using $k = 8$.

As we write in the main article, our algorithm for evaluating the logistic-normal integral exploits a recursive formula developed by Pirjol (2013) that provides an exact solution to the logistic-normal integral on a specifically defined, evenly spaced grid: the integral for $\varphi(\mu, \sigma^2)$ given by (12) in the main article satisfies the recursion given by (13), with initial condition $\varphi(0, \sigma^2) = 1/2$. Combining the recursion with the above approximation of Monahan and Stefanski (1992), to approximate the logistic-normal integral $1 - \varphi(\mu, \sigma^2)$ for $\mu > 0$, where $\varphi(\cdot, \cdot)$ is defined as in (12), we first write $\mu = \mu^* + t\sigma^2$, where $\mu^* \in [0, \sigma^2)$ and t is a nonnegative integer. We then approximate $1 - \varphi(\mu^*, \sigma^2)$ using (S1) with $k = 8$ mixture weights and apply the recursion (13) t times to obtain an approximation for $1 - \varphi(\mu, \sigma^2)$. When $\mu < 0$, the integral of interest is $1 - \varphi(\mu, \sigma^2) = \varphi(|\mu|, \sigma^2)$, and the approximation can still be handled as if the first argument of $\varphi(\cdot, \cdot)$ were positive.

Denoting our approximation of $\varphi(\mu, \sigma^2)$ as $\tilde{\varphi}(\mu, \sigma^2)$, we define the error associated with this approximation as

$$\varepsilon(\mu, \sigma^2) = \varphi(\mu, \sigma^2) - \tilde{\varphi}(\mu, \sigma^2).$$

Pirjol (2013) showed that the error $\varepsilon(\mu, \sigma^2)$ is bounded by

$$|\varepsilon(\mu, \sigma^2)| \leq \exp\left(-\frac{1}{2\sigma^2}\mu^2 + \frac{1}{8}\sigma^2\right)\sup_{z\in[0,\sigma^2)}|\varepsilon(z, \sigma^2)|,$$

which means that $\tilde{\varphi}(\mu, \sigma^2)$ is generally more accurate for larger values of μ and that the error associated with $\tilde{\varphi}(\mu, \sigma^2)$ is never worse than the maximum error of the Monahan-Stefanski approximation (S1) over $[0, \sigma^2)$.

To assess the speed and accuracy of our hybrid approach we compared it to both 30-point Gauss-Hermite quadrature and to a direct application of (S1). Specifically, for each of the 80 values of σ in the set $\{0.05, 0.10, \ldots, 4.00\}$ we evaluated the integral $1-\varphi(\mu, \sigma^2)$ for 1,000 values of μ in each of the four intervals $[0, \sigma^2]$, $[\sigma^2, 2\sigma^2]$, $[2\sigma^2, 3\sigma^2]$, and $[3\sigma^2, 4\sigma^2]$ using the hybrid approach, the Monahan-Stefanski approximation, 30-point Gauss-Hermite quadrature, and 1,000-point Gauss-Hermite quadrature. This required 4,000 integral evaluations for each of the 80 values of σ and each method. These evaluations were completed on a Dual Quad Core Xeon computer with 32 GB of RAM. To ensure a fair comparison of speed, all four approaches were implemented using the `Rcpp` package in `R` (R Core Team, 2018; Eddelbuettel and François, 2011; Eddelbuettel, 2013). Gauss-Hermite quadrature with 1,000 quadrature points was treated as the *gold standard* to which the other three methods were compared to. For each method and each value of σ we computed the maximum "error" relative to 1,000-point quadrature within each of the four intervals for μ. Figure S1 summarizes a proportion of the results. Although 30-point Gauss-Hermite quadrature is the most accurate for small values of σ, the hybrid approach is the most accurate in the majority of cases. Notably, our hybrid approach, combining Pirjol (2013) and Monahan and Stefanski (1992), clearly outperforms a direct application of Monahan and Stefanski (1992).

The 320,000 integral evaluations required for the accuracy assessment took 2.1 seconds for the hybrid approach, 2.1 seconds for the Monahan-Stefanski approximation, 2.2 seconds for 30-point Gauss-Hermite, and 19.4 seconds for 1,000-point Gauss-Hermite. Thus, the computational speed of the hybrid approach is comparable to Monahan-Stefanski and slightly better than 30-point Gauss-Hermite. Since 1,000-point quadrature is considerably slower than the other three methods, we conclude that our hybrid approach is the most efficient, offering the best tradeoff between accuracy and speed among the methods.

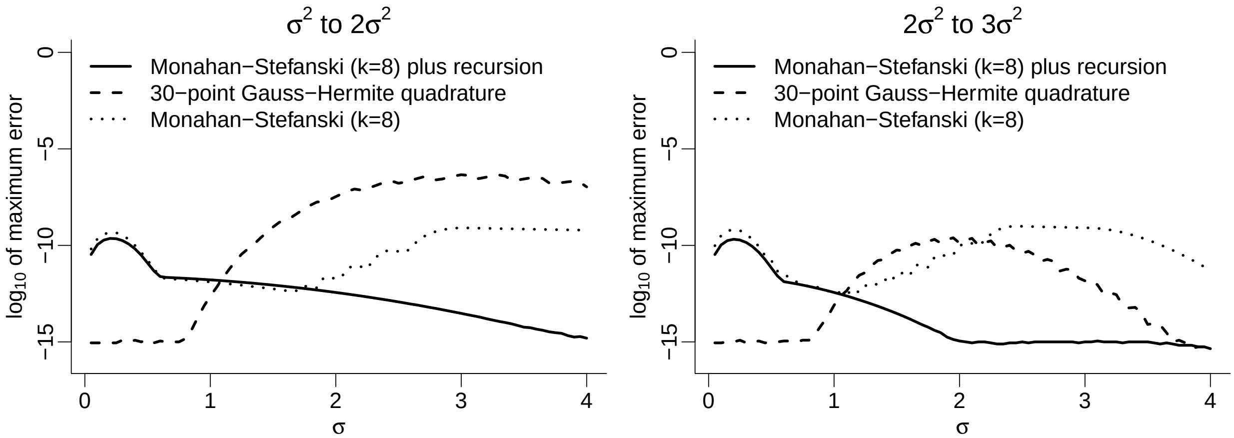

Figure S1: Maximum error relative to 1,000-point Gauss-Hermite quadrature for various approximations of the logistic-normal integral with μ in $[\sigma^2, 2\sigma^2]$ and $[2\sigma^2, 3\sigma^2]$; machine accuracy is approximately 10^{-16}, accounting for the floor in the plots.

