Published in final edited form as:

Stat. 2013 ; 2(1): . doi:10.1002/sta4.22.

Practical Marginalized Multilevel Models

Michael E. Griswold^{a,*}, Bruce J. Swihart^a, Brian S. Caffo^a, and Scott L. Zeger^a

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

Abstract

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 the MMM and approximate MMM approaches on a cerebrovascular deficiency crossover trial using SAS and an epidemiological study on race and visual impairment using R. Datasets, SAS and R code are included as supplemental materials.

Keywords

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), marginal models fit with generalized estimating equations (GEE), and more recently, marginalized multilevel models (MMM) (Heagerty & Zeger (2000)). 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 (Hubbard, Ahern, Fleischer, Laan, Lippman, Jewell, Bruckner & Satariano (2010)). 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 effects. 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 missing completely at random (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 MMMs 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 considerable resources exist 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 and an epidemiological study on race and visual impairment.

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, Heagerty, Liang & Zeger (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 & Zeger (1986)). 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 (and thus the complete likelihood of the parameters) remains unspecified. This lack of specification for the joint distribution has many implications. When estimating functions are used, data that are not missing completely at random must be addressed with inverse probability weighting techniques (Scharfstein et al. (1999)) or a similar strategy. Likelihood-based estimation techniques such as mixed models can accommodate missing at random by default. 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, therefore MMMs are primed for a Bayesian analysis, whereas estimating functions are not. Similarly, evidential methodology relies on likelihood methods, and as such the MMM techniques we propose allow the construction of profile likelihood inference and likelihood intervals whereas GEE is incapable of such constructions (Blume (2002)). While MMMs do not require likelihood intervals, they remain as an option for an alternative inferential technique to standard frequentist and Bayesian methodology. As put forth, we contend that MMMs can handle a wider range of data features with respect to missingness as well as provide opportunity for more estimation paradigms than GEEs.

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 & 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:

$$\begin{array}{ll}
 i) & g\left(\mu_{ij}^m\right)=\mathbf{x}_{ij} \boldsymbol{\alpha}^m \quad \text{Mean Model} \\
 ii) & g\left(\mu_{ij}^c\right)=\Delta_{ij}+\mathbf{z}_{ij} \mathbf{a}_i \quad \text{Association Model} \\
 iii) & \mathbf{a}_i \sim \mathbf{F}_a\left(0, \mathbf{D}\right) \quad \text{Latent Effects Distribution} \\
 iv) & Y_{ij}^c=\left(Y_{ij} \mid \mathbf{a}_i\right) \sim F_{Y^c}\left(\mu_{ij}^c, v\right) \quad \text{Conditional Response Distribution}
 \end{array}$$

where Y_{ij} is the j^{th} observation in the i^{th} cluster, ($j = 1 \dots n_i, i = 1 \dots N$), g is a link function for the marginal and conditional means, $\mu_{ij}^m = E\left(Y_{ij}\right)$ and $\mu_{ij}^c = E\left(Y_{ij} \mid \mathbf{a}_i\right)$, effects of the explanatory variables $\mathbf{x}_{ij}$ are modeled through the $p \times 1$ vector of marginal parameters $\boldsymbol{\alpha}^m$, the vector $\mathbf{a}_i$ is a $q \times 1$ set of cluster-specific latent effects with $q \times q$ covariance matrix $\mathbf{D}$ and distribution $\mathbf{F}_a(\cdot)$, 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 v .

Every conditional model implies a marginal model via integration over the dependence

structure, $\mu_{ij}^m = E\left(Y_{ij}\right) = E_a\left\{E\left(Y_{ij} \mid \mathbf{a}_i\right)\right\} = E_a\left(\mu_{ij}^c\right)$ and thus, Δ_{ij} forms a mapping between the conditional and marginal models as the solution to the integral equation $h\left(\mathbf{x}_{ij} \boldsymbol{\alpha}^m\right) = \int_a h\left(\Delta_{ij} + \mathbf{z}_{ij} \mathbf{a}\right) d\mathbf{F}_a(\mathbf{a})$, 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, $\Delta_{ij} = \Delta_{ij}\left(\mathbf{x}_{ij}, \boldsymbol{\alpha}^m, \mathbf{z}_{ij}, \mathbf{F}_a, \mathbf{D}\right)$, 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:

