An approximate marginal logistic distribution for the analysis of longitudinal ordinal data

Nooraee, Nazanin; Abegaz, Fentaw; Ormel, Johan; Wit, Ernst; van den Heuvel, Edwin R

Published in:
Biometrics

DOI:
10.1111/biom.12414

Biometrics 72, 253–261
March 2016

DOI: 10.1111/biom.12414

An Approximate Marginal Logistic Distribution for the Analysis of Longitudinal Ordinal Data

Nazanin Nooraee,[1,4,*] Fentaw Abegaz,[2] Johan Ormel,[3] Ernst Wit,[2] and Edwin R van den Heuvel[1,4]

[1]University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[2]Johann Bernoulli Institute of Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands
[3]University of Groningen, University Medical Center Groningen, Interdisciplinary Center of Psychopathology and Emotion Regulation, Groningen, The Netherlands
[4]Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands
*email: n.nooraee@tue.nl

Summary. Subject-specific and marginal models have been developed for the analysis of longitudinal ordinal data. Subject-specific models often lack a population-average interpretation of the model parameters due to the conditional formulation of random intercepts and slopes. Marginal models frequently lack an underlying distribution for ordinal data, in particular when generalized estimating equations are applied. To overcome these issues, latent variable models underneath the ordinal outcomes with a multivariate logistic distribution can be applied. In this article, we extend the work of O'Brien and Dunson (2004), who studied the multivariate t-distribution with marginal logistic distributions. We use maximum likelihood, instead of a Bayesian approach, and incorporated covariates in the correlation structure, in addition to the mean model. We compared our method with GEE and demonstrated that it performs better than GEE with respect to the fixed effect parameter estimation when the latent variables have an approximately elliptical distribution, and at least as good as GEE for other types of latent variable distributions.

Key words: Flexible correlation matrix; Generalized estimating equations; Latent variable models; Maximum likelihood; Multivariate logistic distribution; Population-averaged (marginal) models; t-distribution.

1. Introduction

Longitudinal ordinal outcomes are often used to measure and understand the well-being of groups of subjects in the population over time. Both subject-specific and marginal models can be used to analyze this type of data (Fitzmaurice et al., 2009), but subject-specific models seem less suitable. Subject-specific models treat some model parameters as random variables to induce temporal correlations, but the interpretation of the fixed parameters at the population level is not straightforward due to the complication of integrating out these random effects. The so-called bridge distribution for random effects in combination with logit link functions do provide marginal logistic distributions (Wang and Louis, 2003; Parzen et al., 2011), but they are more difficult to interpret than normally distributed random effects and are not yet available as an option in standard software packages. Furthermore, penalized quasi-likelihood (Stiratelli et al., 1984), marginal quasi-likelihood (Breslow and Clayton, 1993) and marginalized subject-specific models (Heagerty, 1999, 2002; Lee and Daniels, 2008) all provide population-averaged interpretations, but they do not necessarily imply a marginal distribution for the longitudinal ordinal data in the population.

Marginal models model the mean levels and the associations of repeated ordinal outcomes at a population level, but the parameters are predominantly estimated with generalized estimating equations (GEE). An important limitation of GEE is that it treats the associations between outcomes only as nuisance parameters. Though it does provide some information on the association, it is statistically weak. This limitation can be overcome by specifying a set of equations for both the fixed parameters and the association parameters (Heagerty and Zeger, 1996), but it does not necessarily lead to a multivariate distribution for the ordinal outcomes and thus complicates the interpretation of the association structure. Even if such a distribution would exist (Chaganty and Joe, 2004), interpretation of the association structure for ordinal outcomes is still difficult. Indeed, the correlation between each pair of ordinal outcomes is evaluated via a matrix rather than a scalar since each level of the ordinal outcome is represented by a vector of binary elements (Clayton, 1992), which are all zero except for one entry that connects to one level of the ordinal outcome. Another drawback of GEE is that it may result in biased estimates when missing data occur and the mechanism is other than missing completely at random (MCAR), see Molenberghs and Kenward (2007). This problem can be solved by use of weighted GEE (Robins et al., 1995), doubly-robust inverse probability weighting (Scharfstein et al., 1999) or a combination of multiple imputation (MI) and GEE (Rubin, 1987). However, it is unknown which of these approaches is best for ordinal outcomes.

Alternatively, several joint distribution functions have been developed to handle repeated ordinal outcomes. Molenberghs and Lesaffre (1994) used joint probabilities via the extended Dale model (Dale, 1986), i.e., the joint probabilities (cell probabilities) are decomposed into main effects and higher order associations using global odds ratios. Unfortunately, this model is not easy to implement, unless the number of repeated measurements is small (Agresti and Natarajan, 2001). Another approach is treating ordinal outcomes as manifest variables for one or more latent variables. This approach is widely used in structural equation modeling, and frequently assume that the latent variables follow multivariate normal distributions (Muthén, 1983, 1984; Kline, 2011). Such a distribution for the latent variable underneath the ordinal outcomes was also used by Li and Schafer (2008). These models result in marginal multivariate distributions only when the link function is the probit. Therefore, an odds ratio interpretation, like GEE, is not guaranteed with these models. On the other hand, multivariate logistic distributions in combination with a logit link function could be used instead. A widely known bivariate logistic distribution was introduced by Gumbel (1961) and further extended to higher order dimensions by Malik and Abraham (1973), (see also Kotz et al. 2000). This multivariate distribution belongs to the class of Archimedean distributions (Nelson, 2006), but they have just one parameter to represent all correlations. Thus Gumbel distributions support only the exchangeable correlation matrix. The (Generalized) Farlie-Gumbel-Morgenstern distribution (Kotz et al., 2000) does have an unstructured correlation matrix, but the correlation coefficients are usually small due to the restrictions on the parameters of the distribution. An alternative, more flexible, approach is the use of the t-copula to construct multivariate distribution with marginal logistics (O'Brien and Dunson, 2004), which has been applied to longitudinal ordinal data using a Bayesian method to estimate the parameters.