S3 Estimating the parameters of a marginally interpretable model: a simulation study

In this section we investigate how well we can recover the model parameters in a marginally interpretable binomial model. For a sample of size n, suppose that

$$Y_i|\boldsymbol{\beta}, U_i \overset{ind}{\sim} \text{Binomial}\{m_i, \text{E}(p_i|\boldsymbol{\beta}, U_i)\}, \quad i = 1, \ldots, n.$$

Letting h denote the inverse logit function we model the conditional mean as

$$\text{E}(p_i|\boldsymbol{\beta}, U_i) = h(\beta_0 + \beta_1 x_i + U_i + a_i), \quad i = 1, \ldots, n,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is the vector of fixed effects parameters and a_i, defined by equation (6), is the adjustment that ensures the model is marginally interpretable. Throughout the simulation study we calculate the adjustment using the method proposed in Section 5.2. To complete the model we assume that the set of random effects $\{U_i : i = 1, \ldots, n\}$ consists of independent N(0, σ^2) random variables. In addition, we set the covariate values to $x_i = (i-1)/n$, $i = 1, \ldots, n$ and fix each $m_i = m$, for some common number of Bernoulli trials m.

We pick the same values of $\boldsymbol{\beta}$ as the illustration in Section 4.1: $\beta_0 = -3$, $\beta_1 = 5$. We set $\sigma = 0.5$ or $\sigma = 1$ in our simulations. As shown in Section 4.1 these values of $\boldsymbol{\beta}$ and σ induce different adjustments a_i for different x_i values. We fix the sample size at $n = 50$, but vary m with values of 1, 5, 10, 25 and 50.

For each combination of values, we simulate 500 realizations from our marginally interpretable model and estimate the parameters using maximum likelihood (ML). We evaluate the integrals in the likelihood function using Gauss-Hermite quadrature, as implemented by the statmod R package. To understand the effect of ignoring the adjustment, we also compare to the case when we fit a traditional GLMM model that does not include an adjustment. (The conventional model still requires us to evaluate the likelihood function using quadrature methods.)

Figure S2 displays the bias, standard deviation (SD), and root mean square error (RMSE) of ML estimates of β_0 (first row) and β_1 (second row) as the value of m increases when $\sigma = 0.5$. In each panel, the black lines show the summary of the ML estimates for the marginally interpretable GLMM, and the gray lines show summary values of the ML estimates for the conventional GLMM. The vertical lines denote 95% bootstrap confidence intervals for each quantity. Figure S3 summarizes the ML estimates when $\sigma = 1$.

Figure S2 demonstrates for the marginally interpretable GLMM that as the number of trials at each covariate value, m, increases, the biases of both the ML estimates of β_0 and β_1 approach zero. In addition the SD and RMSE of both ML estimates also decrease as m increases. Figure S2 shows that the patterns we see for the bias, SD, and RMSE for the $\sigma = 0.5$ case are the same when $\sigma = 1$. This demonstrates that our ML estimates are able to efficiently recover the true values of the $\boldsymbol{\beta}$ fixed effect parameters, regardless of the values of σ and m. (Naturally, estimating these parameters is harder when we have fewer trials.) This also confirms that our method to calculate the adjustments does not introduce significant bias in estimation.

Comparing the ML estimates of the marginally interpretable GLMM to the ML estimates

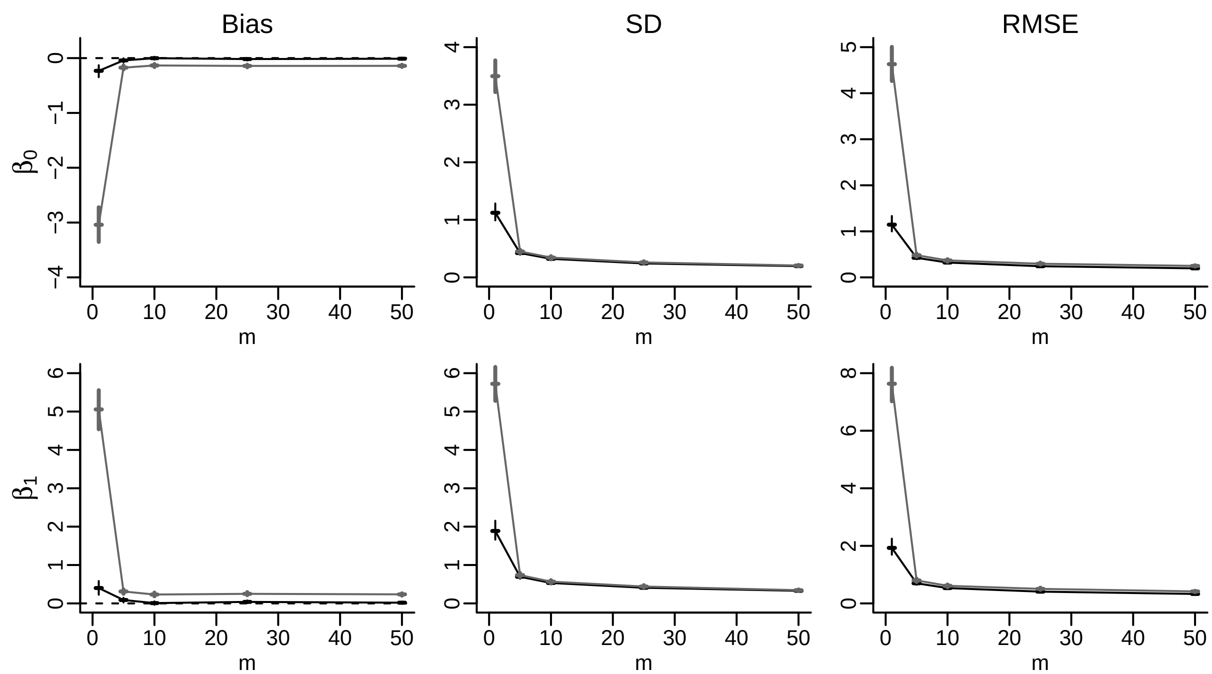

Figure S2: When $\sigma = 0.5$, the bias, SD, and RMSE of ML estimates for a marginally interpretable simple linear logistic GLMM (black lines) and a conventional simple linear logistic GLMM (gray lines). The vertical lines denote 95% bootstrap confidence intervals for each quantity.

of the conventional GLMM, we see significant non-zero biases in the conventional GLMM, especially when m is small and σ is larger. This is due in part to the true model being the marginally interpretable model in this case, but also due to the Kim Paradox where the estimation of $\boldsymbol{\beta}$ and σ is confounded for smaller values of m. In addition, the SD and RMSE of the ML estimates for the conventional model are higher than those for the marginally interpretable model.

We conclude that using the marginally interpretable model leads to more robust inferences without compromising our ability to carry out inference on the model parameters.