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

where $h_m(\cdot)$ and $h_c(\cdot)$ are possibly distinct inverse-link functions for the marginal and conditional means, θ^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) includes all the classes of MMMs contained in the original definition of Heagerty & 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:

$$\mu^m = h_m^{-1}\left(\mu^m\right) = h_m^{-1}\left(\int \mu^c d\mathbf{F}_a\right) = h_m^{-1}\left\{\int h_c\left(\theta^m, \boldsymbol{\psi}, \mathbf{a}\right) d\mathbf{F}_a\right\} \quad (2)$$

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 $\mathbf{F}_a$ we

may examine relative strengths and weaknesses of competing models for estimating θ^n , as well as gauge the sensitivity of our inferences to particular model assumptions. We consider three widely applicable examples for binary data for illustration. MMMs exist for count data and one such example, the Log-Log-Gamma, is included below. Another, the Log-Log-Normal, arises similarly.

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:

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

The conditional parameters α^c represent cluster-specific log-odds-ratios. 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

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

where $\text{expit}(x) = (1 + e^{-x})^{-1}$ is the inverse-logit function and α^m are marginal log-odds-ratios with population-average interpretations. To estimate α^m , we must determine the form of Δ_{ij} that connects the mean and association models. For this we employ the marginalization constraint (2):

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

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, marginal logistic MMMs are 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 and the marginalized adaptation under definition (1). Allowing $\Phi(\cdot)$ as the cumulative normal distribution function:

Probit–Normal GLMM:

- i) $\Phi^{-1}(\pi_{ij}^c) = \mathbf{x}_{ij}'\alpha^c + \mathbf{z}_{ij}'\mathbf{a}_i$
- ii) $\mathbf{a}_i \sim \text{MVN}(0, \mathbf{D})$
- iii) $Y_{ij}|\mathbf{a}_i \sim \text{Binomial}(n_{ij}, \pi_{ij}^c)$

Probit–Probit–Normal MMM:

- i) $\pi_{ij}^m = \Phi(\mathbf{x}_{ij}'\alpha^m)$
- ii) $\pi_{ij}^c = \Phi(\Delta_{ij} + \mathbf{z}_{ij}'\mathbf{a}_i)$
- iii) $\mathbf{a}_i \sim \text{MVN}(0, \mathbf{D})$

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

$$\alpha^m = (\mathbf{x}_{ij}'\mathbf{x}_{ij})^{-1} \mathbf{x}_{ij}' \Phi^{-1} \left\{ \int_a \Phi(\Delta_{ij} + \mathbf{z}_{ij}'\mathbf{a}) d\mathbf{F}_a \right\} = (\mathbf{x}_{ij}'\mathbf{x}_{ij})^{-1} \mathbf{x}_{ij}' \Phi^{-1} \left\{ \Phi \left(\frac{\Delta_{ij}}{\sqrt{1 + \mathbf{z}_{ij}'\mathbf{D}\mathbf{z}_{ij}}} \right) \right\}$$

and thus, $\Delta_{ij} = \left(\sqrt{1 + \mathbf{z}_{ij}'\mathbf{D}\mathbf{z}_{ij}} \right) \mathbf{x}_{ij}'\alpha^m$. In the special case where $\mathbf{z}_{ij}'\mathbf{a}_i = a_i$, (a scalar ‘random intercept’ model), the conditional predictor is a simple rescaling of the marginal predictor,

$\Delta_{ij} = \left(\sqrt{1 + \tau^2} \right) \mathbf{x}_{ij}'\alpha^m = \mathbf{x}_{ij}'\alpha^c$, but this does not hold for general $\mathbf{z}_{ij}'\mathbf{a}_i$.

