Biometrics 63, 322–331
June 2007
DOI: 10.1111/j.1541-0420.2006.00680.x

Marginalized Models for Moderate to Long Series of Longitudinal Binary Response Data

Jonathan S. Schildcrout

Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, U.S.A.
email: jonathan.schildcrout@vanderbilt.edu

and

Patrick J. Heagerty

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

Summary. Marginalized models (Heagerty, 1999, *Biometrics* **55,** 688–698) permit likelihood-based inference when interest lies in marginal regression models for longitudinal binary response data. Two such models are the marginalized transition and marginalized latent variable models. The former captures within-subject serial dependence among repeated measurements with transition model terms while the latter assumes exchangeable or nondiminishing response dependence using random intercepts. In this article, we extend the class of marginalized models by proposing a single unifying model that describes both serial and long-range dependence. This model will be particularly useful in longitudinal analyses with a moderate to large number of repeated measurements per subject, where both serial and exchangeable forms of response correlation can be identified. We describe maximum likelihood and Bayesian approaches toward parameter estimation and inference, and we study the large sample operating characteristics under two types of dependence model misspecification. Data from the Madras Longitudinal Schizophrenia Study (Thara et al., 1994, *Acta Psychiatrica Scandinavica* **90,** 329–336) are analyzed.

Key words: Binary data; Longitudinal data analysis; Marginal models; Marginalized models; Time-dependent covariates.

1. Introduction

The Madras Longitudinal Schizophrenia Study examined the 10-year course of schizophrenia after initial hospitalization for disease (Thara et al., 1994). Since the original publication, several authors (e.g., Heagerty and Zeger, 1998; Diggle et al., 2002) have used a subset of the data to study the 1-year course of recovery from two classes of symptoms with specific interest in the differences in recovery rates for males and females and for subgroups formed on the basis of age at onset. The six symptoms examined in these analyses were hallucinations, delusions, thought disorders, flat affect, apathy, and withdrawal, and their presence or absence was assessed on a monthly basis. Longitudinal logistic regression models have been used to analyze these data, where statistical inference has targeted the coefficients corresponding to age at onset by time and gender by time interactions.

One approach to estimating the coefficients in a longitudinal regression model is with the semiparametric, generalized estimating equations procedure (GEE; Liang and Zeger, 1986). While GEE estimates are robust to dependence model misspecification (with potential losses in efficiency), they can be sensitive to common types of missing data. In addition, GEE does not exploit an objective likelihood function. Likelihoods are critical for model selection among nonnested models, and with nested models, Wald tests (used for model selection with GEE) have been shown to perform poorly with small datasets (e.g., Feng, McLerran, and Grizzle, 1996) and with large effect sizes in logistic regression (Menard, 1995). Likelihoods are also very useful if we want to display uncertainty associated with parameter estimates using, say, profile likelihoods.

Several authors have proposed maximum likelihood procedures for marginal regression models with categorical response data (e.g., Fitzmaurice and Laird, 1993; Lang and Agresti, 1994; Molenberghs and Lesaffre, 1994). However, these methods were generally limited either to equal-sized clusters needed for dependence parameter specification and/or limited to small clusters for computational feasibility. Azzalini (1994) proposed a marginally specified, first-order transition model for binary response data, which was extended to pth-order transition models and to latent variable models in a series of manuscripts (Heagerty, 1999, 2002; Heagerty and Zeger, 2000). These "marginalized" models constitute a highly flexible class of marginal regression models that permit likelihood-based estimation while adopting dependence structures natural for serial measurements with large and possibly unequal numbers of within-subject measurements.

Several authors have contrasted marginal and conditional models for binary response data (e.g., Zeger, Liang, and Albert, 1988; Neuhaus, Kalbfleisch, and Hauck, 1991; Pendergast et al., 1996; Lindsey and Lambert, 1998), as well as the operating characteristics of the associated estimates (e.g., Ten Have, Kunselman, and Tran, 1999; Heagerty and Kurland, 2001; Kurland and Heagerty, 2004). A feature that distinguishes the marginal from the conditional models is the separation of the mean model from the response dependence model. If the goal of an analysis is to examine the relationship between the response and a covariate, the functional form of the response dependence model does not impact marginal mean model parameter interpretation. For example, if the mean model is properly specified, parameter estimates using GEE with various working covariance matrices are consistent for the same quantity. This is not the case for conditional models where the mean and dependence are captured in the same regression.

The primary purpose of this article is to introduce an extension to the marginalized regression model class that can reflect dependence structures typical of longer response series such as those observed in the Madras study, where subjects were observed monthly for a year. In Section 2, we describe the proposed model that we call the marginalized transition and latent variable model (*mTLV*), and we discuss inference based on maximum likelihood and Bayesian procedures. We examine the impact response dependence misspecification can have on marginalized model inference in large samples in Section 3. Because several authors have shown the importance of the covariate distribution on operating characteristics of various estimators (e.g., Fitzmaurice, 1995; Mancl and Leroux, 1996; Schildcrout and Heagerty, 2005), we study covariates with varying degrees of between- versus within-subject variation. Finally, we analyze the Madras study data in Section 4 and provide concluding remarks and a discussion in Section 5.

2. Model and Estimation

In this section, we describe the marginalized model formulation, and introduce an extension for longer series of response data that combines the response dependence models from the first-order marginalized transition model (*mT*) and the marginalized random intercept (*mLV*). Hence, we refer to the extended model as the marginalized transition and latent variable model (*mTLV*). We describe marginalized models in general and specifically the *mTLV* in Section 2.1. In Section 2.2, we discuss estimation with maximum likelihood, and in Section 2.3 we outline a Bayesian approach to inference with the *mTLV*.

