Johns Hopkins University, Dept. of Biostatistics Working Papers

11-15-2004

ON MARGINALIZED MULTILEVEL MODELS AND THEIR COMPUTATION

Michael E. Griswold

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, mgriswol@jhsph.edu

Scott L. Zeger

The Johns Hopkins Bloomberg School of Public Health, szeger@jhsph.edu

On Marginalized Multilevel Models and Their Computation

Michael E. Griswold*

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. 21205, U.S.A.

and

Scott L. Zeger

November 15, 2004

SUMMARY. Clustered data analysis is characterized by the need to describe both systematic variation in a mean model and cluster-dependent random variation in an association model. Marginalized multilevel models embrace the robustness and interpretations of a marginal mean model, while retaining the likelihood inference capabilities and flexible dependence structures of a conditional association model. Although there has been increasing recognition of the attractiveness of marginalized multilevel models, there has been a gap in their practical application arising from a lack of readily available estimation procedures. We extend the marginalized multilevel model to allow for nonlinear functions in both the mean and association aspects. We then formulate marginal models through conditional specifications to facilitate estimation with mixed model computational solutions already in place. We illustrate this approach on a cerebrovascular deficiency crossover trial.

* *email:* mgriswol@jhsph.edu

KEY WORDS: Marginal Model; Generalized Linear Mixed Model; Nonlinear Mixed Model; Latent Variable; Random Effects; Likelihood Inference.

1. Introduction

Models for longitudinal and other clustered data must describe systematic variation in the mean response as well as associations among observations within clusters. Statistical approaches that have received considerable attention in addressing these objectives include generalized linear mixed models (GLMM) (Laird and Ware, 1982; Zeger and Karim, 1991; Breslow and Clayton, 1993; McCulloch and Searle, 2001; Goldstein, 2002; Demidenko, 2004), marginal models fit with generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger, Liang and Albert, 1988), and more recently, marginalized multilevel models (MMM) (Heagerty and Zeger, 2000; Diggle, Heagerty, Liang and Zeger, 2002; Mills, Field and Dupuis, 2002; Miglioretti and Heagerty, 2004). Choice of statistical approach depends on both the primary research question and method availability. While much debate has centered on contrasting the marginal vs conditional approaches, a simple resolution is to use both as appropriate. MMMs demonstrate that latent structures may be used for either purpose. When primary interest focuses on conditional effects, the MMM may be used to examine implications of the assumed latent process on the observable marginal responses. When marginal effects are of primary concern, the MMM may be used for a variety of functions: 1) to define a full joint distribution for likelihood-based inference, 2) to relax the MCAR missing data assumptions of GEE methods, and 3) to investigate underlying contributions to the association structure, which may also be of substantive interest.

The application of marginalized multilevel models has been impeded by a lack of available computational methods for estimation. We reformulate the MMM to make connections between marginal and conditional models transparent, and then construct marginalized models in terms of their conditional model counterparts. Computational techniques devoted to

mixed model estimation have received widespread attention and there exist considerable resources for fitting these models. Constructing marginal models through conditional specifications allows direct estimation of MMMs with powerful mixed model computing solutions already in place. The technique is illustrated on a cerebrovascular deficiency crossover trial. We present three exact and three approximate estimation approaches based on the application.

2. Marginalized Multilevel Models

Random effects models (such as GLMMs) are applied to clustered data by specifying a mean model that is conditioned on a set of latent 'random' effects. The latent effects are conceived as embodying sources from which the within-cluster associations arise. GLMMs have many advantages, including the ability to work within a likelihood framework, having cluster specific regression coefficients, flexibility in specifying within-cluster dependence mechanisms, and valid inferences under missing at random (MAR) dropout mechanisms. Drawbacks of GLMMs include sensitivity of regression coefficients to association structure assumptions and, in many problems, regression parameter interpretations being conditional on unobservable effects. (Diggle et al., 2002).

Marginal models are an alternative in which the mean and association structures are separated. Their regression coefficients have standard generalized linear model (GLM) interpretations and inferences about them are less sensitive to association structure assumptions. Marginal models are often estimated by solving generalized estimating equations (GEE), (Liang and Zeger, 1986; Zeger et al., 1988). While there are many advantages to using estimating functions, one considerable disadvantage is that likelihood-based methods are sacrificed, as the complete joint distribution of the observations remains unspecified. This has many implications. When estimating functions are used, data that are not missing completely at random must be addressed with inverse probability weighting techniques (Robins,

Rotnitzky and Zhao, 1995; Scharfstein, Rotnitzky and Robins, 1999) or a similar strategy. For estimation, likelihood-based methods tend to assure a unique maximum where estimating equations can have multiple roots. Likelihoods are the building blocks of Bayesian methods and are integral to shrinkage estimation. Evidential methodology also relies on likelihood methods (Birnbaum, 1962; Edwards, 1992; Royall, 1999; Blume, 2002).

Marginalized multilevel models embrace the interpretation and robustness of regression coefficients from a marginal model, while retaining the likelihood inference capabilities and flexible dependence specifications from a GLMM. The MMM formulation given in Heagerty and Zeger (2000) uses a standard GLM for the marginal mean, a non-linear mixed model (NLMM) for the within-cluster associations and a specified probability distribution for the underlying latent effects:

$i)$	$g(\mu_{ij}^m) = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$	Mean Model
$ii)$	$g(\mu_{ij}^c) = \Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i$	Association Model
$iii)$	$\boldsymbol{a}_i \sim \boldsymbol{F}_a(\boldsymbol{0}, \boldsymbol{D})$	Latent Effects Distribution
$iv)$	$Y_{ij}^c = (Y_{ij}\vert\boldsymbol{a}_i) \sim F_{Y^c}(\mu_{ij}^c, \upsilon)$	Conditional Response Distribution

where Y_{ij} is the j^{th} observation in the i^{th} cluster, $(j = 1 \ldots n_i, i = 1 \ldots N)$, g is a link function for the marginal and conditional means, $\mu_{ij}^m = E(Y_{ij})$ and $\mu_{ij}^c = E(Y_{ij} \mid \boldsymbol{a}_i)$, effects of the explanatory variables $\boldsymbol{x}_{ij}$ are modeled through the $p \times 1$ vector of marginal parameters $\boldsymbol{\alpha}^m$, the vector $\boldsymbol{a}_i$ is a $q \times 1$ set of cluster-specific latent effects with $q \times q$ covariance matrix $\boldsymbol{D}$ and distribution $\boldsymbol{F}_a(\cdot)$, the function Δ_{ij} connects the marginal and conditional models as described below, and the conditional observations independently follow an exponential family distribution with mean and dispersion parameters μ_{ij}^c and υ.

