BIOMETRICS 55, 688–698
September 1999

# Marginally Specified Logistic-Normal Models for Longitudinal Binary Data

**Patrick J. Heagerty**

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.
*email:* heagerty@biostat.washington.edu

SUMMARY. Likelihood-based inference for longitudinal binary data can be obtained using a generalized linear mixed model (Breslow, N. and Clayton, D. G., 1993, *Journal of the American Statistical Association* **88**, 9–25; Wolfinger, R. and O'Connell, M., 1993, *Journal of Statistical Computation and Simulation* **48**, 233–243), given the recent improvements in computational approaches. Alternatively, Fitzmaurice and Laird (1993, *Biometrika* **80**, 141–151), Molenberghs and Lesaffre (1994, *Journal of the American Statistical Association* **89**, 633–644), and Heagerty and Zeger (1996, *Journal of the American Statistical Association* **91**, 1024–1036) have developed a likelihood-based inference that adopts a marginal mean regression parameter and completes full specification of the joint multivariate distribution through either canonical and/or marginal higher moment assumptions. Each of these marginal approaches is computationally intense and currently limited to small cluster sizes. In this manuscript, an alternative parameterization of the logistic-normal random effects model is adopted, and both likelihood and estimating equation approaches to parameter estimation are studied. A key feature of the proposed approach is that marginal regression parameters are adopted that still permit individual-level predictions or contrasts. An example is presented where scientific interest is in both the mean response and the covariance among repeated measurements.

KEY WORDS: Estimating equation; Marginal model; Quasi-likelihood; Random effects model.

## 1. Introduction

In a prospective study of first-episode schizophrenia patients (Thara et al., 1994), interest is in the description of "patterns of course" and the factors that influence them. Disease measures are monthly indicators of the presence or absence of specific symptoms. Covariates of interest include both time-dependent variables, such as an indicator of whether a patient received pharmacotherapy, and time-independent variables, such as gender and age at onset. For a single symptom the response variable is binary (presence/absence), and categorical data regression methods can be used to characterize the relationship between symptom course and covariates.

Marginal models (Liang and Zeger, 1986) and generalized linear mixed models (Breslow and Clayton, 1993) are two major regression approaches for the analysis of longitudinal data. Generalized linear mixed models account for within-subject dependence in an intuitive fashion by assuming the existence of unobserved random effects. Mean models are constructed conditional on the latent variables, and dependence parameters are given by the random effects variance components. Use of a nonlinear mixed model, such as a logistic-normal model (Stiratelli, Laird, and Ware, 1984), leads to regression parameters that have a simple interpretation for covariates that vary within a subject (i.e., are time dependent). These have been termed *subject-specific* regression coefficients (Zeger, Liang, and Albert, 1988) because with a simple random intercepts model, they measure the change in an individual's odds of response associated with a change in covariates. However, for covariates that do not vary within individuals, the interpretation of subject-specific coefficients can be difficult or misleading because they measure a contrast in covariates when the random effects are controlled for, i.e., are held equal. This quantity is not directly observed for time-independent covariates. With nonlinear models, the magnitude and interpretation of such coefficients depends entirely on the random effects model assumptions.

For analysis with time-independent covariates, a marginal regression model is recommended (Zeger et al., 1988; Neuhaus, Kalbfleish, and Hauck, 1991; Graubard and Korn, 1994). A marginal regression coefficient simply describes how the response average changes across various subsets of the study population where subsets are defined by covariate values. Marginal means can be interpreted as averaging over both measurement error and random interindividual heterogeneity, which is unexplained by measured covariates. Estimation for marginal regression parameters was originally approached using semiparametric estimating equations where only the first and second moments are specified (Liang and Zeger, 1986; Prentice, 1988; Lipsitz, Laird, and Harrington, 1991). Recently, likelihood-based approaches have been proposed, but these methods can be computationally intense even for small cluster sizes (Fitzmaurice and Laird, 1993; Molenberghs and Lesaffre, 1994; Heagerty and Zeger, 1996).

Our approach is to adopt a hierarchical logistic-normal

model (Stiratelli et al., 1984) and to specify a regression model for both the induced marginal mean and the random effects variance components. This model adopts the flexibility and interpretability of random effects models for introducing dependence, but it builds regression structure for the marginal mean, thereby allowing valid application with both time-dependent and time-independent covariates. The marginally specified logistic-normal model permits both likelihood and estimating function approaches to estimation and allows individual-level empirical Bayes predictions. Also, estimates of changes over time for an individual can easily be obtained based on the fitted model parameters. We apply the methods to longitudinal symptom measures, characterizing systematic and random components of variation in these data.

The estimating equations that are adopted in this manuscript are similar to those used in marginal quasi-likelihood (MQL), as introduced by Zeger et al. (1988) and discussed by Breslow and Clayton (1993). However, by using numerical integration to compute the first and second moments, we are able to obtain a consistent estimation of both mean and variance component parameters. The likelihood approach that we adopt is also related to methods described in Drum and McCullagh (1993), who restricted their attention to models for which the marginal mean can be represented as a linear function of the fixed effects. Also, the estimating equations approach is similar to that proposed by Qu et al. (1992), who included random effects on the probit scale.

## 2. Models

In this section, we propose a marginally specified logistic-normal model for longitudinal binary data and contrast it with alternatives currently in use. Each of these models can be viewed as adopting "mixed regressions" in which models for both the marginal mean and a specific conditional mean are adopted. The marginal mean model captures the relationship between the response variables and measured covariates, whereas the various conditional mean models are used to parameterize dependence.

We focus on longitudinal binary response variables $\boldsymbol{Y}_i = \text{vec}(Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$, measured at recorded times $\boldsymbol{t}_i = \text{vec}(t_{i1}, t_{i2}, \ldots, t_{in_i})$, along with possibly time-dependent exogenous covariates $\boldsymbol{X}_i = \text{matrix}(\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{in_i})$, where at each observation the covariate is a $p$-dimensional vector $\boldsymbol{X}_{ij} = \text{vec}(X_{ij1}, X_{ij2}, \ldots, X_{ijp})$, for $i = 1, 2, \ldots, N$ subjects.

### 2.1 *Marginal Mean*

Consider longitudinal binary measurements $\boldsymbol{Y}_i$ and covariates $\boldsymbol{X}_i$. Define the marginal mean as $\mu_{ij} = E(Y_{ij} \mid \boldsymbol{X}_i)$ and define a marginal generalized linear model as $g(\mu_{ij}) = \boldsymbol{X}_i\boldsymbol{\beta}$ for a link function $g(\mu)$. Specific models may assume that $E(Y_{ij} \mid \boldsymbol{X}_i) = E(Y_{ij} \mid \boldsymbol{X}_{ij})$. However, this assumption should be justified scientifically or verified empirically; otherwise, biased estimates may result (Pepe and Anderson, 1994). The term *marginal mean* is used to highlight the fact that we are not conditioning on other response variables or on unobserved random effects. In the next section, we discuss how alternative conditional mean models may be used to augment the marginal mean model as a natural way to model interresponse dependence.

### 2.2 *Odds Ratio Association Models*

In this section we review and connect log-linear models for characterizing dependence among multivariate binary responses.