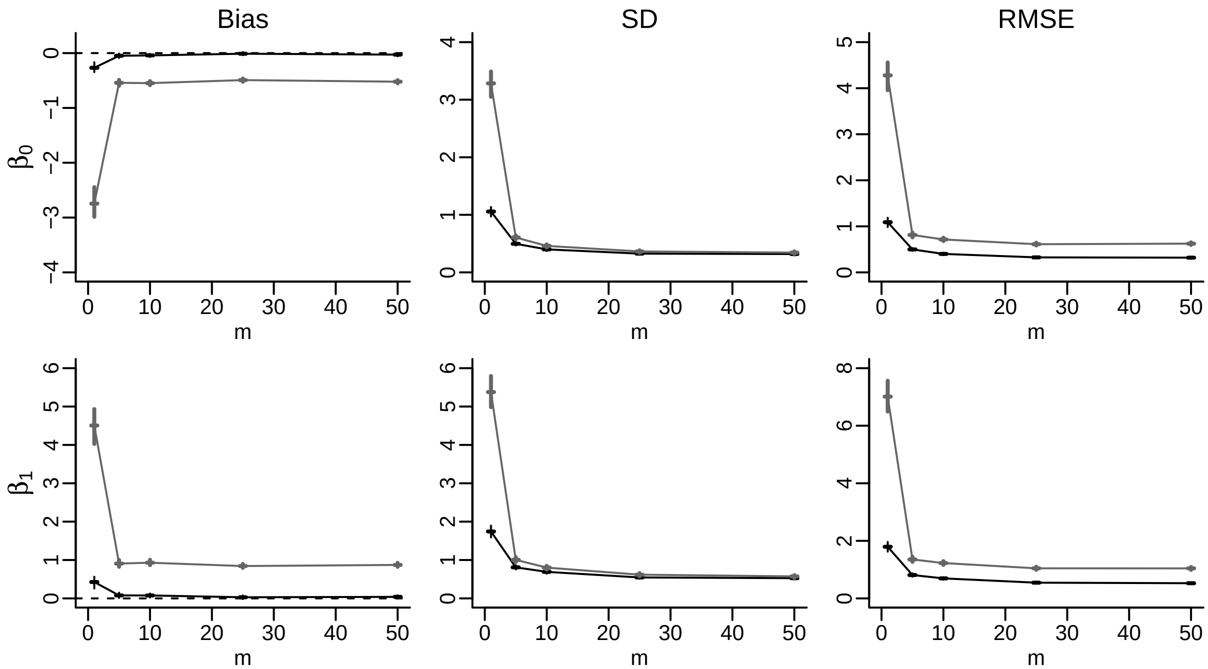

Figure S3: As Figure S2, but with $\sigma = 1$.

S4 Markov chain Monte Carlo algorithm for the rat teratology data

In Section 6, we sample from the posterior distribution of the unknown parameters in our model for the rat teratology data using an MCMC algorithm. Further details are provided below.

We sample from the target posterior via MCMC by iteratively updating the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{U})^T$, where $\boldsymbol{\beta}$ is the vector of fixed effects parameters, $\boldsymbol{\alpha} = (\sigma_1^2, \sigma_2^2)^T$ includes the variance components, and $\mathbf{U} = (U_{1,1}, \ldots, U_{2,16})^T$ includes the 32 latent variables associated with the random effect for litter. Our target posterior is

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{U}|\mathbf{Y}) \propto f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{U}) f_{\mathbf{U}}(\mathbf{U}|\sigma_1^2, \sigma_2^2) \pi_{\beta_0}(\beta_0) \pi_{\beta_1}(\beta_1) \pi_{\sigma_1}(\sigma_1^2) \pi_{\sigma_2}(\sigma_2^2),$$

where $f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{U})$ is a product of binomial densities given by

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{U}) = \prod_{i=1}^{2}\prod_{j=1}^{16} f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta},\sigma_i^2,U_{ij})$$
$$= \prod_{i=1}^{2}\prod_{j=1}^{16}\binom{m_{ij}}{Y_{ij}}(\beta_0+\beta_1 x_i+U_{ij}+a_i)^{Y_{ij}}(1-\beta_0-\beta_1 x_i-U_{ij}-a_i)^{(m_{ij}-Y_{ij})},$$

$f_{\mathbf{U}}(\mathbf{U}|\sigma_1^2,\sigma_2^2)$ is a product of normal densities given by

$$f_{\mathbf{U}}(\mathbf{U}|\sigma_1^2,\sigma_2^2) = \prod_{i=1}^{2}\prod_{j=1}^{16} f_U(U_{ij}|\sigma_i^2) = \prod_{i=1}^{2}\prod_{j=1}^{16}\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{1}{2\sigma_i^2}U_{ij}^2},$$

and $\pi_{\beta_0}(\cdot)$, $\pi_{\beta_1}(\cdot)$, $\pi_{\sigma_1}(\cdot)$, and $\pi_{\sigma_2}(\cdot)$ are the prior densities for β_0, β_1, σ_1^2, and σ_2^2, respectively. Note that we have assumed, a priori, that these parameters are independent of one another.

To sample from the target posterior we use Metropolis steps to iteratively produce draws from the full conditional distributions of the unknown parameters. The conditional posterior of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$, $\mathbf{U}$, and $\mathbf{Y}$ is proportional to

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{U})\pi_{\beta_0}(\beta_0)\pi_{\beta_1}(\beta_1),$$

and the conditional posterior of $\boldsymbol{\alpha} = (\sigma_1^2,\sigma_2^2)^T$ given $\boldsymbol{\beta}$, $\mathbf{U}$, and $\mathbf{Y}$ is proportional to

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{U})f_{\mathbf{U}}(\mathbf{U}|\sigma_1^2,\sigma_2^2)\pi_{\sigma_1}(\sigma_1^2)\pi_{\sigma_2}(\sigma_2^2).$$

Further, for each $i = 1,2$ and $j = 1,\ldots,16$, the conditional posterior of U_{ij} given $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\mathbf{U}_{-(ij)}$, and $\mathbf{Y}$ is proportional to

$$f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta},\sigma_i^2,U_{ij})f_U(U_{ij}|\sigma_i^2),$$

where $\mathbf{U}_{-(ij)}$ represents all elements of $\mathbf{U}$ except U_{ij}. Since the U_{ij} are conditionally independent, $\mathbf{U}_{-(ij)}$ does not enter this expression for the full conditional of U_{ij}. These full conditional distributions lead naturally to an MCMC algorithm that iteratively performs the following steps:

MCMC Algorithm

1. Propose $\boldsymbol{\beta}^{prop}$ given $\boldsymbol{\alpha}^{(t)}$ and $\mathbf{U}^{(t)}$. With probability

$$\min\left(1, \frac{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{prop}, \boldsymbol{\alpha}^{(t)}, \mathbf{U}^{(t)})\pi_{\beta_0}(\beta_0^{prop})\pi_{\beta_1}(\beta_1^{prop})}{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \mathbf{U}^{(t)})\pi_{\beta_0}(\beta_0^{(t)})\pi_{\beta_1}(\beta_1^{(t)})}\right),$$

   set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{prop}$. Otherwise, set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)}$;

2. Propose $\boldsymbol{\alpha}^{prop}$ given $\boldsymbol{\beta}^{(t+1)}$and $\mathbf{U}^{(t)}$. With probability

$$\min\left(1, \frac{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{prop}, \mathbf{U}^{(t)})f_{\mathbf{U}}(\mathbf{U}^{(t)}|\sigma_1^{2(prop)}, \sigma_2^{2(prop)})\pi_{\sigma_1}(\sigma_1^{2(prop)})\pi_{\sigma_2}(\sigma_2^{2(prop)})}{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t)}, \mathbf{U}^{(t)})f_{\mathbf{U}}(\mathbf{U}^{(t)}|\sigma_1^{2(t)}, \sigma_2^{2(t)})\pi_{\sigma_1}(\sigma_1^{2(t)})\pi_{\sigma_2}(\sigma_2^{2(t)})}\right),$$

   set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{prop}$. Otherwise, set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)}$;

3. For $i = 1, 2$ and $j = 1, \ldots, 16$, propose U_{ij}^{prop} given $\boldsymbol{\beta}^{(t+1)}$ and $\boldsymbol{\alpha}^{(t+1)}$. With probability

$$\min\left(1, \frac{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \sigma_i^{2(t+1)}, U_{ij}^{prop})f_U(U_{ij}^{prop}|\sigma_i^{2(t+1)})}{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \sigma_i^{2(t+1)}, U_{ij}^{(t)})f_U(U_{ij}^{(t)}|\sigma_i^{2(t+1)})}\right).$$

   set $U_{ij}^{(t+1)} = U_{ij}^{prop}$. Otherwise, set $U_{ij}^{(t+1)} = U_{ij}^{(t)}$.