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:

- i) $\pi_{ij}^m = \text{expit}(\mathbf{x}_{ij}'\alpha^m)$
- ii) $\pi_{ij}^c = \Phi(\Delta_{ij} + \mathbf{z}_{ij}'\mathbf{a}_i)$
- iii) $\mathbf{a}_i \sim \text{MVN}(0, \mathbf{D})$

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

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

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

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

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

2.3. A Logistic-Logistic-Bridge Model

Wang & Louis (2003) and Wang & Louis (2004) 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:

- i) $\pi_{ij}^m = \text{expit}(\mathbf{x}_{ij}\alpha^m)$
- ii) $\pi_{ij}^c = \text{expit}(\Delta_{ij} + \tilde{a}_i)$
- iii) $\tilde{a}_i \sim B_l(0, \tau^2)$
- iv) $Y_{ij}|\tilde{a}_i \sim \text{Binomial}(n_{ij}, \pi_{ij}^c)$

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

$$\text{expit}(\mathbf{x}_{ij}\alpha^m) = \int_a \text{expit}\left\{\left(\sqrt{1+3\tau^2/\pi^2}\right)\mathbf{x}_{ij}\alpha^m + \tilde{a}\right\} dB_l(\tilde{a})$$

Using the marginalization constraint (2) and the definition of B_l leads to specifying

$\Delta_{ij} = \left(\sqrt{1+3\tau^2/\pi^2}\right)\mathbf{x}_{ij}\alpha^m = \mathbf{x}_{ij}\alpha^c$, and produces logistic regressions in both the marginal and conditional models.

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:

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

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

$$\alpha^m = \left(\mathbf{x}_{ij}'\mathbf{x}_{ij}\right)^{-1} \mathbf{x}_{ij}' \log \left[\int_a \exp\{\Delta_{ij} + \log(a)\} dF_a \right] = \left(\mathbf{x}_{ij}'\mathbf{x}_{ij}\right)^{-1} \mathbf{x}_{ij}' \log [\exp\{\Delta_{ij} + \log(\nu)\}]$$

and thus, $\Delta_{ij} = \mathbf{x}_{ij}\alpha^m - \log(\nu)$. Lee & 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 $\Delta_{ij} = \mathbf{x}_{ij}\alpha^m$.

3. Estimation Via Generalized Nonlinear Mixed Models

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

Goldstein (2003); Demidenko (2004); Sinha (2004) and Skrondal & 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 existing generalized 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.

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

2. Propose a conditional model representing the association structure.

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

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

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

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

LPN Example: Estimating the nonlinear mixed model:

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

produces marginally logistic regression coefficients, α^m .

The α^m estimates obtained from fitting the generalized 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 existing generalized 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 2.3:

Logistic-Logistic-Bridge MMM:

$$\begin{array}{ll} i) & \pi_{ij}^m = \text{expit}(\mathbf{x}_{ij}\alpha^m) \\ ii) & \pi_{ij}^c = \text{expit} \left[\left(\sqrt{1 + 3\tau^2/\pi^2} \right) \mathbf{x}_{ij}\alpha^m + B_l^{-1} \{ \Phi(a_i/\tau) \} \right] \\ iii) & \mathbf{a}_i \sim N(0, \tau^2) \\ iv) & Y_{ij} | \tilde{a}_i \sim \text{Binomial}(n_{ij}, \pi_{ij}^c) \end{array}$$

Since the random effects are Gaussian in this version, we may fit the generalized NLMM defined by *ii*) – *iv*) using widely available techniques to obtain estimates of the desired marginal α^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, $\tilde{\mathbf{a}} \sim F_{\tilde{\mathbf{a}}} \left\{ E(\tilde{\mathbf{a}}) = \tilde{\mu}, \text{var}(\tilde{\mathbf{a}}) = \tilde{\mathbf{D}} \right\}$,

which infers a known distribution on the linear combination, $\mathbf{z}\tilde{\mathbf{a}} \sim F_{\mathbf{z}\tilde{\mathbf{a}}}(\cdot)$, we apply the

appropriate distributional transformation, $\mathbf{z}\tilde{\mathbf{a}} = F_{\mathbf{z}\tilde{\mathbf{a}}}^{-1} \left\{ \Phi \left(\frac{\mathbf{z}\mathbf{a} - \mathbf{z}\tilde{\mu}}{\sqrt{\mathbf{z}'_{ij}\tilde{\mathbf{D}}\mathbf{z}_{ij}}} \right) \right\}$, and are able to use gaussian integration over $\mathbf{a} \sim MVN(0, \tilde{\mathbf{D}})$.