*Pairwise log-odds ratios.* Lipsitz et al. (1991) introduced a semiparametric model for clustered binary data where only marginal means $\mu_{ij}$ and pairwise odds ratios are specified. The present authors use the odds ratio $\Psi_{i(j,k)}$ as a natural and interpretable measure of dependence between the responses $Y_{ij}$ and $Y_{ik}$ and approach estimation through the use of paired estimating equations. Carey, Zeger, and Diggle (1993) introduced an alternative estimator that exploits the relationship

$$\text{logit}\, E(Y_{ij} \mid Y_{ik}, \boldsymbol{X}_i) = \Delta^{**}_{i(j,k)} + \{\log \Psi_{i(j,k)}\} Y_{ik}.$$

They develop estimating equations based on $Y_{ij} - E(Y_{ij} \mid Y_{ik}, \boldsymbol{X}_i)$ for an odds ratio regression model, $\log \Psi_{i(j,k)} = \boldsymbol{Z}_{i(j,k)}\boldsymbol{\alpha}^{**}$, where $\boldsymbol{Z}_{i(j,k)}$ is a subset of $\boldsymbol{X}_i$. The parameters $\Delta^{**}_{i(j,k)}$ are specified only indirectly via the marginal means $\mu_{ij}$, $\mu_{ik}$, and the odds ratio $\Psi_{i(j,k)}$ and are not of direct interest. The model can be viewed as combining the marginal mean model, $E(Y_{ij} \mid \boldsymbol{X}_i)$, used to describe the dependence of the average response on covariates, and the conditional mean model given a single other response, $E(Y_{ij} \mid Y_{ik}, \boldsymbol{X}_i)$, used to measure dependence among response variables.

The advantages of the models and estimators given by Lipsitz et al. (1991) and Carey et al. (1993) are that they use association parameters that are easily interpreted and that mean regression parameters can be consistently estimated even under association model misspecification. For multivariate binary data with $n_i > 2$, higher-order model assumptions are required to specify the full joint distribution. Likelihood inference is possible, but computationally difficult, even for moderate cluster sizes (Lang and Agresti, 1994; Molenberghs and Lesaffre, 1994; Glonek and McCullagh, 1995; Heagerty and Zeger, 1996).

*Canonical association models.* Fitzmaurice and Laird (1993) derived GEE1 equations (Liang and Zeger, 1986) as score equations for the marginal mean regression parameter, $\boldsymbol{\beta}$, under the model that specifies both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, the higher order canonical association parameters. In this approach, the dependence model is given by

$$\text{logit}\, E(Y_{ij} \mid Y_{ik}, k \neq j, \boldsymbol{X}_i) = \Delta^{*}_{ij} + \boldsymbol{\theta}^{\mathrm{T}}_{ij}\boldsymbol{W}_{ij},$$

where $\boldsymbol{W}_{ij}$ contains $Y_{ik}$, $k \neq j$ and pairwise and higher order products of $Y_{ik}$, as in classical log-linear models. The parameter $\boldsymbol{\theta}_{ij}$ can be further assumed to follow a regression structure: $\boldsymbol{\theta}_{ij} = \boldsymbol{Z}_{ij}\boldsymbol{\alpha}^{*}$, where $\boldsymbol{Z}_{ij}$ is a subset of $\boldsymbol{X}_i$. Using iterative proportional fitting (Deming and Stephan, 1940), $\Delta^{*}_{ij}$ can be recovered numerically as a function of the model parameters $\Delta^{*}_{ij} = f^{*}(\boldsymbol{\beta}, \boldsymbol{\alpha}^{*})$ and is used for evaluation of score equations and calculation of the information matrix. Note that $\Delta^{*}_{ij}$ is not of primary interest under the mixed parameter model despite having a direct interpretation under the conditional model as the log odds of success for $Y_{ij}$ given all other responses $Y_{ik} = 0$.

The mixed marginal–canonical approach adopts a model for the marginal expectation, $E(Y_{ij} \mid \boldsymbol{X}_i)$ and a conditional model for the association parameters $E(Y_{ij} \mid Y_{ik}, k \neq j, \boldsymbol{X}_i)$.

Unfortunately, the canonical association parameters need to be specified separately for varying cluster sizes $n_i$, and the fitting algorithm quickly grows computationally impractical for $n_i > 10$.

### 2.3 *Proposed Models*

Similar to the models of Lipsitz et al. (1991) and Fitzmaurice and Laird (1993), we propose adopting a pair of regression models. The first model is a marginal logistic regression for the average response as a function of covariates

$$\text{logit}\, E(Y_{ij} \mid \boldsymbol{X}_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}. \tag{1}$$

The second model is used to describe the dependence among longitudinal measurements and is based on a conditional model where instead of conditioning on other response variables, we condition on a latent variable

$$\text{logit}\, E(Y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{X}_i) = \Delta_{ij} + b_{ij}. \tag{2}$$

Further, we assume that the response vector $\boldsymbol{Y}_i$ is conditionally independent given $\boldsymbol{b}_i = \text{vec}(b_{i1}, b_{i2}, \ldots, b_{in_i})$ and that

$$\boldsymbol{b}_i \mid \boldsymbol{X}_i \sim N(\boldsymbol{0}, \boldsymbol{D}_i). \tag{3}$$

We refer to the model given by (1)–(3) as the *marginally specified logistic-normal model.* We assume that the covariance matrix $\boldsymbol{D}_i$ can be obtained as a function of the observation times $\boldsymbol{t}_i$ and a parameter vector $\boldsymbol{\alpha}$. Thus, the complete parameter for this model is $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Common models for longitudinal data include a subject-level random effects model such that $b_{ij} = b_{i0}$ (scalar) yielding $\text{cov}(b_{ij}, b_{ik}) = D_i(j,k) = \sigma^2$, or a temporal association model, where $b_{ij}$ is assumed to have an autoregressive covariance structure and $\text{cov}(b_{ij}, b_{ik}) = D_i(j,k) = \sigma^2 \exp(-\gamma |t_{ij} - t_{ik}|^{\theta})$, where $1 \le \theta \le 2$ (fixed).

In the marginally specified logistic-normal model, the parameter $\Delta_{ij}$ is a function of both the marginal linear predictor $\eta_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta}$ and the random effects standard deviation $\sigma_{ij} = (\text{var}(b_{ij}))^{1/2}$. $\Delta_{ij}$ can be obtained as the solution to the convolution equation

$$h(\eta_{ij}) = \int h(\Delta_{ij} + \sigma_{ij} z)\phi(z)\, dz, \tag{4}$$

where $h = \text{logit}^{-1}$ and $\phi$ is the standard normal density function. Given $(\eta_{ij}, \sigma_{ij})$, this equation can be solved for $\Delta_{ij}$ using numerical integration and Newton–Raphson iteration.

The marginally specified logistic-normal model is related to the conditionally specified logistic-normal model, where the conditional log odds $\Delta_{ij}$ is directly modeled as a function of covariates. This model was introduced by Pierce and Sands (1975), and estimation methods have been summarized by Breslow and Clayton (1993). The model can be written as

$$\text{logit}\, E(Y_{ij} \mid b_{ij}, \boldsymbol{X}_i) = \boldsymbol{X}_{ij}\boldsymbol{\beta}^C + b_{ij},$$

where $b_{ij}$ is assumed to be normally distributed. Because $\boldsymbol{\beta}^C$ contrasts the log odds for different values of $\boldsymbol{X}_{ij}$, but with equal values of $b_{ij}$, it must be interpreted as controlling for the unobserved variables, and it does not hold the same interpretation as the marginal regression parameter $\boldsymbol{\beta}$. As discussed by Zeger et al. (1988), Neuhaus et al. (1991), and Graubard and Korn (1994), interpretation of the conditionally specified regression parameter can be difficult, particularly for covariates that are not time dependent (do not vary within a cluster). Therefore, although the introduction of dependence through a random effect $b_{ij}$ is often scientifically plausible, the impact on the magnitude and interpretation of $\boldsymbol{\beta}^C$ may preclude direct use of this model. Our proposal is to use the logistic-normal model as a simple, interpretable, dependence model; however, we choose to build a regression model for the induced marginal mean.

*Model interpretation.* For any marginal logistic regression model, the parameter $\boldsymbol{\beta}$ is interpreted as contrasting the log odds of success for subgroups defined by measured covariates. For example, if we consider a single binary covariate $X_{ij1}$, then $\beta_1 = \text{logit}\, E(Y_{ij} \mid X_{ij1} = 1) - \text{logit}\, E(Y_{i'j'} \mid X_{i'j'1} = 0)$ measures the variation in success log odds "between groups." In the logistic-normal model, we explicitly assume that individual heterogeneity exists. For the group contrast $\beta_1$, however, we average over this distribution within each group.