Every conditional model implies a marginal model via integration over the dependence structure, $\mu_{ij}^m = E(Y_{ij}) = E_a\big\{E(Y_{ij} \mid \boldsymbol{a}_i)\big\} = E_a(\mu_{ij}^c)$ and thus, Δ_{ij} forms a mapping between the conditional and marginal models as the solution to the integral equation $h(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) = \int_a h(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\, dF(\boldsymbol{a})$, where h is the inverse link function $h(\cdot) = g^{-1}(\cdot)$. Note that Δ_{ij} is

dependent on the covariates, marginal parameters, and random effect specification, $\Delta_{ij} = \Delta_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{\alpha}^m, \boldsymbol{z}_{ij}, \boldsymbol{F}_a, \boldsymbol{D})$, but this notation is suppressed to simplify the exposition.

To expand the model above, we formally relax the usual assumption that the marginal and conditional link functions are the same and allow possibly nonlinear effects to enter any of the marginal fixed, conditional fixed, or conditional random aspects. The marginalized multilevel model may then be formulated, (dropping subscripts and covariate dependence for brevity) as:

$$\begin{array}{lll} i) & \mu^m = h_m(\boldsymbol{\theta}^m) & \text{Mean Model} \\ ii) & \mu^c = h_c(\boldsymbol{\theta}^m, \boldsymbol{\psi}, \boldsymbol{a}) & \text{Association Model} \\ iii) & \boldsymbol{a} \sim \boldsymbol{F}_a(\boldsymbol{\psi}) & \text{Latent Effects Distribution} \\ iv) & Y^c \sim F_{Y^c}(\mu^c, \upsilon) & \text{Conditional Response Distribution} \end{array} \tag{1}$$

where $h_m(\cdot)$ and $h_c(\cdot)$ are possibly distinct inverse-link functions for the marginal and conditional means, $\boldsymbol{\theta}^m$ are marginal parameters of interest, and the random effects are assumed to follow a distribution indexed by parameters $\boldsymbol{\psi}$. Often the latent effects and conditional response distributions are implicitly stated and the MMM may be specified with $i)$ & $ii)$ alone. Definition (1) has many advantages. It is simple and intuitive, yet flexible. It directly addresses the dual mean and association objectives inherent in a clustered data analysis. It includes all the classes of MMMs contained in the original definition of Heagerty and Zeger (2000), as well as additional classes through extensions of the marginal, conditional and latent distribution specifications. Importantly, definition (1) may be used to identify MMMs that can be estimated using existing mixed model computational procedures. As previously discussed, the marginal and conditional models are tied together via integration over the random effects distribution, thus inducing the marginalization constraint:

$$\begin{aligned} \boldsymbol{\theta}^m &= h_m^{-1}(\mu^m) \\ &= h_m^{-1}\Big(\int \mu^c \, d\boldsymbol{F}_a \Big) \\ &= h_m^{-1}\Big\{ \int h_c(\boldsymbol{\theta}^m, \boldsymbol{\psi}, \boldsymbol{a}) \, d\boldsymbol{F}_a \Big\} \end{aligned} \tag{2}$$

Specific choices for $i) - iv)$ in definition (1) lead to specific model and parameter characteristics established through (2). By varying the definitions of $h_m(\cdot)$, $h_c(\cdot)$, and $\boldsymbol{F}_a$ we may examine relative strengths and weaknesses of competing models for estimating $\boldsymbol{\theta}^m$, as well as gauge the sensitivity of our inferences to particular model assumptions. We consider five widely applicable examples for illustration.

2.1 *A Logistic-Logistic-Normal Model*

The "Logistic-Normal" GLMM is the most common conditional specification for correlated binary data, probably for historical rather than scientific or computational reasons:

Logistic-Normal GLMM:

$$\begin{array}{ll} i) & \text{logit}(\pi_{ij}^c) = \log\left(\frac{\pi_{ij}^c}{1-\pi_{ij}^c}\right) = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^c + \boldsymbol{z}_{ij}\boldsymbol{a}_i \\ ii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \\ iii) & Y_{ij}|\boldsymbol{a}_i \sim \text{Binomial}(n_{ij}, \pi_{ij}^c) \end{array}$$

The conditional parameters $\boldsymbol{\alpha}^c$ represent cluster-specific log-odds-ratios (Zeger et al., 1988; Neuhaus, Kalbfleisch and Hauck, 1991). A marginalized version of this model has been labeled the logistic-normal MMM (Heagerty, 1999), but we term it instead the logistic-logistic-normal MMM, to indicate that logit links are used in both regression aspects:

Logistic-Logistic-Normal MMM

$$\begin{array}{ll} i) & \text{logit}(\pi_{ij}^m) = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m \\ ii) & \text{logit}(\pi_{ij}^c) = \Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i \\ iii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \end{array}$$

The mean and association models may be re-written with definition (1) simply as:

$$\begin{array}{ll} i) & \pi_{ij}^m = \text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\ ii) & \pi_{ij}^c = \text{expit}(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i) \end{array}$$

where $\text{expit}(x) = (1 + e^{-x})^{-1}$ is the inverse-logit function and $\boldsymbol{\alpha}^m$ are marginal log-odds-ratios with population-average interpretations (Zeger et al., 1988; Neuhaus et al., 1991). To estimate $\boldsymbol{\alpha}^m$, we must determine the form of Δ_{ij} that connects the mean and association

models. For this we employ the marginalization constraint (2):

$$\boldsymbol{\alpha}^m = (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\text{logit}\left\{\int_a \text{expit}(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\; d\boldsymbol{F}_a\right\} \tag{3}$$

With the assumed latent Gaussian effects, the integral in (3) is the well-known 'logit-normal' integral and does not have a closed form solution. Thus estimation techniques are required to evaluate this integral and its derivatives. Heagerty (1999) discusses a Newton-Raphson procedure with Gauss-Hermite quadrature. Since (3) requires additional numerical integration for estimation, MMMs with marginal logistic regression structures have been challenging to implement.