2.1 *The Marginalized Transition and Latent Variable Model*

The class of marginalized regression models is defined by at least two regression components. If we let $\mathbf{Y}_i$ denote the n_i-dimensional response vector for the ith subject, $i \in \{1, 2, \ldots, N\}$, $\mathbf{X}_i$ the $n_i \times p$ design matrix, and $g(\cdot)$ the link function, then the marginal mean regression model is given by

$$g\left(\mu_{ij}^{m}\right) \equiv g\{E(Y_{ij} \mid \mathbf{X}_{ij})\} \equiv \eta_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}^{m}, \tag{1}$$

where $\boldsymbol{\beta}^m$ is a p-dimensional parameter vector, $j \in \{1, 2, \ldots, n_i\}$ is the observation number for the ith subject, and η_{ij} is the linear predictor. While this model identifies the first moment of the multivariate distribution of $\mathbf{Y}_i$ given $\mathbf{X}_i$, we construct a second regression model to describe the within-subject dependence among responses as well as higher-order moments. Throughout this article, we refer to the second regression component as either the conditional mean model or the response dependence model. Let $\mathbf{A}_i$ be an $n_i \times q$ response dependence design matrix which could include previous and/or future responses, latent variables, and other covariates, and let $\boldsymbol{\alpha}$ be the associated q-dimensional parameter vector. The conditional mean model is given by

$$g\left(\mu_{ij}^{c}\right) \equiv g\{E(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{A}_{ij})\} = \Delta(\mathbf{X}_{ij}) + \mathbf{A}_{ij}\boldsymbol{\alpha}, \tag{2}$$

where variables in the $1 \times q$ vector $\mathbf{A}_{ij}$ impose structure on the conditional mean. For latent variable models, additional assumptions regarding the distribution of the random components of $\mathbf{A}_{ij}$ are required to establish the multivariate distribution, $\mathbf{Y}_i \mid \mathbf{X}_i$.

In the class of marginalized models, $\Delta(\mathbf{X}_{ij})$, or simply Δ_{ij}, is an implicitly defined value that links the conditional and marginal mean regression models. Given η_{ij}, g, $\boldsymbol{\alpha}$, and the mixing distribution $F_{\mathbf{A}_{ij}}$, Δ_{ij} is the value that satisfies the convolution equation

$$\mu_{ij}^{m} \equiv g^{-1}(\eta_{ij}) = \int_{\mathbf{A}_{ij}} g^{-1}(\Delta_{ij} + \mathbf{A}_{ij}\boldsymbol{\alpha})\, dF_{\mathbf{A}_{ij}}. \tag{3}$$

Because our focus is on binary response data, we let $g = $ *logit*, though other link function choices (e.g., probit) could be adopted. For most link function and mixing distribution combinations, Δ_{ij} is analytically intractable, and we solve for it numerically to evaluate the likelihood. However, for specific combinations, a simple analytical relationship between Δ_{ij} and η_{ij} exists (see Wang and Louis, 2003). For example, with the probit link and $N(0, \sigma^2)$ random intercepts, $\eta_{ij} = \Delta_{ij}/(1 + \sigma^2)$ (Heagerty, 1999).

For long series of longitudinal response data, the *mLV* and *mT* models may not adequately characterize the dependence among outcomes. We now introduce a model that is designed to capture both major components of dependence. The *mTLV* describes serial dependence with a Markov or transition component and long-range or nondiminishing dependence with random intercepts. Thus, using the logit link function, the complete specification of the *mTLV* is

$$\begin{aligned} \text{logit}\left(\mu_{ij}^{m}\right) &= \mathbf{X}_{ij}\boldsymbol{\beta}^{m} \\ \text{logit}\left(\mu_{ij}^{c}\right) &= \Delta_{ij} + \gamma(\mathbf{X}_i)Y_{ij-1} + b_i,\ b_i \sim N\left\{0, \sigma^2(\mathbf{X}_i)\right\}. \end{aligned} \tag{4}$$

Both components of the dependence model, $\gamma \equiv \gamma(\mathbf{X}_i)$ and $\sigma \equiv \sigma(\mathbf{X}_i)$, could depend on covariates in $\mathbf{X}_i$. The value, Δ_{ij}, which will be discussed further in the next section, can be thought of as an offset in the conditional mean logistic regression model that contains a transition plus a random intercept.

2.2 *Maximum Likelihood Estimation*

Maximum likelihood parameter estimation for the *mTLV* can be accomplished with a Newton–Raphson algorithm. At each iteration, the critical step toward evaluation of the likelihood is the calculation of Δ_{ij} in (4). In general, this is done by solving a convolution equation that relates the marginal and conditional means in equation (3). We now describe the calculation of Δ_{ij} when our goal is to base inference on the *mTLV*. The marginal mean is equal to the "marginalized" conditional mean, where marginalization is taken over the joint distribution of the lagged response and the random intercept distribution, $(Y_{ij-1}, b_i \mid \mathbf{X}_i)$. For ease of exposition, we will assume the dependence model parameters are not modified by covariates, and $b_i \sim N(0, \sigma^2)$ which we rewrite, $b_i = \sigma Z_i, Z_i \sim N(0, 1)$. The marginal mean is therefore related to the conditional mean by

$$\mu_{ij}^m = E_{Z_i, Y_{ij-1}}\left(\mu_{ij}^c\right) = E_{Z_i}\left[E_{Y_{ij-1}}\left\{\text{logit}^{-1}(\Delta_{ij} + \gamma Y_{ij-1} + \sigma Z_i)\right\}\right]$$
$$= \int \underbrace{\left\{\text{logit}^{-1}(\Delta_{ij} + \sigma z_i)\left(1 - \mu_{ij-1}^{pc,z}\right) + \text{logit}^{-1}(\Delta_{ij} + \gamma + \sigma z_i)\mu_{ij-1}^{pc,z}\right\}}_{\mu_{ij}^{pc,z}} \phi(z_i)\, dz_i, \tag{5}$$

where $\phi(\cdot)$ is the standard normal probability density function and $\mu_{ij}^{pc,z}$ is a "partly conditional" mean obtained by taking the expectation of the conditional mean μ_{ij}^c over the distribution of the lag-1 response.

To evaluate the distribution $\mathbf{Y}_i \mid \mathbf{X}_i$ and therefore the likelihood contribution by the ith subject, we solve equation (5) for Δ_{ij} for all $j \in (1, 2, \ldots, n_i)$. This convolution equation is a simple extension of those that identify $\boldsymbol{\Delta}_i$ for the *mT* and the *mLV*; however, because marginalization is taken over the joint distribution (Y_{ij-1}, b_i), $\mu_{ij}^{pc,z}$ depends upon $\mu_{ij-1}^{pc,z}$ and sequential updates over repeated within-individual measurements are required (see Web Appendix for details).