Consider the interpretation of the heterogeneity parameter in the simplest marginally specified logistic-normal model: the subject-level effect model that assumes $b_{ij} = b_{i0}$ and $b_{i0} \sim N(0, \sigma^2)$. We can substitute $b_{ij} = \sigma z_i$, where $z_i \sim N(0,1)$, and express the model for random individual variation by the conditional expectation: $\text{logit}\, E(Y_{ij} \mid z_i, \boldsymbol{X}_i) = \Delta_{ij} + \sigma z_i$. This representation shows that the variance component $\sigma$ may be interpreted as a regression coefficient for a standardized omitted covariate, with $\sigma$ contrasting individuals with equal $\Delta_{ij}$, whose $z_i$'s differ by one unit. Because $\Delta_{ij}$ is determined by $\boldsymbol{X}_{ij}$ and the parameters $(\boldsymbol{\beta}, \sigma)$, subgroups defined by $\boldsymbol{X}_{ij}$ are the same as subgroups defined by $\Delta_{ij}$. Therefore, $\sigma$ measures the magnitude of variation in the log odds "between individuals" within a group, where the group is defined by the measured covariates.

For longitudinal data with possibly time-varying $b_{ij}$, the marginal mean captures the systematic variation in the mean that is due to $\boldsymbol{X}_{ij}$, whereas parameters in $\text{cov}(b_{ij})$ provide measures of random variation both across individuals and over time. Consider the general stationary random effects model, $\text{cov}(b_{ij}, b_{ik}) = D_i(j,k) = \sigma^2 \rho(|t_{ij} - t_{ik}|)$, and for simplicity, assume that all subjects are measured at the same set of times, $t_{ij} = t_j$. Consider a fixed time $t_j$ and individuals $i$ and $i'$ such that $\boldsymbol{X}_i = \boldsymbol{X}_{i'}$. The transformed conditional expectations $\text{logit}\, E(Y_{ij} \mid \boldsymbol{X}_i, b_{ij}) = \Delta_{ij} + b_{ij}$ and $\text{logit}\, E(Y_{i'j} \mid \boldsymbol{X}_{i'}, b_{i'j}) = \Delta_{i'j} + b_{i'j}$ differ only in the random effects because the covariates are identical, yielding $\Delta_{ij} = \Delta_{i'j}$. Therefore, variation in the log odds of success across individuals, homogeneous with respect to measured covariates (including time), is given by $b_{ij} - b_{i'j} = \sigma(z_{ij} - z_{i'j}) \sim N(0, 2\sigma^2)$, showing that $\sigma$ measures cross-sectional random variation.

Finally, variation in the log odds within an individual over time that cannot be attributed to variation in measured covariates is given by the contrast $b_{ij} - b_{ik} \sim N(0, 2\sigma^2\{1 - \rho(|t_j - t_k|)\})$, so that $1 - \rho(|t_j - t_k|)$, captures the rate at which variation within a person ("longitudinal") approaches variation between individuals ("cross-sectional").

## 3. Estimation

In this section, we detail both maximum-likelihood and estimating-equation approaches to point estimation and inference for marginally specified logistic-normal model parameters.

For each of the estimation methods, $\Delta_{ij}$ must be computed as a function of the marginal mean parameters $\boldsymbol{\beta}$ and the random effects covariance parameters $\boldsymbol{\alpha}$. This is achieved through numerical solution of the convolution equation (4). Also required for estimation are the partial derivatives of $\Delta_{ij}$; these can be obtained via implicit differentiation of the convolution equation. Details are provided in the Appendix.

To numerically evaluate the convolution equation, we use either a general method such as a 20-point Gauss–Hermite quadrature (Abramowitz and Stegun, 1972) or a specialized method tailored to the logistic-normal integral. Monahan and Stefanski (1992) described a method based on using a mixture of Gaussian distribution functions to approximate the logistic function known as least maximal approximants (LMA). Using LMA with as few as five quadrature locations yields an error bound on the marginal expectation of $10^{-6}$.

### 3.1 *Maximum Likelihood*

The likelihood contribution from measurements $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ can be constructed by the assumption of conditional independence, given $\boldsymbol{b}_i$ and the assumption that $[\boldsymbol{b}_i \mid \boldsymbol{X}_i]$ follows a mean zero Gaussian distribution. We also assume that $\boldsymbol{b}_i$ is a linear transformation of a possibly lower-dimensional random effect $\boldsymbol{b}_i = \boldsymbol{C}_i \boldsymbol{z}_i$ for an $n_i \times q$ matrix $\boldsymbol{C}_i$, and $\boldsymbol{z}_i \in \Re^q$ spherically normal, $\boldsymbol{z}_i \sim N(\boldsymbol{0}, \boldsymbol{I}_{q\times q})$. Examples include a random intercepts model, where $b_{ij} = b_{i0}$, which can be represented as $\boldsymbol{b}_i = \text{vec}(\sigma, \sigma, \ldots, \sigma) z_i$ for $z_i \sim N(0,1)$. Similarly, if $\boldsymbol{b}_i$ is assumed to have an autoregressive structure, then $\boldsymbol{z}_i \in \Re^{n_i}$ and $\boldsymbol{C}_i$ is the Cholesky decomposition of $\text{cov}(\boldsymbol{b}_i)$.

The observed data likelihood for subject $i$ is then a mixture over the random effects distribution and is given by

$$L_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \int \prod_{j=1}^{n_i} P(Y_{ij} = y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{X}_i)\, dF_{b_i},$$

$$L_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \int \prod_{j=1}^{n_i} h(\Delta_{ij} + \boldsymbol{C}_{ij}\boldsymbol{z}_i)^{y_{ij}} \times \left\{1 - h(\Delta_{ij} + \boldsymbol{C}_{ij}\boldsymbol{z}_i)\right\}^{1-y_{ij}} \phi_q(z_i)\, dz_i,$$

where $\phi_q$ represents the unit spherical multivariate Gaussian density and $\phi_q(z_i) = \Pi_{k=1}^{q} \phi(z_{ik})$ for $\phi$, the standard normal density function.

Since $L_i$ cannot be evaluated analytically, numerical methods are required to compute the $q$-dimensional integral. For low-dimensional random effects models, Gauss–Hermite quadrature can be used, requiring $K^q$ total evaluation points when using $K$ points in each dimension ($K$ is typically 5, 10, or 20). We have implemented numerical integration methods for the logistic-normal model only for $q = 1$ (random intercepts, $b_{ij} = b_{i0}$), although allowing $\sigma_i = (\text{var}(b_{i0}))^{1/2}$ to possibly depend on individual level covariates. Extension to higher-dimensional random effects structures is possible. However, as $q$ increases, e.g., with autocorrelated random effects, the computational burden for quadrature methods grows exponentially, and alternative approaches may need to be considered (McCulloch, 1997).

### 3.2 *Quadratic Estimating Equations*

An estimating-equations approach requires the specification of a model for the marginal mean and a "working model" for the marginal covariance, $\text{cov}(\boldsymbol{Y}_i)$. It is possible to use estimating equations (Liang and Zeger, 1986) and a "working" logistic-normal model because we are able to calculate the induced covariance structure. The primary advantage of an estimating-equations approach is that the inference regarding regression parameters $\boldsymbol{\beta}$ can be made robust to incorrect specification of the within-subject dependence model.

