GEE for Multinomial Responses Using a Local Odds Ratios Parameterization

Anestis Touloumis,[1],* Alan Agresti,[2], Maria Kateri[3],***

[1]EMBL-European Bioinformatics Institute, Hinxton, U.K.
[2]Department of Statistics, University of Florida, Gainesville, Florida, U.S.A
[3]Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany
*email: anestis@ebi.ac.uk
**email: aa@stat.ufl.edu
***email: maria.kateri@rwth-aachen.de

Summary. In this article, we propose a generalized estimating equations (GEE) approach for correlated ordinal or nominal multinomial responses using a local odds ratios parameterization. Our motivation lies upon observing that: (i) modeling the dependence between correlated multinomial responses via the local odds ratios is meaningful both for ordinal and nominal response scales and (ii) ordinary GEE methods might not ensure the joint existence of the estimates of the marginal regression parameters and of the dependence structure. To avoid (ii), we treat the so-called "working" association vector α as a "nuisance" parameter vector that defines the local odds ratios structure at the marginalized contingency tables after tabulating the responses without a covariate adjustment at each time pair. To estimate α and simultaneously approximate adequately possible underlying dependence structures, we employ the family of association models proposed by Goodman. In simulations, the parameter estimators with the proposed GEE method for a marginal cumulative probit model appear to be less biased and more efficient than those with the independence "working" model, especially for studies having time-varying covariates and strong correlation.

Key words: Association models; Generalized estimating equations; Local odds ratios; Longitudinal data analysis; Multinomial responses.

1. Introduction

Liang and Zeger (1986) originally proposed the generalized estimating equations (GEE) method as an extension of generalized linear models to handle longitudinal data. In contrast to ordinary maximum likelihood approaches, the GEE method provides consistent estimators of the marginal regression parameter vector β and of the covariance matrix of those estimates even if α, the parameter vector that describes the correlation/association pattern within the subjects, has been misspecified.

Application of the GEE method for correlated multinomial responses with at least three response categories has been in need of further development. One reason relates to difficult issues in parameterizing the association structure in a way that is sensible for categorical response variables, and in particular, is suitable for both nominal and ordinal variables. In the relevant literature, a correlation coefficient (Lipsitz, Kim, and Zhao, 1994; Miller, Davis, and Landis, 1993; Parsons, Edmondson, and Gilmour, 2006) and a global odds ratios (Williamson, Kim, and Lipsitz, 1995; Heagerty and Zeger, 1996; Lumley, 1996) parameterization for α have been proposed. The correlation coefficient parameterization is severely restricted by the marginal model even for bivariate multinomial responses, as we show in Section 2, while the use of a global odds ratios parameterization is limited to ordinal responses. To this end, note that the use of the GEE approach of Parsons et al. (2006) is restricted to ordinal responses under a marginal cumulative logistic model. Another difficult issue is

that mere modeling of the marginal probabilities and the pairwise association pattern might not lead to a proper joint distribution for the correlated multinomial responses (Bergsma and Rudas, 2002). Although existing GEE approaches avoid specification of this distribution by adopting a "working" assumption for α, it is implicit that such a distribution exists with respect to the marginal model specification and the "working" assumption made for the association pattern. In Section 2, we argue that this need not be true, and hence the asymptotic properties of the GEE estimator of β might be affected.

A simple way to deal with these issues is to recognize α as a "nuisance" vector that minimizes/maximizes an objective function (Crowder, 1995) such that larger values of α indicate stronger underlying association patterns (Chaganty and Joe, 2004) regardless of the response scale or the marginal model. The aforementioned problems of the GEE method are avoided because α is not formally defined as an association vector, while a maximization process ensures the existence of α. Also, it is desirable to express α using parameters that measure the association between multinomial responses but are less restrictive than those used in existing GEE approaches. Based on these considerations, we propose a unified GEE approach for correlated nominal or ordinal multinomial responses using a local odds ratios parameterization for α. The local odds ratios have been widely used to summarize association in contingency tables but have never been employed in the GEE context.