Once the complete $\boldsymbol{\Delta}_i$ vector has been evaluated, the calculation of $\boldsymbol{\mu}_i^c$ and the contribution to the likelihood by subject i is straightforward. Because subjects are assumed to be independent of one another, the likelihood for the parameters given in the data is the product of individual contributions,

$$L(\boldsymbol{\beta}^m, \gamma, \sigma \mid \mathbf{Y}, \mathbf{X}) = \prod_{i=1}^N L_i = \prod_{i=1}^N \int \left\{\prod_{j=1}^{n_i} \mu_{ij}^{c\ y_{ij}} \left(1 - \mu_{ij}^c\right)^{(1-y_{ij})}\right\} \phi(z_i)\, dz_i. \tag{6}$$

To obtain maximum likelihood estimates of the parameter vector $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}^m, \gamma, \sigma)$, analytic expressions for the score and Hessian are useful, although not entirely necessary. Let $L_{i,z}$ denote the bracketed portion of the integrand in equation (6), and let θ_1 be any parameter in the *mTLV* regression model, then the first partial derivative of the log likelihood, l, is given by

$$\frac{\partial l}{\partial \theta_1} = \sum_{i=1}^N \frac{\int L_{i,z}\left[\sum_j \left\{\left(y_{ij} - \mu_{ij}^c\right) \cdot \left(\frac{\partial \Delta_{ij}}{\partial \theta_1} + \frac{\partial \gamma}{\partial \theta_1} y_{ij-1} + \frac{\partial \sigma}{\partial \theta_1} z_i\right)\right\}\right] \phi(z_i)\, dz_i}{\int L_{i,z}\phi(z_i)\, dz_i},$$

where $\partial\gamma/\partial\theta_1 = 0$ if $\theta_1 \neq \gamma$, and $\partial\sigma/\partial\theta_1 = 0$ if $\theta_1 \neq \sigma$. The second partial derivative follows directly.

2.3 *mTLV with Bayesian Inference*

Another likelihood-based approach to inference with the *mTLV* is through Bayesian methods. Miglioretti and Heagerty (2004) used Bayesian methods to estimate marginalized model parameters in a setting with crossed random effects. For the *mTLV*, Bayesian methods may be relatively less cumbersome to implement because neither likelihood maximization nor integration of the likelihood function over the random effects distribution is directly required. Note that integration is still involved in Δ_{ij} calculations.

For the *mTLV*, the posterior distribution is proportional to $P(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}^m, \gamma, \sigma, \mathbf{b}) \cdot \pi(\boldsymbol{\beta}^m) \cdot \pi(\gamma) \cdot \pi(\sigma) \cdot \prod_{i=1}^N \times \pi(b_i \mid \sigma)$, where $\mathbf{b}$ represents the vector of random intercepts, $\pi(\cdot)$ is a prior probability distribution function, and σ is a hyperparameter for the distribution of the random intercepts. The contribution to the posterior distribution by the data in the Bayesian analysis is

$$P(\mathbf{Y} \mid \mathbf{X}, \beta^m, \gamma, \sigma, \mathbf{b}) = \prod_{i=1}^N L_i = \prod_{i=1}^N \prod_{j=1}^{n_i} \mu_{ij}^{c\ Y_{ij}} \left(1 - \mu_{ij}^c\right)^{(1-Y_{ij})}.$$

In Section 4, we analyze the Madras schizophrenia data using maximum likelihood and Bayesian procedures.

3. Large Sample Properties of the Maximum Likelihood Estimator under Misspecification

The *mTLV* extends the class of marginalized regression models to accommodate dependence structures with serial and long-range components. We now explore the implications of dependence model misspecification with marginalized model inference. Heagerty and Kurland (2001) showed that, while not directly comparable, mean model parameter estimates for the *mLV* tend to be less sensitive to dependence model misspecification than the analogous estimates from a generalized linear mixed model with random intercepts. Nevertheless, when the modifying effect of covariates on parameters in the dependence model are ignored, marginalized model parameter estimates may be inconsistent. In this section, we continue the examination of dependence model misspecification for inference on mean model parameters, and we focus on the role between- and within-subject covariate variation can play in determining operating characteristics of the estimators. We study large sample bias in parameter and variance estimates as well as the inefficiency associated with dependence model misspecification.

Sources of variation in the target predictor have been shown to play a large role in determining the operating characteristics of regression estimators for longitudinal data (e.g., Fitzmaurice, 1995; Mancl and Leroux, 1996; Schildcrout and Heagerty, 2005). The intracluster correlation in X is the proportion of total variance that occurs among individuals. In between-subject covariates (e.g., demographics), $ICC_x = 1$, and in balanced and mean balanced covariates, $ICC_x = 0$; however, for time-varying covariates, it is often the case that $0 < ICC_x < 1$. Neuhaus and Kalbfleisch (1998) discussed the decomposition of time-varying covariates into between- and purely within-subject components. They showed that when $0 < ICC_x < 1$, an implicit assumption is being made (i.e., the between-subject and purely within-subject covariate effects are equal). We examine the impact of between- and within-individual variation in the target predictor by considering estimation on covariates with $ICC_x \in \{0, 0.5, 1\}$.

Because closed form solutions for bias in parameter and variance estimates due to dependence model misspecification are not available, we rely on Monte Carlo methods for our calculations. We generated a large number of clusters, $N \in \{5{,}000, 10{,}000\}$, with 14 repeated measurements per cluster, $n_i \equiv n = 14$, and fit the data with properly specified and misspecified models to examine large sample operating characteristics of the estimators. Because error is associated with Monte Carlo methods we replicated this process 100 times. Two types of misspecifications were examined: (1) dependence model functional form (FF) misspecification and (2) dependence component modification (DCM) misspecification. With FF misspecification, either the serial or the long-range component of the dependence model is ignored during fitting. For example, the true dependence model may be that of the *mTLV*, and the data are fit assuming the *mT* or *mLV*. With DCM misspecification, the modifying effect of a covariate on a dependence model parameter is ignored. For example, one may assume that a single random intercept distribution is sufficient to account for the heterogeneity in the study sample, when in fact, there may be substantial differences in heterogeneity among subgroups.

3.1 *Inference with FF Misspecification*

To study the impact of FF misspecification, we generated 5000 clusters with 14 repeated measurements per cluster under the marginalized logistic regression model,

$$\text{logit}\left(\mu_{ij}^m\right) = \beta_0^m + \beta_1^m X_{ij},$$

$$\text{logit}\left(\mu_{ij}^c\right) = \Delta_{ij} + \gamma Y_{ij-1} + \sigma Z_i, \ Z_i \sim N(0,1),$$

where $\beta_0^m = -0.25$, $\beta_1^m = 0.25$, $(\gamma, \sigma) \in \{(4, 0), (2, 0), (0, 2), (0, 0.7), (2.75, 1), (1.45, 0.45)\}$. X_{ij} was normally distributed and $ICC_x \in \{0, 0.5, 1\}$.