Using paired estimating equations, as in Prentice (1988), Lipsitz et al. (1991), or Carey et al. (1993), we obtain estimates of the mean regression parameter and the logistic-normal variance components. We can use these model estimates to construct valid covariance weight matrices for an iterative estimation of $\boldsymbol{\beta}$. Define the estimating functions as follows:

$$\boldsymbol{U}_1^N(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_i^N \boldsymbol{D}_{i1}^{\mathrm{T}} \boldsymbol{V}_{i1}^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i), \tag{5}$$

$$\boldsymbol{U}_2^N(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_i^N \boldsymbol{D}_{i2}^{\mathrm{T}} \boldsymbol{V}_{i2}^{-1} (\boldsymbol{S}_i - \boldsymbol{\nu}_i), \tag{6}$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i \mid \boldsymbol{X}_i)$; $\boldsymbol{V}_{i1} = \text{cov}(\boldsymbol{Y}_i)$; $\boldsymbol{D}_{i1} = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$; $\boldsymbol{S}_i = \text{vec}\{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\}$ for $j \neq k$; $\boldsymbol{\nu}_i = E(\boldsymbol{S}_i \mid \boldsymbol{X}_i)$; $\boldsymbol{V}_{i2} \approx \text{cov}(\boldsymbol{S}_i \mid \boldsymbol{X}_i)$; and $\boldsymbol{D}_{2i} = \partial \boldsymbol{\nu}_i / \partial \boldsymbol{\alpha}$. We refer to $\boldsymbol{U}_1^N = \boldsymbol{0}$ and $\boldsymbol{U}_2^N = \boldsymbol{0}$ as quadratic estimating equations (QEE).

Estimates of $\boldsymbol{\beta}$ obtained as the solution to $\boldsymbol{U}_1^N = \boldsymbol{0}$ have been shown to be consistent under quite general conditions, including incorrect specification of $\boldsymbol{V}_{i1}$ (Liang and Zeger, 1986). However, efficient estimation requires that $\boldsymbol{V}_{i1}$ be correctly specified, and poor efficiency may result for grossly misspecified models (Fitzmaurice, Laird, and Rotnitzky, 1993; Mancl and Leroux, 1996). If the expectation $\boldsymbol{\nu}_i$ is also correctly specified, then covariance parameter estimates $\widehat{\boldsymbol{\alpha}}$ obtained by jointly solving $\boldsymbol{U}_1^N = \boldsymbol{0}$ and $\boldsymbol{U}_2^N = \boldsymbol{0}$ will be consistent. Details regarding the asymptotic distribution of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ obtained as the solution to paired estimating equations can be found in Prentice (1988) or Lipsitz et al. (1991).

Use of an estimating-equations approach for the marginally specified logistic-normal model requires that $\text{cov}(Y_{ij}, Y_{ik} \mid \boldsymbol{X}_i) = \pi_{i(j,k)} - \mu_{ij}\mu_{ik}$ be computed. This requires at most a two-dimensional numerical integration over the distribution of $(b_{ij}, b_{ik})$ to compute the pairwise probability $\pi_{i(j,k)}$. We use either a general numerical method, such as Gauss–Hermite quadrature, or a specialized method, such as LMA. Quadratic estimating equations provide feasible estimation for even high-dimensional random effects models, such as latent autoregressive random effects, because evaluation of all possible pairwise probabilities rather than the joint likelihood is required.

### 3.3 *Individual-Level Estimation*

Given point estimates for the mean parameters $\boldsymbol{\beta}$ and the variance components $\boldsymbol{\alpha}$, we can obtain an empirical Bayes estimate of the random effects $b_{ij}$, which we denote $\widetilde{b}_{ij}$. To obtain the empirical Bayes estimates for subject $i$, we solve the posterior score equations for $\boldsymbol{z}_i$ where, as in Section

3.1, $\boldsymbol{b}_i = \boldsymbol{C}_i \boldsymbol{z}_i$. The posterior for $\boldsymbol{z}_i$ is proportional to the conditional distribution $[\boldsymbol{Y}_i \mid \boldsymbol{z}_i]$ times the prior $[\boldsymbol{z}_i]$. Upon taking the logarithm and then the derivative with respect to $\boldsymbol{z}_i$, we obtain the posterior score equations:

$$\mathbf{0} = \sum_j \boldsymbol{C}_{ij}^{\mathrm{T}} \{Y_{ij} - h(\widehat{\Delta}_{ij} + \boldsymbol{C}_{ij}\boldsymbol{z}_i)\} - \boldsymbol{z}_i,$$

where $\widehat{\Delta}_{ij}$ is obtained as a function of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ via equation (4). Defining $\widetilde{\boldsymbol{z}}_i$ as the solution to the preceding equations then yields the empirical Bayes estimate $\widetilde{b}_{ij} = \boldsymbol{C}_{ij}\widetilde{\boldsymbol{z}}_i$. The estimator $\widetilde{b}_{ij}$ can also be motivated as the Laplace approximation to the posterior mean, and an expression for the approximate posterior variance is available (Booth and Hobert, 1998).

One key reason for adopting the conditionally specified logistic-normal model is that subject-specific effects can be estimated (Zeger et al., 1988). This generally refers to the ability to consider changes in an individual's covariates and then to estimate the corresponding change in the individual's odds of response. For a random intercepts model with a single covariate, this amounts to summarizing contrasts of the form $\{\beta_0^C + \beta_1^C(X_{ij}+1) + b_i\} - \{\beta_0^C + \beta_1^C X_{ij} + b_i\} = \beta_1^C$, which can be directly provided via the conditionally specified logistic-normal model. Because we adopt a model for the marginal mean, the analogous $\beta_1$ cannot be given the same individual-level intervention interpretation. However, we are able to compute the corresponding conditional log odds $\Delta_{ij}$ based on the marginally specified logistic-normal model and can therefore provide $\Delta(X_{ij}+1) - \Delta(X_{ij})$ as the corresponding estimate of the change in log odds at the individual level. Thus, although our regression focus is on the marginal mean, the use of an underlying logistic-normal model allows estimates of individual-level effects to be easily obtained as model summaries. We illustrate this aspect in the example presented in Section 5.

### 3.4 *Missing Data*

An important issue with longitudinal data is the ubiquitous phenomenon of missing data. To obtain valid inference using estimating equations (QEE) based on observed cases requires that the data be missing completely at random (MCAR), whereas likelihood inference requires only that observations are missing at random (MAR) (Laird, 1988). Therefore, one advantage of likelihood-based methods is the ability to handle data that may be MAR. The logistic-normal model easily allows analysis with clusters of varying sizes, the natural consequence of missing observations.

## 4. Example

The Madras longitudinal schizophrenia study followed 90 first-episode schizophrenics for 10 years with the primary objective of characterizing the natural history of the disease process (Thara et al., 1994). Each month after initial hospitalization, six symptoms were recorded as either present or absent. Here, we consider two symptoms: *Delusions*, termed a "positive symptom," and *Apathy*, a "negative symptom." Our analysis considers data collected during the first 12 months of the study and focuses on differences with respect to gender and age at onset (age > 20 years). Scientific interest also lies in the hypothesis that the positive symptoms are "states", that is, manifestations of a short-term physiological disturbance, and that the negative symptoms represent "traits" or characteristics that are specific to individual patients and do not vary in time. An assessment of the covariance structure for the longitudinal measurements can be used to distinguish temporal ("state") versus long-range ("trait") dependence.

The analysis of these symptoms illustrates several key points regarding the marginally specified logistic-normal model: (1) use of a marginal mean allows changes in the latent covariance structure to be considered without affecting the meaning of the regression parameters; (2) conditionally specified parameters can be substantially different from the marginally specified parameters, and their estimates can be sensitive to the covariance assumptions; and (3) use of an estimating equations approach permits valid inference for mean parameters by using a "working" random effects covariance model. Finally, we illustrate that, although marginal means are the regression focus, our approach does allow assessment of hypothetical intervention at the individual level.

### 4.1 *Negative Symptom—Apathy*

The prevalence of *Apathy* declines over the first 12 months from a prevalence of 43% in the first month after hospitalization to a prevalence of 7% at the end of the first year. To assess the association between this symptom and age ($X_{ij1}$: 0 = onset age $\leq$20 years, 1 = onset after age 20 years) and gender ($X_{ij2}$: 0 = male, 1 = female), a marginal logistic regression model is constructed using a linear trend in time, with the time-independent binary covariates $X_{ij1}$ (age) and $X_{ij2}$ (gender):

$$\operatorname{logit} E(Y_{ij} \mid \boldsymbol{X}_i) = \beta_0 + \beta_1(t_j - 6) + \beta_2 X_{ij1} + \beta_3 X_{ij2},$$

where $Y_{ij}$ records the presence or absence of *Apathy* for subject $i$ during month $j$ ($t_{ij} = j, j = 1, 2, \ldots, 12$).