At each stage of the MCMC algorithm we propose new values for the parameters given their current values. The proposed values are random draws from the following distributions:

$$\beta_0^{prop} \sim \mathrm{N}(\beta_0^{(t)}, 0.0625); \quad \beta_1^{prop} \sim \mathrm{N}(\beta_1^{(t)}, 0.0625)$$
$$\log(\sigma_1^{2(prop)}) \sim \mathrm{N}\{\log(\sigma_1^{2(t)}), 0.25\}; \quad \log(\sigma_2^{2(prop)}) \sim \mathrm{N}\{\log(\sigma_2^{2(t)}), 0.25\};$$
$$U_{ij}^{prop} \sim \mathrm{N}(U_{ij}^{(t)}, 1).$$

S5 Epileptic seizures

Our second example comes from a clinical trial of 59 epileptics conducted by Leppik et al. (1987). Each subject received either a placebo or the drug progabide and then made four successive follow-up visits to the clinic during which they reported the number of partial seizures they had suffered in the two-week period immediately preceding the visit. We denote these reported counts by Y_{ij}, where $i = 1, \ldots, 59$ indexes the subjects and $j = 1, 2, 3, 4$ indexes the visits. Thall and Vail (1990) used GEE to fit a marginal model to these data. They included as predictors the logarithm of one-fourth of the baseline count of partial seizures suffered by each patient in the eight-week period prior to treatment (denoted BASE_i), a treatment indicator (1 if progabide, 0 if placebo, denoted TRT_i), the interaction between BASE_i and TRT_i, the logarithm of the subject's age in years (denoted AGE_i), and a fourth-visit indicator (1 for the subject's fourth post-treatment visit, 0 otherwise, denoted VISIT4_j). Breslow and Clayton (1993) and Gamerman (1997) fit a GLMM with the same fixed effects and also two levels of random effects. Their model has the form

$$\begin{aligned} \mathrm{E}(Y_{ij}|\boldsymbol{\beta}, \gamma_i, \delta_{ij}) = \exp\big(&\beta_0 + \beta_1 \times (\text{BASE}_i) + \beta_2 \times (\text{TRT}_i) + \beta_3 \times (\text{BASE}_i * \text{TRT}_i) + \\ &\beta_4 \times (\text{AGE}_i) + \beta_5 \times (\text{VISIT4}_j) + \gamma_i + \delta_{ij}\big), \end{aligned} \tag{S2}$$

where the $\gamma_i \overset{ind}{\sim} \mathrm{N}(0, \sigma^2)$ are random subject effects, the $\delta_{ij} \overset{ind}{\sim} \mathrm{N}(0, \tau^2)$ are random effects for visit within subject, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_5)^T$ is the vector of fixed effects parameters. Further, conditional on the random effects γ_i and δ_{ij}, the reported seizure counts are assumed to be independent observations from a Poisson$\{\mathrm{E}(Y_{ij}|\boldsymbol{\beta}, \gamma_i, \delta_{ij})\}$.

We adopt a Bayesian approach and sample from a mixed model analogous to (S2), but include an adjustment to ensure that the model is marginally interpretable. In light of Section 3.2, the adjustment is simply $a_{ij} = -\sigma^2/2 - \tau^2/2$ for all i and j. As with the rats teratology example presented in Section 6 of the main document, to clearly distinguish between different parameterizations, let $\boldsymbol{\beta}^*$ denote the marginal parameters in a marginally interpretable GLMM and $\boldsymbol{\beta}$ denote the cluster-specific parameters in a conventional GLMM. We also code the treatment effect as $\text{TRT}_i = 1$ for progabide and as $\text{TRT}_i = -1$ for placebo to ensure that the two treatment groups are on the same footing in terms of variance. We assume, a priori, that $\boldsymbol{\beta}^*$, σ^2, and τ^2 are independent of one another, and place $\mathrm{N}_6(\mathbf{0}, 4\mathbf{I}_6)$, $\mathrm{N}(-1, 2)$, and $\mathrm{N}(-1, 2)$ prior distributions on $\boldsymbol{\beta}^*$, $\log(\sigma^2)$, and $\log(\tau^2)$, respectively. The priors on the variance components reflect our belief that there is little subject-to-subject and visit-to-visit variation, whereas the priors on the fixed effects parameters are meant to be noninformative while also not putting too much mass on unreasonably large values for the expected seizure count. To sample from our target posterior using MCMC, the vector of parameters we must update is $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta})^T$, where $\boldsymbol{\beta}^*$ is the vector of fixed effects parameters, $\boldsymbol{\alpha} = (\sigma^2, \tau^2)^T$ includes the parameters characterizing the random effects distribution, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{59})^T$ includes the 59 latent variables associated with the subject random effect, and $\boldsymbol{\delta} = (\delta_{1,1}, \ldots, \delta_{59,4})^T$ includes the 236 latent variables associated with the visit random effect. Due to the presence of 295 latent variables in this model, proposals for $\boldsymbol{\beta}^*$ are rarely accepted when we use a basic MCMC algorithm that employs Metropolis steps to update the parameters in blocks.

To address the problem with slow mixing, we simultaneously propose $\boldsymbol{\gamma}^{prop}$ and $\boldsymbol{\delta}^{prop}$ to be consistent with each proposed $\boldsymbol{\beta}^{prop}$ as described in Section 5.3. Specifically, for each $\boldsymbol{\beta}^{prop}$ we also propose the following γ_i^{prop} and δ_{ij}^{prop} for each $i = 1, \ldots, 59$ and $j = 1, 2, 3, 4$:

$$\gamma_i^{prop} = \gamma_i^{(t)} + \mathbf{x}_{i,1}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop}), \tag{S3}$$

$$\delta_{i,4}^{prop} = \delta_{i,4}^{(t)} + (\mathbf{x}_{i,4}^T - \mathbf{x}_{i,1}^T)(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop}), \text{ and } \delta_{ij}^{prop} = \delta_{ij}^{(t)} \text{ for } j = 1, 2, 3. \tag{S4}$$

We then choose to accept or reject $\boldsymbol{\beta}^{prop}$, $\boldsymbol{\gamma}^{prop}$, and $\boldsymbol{\delta}^{prop}$ collectively and set $\boldsymbol{\beta}^{(t+1)}$, $\boldsymbol{\gamma}'$, and $\boldsymbol{\delta}'$ accordingly. The intermediate states $\boldsymbol{\gamma}'$ and $\boldsymbol{\delta}'$ are used in place of $\boldsymbol{\gamma}^{(t)}$ and $\boldsymbol{\delta}^{(t)}$ until $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are formally updated. Proposing random effects to be consistent with the fixed effects in this manner increases the acceptance rate for $\boldsymbol{\beta}^{prop}$ from 12.9% to 50.7%. Using this improved proposal scheme also decreases the integrated autocorrelation time for β_1^*, which represents the effect for the baseline count and has the highest such value among the six fixed effects parameters, from 642.7 to 161.9. We carried out our modified MCMC algorithm both for the marginally interpretable model and the conventional GLMM. We ran each chain for

Table S1: Posterior means of the model parameters for the epileptic seizures data with corresponding posterior standard deviations in parentheses.