In particular, we identify $\boldsymbol{\alpha}$ as the marginalized local odds ratios after cross-classifying the responses at each time pair, with the degree of association in these contingency tables providing information about the true association pattern. We model the association in the marginalized tables using association models (see Goodman, 1985), which enable us to reduce the dimension of $\boldsymbol{\alpha}$ by imposing meaningful local odds ratio structures both for ordinal and nominal response categories. The association models employed are special cases of Goodman's row and column effects (RC) model with homogeneous score parameters. For a pair of responses, consider an $I \times I$ table in which cell (j, j') presents the probability of response outcome at row j and at column j'. Denote by $\theta_{jj'}$ the local odds ratio at the cutpoint (j, j'). The homogeneous RC model assumes that

$$\log \theta_{jj'} = \phi(\mu_j - \mu_{j+1})(\mu_{j'} - \mu_{j'+1}) \tag{1}$$

for $j, j' = 1, \ldots, I - 1$. After imposing identifiability constraints, relation (1) decomposes the local dependency into two parts: the scores $\{\mu_j : j = 1, \ldots, I\}$ for the I response categories, and the intrinsic parameter ϕ that describes the strength of association. For ordinal responses, a fixed unit-spaced scores assignment is often sensible, which implies that the association is characterized in terms of a single parameter (ϕ) which determines a uniform value for all local odds ratios equal to e^ϕ. For nominal response categories, $I - 2$ non-redundant scores are treated as parameters and are estimated, but this still gives a substantial reduction over the $(I - 1)^2$ unrestricted local odds ratios. The parametric scores need not be monotone, providing the flexibility required to handle nominal responses.

An attractive feature of the association models is their ability to capture underlying correlation patterns of many discretized continuous variables in two-way contingency tables. For example, Goodman (1979), Becker (1989), and Wang (1987) showed that the correlation parameter of a bivariate normal distribution can be approximated closely using a scaling of the intrinsic parameter ϕ. Anderson and Vermunt (2000) provided latent continuous variable interpretations for the RC model, with more than two correlated latent variables for both ordered and nominal response categories. Moreover, association models can approximate strong correlation patterns between correlated multinomial responses regardless of the response scale, which makes them appealing for use in the GEE approach.

This article is structured as follows. In Section 2 we introduce the notation and the basic features of the GEE method for correlated multinomial responses, assuming the data arise from a longitudinal study with possibly missing observations. Further, we show that modeling the association structure between correlated multinomial responses via the local odds ratios is more flexible than existing parameterizations (with related material at the journal website). In Section 3, we give details of the proposed GEE approach, focusing on the estimation procedure for $\boldsymbol{\alpha}$ and potential marginalized local odds ratios structures. Section 4 summarizes a simulation study that investigates the performance of the proposed approach. To illustrate our approach, we analyze a rheumatoid arthritis clinical trial in Section 5. Conclusions and proposals for future research are summarized in Section 6.

2. Notation and Basic Structure of GEE Approach

Let $Y_{it} \in \{1, 2, \ldots, I > 2\}$ be the multinomial response for subject i $(i = 1, \ldots, N)$ at time t $(t = 1, \ldots, T_i)$, and suppose that data are missing completely at random (MCAR) as defined in Rubin (1976). Define $Y_{itj} = \mathrm{I}(Y_{it} = j)$ for $j = 1, \ldots, I$, where $\mathrm{I}(A)$ denotes the indicator function of the event A. We convert Y_{it} into the equivalent $(I - 1)$-variate vector $\mathbf{Y}_{it} = (Y_{it1}, \ldots, Y_{it(I-1)})'$ with the response category I omitted since $\sum_{j=1}^{I} Y_{itj} = 1$, and let $\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \ldots, \mathbf{Y}'_{iT_i})'$. We let $\mathbf{x}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT_i})'$ denote the $T_i(I-1) \times p$ matrix of explanatory variable values for subject i, with the $(I-1) \times p$ matrix $\mathbf{x}_{it}$ for time t including intercepts, time-stationary, time-varying, and category-specific covariates. Let $\pi_{itj} = E(Y_{itj}|\mathbf{x}_i) = \Pr(\mathrm{Y}_{itj} = 1|\mathbf{x}_i)$, and let $\boldsymbol{\pi}_{it} = (\pi_{it1}, \ldots, \pi_{it(I-1)})'$ and $\boldsymbol{\pi}_i = E(\mathbf{Y}_i|\mathbf{x}_i) = (\boldsymbol{\pi}'_{i1}, \ldots, \boldsymbol{\pi}'_{iT_i})'$.