Our goal is to generalize the approach of O'Brien and Dunson (2004) to be able to model the temporal correlations with possible time-(in)variant covariates using existing software. Instead of a Bayesian approach, we will apply maximum likelihood. Incomplete outcomes that satisfy MCAR or missing at random (MAR) can then be handled easily without using other methods like MI. We examine the performance of this approach in terms of bias, mean squared error, and the coverage probability of Wald-type confidence intervals using simulation studies. Ordinal outcomes are simulated via moderately to highly correlated latent variables with different multivariate distributions. Furthermore, our approach will be compared with GEE and MI-GEE for full and incomplete data sets, respectively.

A prospective cohort study (TRAILS: TRacking Adolescents' Individual Lives Survey) of Dutch adolescents using bi- or triennial measurements on mental health from age 11 onward motivated our research. A depression subscale of mental health is the Youth Self-Report (YSR) obtained at baseline and two follow-ups. This subscale has good psychometric properties (Achenbach and Rescorla, 2006) and consists of 13 items. Each item has three levels (0, 1, 2) and higher levels indicate more depression. The sum scale of the 13 items is categorized for boys and girls differently. A sum score of [0; 5] or [0; 7] indicate a normal range for boys and girls, respectively. A sum score of [6; 8] or [8; 11] indicate mild depressive symptoms, while a sum score of [9; 11] or [12; 14] indicate subthreshold depressive symptoms. A sum score above 12 and 15 suggest significant depressive symptoms and possibly clinical depression for boys and girls, respectively. Age and gender of adolescents, a history of parental internalizing and externalizing problems, family structure, and social-economic status of family were considered explanatory variables for depression status. Additionally, different temporal correlations in depression scores for boys and girls would indicate that stability of depression symptoms is gender specific. This case study shows the need for a population-averaged longitudinal ordinal logistic regression model that could also address different temporal correlations for certain subpopulations. The threshold values for the ordinal outcome should be considered constant across subpopulations since the ordinal scale is already be corrected for subpopulations.

The rest of the article is organized as follows. Statistical methods (i.e., models and estimation) are considered in Section 2. Simulation studies and our case study are described in Sections 3 and Section 4, respectively. Finally, Section 5 discusses the model and results.

2. Methods

Let O_{ij} represent an ordinal outcome, having C levels, for subject i $(i = 1, \ldots, N)$ at time point j $(j = 1, \ldots, J)$. Corresponding to each subject i, let $\mathbf{X}_i$ be a $J \times p$ matrix containing all time varying and time invariant covariates. The effect of covariates on an ordinal outcome can be evaluated through the proportional odds cumulative model, i.e.,

$$\operatorname{logit}\left\{P(O_{ij} \leq c|\boldsymbol{X}_{ij} = \boldsymbol{x}_{ij})\right\} = \gamma_c - \boldsymbol{x}'_{ij}\boldsymbol{\beta} \qquad (1)$$

with γ_c a threshold parameter for level c $(c = 0, 1, \ldots, C)$, such that $-\infty = \gamma_0 < \gamma_1 < \ldots . < \gamma_C = \infty$ and $\boldsymbol{\beta}$ a p-dimensional vector of regression parameters (McCullagh, 1980). This model may be viewed as a latent variable model in which an unobservable continuous random variable Y lies underneath the ordinal variable, such that it transforms the continuous outcome into an ordinal outcome for each subject i at time j, i.e. $O_{ij} = c$ if and only if $\gamma_{c-1} < Y_{ij} \leq \gamma_c$. To satisfy equation (1) the latent variable Y_{ij} must have a logistic distribution with mean parameter $\boldsymbol{\mu}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ and scale parameter 1. Thus, the multivariate density function for $\boldsymbol{O}_i = (O_{i1}, O_{i2}, \ldots, O_{iJ})$ can be written by

$$g(c_i|\gamma, \boldsymbol{\beta}, \boldsymbol{X}_i, \boldsymbol{R}) = \int_{\gamma_{c_{i1}-1}}^{\gamma_{c_{i1}}} \int_{\gamma_{c_{i2}-1}}^{\gamma_{c_{i2}}} \ldots . \int_{\gamma_{c_{iJ}-1}}^{\gamma_{c_{iJ}}} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{R}) dy_i \qquad (2)$$

with $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iJ})$ a vector of levels, $f(.|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{R})$ a multivariate logistic density, $\boldsymbol{X}_i$ a matrix of p covariates, $\boldsymbol{\beta}$ a vector of regression parameters, and $\boldsymbol{R}$ a correlation matrix.

2.1. *Multivariate Logit Model*

The density function for the t-copula based multivariate distribution used by O'Brien and Dunson (2004) for subject i is given by

$$f(z_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{R}) = \frac{t\left\{h(z_{i1}), h(z_{i2}), \ldots, h(z_{iJ})|\boldsymbol{R}\right\}}{\prod_{j=1}^{J} t_1\{h(z_{ij})\}} \prod_{j=1}^{J} f_1(z_{ij}), \tag{3}$$

with t and t_1 the multivariate and univariate t-densities with ν degrees of freedom, respectively, $h(z) = T_1^{-1}\{e^z/(1+e^z)\}$, with T_1 the univariate t-distribution function with ν degrees of freedom, $f_1(z) = \exp(z)/\left\{1+\exp(z)\right\}^2$ the univariate logistic density, and $z_i = (y_{i1} - \mu_{i1}, y_{i2} - \mu_{i2}, \ldots, y_{iJ} - \mu_{iJ})$.

Albert and Chib (1993) showed that the univariate logistic distribution, f_1, can be approximated by a univariate t-distribution with $\nu = 8$ degrees of freedom. To make the distributions similar, a scale parameter $\pi^2(\nu-2)/3\nu$ in the t-distribution is needed. Using these approximations, equation (3) becomes approximately

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{R}) \approx t\left\{h(z_{i1}), h(z_{i2}), \ldots, h(z_{iJ})|\widetilde{\boldsymbol{R}}\right\}, \tag{4}$$

with $\widetilde{\boldsymbol{R}} = \{\pi^2(\nu-2)/3\nu\}\boldsymbol{R}$, and $\boldsymbol{R}$ the correlation matrix for the continuous latent variables.