To complete the model specification we adopt a second regression model that describes the individual-level heterogeneity: $\operatorname{logit} E(Y_{ij} \mid \boldsymbol{X}_i, b_i) = \Delta_{ij} + b_i$, and $b_i \mid \boldsymbol{X}_i \sim N(0, \sigma_i^2)$, where $\sigma_i$ is allowed to depend on covariates $\boldsymbol{Z}_i$, a subset of $\boldsymbol{X}_i$:

$$\log(\sigma_i) = \boldsymbol{Z}_i \boldsymbol{\alpha}.$$

In Table 1, we present both maximum likelihood (ML) estimates and QEE estimates for two heterogeneity models. The first model assumes that the variance of $b_i$ does not depend on $\boldsymbol{X}_i$, which corresponds to a standard random intercepts model. Both female subjects and those with an older age at onset are estimated to have a higher probability of developing symptoms, although the point estimates are not significant at the nominal 5% level. The ML estimate for the heterogeneity parameter is $\widehat{\sigma} = \exp(0.972) = 2.64$, indicating large subject-to-subject variation in the odds of *Apathy*. The QEE estimate is comparable: $\widehat{\sigma} = \exp(0.884) = 2.42$.

The second model allows different heterogeneity depending on the gender of the subject: $\log(\sigma_i) = \alpha_0 + \alpha_1 X_{ij2}$. Based on either the Wald test or the likelihood ratio test, the two groups have significantly different degrees of between-subject variation, with $\widehat{V}(b_i)$ larger among males than among females, $\widehat{V}(b_i \mid X_{ij2} = 0) = \exp(1.408) = 4.09$ and $\widehat{V}(b_i \mid X_{ij2} = 1) = \exp(1.408 - 0.853) = 1.74$. Allowing different random

**Table 1**
*Marginally and conditionally specified logistic-normal model estimates for* Apathy

| Coefficient | Model 1 Estimate | Model 1 Model SE | Model 1 Empirical SE | Model 2 Estimate | Model 2 Model SE | Model 2 Empirical SE |
|---|---|---|---|---|---|---|
| **Marginal Mean/Maximum Likelihood** | | | | | | |
| Marginal mean | | | | | | |
| Intercept | −2.071 | 0.318 | | −1.945 | 0.242 | |
| Time | −0.217 | 0.029 | | −0.209 | 0.028 | |
| Age | 0.444 | 0.319 | | 0.494 | 0.247 | |
| Gender | 0.255 | 0.314 | | −0.026 | 0.265 | |
| Log($\sigma$) | | | | | | |
| Intercept | 0.9716 | 0.139 | | 1.408 | 0.189 | |
| Gender | | | | −0.853 | 0.267 | |
| Max log L | −301.92 | | | −297.05 | | |
| **Marginal Mean/Quadratic Estimating Equations** | | | | | | |
| Marginal mean | | | | | | |
| Intercept | −2.163 | 0.398 | 0.409 | −2.120 | 0.404 | 0.381 |
| Time | −0.232 | 0.031 | 0.041 | −0.236 | 0.031 | 0.040 |
| Age | 0.544 | 0.383 | 0.362 | 0.505 | 0.357 | 0.347 |
| Gender | 0.141 | 0.358 | 0.337 | 0.089 | 0.354 | 0.339 |
| Log($\sigma$) | | | | | | |
| Intercept | 0.884 | | 0.191 | 1.365 | | 0.505 |
| Gender | | | | −0.996 | | 0.576 |
| **Conditional Mean/Maximum Likelihood** | | | | | | |
| Conditional mean | | | | | | |
| Intercept | −3.909 | 0.640 | | −4.300 | 0.644 | |
| Time | −0.376 | 0.044 | | −0.374 | 0.044 | |
| Age | 0.864 | 0.593 | | 0.820 | 0.482 | |
| Gender | 0.583 | 0.576 | | 1.202 | 0.605 | |
| Log($\sigma$) | | | | | | |
| Intercept | 0.978 | 0.141 | | 1.300 | 0.149 | |
| Gender | | | | −0.727 | 0.235 | |
| Max log L | −303.09 | | | −299.84 | | |

intercept variances results in the point estimate of $\beta_2$ being approximately 0 for either ML or QEE estimation. Thus, the marginally specified logistic-normal models provide the conclusion that the prevalence of symptoms is the same in men and women but that males are a more heterogeneous group than women.

Table 1 also shows the results of ML estimation for comparable conditionally specified logistic-normal models. Here the regression coefficients are defined conditional on the random effect

$$\text{logit}\, E(Y_{ij} \mid \boldsymbol{X}_i, b_i) = \beta_0^C + \beta_1^C(t_j - 6) + \beta_2^C X_{ij1} + \beta_3^C X_{ij2} + b_i,$$

and $b_i$ is assumed to be normally distributed with standard deviation $\sigma_i$ that may depend on covariates. Note that the point estimate $\widehat{\boldsymbol{\beta}}^C$ is nearly twice the marginally specified point estimate, although direct comparison is misleading because these represent different contrasts. Specifically, $\widehat{\beta}_3^C$ contrasts females and males that have identical $b_i$, a comparison that is not directly observed in these data.

Allowing the variance of $b_i$ to depend on gender results in the point estimate $\widehat{\beta}_3^C$ increasing from 0.583 to 1.202. This is particularly striking because with the marginally specified model, the group contrast was reduced ($\widehat{\beta}_3 = -0.026$), an apparently contradictory result. However, we can compute the induced marginal probability of symptoms based on the conditionally specified model as $\int h(-4.300 + 3.699z)\phi(z)\,dz = 0.1460$ among males and $\int h(-4.300+1.202+1.773z)\phi(z)\,dz = 0.1072$ among females. Therefore, although the regression point estimate under the conditionally specified model suggests a higher probability of symptoms for females, this is not what is implied in aggregate due to the difference in the degree of between-subject variation among males, as compared with that among females.

To assess whether there is evidence that the probability of *Apathy* varies stochastically over time within a person we fit the marginally specified logistic-normal model with $\boldsymbol{b}_i \mid \boldsymbol{X}_i \sim N\{\boldsymbol{0}, D(\sigma, \rho, \boldsymbol{t}_i)\}$, where $b_{ij}$ is assumed to be a Gaussian process with covariance $\text{cov}(b_{ij}, b_{ik}) = \sigma^2 \exp(-\gamma|t_j - t_k|^2)$ and $-\gamma = \log(\rho)$. Using QEE to estimate $(\boldsymbol{\beta}, \sigma, \rho)$ resulted in convergence to the boundary $\widehat{\rho} = 1$, implying $b_{ij} = b_{i0}$ or no temporal variation in the random effect. Thus, we conclude