Denote by $\mathbf{g} : (0, 1)^{I-1} \to \Re^{I-1}$ an $(I-1)$-variate vector of link functions and suppose that a multinomial generalized linear model (Fahrmeir and Tutz, 2001) is used to model the marginal expected vector for subject i at time t

$$\mathbf{g}[E(\mathbf{Y}_{it}|\mathbf{x}_i)] = \mathbf{g}(\boldsymbol{\pi}_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{\beta}$ is the p-variate regression vector of interest. The choice of the link vector $\mathbf{g}$ should reflect the response scale. Options include the family of cumulative link models or the adjacent-categories logit model for ordinal response categories, and the baseline-category logit model for nominal response categories.

2.1. Overview of the GEE Approach

The GEE estimator $\widehat{\boldsymbol{\beta}}_G$ is the solution of the estimating equations

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\pi}_i) = \mathbf{0}$$

where $\mathbf{D}_i = \partial \boldsymbol{\pi}_i / \partial \boldsymbol{\beta}$ and where $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is a $T_i(I-1) \times T_i(I-1)$ "weight" matrix that is usually regarded as a "working" covariance matrix. For unknown $\boldsymbol{\alpha}$, Liang and Zeger (1986) defined $\widehat{\boldsymbol{\beta}}_G$ as the solution of $\mathbf{U}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) = \mathbf{0}$, where $\widehat{\boldsymbol{\alpha}}$ is a $\sqrt{N}$-consistent estimator of $\boldsymbol{\alpha}$ given $\boldsymbol{\beta}$. Under mild regularity conditions and correct specification of the marginal model in (2), Liang and Zeger (1986) proved that $\widehat{\boldsymbol{\beta}}_G$ is consistent and $\sqrt{N}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ converges in distribution to a p-variate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\boldsymbol{\Sigma}_G = \lim_{N \to \infty} N \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1} \tag{3}$$

where $\boldsymbol{\Sigma}_0 = \sum_i \mathbf{D}'_i \mathbf{V}_i^{-1} \mathbf{D}_i$, $\boldsymbol{\Sigma}_1 = \sum_i \mathbf{D}'_i \mathbf{V}_i^{-1} \mathrm{cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i$, and $\mathrm{cov}(\mathbf{Y}_i) = \mathrm{cov}(\mathbf{Y}_i|\mathbf{x}_i)$ is the $T_i(I-1) \times T_i(I-1)$ true covariance matrix for subject i. The elements of $\mathrm{cov}(\mathbf{Y}_i)$ are

$$\mathrm{cov}(Y_{itj}, Y_{it'j'}|\mathbf{x}_i) = \begin{cases} \pi_{itj}(1 - \pi_{itj}) & \text{if } t = t' \text{ and } j = j' \\ -\pi_{itj}\pi_{itj'} & \text{if } t = t' \text{ and } j \neq j' \\ \pi_{itjt'j'} - \pi_{itj}\pi_{it'j'} & \text{if otherwise} \end{cases}$$

where $\pi_{itjt'j'} = E(Y_{itj}Y_{it'j'}|\mathbf{x}_i) = \Pr(Y_{itj} = Y_{it'j'} = 1|\mathbf{x}_i) = \Pr(Y_{it} = j, Y_{it'} = j'|\mathbf{x}_i) > 0$. The "robust" covariance matrix $\mathbf{\Sigma}_G$ can be consistently estimated by ignoring the limit and replacing $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ with $(\widehat{\boldsymbol{\beta}}_G, \widehat{\boldsymbol{\alpha}})$ and $\mathrm{cov}(\mathbf{Y}_i)$ with $(\mathbf{Y}_i - \widehat{\boldsymbol{\pi}}_i)(\mathbf{Y}_i - \widehat{\boldsymbol{\pi}}_i)'$.

2.2. *Association Structures for Multinomial Responses and their Limitations*

To solve $\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{0}$, the parameter vector $\boldsymbol{\alpha}$ must be defined, ideally in a way that it defines $\mathbf{V}_i$ as a "working" covariance matrix. If the "working" assumption for $\mathbf{V}_i$ is close to the true covariance matrix, efficiency gains in estimating $\widehat{\boldsymbol{\beta}}_G$ are likely to occur. The bivariate distribution for $(Y_{it}, Y_{it'})$ can be depicted in an $I \times I$ table of cell probabilities that satisfy the Fréchet bounds (Fréchet, 1940)

$$\max\{0, \pi_{itj} + \pi_{it'j'} - 1\} < \pi_{itjt'j'} < \min\{\pi_{itj}, \pi_{it'j'}\} \quad (4)$$