3.1. Practical Approximate Methods for Logistic GLMM to MMM mappings

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. As a practical recommendation, multiplying conditional estimates from standard GLMM procedures by a numerically approximated, constant attenuation factor $\phi^{(n)}$ can often serve well as a first pass analysis. Three approaches have been proposed previously, all of the form $\beta^{(m)} = \phi_{(n)}\beta^{(c)}$

Johnson, Kotz and Balakrishnan, (1995, pg. 113-163)(JKB) give a detailed discussion of the logistic distribution and compare logistic curves $\pi'(x) = \text{expit}(x)$ and Gaussian curves $\pi^p(x) = \Phi(x)$. Using the JKB results leads to a constant multiplication method of $\pi^p(c_1x) \doteq \pi'(x)$

where $c_1 = (15/16) \left(\pi / \sqrt{3} \right) = 1.700437$. 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 $\Delta_{ij} = \left(\sqrt{1 + \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}} \right) \mathbf{x}_{ij}\alpha^m \cdot c_1$ as the conditional predictor to use in a logistic-probit-normal MMM to obtain marginal logistic regression parameters, α^m .

Page (1977) and Tocher (1963) found $\Phi(x) \approx e^{f(x)} / (1 + e^{f(x)})$, where $f(x) = 2[a_1x(1 + a_2x^2)]$ 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 $\Delta_{ij} = \left(\sqrt{1 + \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}} \right) \Delta_{ij}^{page}$ in a logistic-probit-normal MMM, where Δ_{ij}^{page} is the solution to the scaled cubic equation (see Appendix ??).

Zeger et al. (1988) (ZLA) offer the constant multiplication approximation

$\text{logit} \left(\pi_{ij}^m \right) \approx a_l(\mathbf{D}) \cdot \mathbf{x}_{ij}\alpha^c$ where: $a_l(\mathbf{D}) = \left(1 + c_1^{-2} \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij} \right)^{-1/2}$ and c_1 is the multiplication constant from Johnson et al. (1995). The ZLA approach leads to specifying $\Delta_{ij} = \left(\sqrt{1 + c_1^{-2} \mathbf{z}'_{ij}\mathbf{D}\mathbf{z}_{ij}} \right) \mathbf{x}_{ij}\alpha^m$ 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. Examples

4.1. Crossover Trial Models

Diggle et al. (2002) illustrate marginal and conditional models using a subset of the cerebrovascular deficiency crossover trial data from Jones & Kenward (2003) (Table 1). These data provide 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).

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

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

A conditional logistic-normal model and the corresponding logistic-logistic-normal (LLN) marginalized multilevel model can be written as (note: both models assume

$$Y_{ij}|a_i \sim \text{Binary}(\pi_{ij}^c) \text{ and } a_i \sim N(0, \tau^2):$$

Logistic-Normal GLMM:

$$i) \quad \text{logit}(\pi_{ij}^c) = \alpha_0^c + \alpha_1^c \cdot \text{Tr} + \alpha_2^c \cdot \text{Pd} + a_i$$

Logistic-Logistic-Normal MMM:

$$i) \quad \pi_{ij}^m = \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Tr} + \alpha_2^m \cdot \text{Pd})$$

$$ii) \quad \pi_{ij}^c = \text{expit}(\Delta_{ij} + a_i)$$

with Δ_{ij} as in Section 2.1, which will require iterative, numerical integration to obtain. We also consider the two practical, integration-free approximations of Section 3.1. To obtain α^m from the multiplication factors, $(\alpha_0^m, \alpha_1^m) = (\alpha_0^c, \alpha_1^c) \phi^{JKB}$ for the JKB approximation and $(\alpha_0^m, \alpha_1^m) = (\alpha_0^c, \alpha_1^c) \phi^{ZLA}$ for the ZLA approximation, using (α_0^c, α_1^c) and $\mathbf{z}'_{ij} \mathbf{D} \mathbf{z}_{ij} = \tau^2$ from the GLMM output. 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.