2.2 *A Logistic-Probit-Normal Model*

Instead of the logistic-normal conditional model for binary data, consider a probit-normal model, as commonly used in the econometrics literature:

Probit-Normal GLMM:

$i)$ $\quad \Phi^{-1}\left(\pi^c_{ij}\right) = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^c + \boldsymbol{z}_{ij}\boldsymbol{a}_i$
$ii)$ $\quad \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D})$
$iii)$ $\quad Y_{ij}|\boldsymbol{a}_i \sim \text{Binomial}(n_{ij}, \pi^c_{ij})$

where $\Phi(\cdot)$ is the cumulative normal distribution function. A marginalized version of this model may be written with definition (1) as:

Probit-Probit-Normal (PPN) MMM:

$i)$ $\quad \pi^m_{ij} = \Phi(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)$
$ii)$ $\quad \pi^c_{ij} = \Phi(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i)$
$iii)$ $\quad \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D})$

To estimate $\boldsymbol{\alpha}^m$, we again determine the Δ_{ij} connecting the mean and association models using the marginalization constraint (2):

$$\begin{aligned}\boldsymbol{\alpha}^m &= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\Phi^{-1}\left\{\int_a \Phi(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\; d\boldsymbol{F}_{\boldsymbol{a}}\right\} \\ &= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\Phi^{-1}\left\{\Phi\left(\frac{\Delta_{ij}}{\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}}\right)\right\}\end{aligned}$$

and thus, $\Delta_{ij} = \left(\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$. In the special case where $\boldsymbol{z}_{ij}\boldsymbol{a}_i = a_i$, (a scalar 'random intercept' model), the conditional predictor is a simple rescaling of the marginal predictor, $\Delta_{ij} = \left(\sqrt{1+\tau^2}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^c$, but this does not hold for general $\boldsymbol{z}_{ij}\boldsymbol{a}_i$.

A considerable advantage of using the probit link is that the probit-normal marginalization integral has a closed form solution, while the logit-normal integral does not. Suppose we prefer to use a logistic regression structure for the marginal mean model but wish to retain the computational advantages of the probit-normal association model. We use definition (1), relaxing the common assumption that the mean and dependence parameters are on a common scale, and obtain:

Logistic-Probit-Normal MMM:

$$\begin{array}{ll} i) & \pi^m_{ij} = \text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\ ii) & \pi^c_{ij} = \Phi(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i) \\ iii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \end{array}$$

Determining Δ_{ij} with the marginalization constraint (2) we have:

$$\begin{aligned} \boldsymbol{\alpha}^m &= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\ \text{logit}\left\{\int_a \Phi(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\ d\boldsymbol{F}_{\boldsymbol{a}}\right\} \\ &= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\ \text{logit}\left\{\Phi\left(\frac{\Delta_{ij}}{\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}}\right)\right\} \end{aligned}$$

Hence, the non-linear predictor that induces the marginal logistic model is:

$$\Delta_{ij} = \left(\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\Phi^{-1}\Big\{\text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)\Big\} \tag{4}$$

Since a closed form solution exists, we are able to employ standard NLMM computational procedures to estimate marginally logistic MMMs as discussed in Section 3.

2.3 *A Log-Log-Normal Model*

When continuous data are highly skewed, such as patients' monthly medical expenditures, log-linear gamma models are often used to account for the non-normality. A marginal log-

linear model for skewed data arising in clusters may be written as:

Log-Log-Normal MMM:

$$\begin{array}{ll} i) & \mu_{ij}^m = \exp(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\ ii) & \mu_{ij}^c = \exp(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i) \\ iii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \\ iv) & Y_{ij}|\boldsymbol{a}_i \sim \Gamma(\mu_{ij}^c, \upsilon) \end{array}$$

Where the gamma distribution is parameterized such that $E(Y^c) = \mu^c$ and $\text{var}(Y^c) = (\mu^c)^2/\upsilon$. To estimate $\boldsymbol{\alpha}^m$, we determine Δ_{ij} using the marginalization constraint (2):

$$\begin{aligned} \boldsymbol{\alpha}^m &= (\boldsymbol{x}_{ij}'\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}_{ij}'\log\left\{\int_a \exp(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\, d\boldsymbol{F}\boldsymbol{a}\right\} \\ &= (\boldsymbol{x}_{ij}'\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}_{ij}'\log\Big\{\exp(\Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{D}\boldsymbol{z}_{ij}/2)\Big\} \end{aligned}$$

and thus, $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m - \boldsymbol{z}_{ij}'\boldsymbol{D}\boldsymbol{z}_{ij}/2$.