Parameter	Marginally Interpretable GLMM	Conventional GLMM
β_0^*, β_0	-1.15 (1.09)	-1.29 (1.08)
β_1^*, β_1	1.04 (0.10)	1.04 (0.11)
β_2^*, β_2	-0.45 (0.21)	-0.45 (0.21)
β_3^*, β_3	0.16 (0.11)	0.16 (0.10)
β_4^*, β_4	0.33 (0.31)	0.32 (0.31)
β_5^*, β_5	-0.10 (0.09)	-0.10 (0.09)
σ	0.50 (0.07)	0.50 (0.07)
τ	0.37 (0.04)	0.37 (0.04)

2,100,000 steps, discarding the first 100,000 steps as burn-in and retaining every 200^{th} step thereafter to obtain a final sample of 10,000 draws from the posterior distribution for each model.

Table S1 displays posterior means and standard deviations for the parameters in both the marginally interpretable model and the conventional GLMM. With the exception of the intercept (β_0^* or β_0), the two sets of parameter estimates are virtually identical. Breslow and Clayton (1993) noted that the slope parameters in this model have both a marginal and conditional interpretation while Ritz and Spiegelman (2004) stated that this will generally be the case for a model with a log link and a random intercept that is independent of the covariates in the model. The intercept for the marginally interpretable model is greater than the intercept for the conventional GLMM due to the tendency of the convex inverse link function to pull the marginal mean up.

With the marginally interpretable GLMM, we can estimate the marginal mean for the entire population by computing $\exp(\mathbf{x}^T\boldsymbol{\beta}^*)$ and can make a prediction for a new observation on an individual in the sample by computing $\exp(\mathbf{x}^T\boldsymbol{\beta}^* + \gamma + \delta + \mathbf{d}^T\mathbf{a})$. A marginal model fit via GEE should yield fixed effects parameter estimates that are similar to the posterior means reported in Table S1 for the marginally interpretable model, but obtaining subject-specific predictions would be much more difficult with a purely marginal model.

Similarly, the conventional GLMM should yield individual-level predictions identical to those of the marginally interpretable model, but estimating the marginal mean using the conventional GLMM would be more difficult. For a marginally interpretable GLMM, the average expected seizure count across all subjects in the population with a particular set of covariates is simply $\mathrm{E}(Y|\boldsymbol{\beta}^*, \boldsymbol{\alpha}) = \exp(\mathbf{x}^T\boldsymbol{\beta}^*)$, whereas for a conventional GLMM, the marginal mean is $\mathrm{E}(Y|\boldsymbol{\beta}, \boldsymbol{\alpha}) = \int \exp(\mathbf{x}^T\boldsymbol{\beta} + \mathbf{d}^T\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}$. Note that the expression for the marginal mean in the marginally interpretable model does not functionally depend on the random effects distribution while the same expression for the conventional GLMM does. Consequently, the marginally interpretable GLMM should be less sensitive to perturbations of the random effects distribution, and estimates of the fixed effects parameters based on the marginally interpretable model should be more stable across different samples from the same population.

S5.1 Markov chain Monte Carlo for the epilepsy data

We sample from the target posterior via MCMC by iteratively updating the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta})^T$, where $\boldsymbol{\beta}$ is the vector of fixed effects parameters, $\boldsymbol{\alpha} = (\sigma^2, \tau^2)^T$ includes the parameters characterizing the random effects distribution, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{59})^T$ includes the 59 latent variables associated with the subject random effect, and $\boldsymbol{\delta} = (\delta_{1,1}, \ldots, \delta_{59,4})^T$ includes the 236 latent variables associated with the visit random effect. Our target posterior is

$$\pi_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{Y}) \propto f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\sigma^2) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}|\tau^2) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \pi_{\sigma}(\sigma^2) \pi_{\tau}(\tau^2),$$

where $f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ is a product of Poisson densities given by

$$\begin{aligned} f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) &= \prod_{i=1}^{59} \prod_{j=1}^{4} f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}, \sigma^2, \tau^2, \gamma_i, \delta_{ij}) \\ &= \prod_{i=1}^{59} \prod_{j=1}^{4} \frac{1}{Y_{ij}!} e^{-\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \gamma_i + \delta_{ij} - \frac{\sigma^2}{2} - \frac{\tau^2}{2}\right)} \left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \gamma_i + \delta_{ij} - \frac{\sigma^2}{2} - \frac{\tau^2}{2}\right)^{Y_{ij}}, \end{aligned}$$

$f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\sigma^2)$ is a product of normal densities given by

$$f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\sigma^2) = \prod_{i=1}^{59} f_{\gamma}(\gamma_i|\sigma^2) = \prod_{i=1}^{59} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\gamma_i^2},$$

$f_{\boldsymbol{\delta}}(\boldsymbol{\delta}|\tau^2)$ is a product of normal densities given by

$$f_{\boldsymbol{\delta}}(\boldsymbol{\delta}|\tau^2) = \prod_{i=1}^{59}\prod_{j=1}^{4} f_{\delta}(\delta_{ij}|\tau^2) = \prod_{i=1}^{59}\prod_{j=1}^{4} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}\delta_{ij}^2},$$

and $\pi_{\boldsymbol{\beta}}$, π_{σ}, and π_{τ} are the prior densities for $\boldsymbol{\beta}$, σ^2, and τ^2, respectively.