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 LLN are the logistic-probit-normal (LPN) and the logistic-logistic-bridge (LLB):

LPN MMM:

$$i) \quad \pi_{ij}^m = \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Tr} + \alpha_2^m \cdot \text{Pd})$$

$$ii) \quad \pi_{ij}^c = \Phi(\Delta_{ij} + a_i)$$

LLB MM:

$$i) \quad \pi_{ij}^m = \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Tr} + \alpha_2^m \cdot \text{Pd})$$

$$ii) \quad \pi_{ij}^c = \text{expit}\left[\Delta_{ij} + B_l^{-1} \{\Phi(a_i/\tau)\}\right]$$

with Δ s from Sections 2.2 and 2.3 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.

4.2. Crossover Trial Results

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, LLB and ZLA methods use a logit link. The LLB 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 (see web appendix for more about the Page approximation). 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.

An advantage of using a MMM formulation is that we may conduct likelihood-based inference (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 & 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 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. 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. Figure 1 indicates that all six methods provide similar evidence about this marginal treatment effect, despite the variety of assumptions about the association structure.

4.3. Visual Impairment and Race Study

Three conditional models (below, left column) and their corresponding logit-marginalizations (right column) were fitted to visual impairment data (Table 3). The estimates from each appear in Table 4 (note: all models assume $Y_{ij}|a_i \sim \text{Binary}(\pi_{ij}^c)$ and $a_i \sim N(0, \tau^2)$):

LN conditional model:

$$i) \quad \pi_{ij}^c = \text{expit}(\alpha_0^c + \alpha_1^c \cdot \text{Black} + a_i)$$

LLN MMM:

$$\begin{aligned} i) \quad \pi_{ij}^m &= \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Black}) \\ ii) \quad \pi_{ij}^c &= \text{expit}(\Delta_{ij} + a_i) \end{aligned}$$

PN conditional model:

$$i) \quad \pi_{ij}^c = \Phi(\alpha_0^c + \alpha_1^c \cdot \text{Black} + a_i)$$

LPN MMM:

$$i) \quad \pi_{ij}^m = \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Black})$$

$$ii) \quad \pi_{ij}^c = \Phi(\Delta_{ij} + a_i)$$

LB conditional model:

$$i) \quad \pi_{ij}^c = \text{expit}(\alpha_0^c + \alpha_1^c \cdot \text{Black} + B_l^{-1}\{\Phi(a_i / \tau)\})$$

LLB MMM:

$$i) \quad \pi_{ij}^m = \text{expit}(\alpha_0^m + \alpha_1^m \cdot \text{Black})$$

$$ii) \quad \pi_{ij}^c = \text{expit}\{\Delta_{ij} + B_l^{-1}\{\Phi(a_i / \tau)\}\}$$

In the crossover trial example both the conditional and marginal effects are directly observable. For an epidemiological study of race and visual impairment, the conditional effect is not directly observable because race is constant across repeated-measurements for all subjects in the dataset. We note the estimated subject-specific effects are extrapolations of the observed data whenever predictors do not vary within cluster. Therefore, a marginal interpretation is preferred to compare population prevalences of visual impairment. To illustrate marginalizing of conditional models and how the estimates from the marginal and conditional model differ, we use data from Tielsch et al. (1991). Responses are binary eye health indicators for each eye, healthy=0 vs impaired=1. A total of 5199 subjects (2343 black, 2756 white) were in the study. Table 3 replicates the data from Tielsch et al. (1991).

From Table 4 we note that between the conditional and corresponding marginalized model, the log-likelihoods, estimates and standard errors of the variation of the random intercept distribution are exactly the same. The regression coefficients differ among the conditional models, showing the sensitivity to random effects distributional assumptions between LN and LB, as well as interpretational differences between the probit link of PN and logit link of LN and LB. The marginalization to the logit link diminishes the differences, as seen in the regression coefficients of the marginalized models LLN, LPN, and LLB. Attenuation of marginal regression parameters as compared to conditional regression parameters is well known (Neuhaus, Kalbfleisch & Hauck (1991); Zeger, Liang & Albert (1988)).