2.2. *The Covariance Structure*

To improve on this known model, we increase the complexity by implementing covariates into the correlation matrix $\boldsymbol{R}$. This makes it possible to estimate correlation matrices separately for subgroups of subjects (categorical variables) and for individuals (continuous variables). We re-parameterize all elements of $\boldsymbol{R}$ with Fisher's z-transformation, i.e.,

$$\log\left(\frac{1+\rho_{jk}}{1-\rho_{jk}}\right) = \alpha_{0jk} + w_{ijk}\alpha_{jk} \tag{5}$$

with ρ_{jk} the correlation coefficient of latent variable at time j and k $(j \neq k)$, w_{ijk} a vector of covariates, α_{0jk} an intercept and α_{jk} a slope. The structure is flexible in terms of intercepts and slopes. When $\alpha_{0jk} = \alpha_{jk} = 0$, $\boldsymbol{R}$ is the identity matrix. When $\alpha_{jk} = 0$, no covariates influence the correlation matrix, but the matrix can still attain several structures: the exchangeable structure ($\alpha_{0jk} = \alpha_0\ \forall j, k$), the AR(1) structure ($\alpha_{0jk} = \ln\left\{1 + C(\alpha)\right\} - \ln\left\{1 - C(\alpha)\right\}$ with $C(\alpha) = \left[\left\{\exp(\alpha) - 1\right\} / \left\{\exp(\alpha) + 1\right\}\right]^{|j-k|}$), and the unstructured form. When $\alpha_{jk} \neq 0$, covariates in model (5) may explain variation in the temporal correlation coefficients.

In principle, any value in the range $[-1, 1]$ can be substituted for ρ_{jk}, but the correlation matrix must be (semi)positive definite and it leads to constraints on ρ_{jk}'s. These constrains have been discussed for three variables (Olkin, 1981) and four variables (Budden et al., 2007). For example, element ρ_{jk} in a three-dimensional correlation matrix should be in the interval

$$\left[\rho_{js}\rho_{ks} - \sqrt{(1-\rho_{js}^2)(1-\rho_{ks}^2)}; \rho_{js}\rho_{ks} + \sqrt{(1-\rho_{js}^2)(1-\rho_{ks}^2)}\right], \tag{6}$$

for $s \neq j, k$. (Semi)positive definiteness is fulfilled by maximizing the likelihood under these constraints. For higher dimensions, we can apply Higham's algorithm (Higham, 1988) to assure that positive definiteness of the correlation matrix fulfills.

2.3. *Parameter Estimation: Maximum Likelihood Approach*

Instead of a Bayesian approach (O'Brien and Dunson, 2004), we applied maximum likelihood for estimation of the parameters in equation (4). This choice of estimation is computationally less expensive and estimation is easier when covariates are implemented into the correlation matrix, since selecting an appropriate prior on α_{jk} is not required. However, computing a multivariate t-distribution with an arbitrary correlation matrix is still computationally burdensome. To calculate the t-density in (4), we applied the formulation (7) for the multivariate t-distribution introduced by Cornish (1954) in combination with the numerical method of Genz and Bretz (1999). It is just a different form of the same t-density (Genz and Bretz, 1999). The multivariate t-distribution is then given by

$$T(\gamma_{c_i}|\mu, \widetilde{\boldsymbol{R}}) = \frac{2^{1-\nu/2}}{\Gamma(\nu/2)} \int_0^\infty q^{\nu-1} \exp(-\boldsymbol{q}^2/2)\Phi(\frac{\boldsymbol{q}\gamma_{c_i}}{\sqrt{\nu}} - \mu|\widetilde{\boldsymbol{R}})d\boldsymbol{q} \tag{7}$$

with $\Phi(\xi_1, \ldots, \xi_i|\widetilde{\boldsymbol{R}}) = (2\pi|\widetilde{\boldsymbol{R}}|)^{J/2} \int_{-\infty}^{\xi_1} \cdots \int_{-\infty}^{\xi_i} \exp(-\boldsymbol{q}'\widetilde{\boldsymbol{R}}\boldsymbol{q}/2)d\boldsymbol{q}$ the multivariate normal distribution. Genz and Bretz (2002) showed that their numerical method computes the multivariate t-distribution robustly and reliably up to 20 dimensions. This function is implemented in R software, assuming that $\widetilde{\boldsymbol{R}}$ is a scale matrix (see the Web Appendix A).

To maximize the likelihood function based on the proportional odds model (1) using density (2) with approximation (4) in the form of (7), we used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Shanno, 1970) with function *optim* in R package, which is a quasi-Newton method, and computes the optimum value for nonlinear functions. The starting or initial values were selected using univariate proportional odds logistic regression for the regression parameters and polychoric correlations for the correlation coefficients. Continuous covariates are standardized with the associated means and standard deviations and we implemented $\gamma_c = \gamma_{c-1} + d_{c-1}^2$, similar to Li and Schafer (2008), to ensure that the threshold parameters are ordered. To implement a semi-positive definite correlation matrix for three dimensional outcomes, we maximized the likelihood function under constraint (6) by putting the constraint in the likelihood function, see the R codes that is available with this article at the *Biometrics* website on Wiley Online Library. For higher dimensions, we can implement Higham's algorithm (Higham, 1988) at the same place in the likelihood function using function *make.positive.definite* in package corpor of R software. To ob-

Table 1
Bias(MSE) and CP of estimated parameters for 1000 simulated data sets with t-logistic latent variables and 50 subjects. The true correlation coefficients are indicated in the first row.