2.4 *A Log-Log-Gamma Model*

Public health studies frequently involve Poisson processes where counts of incidents in a specified interval are recorded across multiple visits, locations or both. A mixture of the poisson distributions over a gamma process is often used to account for extra variability (overdispersion) observed in count data of this type. The resulting marginal distribution of the mixture is negative-binomial. A log-linear model for such data may be written as:

Log-Log-Gamma MMM:

$$\begin{array}{ll} i) & \lambda_{ij}^m = \exp(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\ ii) & \lambda_{ij}^c = \exp\{\Delta_{ij} + \log(a_i)\} \\ iii) & a_i \sim \Gamma(\nu, k) \\ iv) & Y_{ij}|\boldsymbol{a}_i \sim \text{Poisson}(\lambda_{ij}^c) \end{array}$$

Where the gamma parameterization produces $E(a) = \nu$, $\text{var}(a) = \nu^2/k$. Determining Δ_{ij} with the marginalization constraint (2) we have:

$$\begin{aligned} \boldsymbol{\alpha}^m &= (\boldsymbol{x}_{ij}'\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}_{ij}'\log\left[\int_a \exp\{\Delta_{ij} + \log(a)\}\, d\boldsymbol{F}\boldsymbol{a}\right] \\ &= (\boldsymbol{x}_{ij}'\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}_{ij}'\log\Big[\exp\{\Delta_{ij} + \log(\nu)\}\Big] \end{aligned}$$

and thus, $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m - \log(\nu)$. Lee and Nelder (1996) recommend constraining $E(a) = \nu = 1$, implying that the coefficient of variation in the gamma random effects distribution is constant, $\sqrt{\text{var}(a)}/E(a) = 1/\sqrt{k}$. In this case $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$.

2.5 *A Logistic-Logistic-Bridge Model*

Wang and Louis (2003) take an innovative approach to matching marginal and conditional structures by assuming the random effects follow link-specific "bridge" distributions. Their logistic model may be written as an MMM:

Logistic-Logistic-Bridge MMM:

$$\begin{array}{ll} i) & \pi^m_{ij} = \text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\ ii) & \pi^c_{ij} = \text{expit}(\Delta_{ij} + \widetilde{a}_i) \\ iii) & \widetilde{a}_i \sim B_l(0, \tau^2) \\ iv) & Y_{ij}|\widetilde{a}_i \sim \text{Binomial}(n_{ij}, \pi^c_{ij}) \end{array}$$

where $B_l(0, \tau^2)$ is the logistic bridge distribution defined with $E(\widetilde{a}_i) = 0$, $\text{var}(\widetilde{a}_i) = \tau^2$, and:

$$\text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) = \int_{\widetilde{a}} \text{expit}\left\{\left(\sqrt{1 + 3\tau^2/\pi^2}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m + \widetilde{a}\right\}\, dB_l(\widetilde{a})$$

Using the marginalization constraint (2) and the definition of B_l leads to specifying $\Delta_{ij} = \left(\sqrt{1 + 3\tau^2/\pi^2}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$, and produces logistic regressions in both the marginal and conditional models.

3. Estimation: A Nonlinear Mixed Model Approach

We begin by noting that techniques for nonlinear mixed model estimation have received vigorous attention. See for example, Stiratelli et al. (1984); Beal and Sheiner (1988); Lindstrom and Bates (1990); Zeger and Karim (1991); Breslow and Clayton (1993); Wolfinger and O'Connell (1993); Davidian and Giltinan (1993); Pinheiro and Bates (1995); Lin and Breslow (1996); McCulloch (1997); Chib and Carlin (1999); Wolfinger (1999); Booth et al. (2001); McCulloch and Searle (2001); Booth and Caffo (2002); Diggle et al. (2002); Goldstein (2002); Demidenko (2004); Sinha (2004) and Skrondal and Rabe-Hesketh (2004).

An important contribution of this article is to show that (1) and (2) allow us to identify conditional structures that induce marginal models of interest (as in the examples of Section 2). Once we have a conditional model that produces marginal parameters, we need only estimate the conditional model via standard NLMM techniques in order to obtain the desired MMM. Consider the following algorithm and accompanying logistic-probit-normal example from Section 2.2 to clarify this estimation approach.

1. Define the marginal mean model.

$$\text{LPN Example:} \quad i) \qquad \pi_{ij}^m = \text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)$$

2. Propose a conditional model representing the association structure.

$$\text{LPN Example:} \quad \begin{array}{ll} ii) & \pi_{ij}^c = \Phi\left(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i\right) \\ iii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \end{array}$$

3. Use the marginalization constraint (2) to identify the conditional mean structure in $ii)$ that induces the marginal mean model in $i)$.

$$\text{LPN Example:} \quad \text{From (4): } \Delta_{ij} = \left(\sqrt{1 + \boldsymbol{z}_{ij}'\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\Phi^{-1}\Big\{\text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)\Big\}$$

4. Estimate the mixed model in $ii)$ & $iii)$ using NLMM techniques.

   LPN Example: Estimating the nonlinear mixed model:

$$\begin{array}{ll} ii) & \pi_{ij}^c = \Phi\left[\left(\sqrt{1 + \boldsymbol{z}_{ij}'\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\Phi^{-1}\Big\{\text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)\Big\} + \boldsymbol{z}_{ij}\boldsymbol{a}_i\right] \\ iii) & \boldsymbol{a}_i \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{D}) \end{array}$$

   produces marginally logistic regression coefficients, $\boldsymbol{\alpha}^m$.

The $\boldsymbol{\alpha}^m$ estimates obtained from fitting the NLMM defined by $ii)$ and $iii)$ are in fact the marginal parameters of interest specified in $i)$. Thus, we have recast the difficult *computational* problem of estimating MMMs into the solvable *analytic* problem of specifying convenient conditional models that produce MMMs. We then fit the conditional models with

standard NLMM techniques to arrive at the MMM estimates. For an additional example, the following is equivalent to the logistic-logistic-bridge model given in Section 2.5:

Logistic-Logistic-Bridge MMM:

$i)$ $\quad \pi_{ij}^m = \text{expit}(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)$

$ii)$ $\quad \pi_{ij}^c = \text{expit}\left[\left(\sqrt{1+3\tau^2/\pi^2}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m + B_l^{-1}\left\{\Phi\left(a_i/\tau\right)\right\}\right]$