Table 1 displays percent bias $\{100(\bar{\hat{\beta}} - \beta)/\beta\}$ in parameter and variance estimates under true and misspecified models. Note that negative percent bias implies attenuation (toward 0) on the raw scale. Parameter estimates were robust to dependence model misspecification irrespective of the data generating model and the covariate distribution of the target predictor. However, with *mTLV* and *mLV* generated data, bias in variance estimates resulted from assuming an *mT* dependence structure. The bias was relatively small when ICC_x was 0 (e.g., −7% to 5%) but it grew substantially with proportionately greater between-subject covariate variation. When $ICC_x = 1$ variances were underestimated by 33% and 66%, when the true (and ignored) variance component was equal to 1.0 and 2.0, respectively. There was an indication of biased variance estimates with *mLV* and time-varying covariates, but considering the degree of misspecification, we believe that these biases are small.

Because FF misspecification did not appear to bias parameter estimates, a valid approach on which to base inference could be the use of maximum likelihood parameter estimates

Table 1

Large sample operating characteristics in the presence of FF misspecification: percent bias in parameter (variance) estimates. For variance calculations, the true value was assumed to be the average robust variance estimate over 100 replicates.

True model values			β_0^m			β_1^m		
γ	σ	ICC_x	*mLV*	*mT*	*mTLV*	*mLV*	*mT*	*mTLV*
4	0	0	1 (0)	0 (0)	0 (0)	−2 (−4)	0 (0)	0 (0)
4	0	0.5	2 (0)	0 (0)	0 (0)	−1 (−4)	0 (0)	0 (0)
4	0	1	1 (0)	−1 (0)	−1 (0)	1 (0)	0 (0)	0 (0)
2	0	0	0 (0)	0 (0)	0 (0)	0 (−1)	0 (0)	0 (0)
2	0	0.5	0 (0)	0 (0)	0 (0)	0 (−1)	0 (0)	0 (0)
2	0	1	−1 (0)	−1 (0)	−1 (0)	1 (0)	1 (0)	1 (0)
0	2	0	0 (0)	0 (−66)	0 (0)	0 (0)	0 (−6)	0 (0)
0	2	0.5	1 (0)	1 (−66)	1 (0)	0 (0)	0 (−35)	0 (0)
0	2	1	−1 (0)	−1 (−66)	−1 (0)	0 (0)	0 (−66)	0 (0)
0	0.7	0	0 (0)	0 (−47)	0 (0)	0 (0)	0 (5)	0 (0)
0	0.7	0.5	0 (0)	0 (−47)	0 (0)	0 (0)	0 (−26)	0 (0)
0	0.7	1	−1 (0)	−1 (−47)	−1 (0)	0 (0)	0 (−47)	0 (0)
2.75	1	0	0 (0)	0 (−34)	0 (0)	−1 (−2)	0 (−7)	0 (0)
2.75	1	0.5	1 (0)	1 (−34)	1 (0)	−1 (−3)	0 (−12)	0 (0)
2.75	1	1	1 (0)	1 (−33)	1 (0)	0 (0)	0 (−33)	0 (0)
1.45	0.45	0	−1 (0)	−1 (−25)	−1 (0)	−1 (−1)	0 (0)	0 (0)
1.45	0.45	0.5	0 (0)	0 (−25)	0 (0)	0 (−1)	0 (−8)	0 (0)
1.45	0.45	1	1 (0)	1 (−25)	1 (0)	0 (0)	0 (−24)	0 (0)

Table 2
Relative variance in the presence of FF misspecification: ratio of the average robust variance for each assumed model to that of the (true) mTLV over 100 *replicates*

True model values			β_0^m		β_1^m	
γ	σ	ICC_x	*mLV*	*mT*	*mLV*	*mT*
4	0	0	1.11	1.00	2.02	1.00
4	0	0.5	1.11	1.00	1.90	1.00
4	0	1	1.12	1.00	1.12	1.00
2	0	0	1.03	1.00	1.36	1.00
2	0	0.5	1.03	1.00	1.26	1.00
2	0	1	1.03	1.00	1.03	1.00
0	2	0	1.00	1.03	1.00	1.18
0	2	0.5	1.00	1.03	1.00	1.41
0	2	1	1.00	1.03	1.00	1.03
0	0.7	0	1.00	1.00	1.00	1.02
0	0.7	0.5	1.00	1.00	1.00	1.13
0	0.7	1	1.00	1.00	1.00	1.00
2.75	1	0	1.03	1.01	1.45	1.02
2.75	1	0.5	1.03	1.01	1.40	1.02
2.75	1	1	1.03	1.01	1.03	1.01
1.45	0.45	0	1.01	1.00	1.18	1.00
1.45	0.45	0.5	1.01	1.00	1.14	1.02
1.45	0.45	1	1.01	1.00	1.01	1.00

coupled with the robust or empirical standard errors of White (1982). Table 2 describes how FF misspecification can affect the efficiency of estimates with this approach. In this table we divide the average variance estimate across replications for the assumed model by the average variance estimate for *mTLV*, which is properly specified in all scenarios. For time-invariant covariates and intercepts, misspecification had little impact on efficiency of the estimates relative to the true model. The largest efficiency losses for time-invariant covariate coefficients (12%) were observed when the data were generated with an mT but were fit assuming an *mLV*. Substantial losses incur with dependence model misspecification when $ICC_x < 1$. When mT data were fit assuming an *mLV*, the large sample relative variance for the slope estimator was 2.02, 1.90, and 1.12, when ICC_x equaled 0, 0.5, and 1.0, respectively. Interestingly, for an assumed mT, the greatest efficiency losses occurred when covariate variation emanated from within and between individuals ($ICC_x = 0.5$). The relative variance of the mT was 1.18, 1.41, and 1.03 when the ICC_x equaled 0, 0.5, and 1, respectively, for *mLV* ($\sigma = 2$).