and they are consistent with the marginal model in (2)

$$\sum_{j=1}^{I} \pi_{itjt'j'} = \pi_{it'j'} \text{ and } \sum_{j'=1}^{I} \pi_{itjt'j'} = \pi_{itj} \quad (5)$$

for $j, j' = 1, \ldots, I$. It follows that $\boldsymbol{\alpha}$ is a "working" association vector if $\mathbf{V}_i$ preserves the structure of $\mathrm{cov}(\mathbf{Y}_i)$ and if there exists a distribution for $\mathbf{Y}_i$ subject to the marginal model specification and to the "working" assumption made for the distributions of $(Y_{it}, Y_{it'})$ for all i and $t < t'$. The necessary conditions (4) and (5) have not been stated in the GEE literature, but they are important if $\boldsymbol{\alpha}$ is defined and interpreted as a "working" association vector. For a given $(\widehat{\boldsymbol{\beta}}_G, \widehat{\boldsymbol{\alpha}})$, violation of (4) questions the required correct specification of the marginal model while violation of (5) suggests that the "working" distribution for $(Y_{it}, Y_{it'})$ is ill-defined.

For the possible parameterization of $\boldsymbol{\alpha}$ for bivariate multinomial responses in GEE methods, we next compare the feasible values for the correlation (Lipsitz et al., 1994; Miller et al., 1993; Parsons et al., 2006), the global odds ratio (Williamson et al., 1995; Heagerty and Zeger, 1996; Lumley, 1996) and the local odds ratio. Technical details for our comparison can be found in the Web Appendix A. The local odds ratios parameterization has two important advantages over the competing ones. Unlike a uniform global odds ratios structure, the local odds ratios parameterization is applicable to both ordinal and nominal response scale. Unlike the correlation parameterization or an unstructured global odds ratios parameterization, the local odds ratios parameterization is not restricted by the marginal model specification for $T_i = 2$, and hence it permits parsimonious, flexible, and efficient "working" forms for $\boldsymbol{\alpha}$ that may depend on the category cutpoints and/or the covariates. Given the marginal probabilities and the local odds ratios, the probability distribution can easily be obtained using the iterative proportional fitting (IPF) procedure proposed by Deming and Stephan (1940). For $T_i \geq 3$, the Web Appendix explain why the set of admissible values for $\widehat{\boldsymbol{\alpha}}$ depends on the unknown marginal regression vector $\boldsymbol{\beta}$ for all parameterizations of $\boldsymbol{\alpha}$ considered herein and the possible implications regarding the asymptotic properties of the GEE estimator of $\boldsymbol{\beta}$. From a practical point of view, the above suggest that a local odds ratios parameterization of $\boldsymbol{\alpha}$ is less

restrictive than the other two parameterizations and it ensures valid "working" marginal pairwise distributions of $(Y_{it}, Y_{it'})$ for all $t < t'$, with respect to the marginal model specification in (5).

3. A Local Odds Ratios GEE Approach

To circumvent theoretical problems, we propose the following two-stage GEE approach. In the first stage we define and estimate $\boldsymbol{\alpha}$ by maximizing a pseudo likelihood function. Let $L = T(T - 1)/2$ be the number of time pairs, where $T = \max\{T_1, \ldots, T_N\}$. At each time pair, we aggregate the available response variables across subjects with no covariates adjustment to form L sets of $I \times I$ marginalized contingency tables. Set

$$\boldsymbol{\alpha} = \left(\theta_{1121}, \ldots, \theta_{112(I-1)}, \ldots, \theta_{(T-1)1T1}, \ldots, \theta_{(T-1)(I-1)T(I-1)}\right)'$$

where $\theta_{tjt'j'}$ is the local odds ratio at the cutpoint (j, j') at the marginalized contingency table corresponding to the time pair (t, t') for $t < t'$. Now, $\boldsymbol{\alpha}$ describes the marginalized local odds ratios structure. We treat the L time pairs as levels of a group variable and fit a generalized version of the *RC* model (Becker and Clogg, 1989) simultaneously to the L marginalized contingency tables. The local odds ratios under the generalized *RC* model satisfy

$$\log \theta_{tjt'j'} = \phi_{tt'}(\mu_{tj}^{tt'} - \mu_{t(j+1)}^{tt'})(\mu_{t'j'}^{tt'} - \mu_{t'(j'+1)}^{tt'}). \quad (6)$$