		(0.45, 0.30, 0.70)				(0.85, 0.85, 0.85)			
		ML		GEE		ML		GEE	
Parameter	True value	Bias(MSE)	CP[a]	Bias(MSE)	CP	Bias(MSE)	CP[b]	Bias(MSE)	CP
γ_1	−0.5	0.018 (0.099)	94.91	0.024 (0.101)	94.20	0.037 (0.146)	93.87	0.037 (0.146)	95.10
γ_2	1	−0.017 (0.108)	95.33	−0.035 (0.110)	95.20	−0.019 (0.154)	94.86	−0.050 (0.158)	95.00
γ_3	2	−0.044 (0.187)	94.29	−0.099 (0.199)	94.40	−0.041 (0.265)	95.62	−0.115 (0.289)	95.20
β_S	−1	0.017 (0.191)	95.33	−0.046 (0.209)	94.10	0.044 (0.291)	94.64	−0.068 (0.311)	94.40
β_A	−0.2	0.003 (0.049)	95.85	−0.014 (0.053)	93.5	0.010 (0.067)	95.84	−0.022 (0.073)	94.60
β_T	−0.1	−0.005 (0.018)	95.53	−0.001 (0.021)	95.2	−0.006 (0.006)	94.09	−0.003 (0.008)	95.40
ρ_{12}	-	0.007 (0.026)	93.15	-	-	0.003 (0.005)	91.03	-	-
ρ_{13}	-	0.005 (0.032)	92.21	-	-	0.001 (0.005)	91.47	-	-
ρ_{23}	-	0.011 (0.014)	90.97	-	-	0.005 (0.005)	90.48	-	-

[a] 37 data sets are removed due to convergence issues in computing the standard errors.
[b] 86 data sets are removed due to convergence issues in computing the standard errors.

tain standard errors, we applied numerical differentiation using function *hessian*, implemented in the package *numDeriv* in R. Note that the first and second derivatives of the logarithm of the likelihood function exist, but their computations are cumbersome (see Web Appendix A). Finally, to address missing outcome data the full likelihood function is changed to the same likelihood function with a lower dimension. For instance, if subject i has data only for the first two time points, the full likelihood function $t\left\{h(z_{i1}), h(z_{i2}), \ldots, h(z_{iJ})\,|\widetilde{\boldsymbol{R}}\right\}$ for a complete set of outcomes is replaced by $t\left\{h(z_{i1}), h(z_{i2})\,|\widetilde{\boldsymbol{R}}_{12}\right\}$, with $\widetilde{\boldsymbol{R}}_{12}$ the partition of $\widetilde{\boldsymbol{R}}$ corresponding to the (scaled) correlation for the first two time points.

3. Simulation Studies

The aim of the simulation study is to investigate four aspects of our approach. Firstly, we study the bias, mean squared error (MSE), and coverage probabilities of Wald-type confidence intervals of the ML estimates when we simulate small sample sizes and complete data from density (3). The performance is compared with GEE. Furthermore, the likelihood ratio test (LRT) is studied for selecting correlation structures. The second part investigates the performance of our approach when we implement covariate(s) in the correlation coefficients. The LRT was used for testing their effects. The third part presents results for latent variables with other multivariate (logistic) distributions. Finally, the last part explores the performance of our approach in the presence of missing outcomes. These results are compared to GEE with multiple imputation (MI-GEE) using predictive mean matching (PMM).

In all settings, we generated three covariates independently: a Bernoulli random variable x_1 with parameter 0.49, representing gender, a normally distributed random variable x_2 from $N(11, 0.5^2)$, representing age at baseline, time at baseline $x_{30} = 0$, and follow-up times x_{31} and x_{32} with distributions $N(2.5, 0.4^2)$ and $N(5, 0.6^2)$, respectively. Three-dimensional continuous latent variables (Z_1, Z_2, Z_3,) were generated via t-, Gumbel, or normal copula functions (Nelson, 2006). The latent variable Y_{ij} was then set to $Y_{ij} = \beta_S X_{i1} + \beta_A X_{i2} + \beta_T X_{i3j} + Z_{ij}$ with $(\beta_S, \beta_A, \beta_T) = (-1, -0.2, -0.1)$ or $(-1, -0.5, -0.5)$. The 4-level ordinal outcomes were produced by fixing the threshold parameters such that about 50, 25, 15, and 10% of the observations attain the first, second, third, and fourth level, respectively. The cut-point values varied with the distribution of the latent variables but they were fixed when different correlation matrices were selected. We repeated simulations 1000 times in all settings.

3.1. *Setting I: Accuracy of the Approximate t-Copula*

Using the t-copula, we generated (u_{i1}, u_{i2}, u_{i3}) simultaneously having a marginal uniform distribution. We transformed u_{ij} to z_{ij} using $F_1^{-1}(u_{ij})$, to obtain marginally logistic distributed data (with F_1 the standard univariate logistic distribution). Note that this transformation led to the multivariate logistic distribution with density (3). Three different correlation matrices were simulated: $(\rho_{12}, \rho_{13}, \rho_{23})$ is $(0.45, 0.30, 0.70)$, $(0.85, 0.85, 0.85)$, and $(0.80, 0.70, 0.90)$, respectively. Table 1 presents the results for $(0.45, 0.30, 0.70)$ and $(0.85, 0.85, 0.85)$, but the results of $(0.80, 0.70, 0.90)$ are not shown since they were similar to the results of $(0.85, 0.85, 0.85)$.

Bias in the ML-estimates is limited to 2% for correlation coefficients and 7% for the regression parameters. A comparison of ML and GEE demonstrates that ML has less absolute bias. Additionally, GEE has a slightly higher MSE than ML (Table 1). Coverage probabilities (CPs) are slightly liberal for the correlation coefficients, but they are quite close to the 95% nominal level for all other parameters. GEE and ML seem to provide comparable coverages for the regression parameters.

Applying the likelihood ratio test for selecting the correlation structure, we first investigated the type I error rate for the setting with correlations (ρ12, ρ13, ρ23) = (0.85, 0.85, 0.85) by comparing the exchangeable structure with unstructured. The type I error rate was 0.05, equal to the significant level. Comparing the same structures but now for setting (ρ12, ρ13, ρ23) = (0.45, 0.30, 0.70) gave a power of 0.53, which is considered acceptable for a sample size of only 50 individuals.