3.2 *Inference with DCM Misspecification*

DCM misspecification occurs when parameters in the *mTLV* dependence model differ according to the value of a cluster level covariate but the modifying impact is ignored. Heagerty and Kurland (2001) showed that in such a circumstance biased parameter estimates may result. To study the impact of DCM misspecification on *mTLV* inference, we generated 10,000 clusters with 14 repeated measurements per cluster under the model,

$$\begin{aligned}\operatorname{logit}\left(\mu_{ij}^m\right) &= \beta_0^m + \beta_1^m X_{ij} + \beta_2^m G_i + \beta_3^m X_{ij} \cdot G_i,\\ \operatorname{logit}\left(\mu_{ij}^c\right) &= \Delta_{ij} + (\gamma_1 + \gamma_2 G_i) Y_{ij-1}\\ &\quad + (\sigma_1 + \sigma_2 G_i) Z_i,\ Z_i \sim N(0,1),\end{aligned}$$

where $(\beta_0^m,\ \beta_1^m,\ \beta_2^m,\ \beta_3^m) = (-0.25,\ 0.25,\ 0.50,\ -0.25)$, $(\gamma_1,\ \gamma_2,\ \sigma_1,\ \sigma_2) \in \{(4.5,\ -4,\ 0.5,\ 0),\ (2.5,\ -2,\ 0.5,\ 0)\}$, X_{ij} was normally distributed with $ICC_x \in \{0,\ 0.5,\ 1\}$, and G_i was a time-invariant dichotomous indicator of group membership, 0 or 1, with equal probabilities of being in each. With an $X_{ij} \cdot G_i$ interaction we effectively have two distinct linear predictors: $\beta_0^m + \beta_1^m X_{ij} = -0.25 + 0.25 X_{ij}$ for group 0 and $(\beta_0^m + \beta_2^m) + (\beta_1^m + \beta_3^m) X_{ij} = 0.25 + 0.0 X_{ij}$ for group 1.

With the above dependence models, misspecification occurs when the γ_1 is estimated and $\gamma_2 = 0$ is forced. For ease of exposition, we only consider misspecification in transition components and not variance components. Note that because $\gamma_2 < 0$, we underestimate the true within-individual dependence among responses in group 0 and we overestimate it in group 1. Heuristically, when $(\gamma_1, \gamma_2, \sigma_1, \sigma_2) = (4.5, -4, 0.5, 0)$, the true transition component parameter for group 0 (1) is 4.5 (0.5), but with misspecification, we assume $\gamma_1 \equiv \gamma_1 + \gamma_2 \approx 2.5$ for all subjects.

Table 3 displays the impact of DCM misspecification on parameter and variance estimates. Small to negligible biases in parameter estimates of up to 2% were observed when $\gamma_2 = -2$ and slightly larger biases of up to 10% were observed when $\gamma_2 = -4$. With $\gamma_2 = -4$, β_0^m estimates were biased by 6–10%, and β_2^m estimates were unbiased. Thus, in both groups, intercepts were overestimated. Biased estimates incurred when slopes were non-zero as estimates of the slope coefficient for group 0, β_1^m, were biased; however, estimates were unbiased for zero slope coefficients as evidenced by noting that percent biases in β_1^m and β_3^m were approximately equal in the scenarios studied (e.g., 8–9% when $\gamma_2 = -4$), and the true values of these parameters were 0.25 and -0.25, respectively. The distribution of the time-varying covariate did not appear to impact percent bias calculations. Overall, we studied what we consider to be substantial misspecifications, and because

Table 3

Large sample operating characteristics in the presence of DCM misspecification: percent bias in parameter (variance) estimates from the assumed mTLV. For variance calculations, the true value was assumed to be the average robust variance estimate over 100 *replicates.*

True model values					Estimates from the $mTLV$				Estimates from the (true) $mTLV_\gamma$			
γ_1	γ_2	σ_1	σ_2	ICC_x	β_0^m	β_1^m	β_2^m	β_3^m	β_0^m	β_1^m	β_2^m	β_3^m
4.5	−4	0.5	0	0	6 (−37)	8 (69)	0 (−4)	9 (4)	−1 (0)	0 (0)	0 (0)	0 (0)
4.5	−4	0.5	0	0.5	8 (−37)	8 (35)	0 (−4)	8 (1)	1 (0)	0 (0)	0 (0)	0 (0)
4.5	−4	0.5	0	1	10 (−40)	9 (−40)	0 (−6)	9 (−7)	0 (0)	−1 (0)	0 (0)	0 (0)
2.5	−2	0.5	0	0	2 (−25)	2 (29)	0 (−1)	2 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2.5	−2	0.5	0	0.5	1 (−25)	2 (13)	0 (−1)	2 (0)	−1 (0)	0 (1)	0 (0)	0 (1)
2.5	−2	0.5	0	1	1 (−26)	2 (−26)	0 (−1)	2 (−1)	−1 (0)	−1 (0)	0 (0)	−1 (0)

percent biases were at most 10%, we believe these estimators are reasonably robust to misspecification.

Though parameter estimates were insensitive to misspecification, inference will likely be invalid if estimates of uncertainty are based on observed or model-based information. Table 3 shows that variance estimates for β_0^m and β_1^m were most susceptible to misspecification. The magnitude and direction of bias in $\text{Var}(\hat{\beta}_1^m)$ depended largely on the distribution of X_{ij}. With $\gamma_2 = -4$ and -2, β_1^m variances were underestimated by approximately 40% and 26%, respectively, when $ICC_x = 1$; however, variances were overestimated when $ICC_x \in \{0, 0.5\}$.

4. Example: The Madras Longitudinal Schizophrenia Study

In this section, we use the proposed model, the $mTLV$, to examine the 1-year course of positive and negative symptoms in subjects participating in the Madras Longitudinal Schizophrenia Study. We examine the differences in recovery rates between males and females as well as between subgroups based on younger (<20 years) and older (≥20 years) age at onset. Thus, statistical inference focuses on the $\text{I}(\text{male})_i \cdot t$ and $\text{I}(\text{age} < 20)_i \cdot t$ interactions in the marginal mean model,

$$\text{logit}\left(\mu_{it}^m\right) = \beta_0 + \beta_1 t + \beta_2 \text{I}(\text{age} < 20)_i + \beta_3 \text{I}(\text{male})_i + \beta_4 \text{I}(\text{age} < 20)_i \cdot t + \beta_5 \text{I}(\text{male})_i \cdot t,$$

where subjects were observed on a monthly basis, $t \in \{0, 1, \ldots, 11\}$, and $\text{I}(\cdot)$ is 1 if $\cdot$ is true and 0 if false. To be included in the analyses presented here, symptom presence or absence at months 0 and 1 must have been recorded. Depending upon the symptom (hallucinations, delusions, thought disorders, flat affect, apathy, and withdrawal), 85 to 89 subjects were considered for analysis. Approximately half of the subjects were female and 35% were less than 20 years old. For the purpose of the current research, if a symptom response was missing during followup, all subsequent responses for that symptom were set to missing as updating the Δ_{it} values requires response data from time $t - 1$. For positive symptoms (hallucinations, delusions, and thought disorders) and negative symptoms (flat affect, apathy, and withdrawal), approximately 90% and 50% of subjects, respectively, had complete followup information, and 97% and 80% of all response values were available. We acknowledge that due to missing response values occurring in these data, results should be interpreted with caution. Although not considered here, the dependence model could be adapted to missing at random data or to unequally spaced followup by incorporating a lagged response by time since lagged response interaction.