$iii)$ $\quad a_i \sim N(0,\tau^2)$

$iv)$ $\quad Y_{ij}|\widetilde{a}_i \sim \text{Binomial}(n_{ij}, \pi_{ij}^c)$

Since the random effects are Gaussian in this version, we may fit the NLMM defined by $ii) - iv)$ using widely available techniques to obtain estimates of the desired marginal $\boldsymbol{\alpha}^m$ parameters in $i)$. This example also shows that the random effects distributional assumption can be relatively flexible in terms of model estimation. When we wish to investigate a non-normal distribution for the latent effects, $\widetilde{\boldsymbol{a}} \sim \boldsymbol{F}_{\widetilde{\boldsymbol{a}}}\left\{\text{E}(\widetilde{\boldsymbol{a}}) = \widetilde{\boldsymbol{\mu}},\ \text{var}(\widetilde{\boldsymbol{a}}) = \widetilde{\boldsymbol{D}}\right\}$, which infers a known distribution on the linear combination, $\boldsymbol{z}\widetilde{\boldsymbol{a}} \sim \boldsymbol{F}_{\boldsymbol{z}\widetilde{\boldsymbol{a}}}(\cdot)$, we apply the appropriate distributional transformation, $\boldsymbol{z}\widetilde{\boldsymbol{a}} = \boldsymbol{F}^{-1}_{\boldsymbol{z}\widetilde{\boldsymbol{a}}}\left\{\Phi\left(\frac{\boldsymbol{z}\boldsymbol{a}-\boldsymbol{z}\widetilde{\boldsymbol{\mu}}}{\sqrt{\boldsymbol{z}'_{ij}\widetilde{\boldsymbol{D}}\boldsymbol{z}_{ij}}}\right)\right\}$, and are able to use gaussian integration over $\boldsymbol{a} \sim MVN(\boldsymbol{0}, \widetilde{\boldsymbol{D}})$. We apply this approach in the application section below.

3.1 *Some Approximate Methods for Logistic Regression*

Marginal logistic models present some of the most difficult computational challenges. As discussed in Section 2.1, the logistic-logistic-normal MMM requires extra numerical integration. The logistic-probit-normal and logistic-logistic-bridge models have closed forms but complex association models. Approximate methods that provide simpler forms for the conditional means may be more numerically stable and have better statistical properties in certain situations. In this section, we briefly discuss three approximate methods for estimating marginally logistic regression models.

Johnson, Kotz and Balakrishnan, (1995, pg. 113-163)(JKB) give a detailed discussion of the logistic distribution and compare logistic curves $\pi^l(x) = \text{expit}(x)$ and Gaussian curves $\pi^p(x) = \Phi(x)$. Using the JKB results leads to a constant multiplication method

of $\pi^p(c_1 x) \doteq \pi^l(x)$ where $c_1 = (15/16)(\pi/\sqrt{3}) = 1.700437$. This approximation has a maximum reported difference of about 0.0095 at $x = 0.7$. Other possible values for c_1 are discussed by Volodin (1994 personal communication with JKB) $c_1 = 1.7017456$; Liao (1994) $c_1 = \pi/\sqrt{3} = 1.813799$; and Amemiya (1981) $c_1 = 1.6$. The JKB approach leads to specifying $\Delta_{ij} = \left(\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m \cdot c_1$ as the conditional predictor to use in a logistic-probit-normal MMM to obtain marginal logistic regression parameters, $\boldsymbol{\alpha}^m$.

Page (1977) and Tocher (1963) found $\Phi(x) \cong e^{f(x)}/(1 + e^{f(x)})$, where $f(x) = 2[a_1 x(1 + a_2 x^2)]$ and the constants $a_1 = 0.7988$ and $a_2 = 0.04417$ provide an approximation with a maximum difference about 0.00014 at $x = 1.476078$. Notice that the form of this approximation is the inverse logit function; providing a direct mapping from logistic regression to probit regression and vice-versa. The Page approach leads to using $\Delta_{ij} = \left(\sqrt{1 + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\Delta_{ij}^{page}$ in a logistic-probit-normal MMM, where Δ_{ij}^{page} is the solution to the scaled cubic equation (see Appendix A).

Zeger et al. (1988) (ZLA) offer the constant multiplication approximation $\text{logit}(\pi_{ij}^m) \approx a_l(\boldsymbol{D}) \cdot \boldsymbol{x}_{ij}\boldsymbol{\alpha}^c$ where: $a_l(\boldsymbol{D}) = (1 + c_1^{-2}\boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij})^{-1/2}$ and c_1 is the multiplication constant from Johnson et al. (1995). The ZLA approach leads to specifying $\Delta_{ij} = \left(\sqrt{1 + c_1^{-2}\boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}}\right)\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$ as the conditional predictor in a logit-logit-normal MMM. The ZLA mean and dependence models are on the same (logit) scale at the expense of an extra probit-logit approximation.

4. Example: Crossover Trial

Diggle et al. (2002) illustrate marginal and conditional models using a subset of the cerebrovascular deficiency crossover trial data from Jones and Kenward (1989). This data provides an extreme test of likelihood methods since the data set is small and the random effects variance large. Responses are binary electrocardiogram reading indicators, abnormal=0 vs normal=1. Thirty-four subjects received the active drug (A) followed by the placebo (B) (group AB) and an additional thirty-three subjects received the placebo followed by the

active drug (group BA). Table 1 replicates the data from Diggle et al. (2002, p.148).

[Table 1 about here.]