The marginal mean parameters have population-average interpretations, with the odds ratio $\exp(\alpha_1^m)$ contrasting the odds of eye disease in the black population to odds of eye disease in the white population. While the marginal race contrast estimates the directly observable effect of a group of black subjects compared to a group of white subjects, the conditional contrast is not directly observable. As the interpretation of the $\exp(\alpha_1^c)$ requires subjects themselves to be able to change race, presentation imbues a counterfactual meaning.

5. Discussion

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 UN, LPN and LLB models. The LPN and PPN models are instances of a “link-probit-normal” model, which is computationally advantageous because the marginalization integral is known (Caffo & Griswold (2006)). The LPN gives marginal log-odds interpretation, while the PPN provides

a unifying family for the marginal, conditional, and random intercept distribution (Caffo, An & Rohde (2007)).

We agree with Neuhaus (2000) that forcing all aspects of an analysis into a single model can be 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 existing computing procedures. We show obtaining marginal parameters is also straightforward by simply re-specifying selected conditional model components and using the same computing procedures. 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 method for exploring the relationship between the answers of these questions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Amemiya T. Qualitative response models: a survey. *Journal of Economic Literature*. 1981; 19(4): 1483–1536.
- Beal S, Sheiner L. Heteroscedastic nonlinear regression. *Technometrics*. 1988; 30(3):327–338.
- Blume J. Likelihood methods for measuring statistical evidence. *Statistics in medicine*. 2002; 21(17): 2563–2599. [PubMed: 12205699]
- Booth J, Caffo B. Unequal sampling for Monte Carlo EM algorithms. *Computational statistics & data analysis*. 2002; 39(3):261–270.
- Booth J, Hobert J, Jank W. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modelling*. 2001; 1(4):333.
- Breslow N, Clayton D. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88(421):9–25.
- Caffo B, An M, Rohde C. Flexible random intercept models for binary outcomes using mixtures of normals. *Computational Statistics and Data Analysis*. 2007; 51(11):5220–5235. [PubMed: 18628822]
- Caffo B, Griswold M. A user-friendly tutorial on link-probit-normal models. *The American Statistician*. 2006; 60:139–145.
- Chib S, Carlin B. On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*. 1999; 9(1):17–26.
- Davidian M, Giltinan D. Some general estimation methods for nonlinear mixed-effects model. *Journal of Biopharmaceutical Statistics*. 1993; 3(1):23–55. [PubMed: 8485535]
- Demidenko, E. *Mixed models: theory and applications*. Wiley-IEEE; 2004.
- Diggle, P.; Heagerty, P.; Liang, K.; Zeger, S. *Analysis of longitudinal data* (2002). Vol. 8. Oxford University Press; Oxford, England: 2002. Section
- Goldstein, H. *Multilevel statistical models*. 3rd. Kendall's library of statistics; 3. E. Arnold, Oxford University Press; London New York: 2003. [Distributor]
- Heagerty P. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*. 1999; 55(3):688–698. [PubMed: 11314994]
- Heagerty P, Zeger S. Marginalized multilevel models and likelihood inference. *Statistical Science*. 2000; 15(1):1–19.
- Hubbard A, Ahern J, Fleischer N, Laan M, Lippman S, Jewell N, Bruckner T, Satariano W. To GEE or not to GEE: comparing population average and mixed models for estimating the associations

between neighborhood risk factors and health. *Epidemiology*. 2010; 21(4):467–474. [PubMed: 20220526]