Table 2

Bias(MSE) and coverage probability (CP) of parameter estimates, considering the impact of covariate on the correlation over time, from 1000 simulated data sets containing 50 individuals

	Sex				Age at baseline	
True value	(1.0, 0.5, 1.0, 0.5, 1.0, 0.5)		(0.4, 0.3, 0.3, 0.1, 0.5, 0.1)		(0.5, 0.07, 0.5, 0.07,0.5, 0.07)	
Parameters	Bias(MSE)	CP[a]	Bias(MSE)	CP[**]	Bias(MSE)	CP[b]
α_{012}	−0.009 (0.062)	96.21	0.009 (0.056)	95.75	−0.002 (0.001)	1.00
α_{12}	−0.009 (0.100)	97.69	0.015 (0.110)	97.57	−0.001 (0.001)	99.49
α_{013}	−0.013 (0.060)	97.20	0.022 (0.061)	93.45	0.002 (0.001)	1.00
α_{13}	−0.013 (0.120)	96.87	−0.111 (0.137)	95.63	−0.002 (0.001)	99.49
α_{023}	−0.016 (0.066)	96.05	0.009 (0.590)	95.87	0.002 (0.001)	1.00
α_{23}	−0.002 (0.115)	95.88	0.012 (0.123)	97.45	−0.000 (0.001)	98.98

[a] 393 data sets are removed due to convergence issues in computing the standard errors.
[**] 176 data sets are removed due to convergence issues in computing the standard errors.
[b] Only 196 data sets provide valid the standard errors.

3.2. *Setting II: Heterogeneity of Subpopulations*

The data were generated similar to the steps in setting I, but now the correlations (ρ_{12}, ρ_{12}, ρ_{23}) were generated using the equation (5). For the binary covariate (gender), we selected $\alpha_{0jk} = 1$ and $\alpha_{jk} = 0.5$ for all j and k. This choice imposes an exchangeable correlation for males ($\rho = 0.76$) and females ($\rho = 0.90$) separately. To simulate an unstructured correlation matrix for two genders, we set $(\alpha_{012}, \alpha_{013}, \alpha_{023}) = (0.4, 0.3, 0.5)$ and $(\alpha_{12}, \alpha_{13}, \alpha_{23}) = (0.3, 0.1, 0.1)$. Yet in another simulation setting, we considered the continuous covariate (age at baseline) in the correlation matrix, and introduced an exchangeable form by selecting $\alpha_{0jk} = 0.5$ and $\alpha_{jk} = 0.07$, for all $j \neq k$. Due to a positive slope ($\alpha_{ijk} > 0$), older individuals at baseline have stronger correlations.

The absolute bias of the association parameters α's is at most 3% (Table 2). In these analyses we applied an unstructured association, even though we simulated exchangeable correlations for two settings. The MSE for the association parameters is comparable with the MSE results for settings that do not include covariates in the correlation matrix. This is also true for the bias and the MSE of the regression parameters (results not shown). The coverage probabilities on the correlation coefficients is close to nominal or somewhat conservative. It should be noted though, that not all data sets provided appropriate standard errors due to eigenvalue problems in the inverse of the Hessian matrix. This issue does not happen with the regression parameters, but seems to be a problem with the estimation of the correlation coefficients. However, this issue appears less frequent when the sample size is increased from 50 subjects to 100 subjects (results are not shown).

The power of the likelihood ratio test was also examined for covariates that were implemented in the correlation coefficients. For the setting with age in the correlation structure, the LRT detected an effect of age in the correlation matrix with power 1 when we compared it with a model that used an unstructured correlation without age. On the other hand, when the binary covariate gender was used in the correlation structure, the power of the LRT was about 0.50. The power is still acceptable considering the sample size that was applied in relation to the effect size. For a comparison of the individual time related correlations between males and females, the power would range from 0.094 to 0.257 (Graybill, 1961). The type I error rate for testing the presence of covariates in the correlation matrix was 0.06 when we simulated data with correlation coefficients $(\rho_{12},\ \rho_{13},\ \rho_{23},) = (0.45, 0.30, 0.70)$ independent of covariates.

3.3. *Setting III: Sensitivity Analysis*

The multivariate Gumbel distribution was generated via the Gumbel copula, given by

$$C_G\left(u_1, \ldots, u_k\right) = \exp\left[-\left\{\sum\left(-\log u_i\right)^{\theta}\right\}^{\frac{1}{\theta}}\right], \qquad (8)$$

with θ the association parameter. We transformed u_{ij} to $z_{ij} = F_1^{-1}(u_{ij})$ again to obtain marginally logistic distributed latent variables, and selected θ equal to 2 and 5. These choices generated a strong ($\simeq 0.70$), and an extremely strong ($\simeq 0.95$) Pearson correlation coefficient, respectively (Nooraee et al., 2014). Moreover, we generated multivariate normally distributed random variables using the normal-copula, with a standard deviation equal to 1.7 and correlation coefficients $(0.45, 0.30, 0.70)$. The value 1.7 makes the variance of the normal distribution close to the logistic distribution.

For the Gumbel distributed data with small sample sizes, estimation of the model parameters using exchangeability in our approach resulted into a bias for a few regression parameters (Table 3). Increasing the sample size reduced the bias, although it did not completely disappear. On the other hand, the bias was almost eliminated when the unstructured association matrix was applied, even for the small sample size (Table 3). The observed bias with the exchangeable association, might be the result of a difference in the Archimedean copula of the Gumbel distribution and the elliptical copula of the t-distribution. By choosing the unstructured association, apparently the fit of the t-distribution seems to get closer to the Gumbel distribution. However, a small bias in the association parameters seems to remain present, also for larger sample sizes (data not shown). The coverage probabilities be-

Table 3

Bias(MSE) and CP in the parameter estimates for 1000 simulated data sets from the Gumbel distribution with parameter $\theta = 2$, and 50 subjects