We have shown that dependence model misspecification can lead to invalid inference through biases in parameter and variance estimates. In the analysis of the Madras study data, we compare six models, mLV, mT, $mTLV$, $mTLV_\sigma$, $mTLV_\gamma$, and $mTLV_{\gamma,\sigma}$. In the $mTLV_\sigma$, $mTLV_\gamma$, and $mTLV_{\gamma,\sigma}$ dependence models, we increase the flexibility of the $mTLV$, by allowing the variance component, the transition component, and both components, to be modified by the indicator $\text{I}(\text{age} < 20)_i$, respectively. Table 4 shows dependence model parameter estimates and associated log likelihoods for all six symptoms using the mean model described above. According to the log likelihood, there is a clear need to incorporate serial dependence among repeated measurements as the mLV log likelihood was, by far, the smallest in all six scenarios. Though likelihood ratio tests for variance components are known to be non-standard (Self and Liang, 1987), there is evidence that for all symptoms except delusions, a variance component improved model fit as the $mTLV$ outperformed the mT. In models for negative symptoms, variance components were modified by $\text{I}(\text{age} < 20)_i$. For example, considering flat affect response, -2 times the difference in the log likelihood between the $mTLV_\sigma$ and the $mTLV$ was $-2 \cdot (-267.130 + 263.338) = 7.584$, $p = 0.006$. The estimated variance component for older subjects was $\hat{\sigma}_1 = 0.58$, while for younger subjects it was $\hat{\sigma}_1 + \hat{\sigma}_2 = 0.58 + 1.65 = 2.23$. The estimated transition component was $\hat{\gamma}_1 = 2.812$, indicating strong serial dependence.

Figure 1 displays the lorelograms (Heagerty and Zeger, 1998) of the dependence models for thought disorders and flat affect. The line denoted by ns is based on a paired estimating equations fit where the dependence model is represented as a natural spline function of time separations with knots fixed at the 33rd and 67th percentiles of the time separation distribution. To produce the remaining lines, we simulated 5000 subjects using the estimated dependence models shown in Table 4, calculated the $\binom{12}{2}$ log odds ratios for all pairwise time combinations, and averaged over the combinations corresponding to equal time separations. Marginal means were fixed at 0.5. For thought disorders, the added flexibility of the $mTLV$ over the mLV is evident, as it provides a closer fit to

Table 4
Log likelihoods and dependence model parameter estimates for each of the six symptoms collected in the Madras ancillary study under six models

Symptom		mLV	mT	$mTLV$	$mTLV_\sigma$	$mTLV_\gamma$	$mTLV_{\gamma,\sigma}$
Hallucinations	γ_1	–	3.616	3.118	3.054	3.037	3.225
	γ_2	–	–	–	–	0.179	−0.387
	σ_1	1.989	–	0.785	0.467	0.772	0.409
	σ_2	–	–	–	0.703	–	0.849
	Log L	−374.706	−322.043	−318.516	−316.963	−318.464	−316.778
Delusions	γ_1	–	3.07	2.98	2.907	2.684	2.731
	γ_2	–	–	–	–	0.656	0.49
	σ_1	1.653	–	0.327	0	0.263	0.001
	σ_2	–	–	–	0.634	–	0.477
	Log L	−461.311	−390.947	−390.709	−389.862	−389.577	−389.366
Thought disorders	γ_1	–	3.164	2.508	2.513	2.628	2.803
	γ_2	–	–	–	–	−0.191	−0.439
	σ_1	2.25	–	1.082	0.971	1.079	0.818
	σ_2	–	–	–	0.18	–	0.396
	Log L	−367.667	−336.664	−331.505	−331.431	−331.442	−331.200
Flat affect	γ_1	–	3.703	2.928	2.812	2.672	2.98
	γ_2	–	–	–	–	0.765	−0.502
	σ_1	2.864	–	1.156	0.578	1.038	0.495
	σ_2	–	–	–	1.649	–	1.979
	Log L	−300.809	−271.533	−267.13	−263.338	−266.401	−263.126
Apathy	γ_1	–	3.196	2.111	2.113	1.99	2.352
	γ_2	–	–	–	–	0.334	−0.598
	σ_1	2.668	–	1.519	0.913	1.476	0.794
	σ_2	–	–	–	1.368	–	1.703
	Log L	−262.864	−253.889	−246.439	−244.141	−246.311	−243.842
Withdrawal	γ_1	–	3.356	2.864	2.85	2.795	3.101
	γ_2	–	–	–	–	0.137	−0.591
	σ_1	2.072	–	0.799	0	0.788	0.001
	σ_2	–	–	–	1.189	–	1.341
	Log L	−274.602	−246.71	−244.615	−242.661	−244.592	−242.293

the empirical *ns* lorelogram. The mT also appears to follow the empirical curve reasonably well, although dependence is constrained to be near 0 at large time separations. The use of the mT or the mLV in this setting is an example of FF misspecification. We display DCM misspecification with the flat affect lorelogram. The estimated dependence models for the $mTLV$ and for young and older subjects estimated from the $mTLV_\sigma$ are provided. Subjects less than 20 years old exhibited far greater response dependence than older subjects. Such a difference is not captured by the $mTLV$.

We now focus on the course of recovery from the thought disorders symptom. Table 5 reports maximum likelihood parameter estimates, and model-based and robust standard error estimates, as well as the mean and standard deviation of the posterior distribution for the $mTLV$ from a Bayesian analysis. According to all models, males tended to recover from thought disorders more slowly than females (e.g., $\mathrm{I(male)}_i \cdot t$ coefficient estimates were positive), and younger subjects recovered more quickly than older subjects (e.g., $\mathrm{I(age} < 20)_i \cdot t$ coefficient estimates were negative). While for most models, these effects were not statistically significant at the 0.05 significance level, the result of dependence model misspecification was apparent using the mLV. With model-based standard errors used for inference, the Wald statistic for the $\mathrm{I(age} < 20)_i \cdot t$ effect was $-0.142/0.063 = 2.253$, and with robust standard errors it was $-0.142/0.092 = 1.543$. Thus, incorrect conclusions could have been drawn with severe dependence model misspecification. The model-based and robust standard errors for the $\mathrm{I(age} < 20)_i \cdot t$ interaction were more similar in the other, better fitting models. A similar pattern was also observed in the $\mathrm{I(male)}_i \cdot t$ interaction estimates.