4.1 *Crossover Trial Models*

The explanatory variables used in the analysis are indicators for the active drug and the second time period:

$$\text{Treatment} = \begin{cases} 0 & \text{Placebo (B)} \\ 1 & \text{Active Drug (A)} \end{cases} \qquad \text{Period} = \begin{cases} 0 & \text{Period 1} \\ 1 & \text{Period 2} \end{cases}$$

A conditional logistic-normal model can be written as:

$$\begin{array}{ll} i) & \text{logit}(\pi_{ij}^c) = \alpha_0^c + \alpha_1^c \cdot \text{Treatment} + \alpha_2^c \cdot \text{Period} + a_i \\ ii) & a_i \sim N(0, \tau^2) \\ iii) & Y_{ij} \,|\, a_i \sim \text{Binary}(\pi_{ij}^c) \end{array}$$

The conditional mean parameters have subject-specific interpretations, with the odds ratio $exp(\alpha_1^c)$ contrasting normal readings for a subject who is on treatment and has underlying normal-reading propensity a^*, to the same subject when they are on placebo, (or to another subject on placebo with the same latent a^*). The random intercept standard deviation τ represents heterogeneity among subject responses with larger values of τ indicating larger correlations for within-subject responses. Note that in the special case of a crossover trial we have measurements for each subject on both treatment and placebo. Thus, the crossover data contains directly observable information concerning the subject-specific contrast. When effects of interest do not vary within a cluster, such as the treatment effect in a standard prospective clinical trial, subject-specific contrasts are extrapolations from within-cluster associations and differences between clusters. Inferences in these situations may be highly sensitive to random effects assumptions.

A corresponding logistic-logistic-normal (LLN) marginalized multilevel model is:

$$\begin{array}{ll} i) & \pi_{ij}^m = \text{expit}\left(\alpha_0^m + \alpha_1^m \cdot \text{Treatment} + \alpha_2^m \cdot \text{Period}\right) \\ ii) & \pi_{ij}^c = \text{expit}\left(\Delta_{ij} + a_i\right) \\ iii) & a_i \sim N(0, \tau^2) \\ iv) & Y_{ij} \,|\, a_i \sim \text{Binary}(\pi_{ij}^c) \end{array}$$

with Δ_{ij} as in Section 2.1. The marginal mean parameters have population-average interpretations, with the odds ratio $exp(\alpha_1^m)$ contrasting normal readings for subjects on treatment to normal readings for subjects on placebo. Similarly, $exp(\alpha_2^m)$ is the odds ratio contrasting normal readings for subjects at time periods 1 & 2. Associations among subject responses are again represented by τ. Since the marginal treatment contrast compares groups of subjects on active drug to groups of subjects on placebo, the effect would be directly observable even if a crossover trial design had not been used.

Two alternative formulations for the same marginal logistic regression parameters are the logistic-probit-normal (LPN) model:

$$\begin{array}{ll} i) & \pi_{ij}^m = \text{expit}\left(\alpha_0^m + \alpha_1^m \cdot \text{Treatment} + \alpha_2^m \cdot \text{Period}\right) \\ ii) & \pi_{ij}^c = \Phi\left(\Delta_{ij} + a_i\right) \end{array}$$

and the logistic-logistic-bridge (LLB) model:

$$\begin{array}{ll} i) & \pi_{ij}^m = \text{expit}\left(\alpha_0^m + \alpha_1^m \cdot \text{Treatment} + \alpha_2^m \cdot \text{Period}\right) \\ ii) & \pi_{ij}^c = \text{expit}\left[\Delta_{ij} + B_l^{-1}\left\{\Phi(a_i/\tau)\right\}\right] \end{array}$$

with Δs from Sections 2.2 and 2.5 respectively. Although the MMMs have very different assumptions on the association models, they all estimate the same objects of interest: marginal odds ratios. Estimates and standard errors for these models are given in Table 2. Also given are results from the ZLA, JKB, and Page approximation techniques discussed in section 3.1. The constant $c_1 = (15/16)(\pi/\sqrt{3})$ is used for the ZLA and JKB approximations.

4.2 *Crossover Trial Results*

[Table 2 about here.]

Despite the wide range of association model assumptions, results for the marginal mean parameters are similar across estimation methods. Differences in association parameters are expected since the LPN, Page and JKB methods use a probit link for the association model whereas the LLN, Bridge and ZLA methods use a logit link. The Bridge method also

has a non-Gaussian latent effects distribution. The Page approximation to the LPN model is precise, at the expense of a more complicated Δ function than the JKB method. The JKB and ZLA approximations give marginal results comparable with the exact methods, while retaining GLMMs for their association models. The LLN results presented here differ slightly from those in Diggle et al. (2002) since we used a greater number of quadrature points to evaluate the logit-normal integral. The standard error of the parameter estimate $\hat{\tau}$ converged at around 175 quadrature points and we used 200 points to ensure accuracy. The LLN model took 77 seconds to run on a 1.80GHz CPU while the other methods took less than one second. The difference is attributable to the extra Gauss-Hermite quadrature necessary to evaluate the logit-normal marginalization integral. We used adaptive quadrature to evaluate the nonlinear mixed model integrals and a quasi-Newton-Raphson line-search optimizer as implemented in standard statistical packages. A Monte Carlo simulation study indicated our chosen methods performed well in terms of accuracy and precision (results available upon request). Example SAS code for the LPN model is given in Appendix B and additional examples with more complicated random effects structures and alternative optimization techniques are available from the authors.

An advantage of using a MMM formulation is that we may conduct likelihood-based inference (Birnbaum, 1962; Royall, 1999; Blume, 2002). The support for any parameter of interest may be represented through a graph of the likelihood function. When there are nuisance parameters, (latent effect variances, etc.), the profile likelihood function (Kalbfleisch and Sprott, 1970) provides an analogous characterization of support (Royall, 1999). To further compare the six estimation methods, we plot the profile likelihood functions for the treatment parameter α_1 in Figure 1. The horizontal lines define support intervals (SI) where parameter values are 'consistent with the observations' at $k = 8$ and $k = 32$ benchmark support levels (Royall, 1999 Blume, 2002). Values outside the 1/8 (1/32) support intervals

are fairly (very) inconsistent with the data, (at least 8 (32) times as unlikely as the MLE). For the LPN model, the crossover data represent evidence supporting the MLE of $\alpha_1 = 0.587$ ($OR = e^{0.587} = 1.8$) over the value of $\alpha_1 = 0$ ($OR = 1$) by a likelihood ratio of $LR_0 = L(0.587)/L(0) = e^{-68.11}/e^{-71.36} = 25.7$. This indicates reasonably strong support for a doubling in the odds of a normal electrocardiogram reading comparing active drug to placebo. It is easy to see from the profile likelihood functions in Figure 1 that all six methods provide similar evidence about this marginal treatment effect, despite the variety of assumptions about the association structure.