		ML estimates fitting an					
		Exchangeable correlation		Unstructured correlation		GEE	
Parameter	True value	Bias(MSE)	CP[a]	Bias(MSE)	CP[b]	Bias(MSE)	CP
γ_1	−0.5	−0.056 (0.090)	97.02	0.038 (0.110)	95.12	−0.027 (0.111)	95.00
γ_2	0.75	−0.089 (0.164)	95.99	0.019 (0.121)	95.33	−0.032 (0.119)	95.70
γ_3	2	−0.311 (1.310)	93.73	−0.100 (0.290)	94.29	−0.135 (0.305)	93.30
β_S	−1	−0.127 (0.161)	98.36	0.027 (0.233)	95.64	−0.055 (0.252)	95.70
β_A	0.2	0.004 (0.054)	96.30	0.018 (0.056)	96.68	−0.024 (0.063)	95.10
β_T	−0.1	−0.003 (0.009)	95.48	0.000 (0.009)	93.98	−0.009 (0.011)	94.20
ρ_{12}	0.7	} 0.038 (0.009)	81.91	−0.043 (0.013)	81.41	-	-
ρ_{13}	0.7			−0.051 (0.013)	80.48	-	-
ρ_{23}	0.7			−0.043 (0.013)	82.04	-	-

[a]27 data sets are removed due to convergence issues in computing the standard errors.
[b] 37 data sets are removed due to convergence issues in computing the standard errors.

came slightly conservative (up to 98.4%) for the regression parameters and substantially liberal for the correlation coefficients. It should be noted though that the performance of ML for the regression parameters is not worse than GEE for the Gumbel distribution (Table 3).

For normally distributed latent variables, the bias and MSE of the parameter estimates are comparable with the bias and MSE of the parameter estimates for t-logistic distributed latent variables. Similarly, results show smaller MSE with ML than with GEE. While CPs for all the regression parameters are close to or slightly above the nominal level for ML, they are all below the nominal level when GEE was applied (Table 4).

3.4. *Setting IV: Incomplete Outcome Data*

For the simulated data with the t-copula, we also simulated missing indicators, D_{ij}, using the following probability

$$\text{logit} P(D_{ij} = 1|\boldsymbol{X}_i, Z_{i1}) = (\psi_{0j} + \psi_{1j}x_{i1} + \psi_{2j}x_{i2} + \psi_{3j}x_{i3j} + \psi_{4j}Z_{i1}), \quad (9)$$

with $(\psi_{0j}, \psi_{1j}, \psi_{2j}, \psi_{3j}, \psi_{4j})$ given by (−2.0, −2.0, 0.5, 1.0, 0.5), and (−0.8, −1.5, 1.5, −2.0, 1.5) for $j = 2$, and $j = 3$, respectively, that results in approximately 9.5 and 18% missingness at the second and third time points, respectively. Note that we did not simulate any missingness at baseline. The missingness fulfills MAR on the latent variable, but not necessarily on the ordinal level. Generating missing data in this way is more realistic since Z_{i1} represents the true ability or liability of a subject at baseline.

Since GEE may lead to biased estimates under MAR (Molenberghs and Kenward, 2007), we employed predictive mean matching (PMM) multiple imputation. We conducted 20 imputations and pooled the results via Rubin's rule (Rubin, 1976) to obtain relative efficiency of at least 95%.

Biases in the regression parameters are limited to at most 2% in all settings (Table 5). However, some larger biases were observed in the correlation coefficients for small sample sizes. By increasing the sample size to 100 subjects, the bias diminishes to at most 4% for the correlation parameters. Both ML and MI-GEE provide similar biases in parameter estimates, however the ML approach indicates slightly better MSE's and

Table 4

Bias(MSE) and CP of the estimated parameters for 1000 simulated data sets with multivariate normal latent variables

		ML		GEE	
Parameter	True value	Bias(MSE)	CP[a]	Bias(MSE)	CP
γ_1	0	0.010 (0.089)	96.10	0.021 (0.186)	94.70
γ_2	1	0.004 (0.107)	95.24	−0.014 (0.107)	94.40
γ_3	2.5	−0.197 (0.351)	96.75	−0.798 (0.417)	93.90
β_S	−1	−0.016 (0.196)	96.86	−0.014 (0.214)	94.60
β_A	−0.2	−0.013 (0.048)	95.56	0.001 (0.056)	93.20
β_T	−0.1	0.004 (0.020)	95.13	−0.003 (0.024)	93.90
ρ_{12}	0.45	0.016 (0.031)	91.99	-	-
ρ_{13}	0.30	0.013 (0.037)	92.42	-	-
ρ_{23}	0.70	0.015 (0.016)	91.61	-	-

[a] 76 data sets are removed due to convergence issue in computing the standard errors.

Table 5

Bias(MSE) and CP of parameter estimates using ML and MI-GEE, from 1000 simulated incomplete data sets

$(\beta_S, \beta_A, \beta_T)$		(-1,-0.2,-0.1)								(-1,-0.5,-0.5)			
	N	50				100				50			
		ML		GEE		ML		GEE		ML		GEE	
Parameter	true value	Bias(MSE)	CP[a]	Bias(MSE)	CP	Bias(MSE)	CP[b]	Bias(MSE)	CP	Bias(MSE)	CP[c]	Bias(MSE)	CP
	−0.5	0.032 (0.115)	95.24	0.01 (0.10)	94.50	0.011 (0.050)	95.57	0.00 (0.05)	94.30	0.011(0.117)	96.07	0.01 (0.11)	94.70
γ_2	0.5	−0.009 (0.119)	95.35	−0.04 (0.11)	94.40	−0.004 (0.052)	95.26	−0.02 (0.05)	94.90	−0.022 (0.118)	94.94	−0.04 (0.12)	94.70
γ_3	1.65	−0.031 (0.117)	95.57	−0.12 (0.17)	93.10	−0.011 (0.084)	94.64	−0.07 (0.09)	93.80	−0.059 (0.176)	95.73	−0.12 (0.17)	92.70
β_S	-	0.018 (0.212)	96.12	−0.03 (0.22)	94.10	0.003 (0.093)	95.88	−0.08 (0.10)	94.50	−0.009 (0.213)	96.29	0.02 (0.23)	94.10
β_A	-	0.006 (0.053)	97.34	−0.01 (0.05)	93.10	0.004 (0.025)	95.88	−0.00 (0.02)	94.90	0.011 (0.057)	97.19	0.01 (0.06)	95.00
β_T	-	0.003 (0.020)	96.90	−0.01 (0.02)	94.50	−0.001 (0.010)	96.19	−0.01 (0.01)	95.50	0.00 (0.024)	96.97	0.03 (0.03)	93.20
ρ_{12}	0.45	0.037 (0.035)	90.25	-	-	0.015 (0.015)	92.99	-	-	0.042 (0.035)	90.11	-	-
ρ_{13}	0.30	0.032 (0.044)	90.59	-	-	0.011 (0.021)	91.65	-	-	0.046 (0.053)	88.65	-	-
ρ_{23}	0.70	0.024 (0.019)	89.15	-	-	0.012 (0.009)	91.75	-	-	0.029 (0.021)	89.21	-	-