According to the balance between parsimony and maximization of the likelihood, the $mTLV$ is possibly the optimal model choice for the thought disorders response. We therefore conducted the analogous Bayesian analysis for this model. To sample from the posterior distributions of the parameters, we used Markov chain Monte Carlo, and for the Metropolis algorithm, we implemented a modified version of the three-simulation strategy of Raftery and Lewis (1996). The prior distributions used were reasonably uninformative with $\boldsymbol{\beta}^m \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ was a diagonal matrix of nines, $\gamma_1 \sim N(2.5, 7)$, and $\log(\sigma_1) \sim N(0, 7)$. Results from the Bayesian analysis were qualitatively similar to those based on maximum likelihood. The means (standard deviations) of the marginal posterior distribution for the $\mathrm{I(male)}_i \cdot t$ and $\mathrm{I(age} < 20)_i \cdot t$ coefficients, −0.130 (0.081) and 0.108 (0.076), were similar to the maximum likelihood estimates (model-based standard errors), −0.121 (0.081) and 0.107 (0.078),

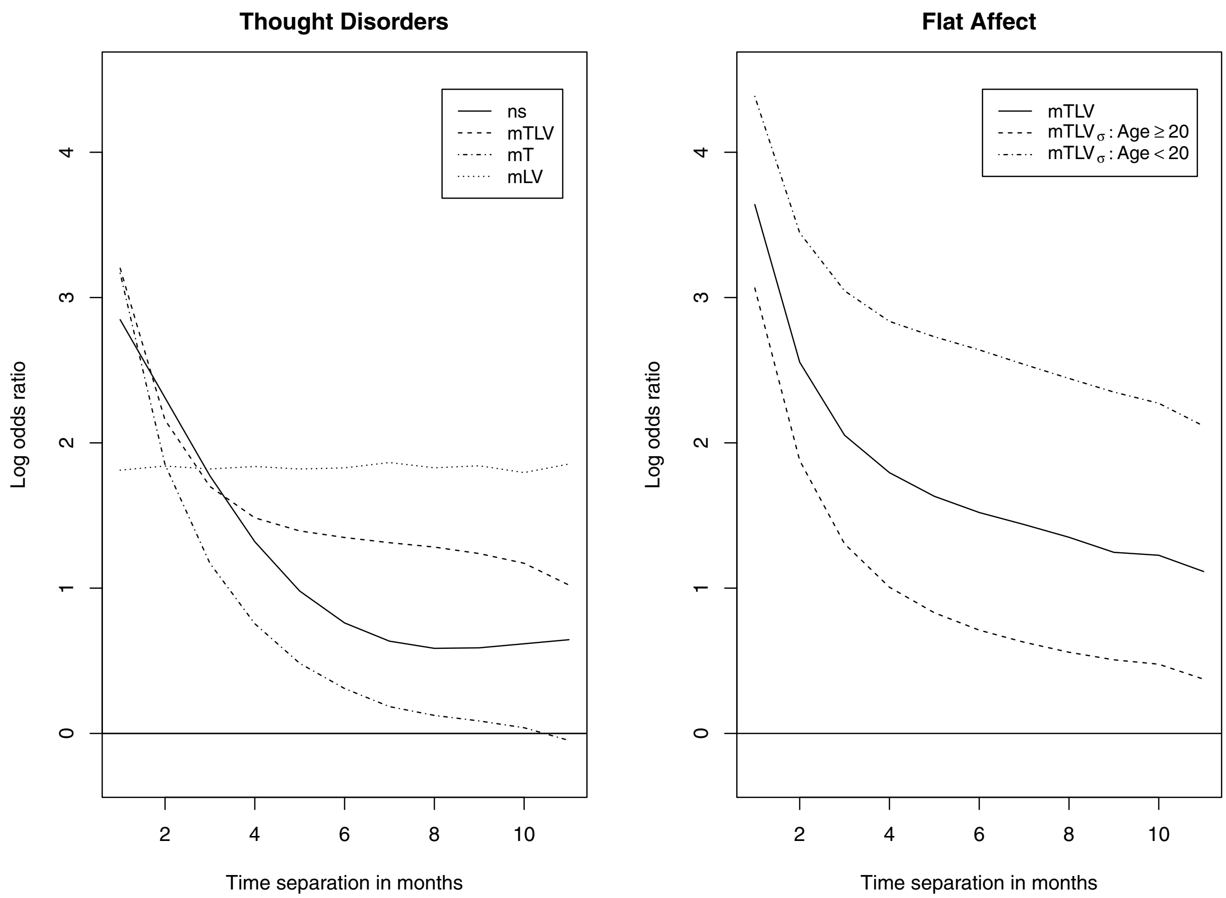

Figure 1. Lorelograms describing within-person response dependence as a function of time separations for the thought disorders and flat affect symptoms. In the thought disorders panel, *ns* denotes the natural spline fitted dependence model (knots at the 33rd and 67th percentiles) based on paired estimating equations as described in Heagerty and Zeger (1998). To produce the lorelograms for the marginalized models, we simulated 5000 subjects using the estimated conditional mean models shown in Table 4, calculated the $\binom{12}{2}$ log odds ratios for all pairwise time combinations, and then averaged over the combinations corresponding to equal time separations. Marginal means were fixed at 0.5.

respectively. In addition, summary statistics from the dependence model parameter posterior distributions were also similar to the maximum likelihood estimates, with mean (standard deviation) of γ_1 and σ_1 equal to 2.585 (0.308) and 1.034 (0.266), respectively.

5. Discussion

In longitudinal studies, repeated measurements taken on the same individual at short time separations are likely to be more highly correlated with one another than those taken at distantly separated time points. In addition, study subjects are generally heterogeneous in their underlying predispositions for positive response values (above and beyond that which can be captured with a mean model). With short response series, say $n_i \leq 5$, it may not be possible or necessary to distinguish serial from long-range dependence; however, when the longitudinal series are relatively long, then empirically separating both components will likely lead to more accurate and efficient estimation. In this article, we extended the flexibility of marginalized models to allow for valid and efficient inference with dependence models that exhibit short-range serial decay and long-range dependence. We discussed maximum likelihood and Bayesian inference and large sample operating characteristics of estimators in the presence of dependence model misspecification while focusing on the nature of the misspecification and on the covariate distribution of the target predictor. We also examined the implications of misspecification in an analysis of the Madras Longitudinal Schizophrenia Study.