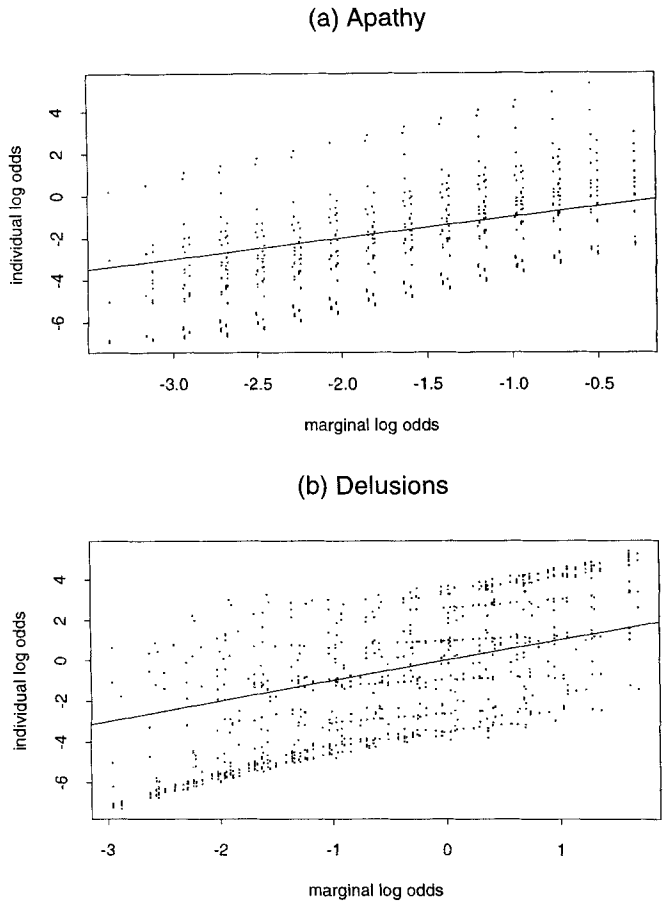


**Figure 1.** Components of variation plots. The vertical axis plots the empirical Bayes estimates of individual log odds $\widehat{\Delta}_{ij} + \widehat{b}_{ij}$, whereas the horizontal axis represents the marginal linear predictor $\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}$. These plots allow direct comparsion of between-group (horizontal axis) and between-individual (vertical axis) variation.

that *Apathy* appears to be a "trait" because the propensity for this symptom varies from person to person ($\sigma$) but does not appear to vary over time ($\widehat{\rho} = 1$).

An advantage of the logistic-normal model is that individual empirical Bayes estimates can be calculated. Figure 1a plots the marginal linear predictor $\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}$ on the horizontal axis and the individual posterior mode $\widehat{\Delta}_{ij} + \widehat{b}_i$ on the vertical axis, both obtained from the ML estimates assuming a common heterogeneity parameter. The horizontal axis displays differences in log odds attributed to covariate differences (time, gender, and age), whereas the vertical differences capture both covariate differences and individual heterogeneity. For a given ordinate, the vertical distribution displays cross-sectional variation among individuals that are otherwise homogeneous based on measured covariates. The $R^2$ statistic for this plot is 34% and can be interpreted as measuring the proportion of total variation in estimated individual log odds that can be attributed to covariate differences. This summary is based on an underestimate of magnitude of the individual variation because empirical Bayes estimates $\widetilde{b}_i$ have less variability than the true random effects $b_i$ (Louis, 1984).

### 4.3 *Positive Symptom—Delusions*

Table 2 shows the results of fitting both a simple random intercepts model and a latent Gaussian autocorrelated random effects model to *Delusions*. Assuming random intercepts results in $\widehat{\sigma} = \exp(0.426) = 1.531$ based on the ML fit. However, fitting a marginally specified logistic-normal model that permits time decay in the dependence, $\mathrm{cov}(b_{ij}, b_{ik}) = \sigma^2 \exp(-\gamma|t_j - t_k|^2)$ and $-\gamma = \log(\rho)$, yields $\widehat{\sigma} = 3.82$ and a correlation estimate of $\widehat{\rho} = 0.92$. This implies that observations that are closer in time are more strongly dependent than observations that are widely separated (note that the latent correlation is 0.005 for $|t_j - t_k| = 8$). Therefore, this analysis suggests that *Delusions* represent a "state," a phenomenon with strong but decaying serial dependence.

The interpretation of the marginally specified regression parameter $\beta$ does not depend on the within-subject dependence assumptions. However, the point estimate obtained via QEE will depend on the covariance assumptions through the weighting $\boldsymbol{V}_i^{-1}$ used to obtain the solution to the estimating equations. In addition, the model-based standard errors reflect the assumptions made regarding $b_{ij}$.

An intermediate step in the QEE estimation algorithm requires that $\Delta_{ij}$ be calculated from the marginal mean and the covariance parameter $\sigma$. A conditionally specified logistic-normal model directly parameterizes $\Delta_{ij}$ as $\boldsymbol{X}_{ij}\boldsymbol{\beta}^C$. Therefore, an approximation to the ML estimate of $\boldsymbol{\beta}^C$ can be obtained by a simple regression of $\widehat{\Delta}_{ij}$ on $\boldsymbol{X}_{ij}$. Using the fitted $\Delta_{ij}$ from the simple random intercepts model presented in Table 2 yields the approximation $\widetilde{\boldsymbol{\beta}}^C = (-0.658, -0.477, 0.405, -0.604)$. This corresponds closely to the actual ML estimate $\widehat{\boldsymbol{\beta}}^C = (-0.640, -0.472, 0.492, -0.651)$. Because the ML estimation is difficult for the time-varying random effects model, we only have the approximate estimate obtained from the QEE fit, $\widetilde{\boldsymbol{\beta}}^C = (-0.983, -0.826, 0.945, -1.334)$. This approximation suggests that the conditionally specified model estimates may be quite different depending on whether the random effects are assumed to be constant or stochastic in time.

We can use the marginal linear predictors $\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}$ and the empirical Bayesian individual estimates $\widehat{\Delta}_{ij} + \widehat{b}_{ij}$ to graphically display components of variance. The horizontal axis in Figure 1b shows the difference in log odds attributable to measured covariates, and the vertical axis shows the individual estimates, reflecting both fixed covariate differences and estimated stochastic differences. Because we are using a time-varying random effect $b_{ij}$, the plots for *Delusions* and for *Apathy* are different. With a simple random intercept model, an individual's relative ranking remains constant (unless covariates change), as in the analysis for *Apathy*, but with a latent autocorrelated Gaussian process, an individual's ranking may change over time.

We also considered analysis with the covariate *drug*, $X_{ij3}$, which is a time-dependent binary covariate denoting whether the patient was on medication during the month. Table 2 gives the estimated marginal log odds associated with medication as $-0.074$. To assess the individual-level effect of beginning drug therapy, we computed the implied conditional linear predictor

**Table 2**
*Marginally specified logistic-normal model estimates for* Delusions *with either a random intercepts model or a latent AR process model*

| | Model 1 | | | | |
|---|---|---|---|---|---|
| | MLE | | QEE | | |
| Coefficient | Estimate | Model SE | Estimate | Model SE | Empirical SE |
| **Random Intercepts Model** | | | | | |
| Marginal mean | | | | | |
| Intercept | −0.488 | 0.264 | −0.500 | 0.273 | 0.240 |
| Time | −0.350 | 0.026 | −0.352 | 0.026 | 0.043 |
| Age | 0.301 | 0.274 | 0.295 | 0.279 | 0.252 |
| Gender | −0.449 | 0.267 | −0.443 | 0.270 | 0.256 |
| Log($\sigma$) | | | | | |
| Intercept | 0.426 | 0.119 | 0.389 | | 0.158 |

| | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|
| Coefficient | Estimate | Model SE | Empirical SE | Estimate | Model SE | Empirical SE |
| **Latent AR Process Model/QEE** | | | | | | |
| Marginal mean | | | | | | |
| Intercept | −0.401 | 0.261 | 0.220 | −0.350 | 0.286 | 0.246 |
| Time | −0.330 | 0.035 | 0.039 | −0.331 | 0.035 | 0.040 |
| Age | 0.385 | 0.269 | 0.244 | 0.386 | 0.270 | 0.244 |
| Gender | −0.543 | 0.260 | 0.247 | −0.549 | 0.261 | 0.248 |
| Drug | | | | −0.074 | 0.171 | 0.167 |
| Variance components (95% confidence interval) | | | | | | |
| $\sigma$ | 3.819 | (2.978, 4.896) | | 3.842 | (2.995, 4.930) | |
| $\rho$ | 0.920 | (0.854, 0.957) | | 0.920 | (0.855, 0.956) | |

at month 6 for a male with late age at onset ($X_{ij1} = 0, X_{ij2} = 0$) and currently off medication ($X_{ij3} = 0$). We let $b_{i6} = 0$ represent a "typical" individual at month 6, allowing us to compute the conditional probability $h\{\Delta_{i6}(X_{i6}) + b_{i6}\} = P\{Y_{i6} = 1 \mid X_{i6} = (0,0,0), b_{i6} = 0\} = 0.282$. We use the notation $\Delta(X)$ to signify that the conditional log odds depends on the covariate value. By changing $X_{ij3}$ to 1, we can recompute the conditional log odds $\Delta(X)$ and evaluate $P\{Y_{i6} = 1 \mid X_{i6} = (0,0,1), b_{i6} = 0\} = 0.244$, thereby allowing us to estimate the individual-level effect of adopting drug therapy as logit(0.244) − logit(0.282) = −0.196. This shows that, although we adopt a marginal regression model, we are still able to obtain estimates of the effect of intervention at the individual level.