[a] 97 data sets are removed due to convergence issues in computing the standard errors.
[b] 30 data sets are removed due to convergence issues in computing the standard errors.
[c] 110 data sets are removed due to convergence issues in computing the standard errors.

CP's compared to the MI-GEE approach, particularly for the regression coefficients.

4. Application: The TRAILS Analysis

The design, methods, and response rates and bases of our prospective cohort study TRAILS has been described in detail elsewhere (Ormel et al., 2012). Briefly, participants were selected from five municipalities (urban and rural) in the North of the Netherlands. Children born between October 1, 1989 and September 30, 1991 were eligible for inclusion. A total of 2935 eligible children agreed to participate in the study. Through extended efforts, 76% of these children and their parents consented to participate (T1: $n = 2230$, mean age $= 11.1 \pm 0.6$ years, 50.8% girls). Response rates at the first two follow-ups are 96.4% (T2: $n = 2149$, mean age $= 13.6 \pm 0.5$, 51.0% girls) and 81.4% (T3: $n = 1816$, mean age $= 16.3 \pm 0.7$, 52.3% girls).

We investigated a gender specific model in the mean and correlation of the latent variable underneath the ordinal depression scale. As mentioned before, different correlations would indicate that differential stability of depression symptoms are gender specific. In total 71 individuals were omitted from the analysis due to the missingness in internalizing and/or externalizing behavior of parents. Table 6 gives the estimates of the regression parameters and correlation coefficients together with their 95% confidence intervals of our full model.

These results cannot demonstrate dissimilarity between the correlations for boys and girls. This was shown by the likelihood ratio test ($\mathrm{LRT} = 6.48; \mathrm{df} = 3; \mathrm{P} = 0.090$). We also investigated whether boys and girls would have the same mean model with respect to the selected variables. The likelihood ratio test rejected this hypothesis ($\mathrm{LRT} = 40.6; \mathrm{df} = 6; \mathrm{P} < 0.001$), indicating gender specific models on the regression parameters for depression status. Results in Table 6 shows that internalizing problems of parents influence significantly on girls. This finding is in line with Bouma et al. (2008).

5. Discussion

We utilized and extended the multivariate logistic distribution, introduced by O'Brien and Dunson (2004), for analysis of longitudinal ordinal data. This resulted in a population-averaged interpretation for the odds ratios of the regression parameters. We applied maximum likelihood instead of a Bayesian approach and we provided the opportunity to include covariates in the time-related correlations. We simulated different distributions for the latent variables underneath the ordinal data via different copula functions to assess the performance of our approach in terms of bias, MSE, and coverage probabilities for Wald-type confidence intervals; and further compared them with the results of GEE. The main focus of this article was on small sample sizes (N = 50).

The results demonstrated ignorable bias, small MSEs and nominal coverage probabilities when the multivariate distribution of the latent variables has an elliptical shape, even though the frequencies of ordinal outcomes were skewed. For the exchangeable Gumbel copula with an Archimedean shape, a bias in the regression parameters was obtained when an exchangeable correlation matrix was assumed in our approach.

Table 6
Estimation and confidence intervals for the TRAILS data set.

	boys	girls
Regression Parameter		
Intercept	0.506 [0.297; 0.715]	-
Age	−0.038 [−0.096; 0.021]	−0.068 [−0.131; −0.006]
Externalizing	0.023 [−0.287; 0.333]	−0.085 [−0.372; 0.202]
Internalizing	0.084 [−0.062; 0.230]	0.230 [0.094; 0.366]
Social-economic status	−0.004 [−0.068; 0.059]	0.045 [−0.008; 0.098]
Structure of family	0.038 [−0.261; 0.337]	0.133 [−0.153; 0.420]
Follow-up time	−0.134 [−0.168; −0.100]	0.019 [0.002; 0.037]
Correlation coefficient		
ρ_{12}	0.486 [0.417; 0.550]	0.522 [0.452; 0.585]
ρ_{13}	0.393 [0.302; 0.478]	0.347 [0.252; 0.435]
ρ_{23}	0.564 [0.448; 0.599]	0.488 [0.403; 0.564]

Using the unstructured correlation matrix, biases in the estimates seem to become smaller and disappeared almost fully. Unfortunately, coverage probabilities for the Wald-type confidence intervals on the coefficients in the correlation matrix were liberal, and in some settings substantially lower than nominal. On the other hand, the coverage probabilities for the regression parameters of the mean model were close to nominal, and the likelihood ratio test was able to detect different correlation structures with reasonable power (considering the effect and sample sizes). Furthermore, the type I error rate for the likelihood ratio test was at the significance level. In all our simulations, ML either performed better than GEE or it performed similar.

The advantage of our approach is that it provides a similar interpretation as GEE with an option to model the temporal correlations with currently available software. Furthermore, for incomplete outcome data that fulfills the MAR assumption and for sample sizes that are not too small (say at least 100 individuals), our approach can be simply applied without having to use additional analysis such as multiple imputation. It should be noted that when covariates are incomplete, maximum likelihood methods alone is not satisfactory since subjects with missing covariates are ignored in the analysis. We believe that our approach is also more appropriate than the method of Li and Schafer (2008). They considered the multivariate normal distribution for the latent variable, but this choice does not support an odds ratio interpretation of the parameter estimates. Furthermore, fitting a normally distributed latent variable may lead to bias when the distribution of the latent variables have a heavier tail. Although Tan et al (1999) demonstrated robustness with respect to the normal distribution assumption for correlated binary data via a simulation study, we are uncertain that this conclusion remains true for ordinal outcomes. Molenberghs and Lesaffre (1994) extended the Dale model, which provides a joint distribution of the ordinal data, but our approach is simpler to apply and can be extended to larger dimensions.