[Figure 1 about here.]

5. Discussion

Conditional model (GLMM) estimates and interpretations can be heavily dependent on assumed variance structures, as shown in Heagerty and Zeger (2000). Marginal models estimate effects that are directly observable in the data and are more robust to the chosen dependence model. This is illustrated by the similarity in the marginal mean parameter estimates across the range of association assumptions in the crossover trial example of Section 4. While alternative approaches to estimating marginal models, such as a GEE approach, avoid specifying the complete joint distribution of the responses, the MMM approach retains the capability of likelihood inference and the consequent benefits therein.

In this article we have reformulated the MMM to include additional classes and provided connections between marginal and conditional models. Using these connections we have shown how to formulate marginal models by inducing them from conditional model specifications. Since there are many conditional models that can produce the same marginal model, there will often exist a subset of conditionally specified marginal models with superior properties, such as practical computation. Constructing marginal models in terms of their

conditional model counterparts facilitates the identification of MMMs that have substantial computational advantages, as shown in the contrast of the LLN, LPN and LLB models.

We agree with Neuhaus (2000) that forcing all aspects of an analysis into a single model is usually suboptimal. Providing connections between complementary analysis aspects however, can be helpful. When conditional inferences are warranted it is relatively straightforward to obtain conditional parameters with standard computing procedures. We show here that is also straightforward to obtain marginal parameters from the same procedures by simply re-specifying selected conditional model components. When appropriate, both marginal and conditional parameters may be presented from fitting both the MMM and the GLMM, allowing inferences to be drawn on the aspect of central scientific interest. The clinician may wish to know how well they can expect an intervention to work for a particular patient, while the public health official may ask what population effects they can expect to see if the intervention is effective. Marginalized multilevel models provide the bridge between these questions.

There are a variety of opportunities for further research. Identifying additional conditional structures that produce commonly used marginal models will extend the availability of the marginalization approach. As further computational advances are made in mixed model estimation, the connections shown here allow them to translate to advances in marginal model estimation as well. Performing sensitivity analyses towards latent variable assumptions will be furthered by having a wide range of easily implemented random effect distributions. Thus, investigations of alternate latent variable constructions via distributional transformations (as used in the LLB example) will be beneficial.

Acknowledgements

Special thanks to Drs. Charles Rohde and Tom Louis for invaluable discussions regarding

this manuscript.

References

Amemiya, T. (1981). Qualitative response models: a survey. *Journal of Economic Literature* **19**, 1483–1536.

Beal, S. L. and Sheiner, L. B. (1988). Heteroscedastic nonlinear regression. *Technometrics* **30**, 327–338.

Birnbaum, A. (1962). On the foundations of statistical inference (Com: p307-326). *Journal of the American Statistical Association* **57**, 269–306.

Blume, J. D. (2002). Likelihood methods for measuring statistical evidence. *Statistics in Medicine* **21**, 2563–2599.

Booth, J. G. and Caffo, B. S. (2002). Unequal sampling for Monte Carlo EM algorithms. *Computational Statistics and Data Analysis* **39**, 261–270.

Booth, J. G., Hobert, J. P. and Jank, W. (2001). A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modelling: An International Journal* **1**, 333–349.

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

Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing* **9**, 17–26.

Davidian, M. and Giltinan, D. M. (1993). Some general estimation methods for nonlinear mixed-effects models. *Journal of Biopharmaceutical Statistics* **3**, 23–55.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. John Wiley & Sons.

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford.

Edwards, A. W. F. (1992). *Likelihood*. Johns Hopkins University Press.

Goldstein, H. (2002). *Multilevel statistical models*. Oxford University Press, 3 edition.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55**, 688–698.

Heagerty, P. J. and Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science* **15**, 1–26.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous univariate distributions. Volume 2 (Second edition) (ISBN 0471584940)*. John Wiley & Sons.

Jones, B. and Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman & Hall Ltd.

Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters (with discussion) (Corr: 72V34 p124-125). *Journal of the Royal Statistical Society, Series B, Methodological* **32**, 175–208.

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

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (Disc: p656-678). *Journal of the Royal Statistical Society, Series B, Methodological* **58**, 619–656.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Liao, T. F. (1994). *Interpreting probability models: logit, probit, and other generalized linear models*. Sage Publications Inc.

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.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated

measures data. *Biometrics* **46**, 673–687.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, linear, and mixed models.* John Wiley & Sons.

Miglioretti, D. L. and Heagerty, P. J. (2004). Marginal modeling of multilevel binary data with time-varying covariates. *Biostatistics* **5**, 381–398.

Mills, Field and Dupuis (2002). Marginally specified generalized linear mixed models: A robust approach. *Biometrics* **58**, 4.

Neuhaus, J. M. (2000). comment on marginalized multilevel models and likelihood inference. *Statistical Science* **15**, 21–22.

Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* **59**, 25–35.

Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator. *Applied Statistics* **26**, 75–76.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

Royall, R. M. (1999). *Statistical evidence: a likelihood paradigm.* Chapman & Hall Ltd.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (C/R: p1121-1146). *Journal of the*

American Statistical Association **94**, 1096–1120.

Sinha, S. K. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association* **99**, 451–460.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable, Multilevel and Panel Modelling*. CRC Press.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40**, 961–971.

Tocher, K. (1963). *The art of simulation*. English universities Press.

Wang, Z. and Louis, T. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika* **90**, 765–775.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.

Wolfinger, R. D. (1999). Fitting nonlinear mixed models with the new NLMIXED procedure. In *Proceedings of the Twenty-Fourth Annual SAS Users Group International Conference*, pages 1666–1675.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86.

Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach (Corr: V45 p347). *Biometrics* **44**, 1049–1060.

Appendix A

The Page Approximation

Solving: $\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m \doteq 2a_1\Delta_{ij}^{page} + 2a_1a_2(\Delta_{ij}^{page})^3$ yields:

$$\begin{aligned} \Delta_{ij}^{page} &= \sqrt[3]{t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) + u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)} - \sqrt[3]{u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)} \\ \text{where:} \quad t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) &= \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m/(2a_1a_2) \\ u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) &= \tfrac{1}{2}\left\{-t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) + \sqrt{t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)^2 - 4/(3a_2)^3}\right\} \end{aligned}$$

Appendix B

Crossover Trial SAS Code: LPN Example

```
data xover;
  input id Period Treatment y count @@;  cards;
  1 0 1 1 22   1 1 0 1 22    2 0 1 0  0  2 1 0 1  0
  3 0 1 1  6   3 1 0 0  6    4 0 1 0  6  4 1 0 0  6
  5 0 0 1 18   5 1 1 1 18    6 0 0 0  4  6 1 1 1  4
  7 0 0 1  2   7 1 1 0  2    8 0 0 0  9  8 1 1 0  9
  ;
run;

TITLE "Logit-Probit-Normal MMM";
PROC NLMIXED data=xover qpoints=100;
   PARMS alpha0_m=.6 alpha1_m=.6 alpha2_m=-.3 tau=3;
   eta_m = alpha0_m + alpha1_m*Treatment + alpha2_m*Period;
   pi_m = 1 / (1 + exp(-eta_m));
   delta = sqrt(1+(tau*tau)) * probit(pi_m);
   eta_c  =  delta + a;
   pi_c = probnorm(eta_c);
   MODEL y ~ binary(pi_c);
   RANDOM a ~ NORMAL(0,tau*tau) SUBJECT=id;
   REPLICATE count;
run;
```

Note: it is generally preferred to model the log of the variance parameters. The above program specifies the variance parameters directly to make the code more transparent.

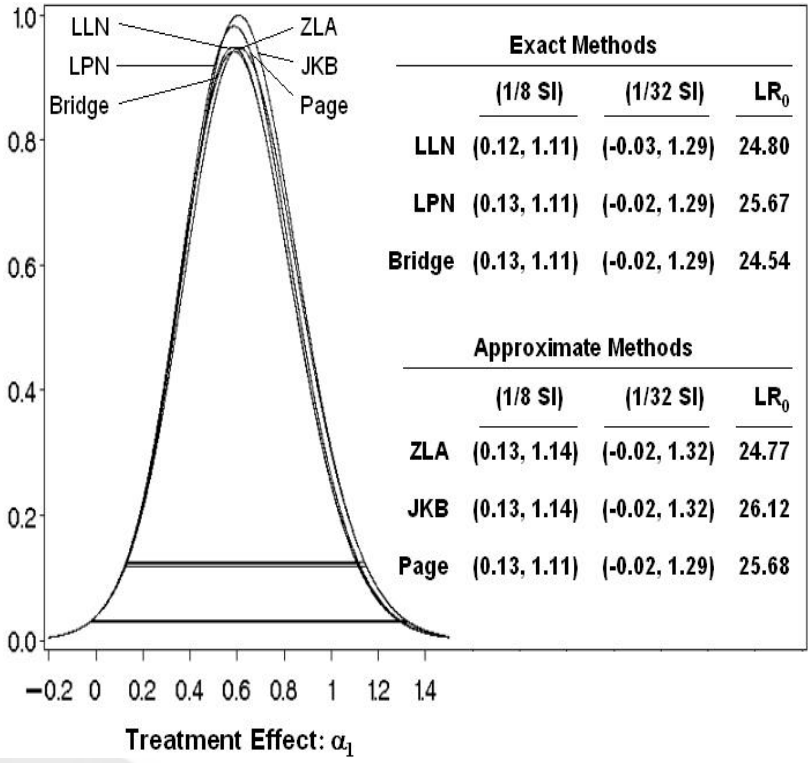

Figure 1. Crossover Trial Profile Likelihoods: All methods provide similar evidence about the marginal treatment effect. Tabled values are the $k = 8, 16$ Support Intervals (SI) and the Likelihood Ratio comparing the MLE for the treatment effect $\hat{\alpha_1}$ to no treatment effect, $\alpha_1 = 0$ (LR_0).

Table 1

Crossover Trial Data *Data from a 2×2 crossover trial on cerebrovascular deficiency. Responses are electrocardiogram readings (abnormal=0, Normal=1) and treatments are active drug (A) and placebo (B.)*

	"Normal" Responses				
Group	(1,1)	(0,1)	(1,0)	(0,0)	Total
AB	22	0	6	6	34
BA	18	4	2	9	33

Table 2

Crossover Trial Results: *Logistic MMM Estimates and Standard Errors for six estimation methods. Results for marginal parameters are similar across methods.*

Parameter		Exact Methods			Approximate Methods		
		LLN	LPN	Bridge	JKB	ZLA	Page
α_0: Intercept	est.	0.681	0.682	0.681	0.719	0.722	0.682
	s.e.	(0.277)	(0.277)	(0.277)	(0.287)	(0.289)	(0.277)
α_1: Treatment	est.	0.584	0.587	0.584	0.606	0.606	0.587
	s.e.	(0.233)	(0.233)	(0.233)	(0.238)	(0.240)	(0.233)
α_2: Period	est.	-0.323	-0.329	-0.324	-0.341	-0.338	-0.329
	s.e.	(0.230)	(0.230)	(0.230)	(0.236)	(0.238)	(0.230)
τ : Association	est.	4.945	2.799	5.480	2.797	4.943	2.799
	s.e.	(1.908)	(1.055)	(2.088)	(1.054)	(1.907)	(1.055)
log-likelihood		-68.19	-68.11	-68.14	-68.09	-68.13	-68.11