## 5. Discussion

In this manuscript a logistic-normal model is used, but a regression structure is adopted for the marginal mean as opposed to the conditional mean. Several authors have discussed the relative merits of marginal, or population-averaged, coefficients versus conditional, or subject-specific, coefficients (Zeger et al., 1988; Neuhaus et al., 1991; Graubard and Korn, 1994; Pendergast et al., 1996). A general conclusion from these papers is that conditional regression coefficients have limited utility. In particular, with time-independent covariates such as gender, the conditional regression coefficient "almost invites an unjustified causal statement about the change in odds" (Neuhaus et al., 1991, p. 20). In contrast, the conditional model can directly address hypothetical individual-level interventions. By adopting a marginal mean regression model with an underlying logistic-normal random effects structure, we are able to report group contrasts via $\beta$ and individual-level contrasts, matching on the random effects, by calculating the implied conditional linear predictor as a function of covariates $\Delta(X)$.

This manuscript has introduced both likelihood-based and estimating equation methods for parameter estimation. A primary advantage of using estimating equations is the resulting robustness of inference regarding $\beta$ when dependence model assumptions are violated. Future research that assesses the impact of random effects model violations on the MLE $\hat{\beta}$ is warranted. Limited simulation studies (not shown) have investigated the potential bias due to maximum likelihood estimation using marginally specified random intercepts models when either $\sigma$ is a function of subject-level covariates or random effects are autocorrelated. Bias may result, but it has been found to be slight (relative bias <15%) in the scenarios considered.

Although this manuscript has focused on the logistic regression model with normally distributed random effects, the ideas herein can be used for other generalized linear mixed models. With certain link functions and mixing distributions, the transformation given in equation (4) may be made analytically (i.e., probit-normal, log-normal); however, likelihood evaluation will often remain computationally demanding.

Finally, we must include the caveat that variance component estimation for categorical data is generally difficult and may require large sample sizes, both in terms of the number of observations per subject and the number of subjects. Also, the identifiability of the variance component parameters depends entirely on the covariance model, as the marginal variance contributes no information regarding the latent covariance structure. This limits the logistic-normal models that may be considered because the latent variance $V(b_{ij})$ must be a function of the covariance $\text{cov}(b_{ij}, b_{ik})$, as is the case for the random intercepts and latent autocorrelated Gaussian process models that we have considered. Therefore, extension of the models to nonlongitudinal multivariate binary data may be limited, although nested data could be handled quite naturally and will be a topic for future research.

## Acknowledgements

The data on schizophrenia originate from the Madras Longitudinal Study, conducted by R. Thara. The Madras Longitudinal Study was supported by the National Institute of Mental Health grant 44653 (W. Eaton, principal investigator), the India Council on Medical Research, and the Schizophrenia Research Foundation (India). Funding for this research was provided by NIH grant PO1 CA76466-01 and a grant from the Cystic Fibrosis Foundation, CF R565.

## Résumé

L'inférence basée sur la vraisemblance pour des données longitudinales binaires peut être obtenue en utilisant un modèle linéaire généralisé à effets mixtes (Breslow et Clayton, 1993; Woldfinger et O'Connell, 1993) étant donné les améliorations récentes dans les approches calculatoires. D'un autre coté, Fitzmaurice et Laird (1993), Molenberghs et Lesaffre (1994), et Heargerty et Zeger (1996) ont développé une inférence basée sur la vraisemblance en utilisant un paramètre de régression de la moyenne marginale et en spécifiant complètement la distribution multivariée conjointe grâce à des hypothèses sur les moments d'ordre supérieur canoniques et/ou marginaux. Toutes ces approches marginales sont intensives d'un point de vue calculatoire et actuellement limitées à des grappes de petite taille. Dans ce manuscrit, une paramétrisation alternative du modèle logistique à effets aléatoires gaussiens est adoptée, et les approches de vraisemblance et des équations d'estimation pour l'estimation des paramètres sont étudiées. Une propriété clé de l'approche proposée est que les paramètres de régression marginale sont obtenus tout en permettant encore des prédictions au niveau individuel ou des contrastes. Un exemple est présenté où l'intérêt scientifique est à la fois dans la réponse moyenne et dans la covariance entre les mesures répétées.

## References

Abramowitz, K. M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover Publications.

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

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

Carey, V. J., Zeger, S. L., and Diggle, P. J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80,** 517–526.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* **11,** 427–444.

Drum, M. L. and McCullagh, P. (1993). REML estimation with exact covariance in the logistic mixed model. *Biometrics* **49,** 677–689.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80,** 141–151.

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statistical Science* **8,** 284–309.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* **57,** 533–546.

Graubard, B. I. and Korn, E. L. (1994). Regression analysis with clustered data. *Statistics in Medicine* **13,** 509–522.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* **91,** 1024–1036.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7,** 305–315.

Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* **89,** 625–632.

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

Lipsitz, S., Laird, N., and Harrington, D. (1991). Generalized estimating equations for correlated binary data: Using odds ratios as a measure of association. *Biometrika* **78,** 153–160.

Louis, T. A. (1984). Estimating a population of parameter values using Bayesian and empirical Bayesian methods. *Journal of the American Statistical Association* **79,** 393–398.

Mancl, L. and Leroux, B. (1996). Efficiency of regression estimates for clustered data. *Biometrics* **52,** 500–511.

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

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89,** 633–644.

Monahan, J. F. and Stefanski, L. A. (1992). Normal scale mixture approximations to $F^*(x)$ and computation of the logistic-normal integral. In *Handbook of the Logistic Distribution*, N. Balakrishnan (ed), 529–540. New York: Marcel Dekker.

Neuhaus, J. M., Kalbfleish, 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.

Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M., and Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review* **64,** 89–118.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics* **23,** 939–951.

Pierce, D. A. and Sands, B. R. (1975). Extra-Bernoulli variation in binary data. Technical Report, Oregon State University.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.

Qu, Y., Williams, G. W., Beck, G. J., and Medendorp, S. V. (1992). Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics* **48,** 1095–1102.

Stiratelli, R., Laird, N., and Ware, J. (1984). Random effects models for serial observations with binary responses. *Biometrics* **40,** 961–970.

Thara, R., Henrietta, M., Joseph, A., Rajkumar, S., and Eaton, W. (1994). Ten year course of schizophrenia—The Madras longitudinal study. *Acta Psychiatrica Scandinavica* **90,** 329–336.

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

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

*Received January* 1998. *Revised August* 1998.
*Accepted September* 1998.

## Appendix

### *Calculation of $\Delta_{ij}$ and Derivatives*

Numerical evaluation of the integral in equation (4) can be accomplished with excellent accuracy for a wide range of parameter values using either a general numerical integration method, such as Gauss–Hermite quadrature, or a specialized method, such as LMA (Monahan and Stefanski, 1992). The present authors show that a 20-point Gauss–Hermite quadrature will have a maximal error $< 10^{-5}$ for $\sigma_{ij} < 2.0$, but only $10^{-3}$ for $\sigma_{ij} = 4.0$, whereas LMA using five mixture points yields a maximal error of $10^{-6}$ for $\sigma_{ij} \leq 5$.