A standard approach for sampling from this posterior involves using Metropolis steps to iteratively produce draws from the full conditional distributions of the unknown parameters. The conditional posterior of $\boldsymbol{\alpha} = (\sigma^2, \tau^2)^T$ given $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\mathbf{Y}$ is proportional to

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\sigma^2) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}|\tau^2) \pi_{\sigma}(\sigma^2) \pi_{\tau}(\tau^2),$$

and the conditional posterior of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\mathbf{Y}$ is proportional to

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}).$$

Further, for each $i = 1, \ldots, 59$, if we define $\boldsymbol{\gamma}_{-i}$ as all elements of $\boldsymbol{\gamma}$ except γ_i, then the conditional posterior of γ_i given $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma_{-i}}$, $\boldsymbol{\delta}$, and $\mathbf{Y}$ is proportional to

$$f_{\gamma}(\gamma_i|\sigma^2) \prod_{j=1}^{4} f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}, \sigma^2, \tau^2, \gamma_i, \delta_{ij}),$$

Since the γ_i are conditionally independent, $\boldsymbol{\gamma}_{-i}$ does not enter this expression for the full conditional of γ_i. Finally, for each $i = 1, \ldots, 59$ and $j = 1, 2, 3, 4$ the conditional posterior of δ_{ij} given $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}_{-(ij)}$, and $\mathbf{Y}$ is proportional to

$$f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}, \sigma^2, \tau^2, \gamma_i, \delta_{ij}) f_{\delta}(\delta_{ij}|\tau^2),$$

where $\boldsymbol{\delta}_{-(ij)}$ represents all elements of $\boldsymbol{\delta}$ except δ_{ij}. Since the δ_{ij} are conditionally independent, $\boldsymbol{\delta}_{-(ij)}$ does not enter this expression for the full conditional of δ_{ij}. These full conditional distributions lead naturally to the MCMC algorithm on the next page:

Basic MCMC Algorithm

1. Propose $\boldsymbol{\alpha}^{prop}$ given $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\gamma}^{(t)}$, and $\boldsymbol{\delta}^{(t)}$. With probability

$$\min\left(1, \frac{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{prop}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(t)}|\sigma^2_{prop}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{(t)}|\tau^2_{prop}) \pi_\sigma(\sigma^2_{prop}) \pi_\tau(\tau^2_{prop})}{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(t)}|\sigma^2_{(t)}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{(t)}|\tau^2_{(t)}) \pi_\sigma(\sigma^2_{(t)}) \pi_\tau(\tau^2_{(t)})}\right)$$

set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{prop}$. Otherwise, set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)}$;

2. Propose $\boldsymbol{\beta}^{prop}$ given $\boldsymbol{\alpha}^{(t+1)}$,$\boldsymbol{\gamma}^{(t)}$, $\boldsymbol{\delta}^{(t)}$. With probability

$$\min\left(1, \frac{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{prop}, \boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{prop})}{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{(t)})}\right)$$

set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{prop}$. Otherwise, set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)}$;

3. For $i = 1, \ldots, 59$, propose γ_i^{prop} given $\boldsymbol{\beta}^{(t+1)}$, $\boldsymbol{\alpha}^{(t+1)}$, and $\boldsymbol{\delta}^{(t)}$. With probability

$$\min\left(1, \frac{f_\gamma(\gamma_i^{prop}|\sigma^2_{(t+1)}) \prod_{j=1}^4 f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{prop}, \delta_{ij}^{(t)})}{f_\gamma(\gamma_i^{(t)}|\sigma^2_{(t+1)}) \prod_{j=1}^4 f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{(t)}, \delta_{ij}^{(t)})}\right)$$

set $\boldsymbol{\gamma}_i^{(t+1)} = \boldsymbol{\gamma}_i^{prop}$. Otherwise, set $\boldsymbol{\gamma}_i^{(t+1)} = \boldsymbol{\gamma}_i^{(t)}$;

4. For $i = 1, \ldots, 59$ and $j = 1, 2, 3, 4$, propose δ_{ij}^{prop} given $\boldsymbol{\beta}^{(t+1)}$, $\boldsymbol{\alpha}^{(t+1)}$, and $\boldsymbol{\gamma}^{(t+1)}$. With probability

$$\min\left(1, \frac{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{(t+1)}, \delta_{ij}^{prop}) f_\delta(\delta_{ij}^{prop}|\tau^2_{(t+1)})}{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{(t+1)}, \delta_{ij}^{(t)}) f_\delta(\delta_{ij}^{(t)}|\tau^2_{(t+1)})}\right)$$

set $\boldsymbol{\delta}_{ij}^{(t+1)} = \boldsymbol{\delta}_{ij}^{prop}$. Otherwise, set $\boldsymbol{\delta}_{ij}^{(t+1)} = \boldsymbol{\delta}_{ij}^{(t)}$.

For each step in this algorithm, we propose new values for the parameters given their current values. The proposed values are random draws from the following distributions:

$$\boldsymbol{\beta}^{prop} \sim \mathrm{N}_6(\boldsymbol{\beta}^{(t)}, \mathbf{V});$$
$$\log(\sigma^2_{prop}) \sim \mathrm{N}\big(\log(\sigma^2_{(t)}), 0.16\big); \quad \log(\tau^2_{prop}) \sim \mathrm{N}\big(\log(\tau^2_{(t)}), 0.0625\big);$$
$$\gamma_i^{prop} \sim \mathrm{N}(\gamma_i^{(t)}, 0.25); \quad \delta_{ij}^{prop} \sim \mathrm{N}(\delta_{ij}^{(t)}, 0.25).$$

The covariance matrix $\mathbf{V}$ for proposing $\boldsymbol{\beta}^{prop}$ is calculated using the weight matrix obtained from fitting an analogous fixed effects model with iteratively reweighted least squares. Specifically,

$$\mathbf{V} = \begin{bmatrix} 0.1735 & -0.0061 & 0.0068 & -0.0041 & -0.0476 & -0.0007 \\ -0.0061 & 0.0010 & -0.0002 & 0.0001 & 0.0011 & 0.0000 \\ 0.0068 & -0.0002 & 0.0061 & -0.0024 & -0.0019 & 0.0000 \\ -0.0041 & 0.0001 & -0.0024 & 0.0010 & 0.0012 & 0.0000 \\ -0.0476 & 0.0011 & -0.0019 & 0.0012 & 0.0136 & 0.0000 \\ -0.0007 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0030 \end{bmatrix}.$$

Since our model includes 295 latent variables, this basic MCMC algorithm is plagued by the issues with slow mixing discussed in Section 5.3. We ran this algorithm for 2,100,000 steps, discarding the first 100,000 steps as burn-in. Only 12.9% of the proposals for $\boldsymbol{\beta}$ were accepted, and there is a high degree of autocorrelation in the Markov chains for the fixed effects parameters. Autocorrelation plots for $\beta_0^*, \ldots, \beta_5^*$ are shown in Figure S4. To address this problem, we simultaneously propose $\boldsymbol{\gamma}^{prop}$ and $\boldsymbol{\delta}^{prop}$ to be consistent with each proposed $\boldsymbol{\beta}^{prop}$. We define our proposals for the random effects in (S3) and (S4).