With dependence model misspecification, consistency and the degree of inconsistency in the parameter estimates depended upon the type and the degree of misspecification, but not on ICC_x in the predictor of interest. Large sample parameter estimates were approximately unbiased in the presence of FF misspecification though small biases could be observed with DCM misspecification. Though the covariate distribution did not impact the accuracy of parameter estimates, it affected the accuracy of model-based variance estimates and the relative efficiency of regression estimates. In the settings we studied, model-based variance estimators were least biased when $ICC_x = 0$, and the bias grew with increasing percentage of between-subject covariate variation. Efficiency losses can be severe with dependence model misspecification when $ICC_x < 1$, but only minor losses were observed when

Table 5

Analyses of the thought disorders symptom under six models; parameter estimates (model-based standard errors, robust standard errors) based on maximum likelihood and the mean (standard error) of the Bayesian posterior distribution are displayed

Parameter	mLV	mT	$mTLV$	$mTLV_\sigma$	$mTLV_\gamma$	$mTLV_{\gamma,\sigma}$	Bayes $mTLV$
Intercept	0.078	0.344	0.266	0.273	0.272	0.296	0.315
	(0.302, 0.379)	(0.344, 0.339)	(0.334, 0.35)	(0.334, 0.353)	(0.335, 0.346)	(0.339, 0.35)	(0.318)
t (months)	−0.338	−0.368	−0.356	−0.361	−0.355	−0.364	−0.356
	(0.059, 0.103)	(0.072, 0.086)	(0.072, 0.092)	(0.072, 0.09)	(0.072, 0.091)	(0.073, 0.088)	(0.067)
I(male)	0.567	0.182	0.322	0.309	0.324	0.297	0.256
	(0.361, 0.46)	(0.419, 0.431)	(0.405, 0.438)	(0.406, 0.444)	(0.405, 0.439)	(0.407, 0.438)	(0.389)
I(age < 20)	0.943	0.681	0.763	0.774	0.75	0.755	0.767
	(0.396, 0.495)	(0.451, 0.48)	(0.438, 0.476)	(0.439, 0.477)	(0.44, 0.475)	(0.441, 0.477)	(0.445)
I(male) · t	0.079	0.138	0.107	0.113	0.104	0.115	0.108
	(0.06, 0.107)	(0.082, 0.092)	(0.078, 0.095)	(0.08, 0.094)	(0.079, 0.096)	(0.08, 0.094)	(0.076)
I(age < 20) · t	−0.142	−0.116	−0.121	−0.121	−0.12	−0.118	−0.130
	(0.063, 0.092)	(0.086, 0.089)	(0.081, 0.086)	(0.081, 0.086)	(0.081, 0.086)	(0.081, 0.085)	(0.081)
γ_1	–	3.164	2.508	2.513	2.628	2.803	2.585
	(–)	(0.23, 0.242)	(0.304, 0.322)	(0.305, 0.327)	(0.457, 0.441)	(0.534, 0.602)	(0.308)
γ_2	–	–	–	–	−0.191	−0.439	–
	(–)	(–)	(–)	(–)	(0.538, 0.497)	(0.652, 0.724)	(–)
σ_1	2.25	–	1.082	0.971	1.079	0.818	1.034
	(0.282, 0.304)	(–)	(0.265, 0.286)	(0.385, 0.377)	(0.264, 0.286)	(0.468, 0.513)	(0.266)
σ_2	–	–	–	0.18	–	0.396	–
	(–)	(–)	(–)	(0.464, 0.432)	(–)	(0.579, 0.637)	–
Log L	−367.667	−336.664	−331.505	−331.431	−331.442	−331.200	–

$ICC_x = 1$. Analysis of the Madras data revealed that care should be taken to model response dependence, and reinforced the conclusion that severe misspecification can lead to invalid inference.

Marginalized models are a relatively new class of models that permit likelihood-based inference for marginal models. They have the potential to play an important role in the analysis of correlated binary response data as they allow us to choose, independently, the regression model, the dependence model, and the estimation procedure. This article is intended to expand the class of appropriate longitudinal models and to further characterize operating characteristics under the commonly faced challenge of a misspecified dependence model.

6. Supplementary Materials

The Web Appendix referenced in Section 2.2 is available under the Paper Information link at the *Biometrics* website `http://www.tibs.org/biometrics`.

Acknowledgements

The authors would like to thank the editor, associate editor, and referee for their insightful comments that greatly improved the content and exposition of this article.

References

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures (Corr: 1997, **84,** 989). *Biometrika* **81,** 767–775.

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

Feng, Z., McLerran, D., and Grizzle, J. (1996). A comparison of statistical methods for clustered data analysis with Gaussian error. *Statistics in Medicine* **15,** 1793–1806.

Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* **51,** 309–317.

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

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

Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58,** 342–351.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88,** 973–985.

Heagerty, P. J. and Zeger, S. L. (1998). Lorelogram: A regression approach to exploring dependence in longitudinal categorical responses. *Journal of the American Statistical Association* **93,** 150–162.

Heagerty, P. J. and Zeger, S. L. (2000). Multivariate continuation ratio models: Connections and caveats. *Biometrics* **56,** 719–732.

Kurland, B. F. and Heagerty, P. J. (2004). Marginalized transition models for longitudinal binary data with ignorable and nonignorable dropout. *Statistics in Medicine* **23,** 2673–2695.

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.

Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* **17,** 447–469.

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

Menard, S. W. (1995). *Applied Logistic Regression Analysis*. Newbury Park, California: Sage Publications.

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

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.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **54,** 638–645.

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

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.

Raftery, A. E. and Lewis, S. M. (1996). Implementing MCMC. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, D. J. Spiegelhalter, and S. Richardson (eds), 115–130. London: Chapman and Hall.

Schildcrout, J. S. and Heagerty, P. J. (2005). Regression analysis of longitudinal binary response data with time-dependent environmental covariates: Bias and efficiency. *Biostatistics (Oxford)* **6,** 633–652.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82,** 605–610.

Ten Have, T. R., Kunselman, A. R., and Tran, L. (1999). A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine* **18,** 947–960.

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

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

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50,** 1–26.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach (Corr: 1989, **45,** 347). *Biometrics* **44,** 1049–1060.

Received December 2005. *Revised June* 2006.
Accepted June 2006.