Given $(\eta_{ij}, \sigma_{ij})$, we use Newton–Raphson to solve for the implied conditional parameter $\Delta_{ij}$. For this we require

$$A_{ij} = \frac{\partial}{\partial \Delta_{ij}} h(\eta_{ij}) = \int h(\Delta_{ij} + \sigma_{ij} z)\{1 - h(\Delta_{ij} + \sigma_{ij} z)\}\phi(z)\, dz,$$

which we also obtain numerically.

For quasi-likelihood and maximum likelihood estimation, we require derivatives of the deconvolution solution $\Delta_{ij}$ with respect to $\eta_{ij}$ and $\sigma_{ij}$. Use of the chain rule then yields derivatives with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, where $g_2(\sigma_{ij}) = \boldsymbol{Z}_i\boldsymbol{\alpha}$. Necessary derivatives can be obtained via implicit differentiation of the convolution equation. Define

$$B_{ij} = \int h_{ij}(1 - h_{ij}) z\phi(z)\, dz,$$

$$C_{ij} = \int (1 - 2h_{ij}) h_{ij}(1 - h_{ij})\phi(z)\, dz,$$

$$D_{ij} = \int (1 - 2h_{ij}) h_{ij}(1 - h_{ij}) z\phi(z)\, dz,$$

$$E_{ij} = \int (1 - 2h_{ij}) h_{ij}(1 - h_{ij}) z^2\phi(z)\, dz,$$

where $h_{ij} = h(\Delta_{ij} + \sigma_{ij} z)$. Using these expressions, we can write the required derivatives as

$$\frac{\partial \Delta_{ij}}{\partial \eta_{ij}} = \mu_{ij}(1 - \mu_{ij})/A_{ij},$$

$$\frac{\partial \Delta_{ij}}{\partial \sigma_{ij}} = -B_{ij}/A_{ij},$$

$$\frac{\partial^2 \Delta_{ij}}{\partial \eta_{ij}^2} = \left[\mu_{ij}(1 - \mu_{ij})(1 - 2\mu_{ij}) - \left(\frac{\partial \Delta_{ij}}{\partial \eta_{ij}}\right)^2 C_{ij}\right] \Big/ A_{ij},$$

$$\frac{\partial^2 \Delta_{ij}}{\partial \eta_{ij} \partial \sigma_{ij}} = -\left[\left(\frac{\partial \Delta_{ij}}{\partial \eta_{ij}}\right)\left(\frac{\partial \Delta_{ij}}{\partial \sigma_{ij}}\right) C_{ij} + \left(\frac{\partial \Delta_{ij}}{\partial \eta_{ij}}\right) D_{ij}\right] \Big/ A_{ij},$$

$$\frac{\partial^2 \Delta_{ij}}{\partial \sigma_{ij}^2} = -\left[\left(\frac{\partial \Delta_{ij}}{\partial \sigma_{ij}}\right)^2 + 2\left(\frac{\partial \Delta_{ij}}{\partial \sigma_{ij}}\right) D_{ij} + E_{ij}\right] \Big/ A_{ij}.$$

### *Maximum Likelihood Estimation*

For the marginally specified logistic-normal model with a scalar random effect (random intercept), the likelihood contribution from subject $i$ is given in Section 3.1 and can be approximated numerically using 20-point Gauss–Hermite quadrature as

$$L_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_k w_k \prod_{j=1}^{n_i} \exp[\{\Delta_{ij} + \sigma_{ij} z_k\} Y_{ij} + \log\{1 - h(\Delta_{ij} + \sigma_{ij} z_k)\}]$$

for $(w_k, z_k)_{k=1}^{20}$, given in Abramowitz and Stegun (1972). We also use Gauss–Hermite quadrature to obtain the score equations and the Fisher information matrix.

### *Estimating Equations*

Estimates of the marginally specified logistic-normal model parameters can be obtained as roots of the estimating functions, given as equations (5) and (6) in Section 3.2. To solve the equations and compute asymptotic standard errors, we need to calculate the pairwise covariance $\nu_{i(j,k)} = \pi_{i(j,k)} - \mu_{ij}\mu_{ik}$, where

$$\pi_{i(j,k)} = \int h(\Delta_{ij} + b_{ij}) h(\Delta_{ik} + b_{ik})\, dF(b_{ij}, b_{ik})$$

and $\partial \boldsymbol{\nu}_i / \partial \boldsymbol{\beta}$, $\partial \boldsymbol{\nu}_i / \partial \boldsymbol{\alpha}$.

For any general bivariate Gaussian covariance, $\text{cov}(b_{ij}, b_{ik})$, we can obtain the Cholesky decomposition:

$$C_{i(j,k)} = \begin{bmatrix} \sigma_{ij} & 0 \\ \sigma_{ik}\rho_{i(j,k)} & \sigma_{ik}\sqrt{1-\rho^2_{i(j,k)}} \end{bmatrix},$$

where $\sigma^2_{ij} = V(b_{ij})$, $\sigma^2_{ik} = V(b_{ik})$, and $\rho_{i(j,k)} = \text{corr}(b_{ij}, b_{ik})$. Using this decomposition allows direct use of bivariate quadrature methods for the numerical evaluation of $\pi_{i(j,k)}$ and derivatives with respect to the following model parameters:

$$\begin{aligned}\pi_{i(j,k)} &= \int\int h(\Delta_{ij} + \sigma_{ij} z_1) \\ &\quad \times h\Big(\Delta_{ik} + \sigma_{ik}\rho_{i(j,k)} z_1 + \sigma_{ik}\sqrt{1-\rho^2_{i(j,k)}}\, z_2\Big) \\ &\quad \times \phi(z_1)\phi(z_2)\, dz_1\, dz_2,\end{aligned}$$

$$\begin{aligned}\frac{\partial \pi_{i(j,k)}}{\partial \boldsymbol{\beta}} &= \int\int \left[ h_{ij}(1-h_{ij})h_{ik}\frac{\partial \Delta_{ij}}{\partial \boldsymbol{\beta}} + h_{ij}h_{ik}(1-h_{ik})\frac{\partial \Delta_{ik}}{\partial \boldsymbol{\beta}} \right] \\ &\quad \times \phi(z_1)\phi(z_2)\, dz_1\, dz_2,\end{aligned}$$

$$\begin{aligned}\frac{\partial \pi_{i(j,k)}}{\partial \boldsymbol{\alpha}} &= \int\int \left\{ h_{ij}(1-h_{ij})h_{ik}\left[ \frac{\partial \Delta_{ij}}{\partial \boldsymbol{\alpha}} + \frac{\partial \sigma_{ij}}{\partial \boldsymbol{\alpha}} z_1 \right] \right\} \\ &\quad \times \phi(z_1)\phi(z_2)\, dz_1\, dz_2 \\ &\quad + \int\int \left\{ h_{ij}h_{ik}(1-h_{ik})\left[ \frac{\partial \Delta_{ik}}{\partial \boldsymbol{\alpha}} + \frac{\partial(\sigma_{ik}\rho_{i(j,k)})}{\partial \boldsymbol{\alpha}} z_1 \right.\right. \\ &\qquad \left.\left. + \frac{\partial\Big(\sigma_{ik}\sqrt{1-\rho^2_{i(j,k)}}\Big)}{\partial \boldsymbol{\alpha}} z_2 \right] \right\} \\ &\quad \times \phi(z_1)\phi(z_2)\, dz_1\, dz_2,\end{aligned}$$

where $h_{ij} = h(\Delta_{ij} + \sigma_{ij} z_1)$ and $h_{ik} = h(\Delta_{ik} + \sigma_{ik}\rho_{i(j,k)} z_1 + \sigma_{ik}(1-\rho^2_{i(j,k)} z_2))^{1/2}$.

An alternative numerical method tailored to the evaluation of the logistic-normal integral is given in Monahan and Stefanski (1992) and is used in a related context by Drum and McCullagh (1993), who provide details regarding the required derivatives.