The simultaneous proposal of the fixed and random effects has no net impact on the likelihood. The conditional density of Y_{ij} given μ_{ij} is Poisson(μ_{ij}), where

$$\mu_{ij} = \mathrm{E}(Y_{ij}|\gamma_i, \delta_{ij}) = \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \gamma_i + \delta_{ij} + a).$$

Further, $\mathbf{x}_{i,1} = \mathbf{x}_{i,2} = \mathbf{x}_{i,3}$ because only the fourth-visit indicator varies within a subject. Defining γ_i^{prop} and δ_{ij}^{prop} as in (S3) and (S4), for $j = 1, 2, 3$

$$\begin{aligned} & \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \gamma_i^{prop} + \delta_{ij}^{prop} + a \\ &= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \{\gamma_i^{(t)} + \mathbf{x}_{ij}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^*)\} + \delta_{ij}^{(t)} + a \\ &= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} - \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \mathbf{x}_{ij}^T\boldsymbol{\beta}^{(t)} + \gamma_i^{(t)} + \delta_{ij}^{(t)} + a \\ &= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{(t)} + \gamma_i^{(t)} + \delta_{ij}^{(t)} + a, \end{aligned}$$

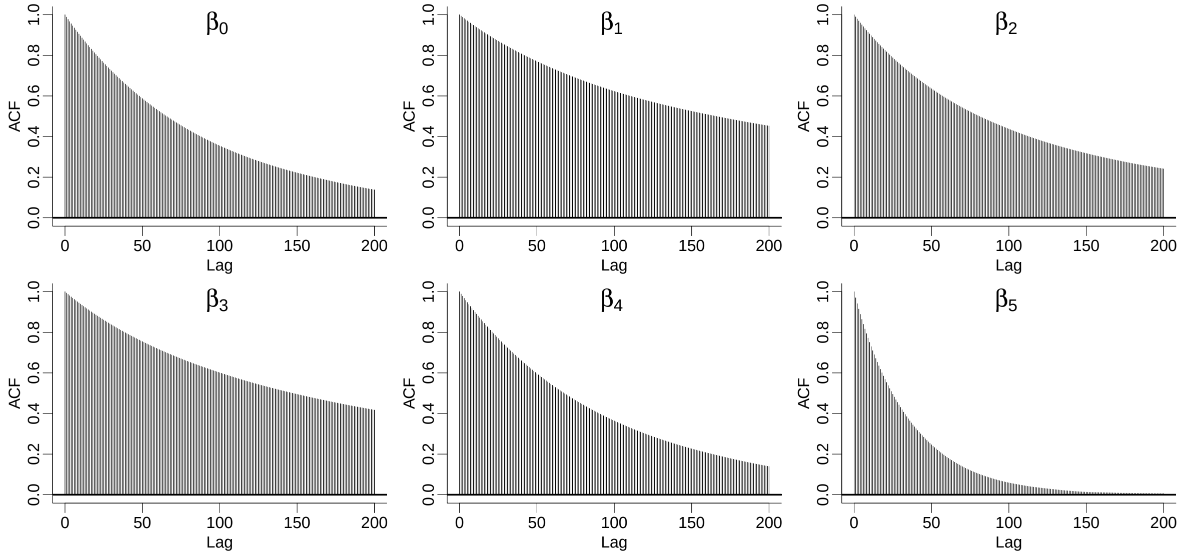

Figure S4: Autocorrelation plots for $\beta_0^*, \ldots, \beta_5^*$ for a basic MCMC algorithm for the epileptic seizures data.

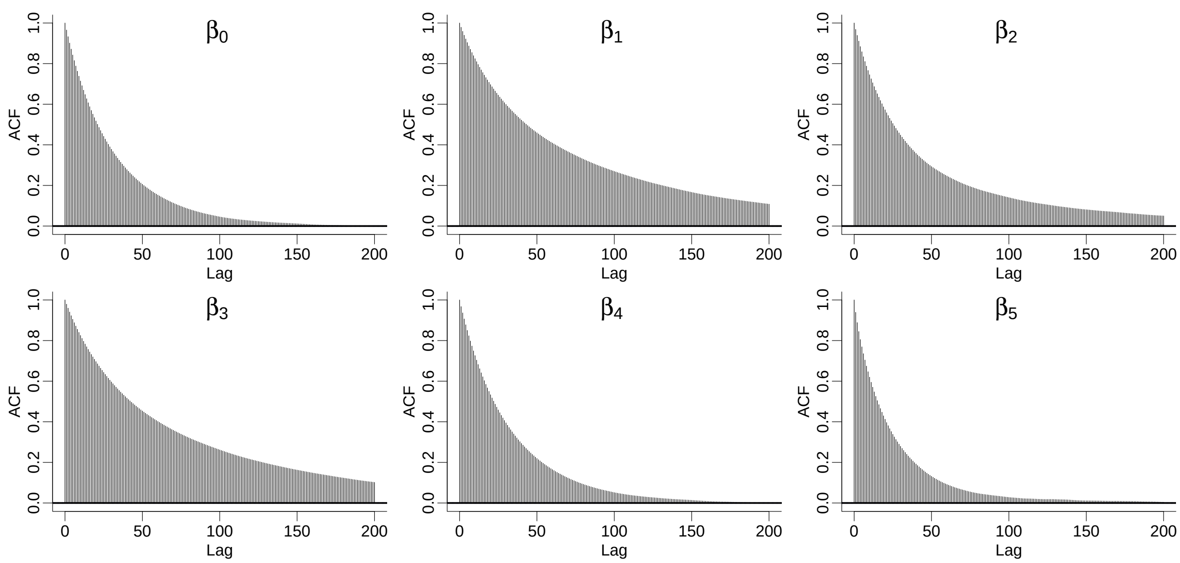

Figure S5: Autocorrelation plots for $\beta_0^*, \ldots, \beta_5^*$ for the modified MCMC algorithm for the epileptic seizures data.

and for $j = 4$

$$
\begin{aligned}
&\mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \gamma_i^{prop} + \delta_{ij}^{prop} + a \\
&= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \{\gamma_i^{(t)} + \mathbf{x}_{i,1}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop})\} + \{\delta_{ij}^{(t)} + (\mathbf{x}_{ij}^T - \mathbf{x}_{i,1}^T)(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop})\} + a \\
&= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \mathbf{x}_{i,1}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop}) - \mathbf{x}_{i,1}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop}) + \mathbf{x}_{ij}^T(\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{prop}) + \gamma_i^{(t)} + \delta_{ij}^{(t)} + a \\
&= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} - \mathbf{x}_{ij}^T\boldsymbol{\beta}^{prop} + \mathbf{x}_{ij}^T\boldsymbol{\beta}^{(t)} + \gamma_i^{(t)} + \delta_{ij}^{(t)} + a \\
&= \mathbf{x}_{ij}^T\boldsymbol{\beta}^{(t)} + \gamma_i^{(t)} + \delta_{ij}^{(t)} + a.
\end{aligned}
$$

Consequently, the conditional mean $\mathrm{E}(Y_{ij}|\boldsymbol{\beta}, \gamma_i, \delta_{ij})$ is the same for both the current state and the proposed state. In turn, the conditional density $f_{\mathbf{Y}|\boldsymbol{\theta}}$ is also the same for both states. Noting that the conditional posterior of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})^T$ given $\boldsymbol{\alpha}$ and $\mathbf{Y}$ is proportional to

$$f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\sigma^2) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}|\tau^2) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}),$$

the acceptance probability for $(\boldsymbol{\beta}^{prop}, \boldsymbol{\gamma}^{prop}, \boldsymbol{\delta}^{prop})^T$ depends entirely on $\pi_{\boldsymbol{\beta}}$, $f_{\boldsymbol{\gamma}}$, and $f_{\boldsymbol{\delta}}$ because $f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{prop}, \boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\gamma}^{prop}, \boldsymbol{\delta}^{prop}) = f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)})$. This joint proposal scheme leads to the modified MCMC algorithm on the next page:

Modified MCMC Algorithm

1. Propose $\boldsymbol{\alpha}^{prop}$ given $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\gamma}^{(t)}$, and $\boldsymbol{\delta}^{(t)}$. With probability

$$\min\left(1, \frac{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{prop}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(t)}|\sigma^2_{prop}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{(t)}|\tau^2_{prop}) \pi_\sigma(\sigma^2_{prop}) \pi_\tau(\tau^2_{prop})}{f_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}|\boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\delta}^{(t)}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(t)}|\sigma^2_{(t)}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{(t)}|\tau^2_{(t)}) \pi_\sigma(\sigma^2_{(t)}) \pi_\tau(\tau^2_{(t)})}\right)$$

   set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{prop}$. Else, set $\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)}$;

2. Propose $(\boldsymbol{\beta}^{prop}, \boldsymbol{\gamma}^{prop}, \boldsymbol{\delta}^{prop})^T$ given $\boldsymbol{\alpha}^{(t+1)}$, with $\boldsymbol{\beta}^{prop}$ being drawn from a proposal distribution and $\boldsymbol{\gamma}^{prop}$ and $\boldsymbol{\delta}^{prop}$ defined as in (S3) and (S4). With probability

$$\min\left(1, \frac{f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{prop}|\sigma^2_{(t+1)}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{prop}|\tau^2_{(t+1)}) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{prop})}{f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(t)}|\sigma^2_{(t+1)}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}^{(t)}|\tau^2_{(t+1)}) \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{(t)})}\right)$$

   set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{prop}$, $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^{prop}$, and $\boldsymbol{\delta}' = \boldsymbol{\delta}^{prop}$. Else, set $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)}$, $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^{(t)}$, and $\boldsymbol{\delta}' = \boldsymbol{\delta}^{(t)}$;

3. For $i = 1, \ldots, 59$, propose γ_i^{prop} given $\boldsymbol{\beta}^{(t+1)}$, $\boldsymbol{\alpha}^{(t+1)}$, and $\boldsymbol{\delta}'$. With probability

$$\min\left(1, \frac{f_\gamma(\gamma_i^{prop}|\sigma^2_{(t+1)}) \prod_{j=1}^4 f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{prop}, \delta'_{ij})}{f_\gamma(\gamma'_i|\sigma^2_{(t+1)}) \prod_{j=1}^4 f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma'_i, \delta'_{ij})}\right)$$

   set $\boldsymbol{\gamma}_i^{(t+1)} = \boldsymbol{\gamma}_i^{prop}$. Else, set $\boldsymbol{\gamma}_i^{(t+1)} = \boldsymbol{\gamma}'_i$;

4. For $i = 1, \ldots, 59$ and $j = 1, 2, 3, 4$, propose δ_{ij}^{prop} given $\boldsymbol{\beta}^{(t+1)}$, $\boldsymbol{\alpha}^{(t+1)}$, and $\boldsymbol{\gamma}^{(t+1)}$. With probability

$$\min\left(1, \frac{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{(t+1)}, \delta_{ij}^{prop}) f_\delta(\delta_{ij}^{prop}|\tau^2_{(t+1)})}{f_{Y|\boldsymbol{\theta}}(Y_{ij}|\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\alpha}^{(t+1)}, \gamma_i^{(t+1)}, \delta'_{ij}) f_\delta(\delta'_{ij}|\tau^2_{(t+1)})}\right)$$

   set $\boldsymbol{\delta}_{ij}^{(t+1)} = \boldsymbol{\delta}_{ij}^{prop}$. Else, set $\boldsymbol{\delta}_{ij}^{(t+1)} = \boldsymbol{\delta}'_{ij}$.

The key difference between this modified algorithm and the basic algorithm is in Step 2. Here, in addition to updating $\boldsymbol{\beta}^{(t)}$ to $\boldsymbol{\beta}^{(t+1)}$, we update $\boldsymbol{\gamma}^{(t)}$ and $\boldsymbol{\delta}^{(t)}$ to the intermediate states $\boldsymbol{\gamma}'$ and $\boldsymbol{\delta}'$. Steps 3 and 4 then update $\boldsymbol{\gamma}'$ and $\boldsymbol{\delta}'$ to $\boldsymbol{\gamma}^{(t+1)}$ and $\boldsymbol{\delta}^{(t+1)}$ in a manner similar to Steps 3 and 4 in the basic MCMC algorithm.

Table S2: Integrated autocorrelation times (before thinning) for the fixed effects parameters in the marginally interpretable model for the epileptic seizures data for both the basic MCMC algorithm and the modified MCMC algorithm

Parameter	Basic MCMC	Modified MCMC	Ratio
β_0^*	205.3	62.8	3.27
β_1^*	642.7	161.9	3.97
β_2^*	355.2	105.0	3.38
β_3^*	565.6	168.0	3.37
β_4^*	204.2	65.6	3.11
β_5^*	71.5	51.0	1.40

Using the modified MCMC algorithm instead of the basic MCMC algorithm increases the acceptance rate for $\boldsymbol{\beta}$ from 12.9% to 50.7%. It also decreases the *integrated autocorrelation times* for the fixed effects parameters, as summarized in Table S2. The integrated autocorrelation time of a parameter provides a measure of the average number of iterations required to obtain approximately independent draws from the posterior distribution of that parameter. For $\beta_0^*, \ldots, \beta_4^*$, this quantity is more than three times larger using the basic MCMC algorithm versus using the modified MCMC algorithm. Thus, the strategy described in Section 5.3 successfully reduces the autocorrelation and allows us to obtain a representative sample from the target posterior with fewer steps of the Markov chain. As further illustration of the improved mixing, Figure S5 (displayed below Figure S4 for comparison) shows autocorrelation plots for $\beta_0^*, \ldots, \beta_5^*$ using the modified MCMC algorithm. Although some autocorrelation remains, even at a lag of 200, it is not as strong as the autocorrelation observed in Figure S4 for the basic MCMC algorithm.

References

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

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York, NY.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40:1–18.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7:57–68.

Leppik, I., Dreifuss, F., Porter, R., Bowman, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J., Stackman, J., Graves, N., Sutula, T., Welty, T., Vickery, J., Brundage, R., Gates, J., Gumnit, R., and Gutierrez, A. (1987). A controlled study of progabide in partial seizures: Methodology and results. *Neurology*, 37:963–968.

Monahan, J. F. and Stefanski, L. A. (1992). Normal scale mixture approximations to f*(z) and computation of the logistic-normal integral. In Balakrishnan, N., editor, *Handbook of the Logistic Distribution*, pages 529–540. Marcel Dekker, New York, NY.

Pirjol, D. (2013). The logistic-normal integral and its generalizations. *Journal of Computational and Applied Mathematics*, 237:460–469.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ritz, J. and Spiegelman, D. (2004). Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research*, 13:309–323.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46:657–671.