A limitation of our approach is that it is sensitive to strong deviations from the multivariate t-distribution for the latent variables, in particular for estimation of the correlation coefficients. The Gumbel distribution seems to suggest this. Another limitation is that the standard errors of the estimators of the correlation coefficients are underestimated in some settings. A bootstrap approach may overcome this, but bootstrapping on correlated data is not straightforward. Another possible limitation is the difficulty in obtaining (semi-)positive definite correlation matrices for the underlying latent variables when the number of repeated measures is substantial. Although good approaches have been developed and implemented in R software, we did not explore if high dimensions lead to numerical issues. An alternative to these numerical approaches is the reparameterization of the correlation matrix into partial autocorrelations to deal with the possible issues of (semi-)positive definiteness (Daniels and Pourahmadi, 2009). Additionally, we did not examine any other issues of the proposed analysis for higher dimensions. We know from literature that the selected estimation method of the multivariate t-distribution is stable and reliable up to 20 dimensions (Genz and Bretz, 2002), but future studies are still needed.

To summarize, we used multivariate logistic distributed latent variables to introduce a joint logistic distribution for longitudinal ordinal outcomes. The dependency between ordinal outcomes can be measured via Pearson correlation coefficients and we could model temporal correlations for different subpopulations. The association of ordinal outcomes with covariates can be modeled with parameters having a population-averaged odds ratio interpretation. Simulation studies demonstrate superiority of this approach over GEE in terms of bias, MSE, and coverage probabilities for several settings. Finally, implementation of this approach can be performed with existing packages in R due to an approximation (see Web Appendix A).

6. Supplementary Materials

Web Appendix A referenced in Sections 2.3 and 5 is available with this article at the *Biometrics* website on Wiley Online Library.

References

Achenbach, T. M. and Rescorla, L. A. (2006). The Achenbach System of Empirically Based Assessment. *ed. Archer, R.P.*

Lawrence Erlbaum Associates Publishers. Mahwah: New Jersey, USA. 229–262.

Agresti, A. and Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review/ Revue Internationale de Statistique* **69**, 345–371.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Bouma, E. M. C., Ormel, J., Verhulst, F. C., and Oldehinkel, A. J. (2008). Stressful life events and depressive problems in early adolescent boys and girls: The influence of parental depression, temperament and family environment. *Journal of Affective Disorders* **105**, 185–193.

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

Budden, M., Hadavas, P., Hoffman, L., and Pretz, C. (2007). Generating Valid 4×4 Correlation Matrices. *Applied Mathematics E-notes* **7**, 53–59.

Clayton, D. (1992). Repeated ordinal measurements: A generalised estimating equation approach. *Medical Research Council Biostatistics Unit Technical Reports* Cambridge, England.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society B* **66**, 851–860.

Cornish, E. A. (1954). The Multivariate t-distribution associated with a set of normal sample deviates. *Australian Journal of Physics* **7**, 531–542.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42**, 909–917.

Daniels, M. J. and Pourahmadi, M. (2009). Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis* **100**, 2352–2363.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal Data Analysis.* New York: Chapman & Hall/CRC Press.

Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* **63**, 361–378.

Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* **11**, 950–971.

Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association* **56**, 335–349.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models.* Volume 1. New York: McGrawHill.

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 Zeger, S. C. (1996). Marginal regression models for clustered ordinal measurements. *Journal of American Statistical Association* **19**, 1024–1036.

Higham, N. J. (1961). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications* **103**, 103–118.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling.* 3nd edition. New York: The Guilford Press.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Models and Applications.* 2nd edition. New York: John Wiley & Sons.

Lee, K. and Daniels, M. J. (2008). Marginalized models for longitudinal ordinal data with application to quality of life studies. *Statistics in Medicine* **27**, 4359–4380.

Li, Y. and Schafer, D. W. (2008). Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. *Computational Statistics & Data Analysis* **52**, 3474–3492.

Malik, H. J. and Abraham, B. (1973). Multivariate logistic distributions. *The Annals of Statistics* **1**, 588–590.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society B* **42**, 109–142.

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.

Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies.* England: John Wiley & Sons.

Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics* **22**, 43–65.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika* **49**, 115–132.

Nelson, R. B. (2006). *An introduction to copula.* 2nd edition, Portland: Springer-Verlag.

Nooraee, N., Molenberghs, G., and van den Heuvel, E. R. (2014). GEE for longitudinal ordinal data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Computational Statistics & Data Analysis* **77**, 70–83.

O'Brien, S. M. and Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics* **60**, 739–746.

Ormel, J., Oldehinkel, A. J., Sijtsema, J., van Oort, F., Raven, D., Veenstra, R., et al. (2012). The TRacking Adolescents' Individual Lives Survey (TRAILS): Design, Current Status, and Selected Findings. *Journal of the American Academy of Child and Adolescent Psychiatry* **51**, 1020–1036.

Olkin, I. (1981). Range restrictions for product-moment correlation matrices. *Psychometrika* **46**, 469–472.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G. M., Mallick, B. K., et al. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *The Annals of Applied Statistics* **5**, 449–467.

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

Rubin, D. B. (1976). Inference and missing data (with discussion). *Biometrika* **63**, 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* NewYork: John Wiley & Sons.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semi-parametric nonresponse models (with comments). *American Statistical Association* **94**, 1096–1146.

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation* **24**, 647–656.

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

Tan, M., Qu, Y., and Suni Rao, J. (1999). Robustness of the Latent Variable Model for Correlated Binary Data. *Biometrics* **55**, 1541–0420.

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.

Received May 2014. *Revised July* 2015. *Accepted August* 2015.