- Johnson N, Kotz S, Balakrishnan N. Continuous Univariate Distributions. 1995; Volume 2
- Jones, B.; Kenward, M. Design and analysis of cross-over trials. Chapman & Hall; 2003.
- Kalbfleisch J, Sprott D. Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1970; 32(2):175–208.
- Lee Y, Nelder J. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996; 58(4):619–678.
- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
- Liao, T. Interpreting probability models: Logit, probit, and other generalized linear models. Sage Publications, Inc; 1994.
- Lin X, Breslow N. Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion. *Journal of the American Statistical Association*. 1996; 91(435):1007–1016.
- Lindsey, J. Nonlinear models in medical statistics. Oxford University Press; USA: 2001.
- Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46(3):673–687. [PubMed: 2242409]
- McCulloch C. Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American statistical Association*. 1997; 92(437):162–170.
- McCulloch, C.; Searle, S. Generalized, linear, and mixed models. Wiley-Interscience; 2001.
- Neuhaus J. [Marginalized Multilevel Models and Likelihood Inference]: Comment. *Statistical Science*. 2000; 15(1):21–22.
- Neuhaus J, Kalbfleisch J, Hauck W. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review/Revue Internationale de Statistique*. 1991; 59(1):25–35.
- Page E. Approximations to the cumulative normal function and its inverse for use on a pocket calculator. *Applied Statistics*. 1977; 26(1):75–76.
- Pinheiro J, Bates D. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*. 1995; 4(1):12–35.
- Royall, R. Statistical evidence: a likelihood paradigm. Chapman & Hall/CRC; 1999.
- Scharfstein D, Rotnitzky A, Robins J. Adjusting for nonignorable drop-out using semiparametric nonresponse models (c/r: P1121-1146). *Journal of the American Statistical Association*. 1999; 94(448):1096–1120.
- Sinha S. Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*. 2004; 99(466):451–460.
- Skrondal, A.; Rabe-Hesketh, S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. CRC Press; 2004.
- Stratelli R, Laird N, Ware J. Random-effects models for serial observations with binary response. *Biometrics*. 1984:961–971. [PubMed: 6534418]
- Tielsch J, Sommer A, Katz J, Quigley H, Ezrine S. Socioeconomic status and visual impairment among urban Americans. *Archives of ophthalmology*. 1991; 109(5):637. [PubMed: 2025164]
- Tocher, K. The art of simulation. English Universities Press; 1963.
- Wang Z, Louis T. Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*. 2003; 90(4):765–775.
- Wang Z, Louis T. Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. *Biometrics*. 2004; 60(4):884–891. [PubMed: 15606408]
- Wolfinger R, O'Connell M. Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation*. 1993; 48(3):233–243.
- Wolfinger, RD. Fitting nonlinear mixed models with the new NLMIXED procedure. *Proceedings of the Twenty-Fourth Annual SAS Users Group International*; Citeseer; 1999. p. 1666-1675.

- Zeger S, Karim M. Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American statistical association*. 1991; 86(413):79–86.
- Zeger S, Liang K, Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988:1049–1060. [PubMed: 3233245]

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, $\alpha_1 = 0$ (LR₀).

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.)

Group	“Normal” Responses				Total
	(1,1)	(0,1)	(1,0)	(0,0)	
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	LLB	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

Table 3

Visual Impairment and Race Study Data from an epidemiological study on visual impairment. Responses are visual impairment of the eye (healthy=0, impaired=1) and the demographic covariates are black (B) and white (W).

Group	<u>“Visual Impairment” Responses</u>				Total
	(1,1)	(0,1)	(1,0)	(0,0)	
B	108	102	94	2039	2343
W	103	135	117	2501	2756

Table 4

Visual Impairment Results: Logistic MMM and the corresponding conditional Estimates and Standard Errors. Results for marginal parameters are similar across different methods. Results for conditional models differ among themselves and from the estimates of the marginal models.

Parameter		Marginalized Models			Conditional Models		
		LLN	LPN	LLB	LN	PN	LB
α_0 : Intercept	est.	-2.431	-2.430	-2.421	-4.941	-2.755	-4.307
	s.e.	(0.058)	(0.058)	(0.057)	(0.212)	(0.118)	(0.179)
α_1 : Black	est.	0.082	0.079	0.060	0.146	0.078	0.108
	s.e.	(0.084)	(0.084)	(0.082)	(0.150)	(0.083)	(0.146)
$\varphi : \alpha_1^M / \alpha_1^C$		0.562	1.013	0.556	-	-	-
$\log(\tau^2)$: Association	est.	2.226	1.057	1.963	2.226	1.057	1.963
	s.e.	(0.106)	(0.109)	(0.115)	(0.106)	(0.108)	(0.115)
log-likelihood		-2699.77	-2699.80	-2699.97	-2699.77	-2699.80	-2699.97