For an ordinal response scale, as a default choice we recommend assigning fixed unit-spaced values to the score parameters (e.g., $\mu_{tj}^{tt'} = \mu_{t'j}^{tt'} = j$ for all $t < t'$ and j), in which case the model reduces to the heterogeneous linear-by-linear association model (Agresti, 2013); further simplicity can be obtained by taking $\phi_{tt'} = \phi$ for each pair. For a nominal response scale, the score parameters must be estimated and additional constraints are needed. For simplicity, as a default choice we recommend using homogeneous score parameters across the cross-classification variables ($\mu_{tj}^{tt'} = \mu_j$) and employing the constraints

$$\sum_{j=1}^{I} \mu_j = 0 \text{ and } \sum_{j=1}^{I} \mu_j^2 = 1.$$

Since the log local odds ratios in (6) do not depend on the marginal model specification in (2), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are variation independent.

To estimate $\boldsymbol{\alpha}$, we treat the L marginalized contingency tables as independent. Since we are merely interested in estimating the marginalized local odds ratios and not their standard errors, treating the tables as independent even though they are clearly not so facilitates the estimating procedure without affecting the consistency of the estimates. We utilize maximum likelihood methods, assuming independent Poisson sampling for the observed counts. If zero observed counts occur in such a pattern to cause non-existence of estimates, then one can add a very small positive constant to each cell of the L marginalized contingency tables. The maximization of the Poisson likelihood function ensures the existence of $\widehat{\boldsymbol{\alpha}}$ and

standard asymptotic arguments combined with the continuity mapping theorem imply the $\sqrt{N}$-consistency of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}$.

At the second stage, we solve the estimating equations $\mathbf{U}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) = \mathbf{0}$ to estimate $\boldsymbol{\beta}$. Since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are variation independent and $\widehat{\boldsymbol{\alpha}}$ is a $\sqrt{N}$-consistent estimator of $\boldsymbol{\alpha}$, the estimate of $\boldsymbol{\beta}$ maintains the asymptotic properties presented in Section 2. The IPF procedure makes feasible the calculation of $\mathbf{V}_i$ using $\{\widehat{\theta}_{tjt'j'} : t < t' \text{ and } j, j' = 1, \ldots, I-1\}$ as the common local odds ratios structure across subjects.

This approach allows us to borrow "strength from the strata" and estimate the dependency between the marginalized responses at each time point. If a similar underlying correlation pattern holds at a subject level, then $\mathbf{V}_i \approx \text{cov}(\mathbf{Y}_i)$ for all i and the efficiency of the GEE estimator $\widehat{\boldsymbol{\beta}}_G$ should increase. Despite this, we emphasize that $\boldsymbol{\alpha}$ is a "nuisance" vector which enables the "weight" matrix $\mathbf{V}_i$ to preserve the structure of a potential "working" covariance matrix.

3.1. *Local Odds Ratio Structures*

Since the L marginalized tables are square, it would usually be sensible to increase parsimony by using homogeneous score parameters for the response categories at each time pair. Further simplifications can be specified in terms of the $\log \theta_{tjt'j'}$ used to specify the association model:

(i) The uniform structure, $\log \theta_{tjt'j'} = \phi$, can be obtained by fitting the homogeneous linear-by-linear association model (Agresti, 2013). This structure assumes exchangeability of time pairs and of adjacent-category pairs.

(ii) The category exchangeability structure, $\log \theta_{tjt'j'} = \phi_{tt'}$, can be obtained by fitting the heterogeneous linear-by-linear association model. This structure assumes a common local odds ratio at each time pair, but permits different pairs to have differing associations.

(iii) The time exchangeability structure, $\log \theta_{tjt'j'} = \phi(\mu_j - \mu_{j+1})(\mu_{j'} - \mu_{j'+1})$, is obtained by setting $\phi_{tt'} = \phi$ and $\mu_{tj}^{tt'} = \mu_{t'j}^{tt'} = \mu_j$ for all j and $t < t'$ but treats the category scores as unknown parameters rather than fixed, equally spaced values. This structure does not assume any time dependency and it implies equal local odds ratios at the category cutpoints (j, j') and (j', j) when $j \neq j'$.

(iv) The RC structure, $\log \theta_{tjt'j'} = \phi_{tt'}(\mu_j^{tt'} - \mu_{j+1}^{tt'})(\mu_{j'}^{tt'} - \mu_{j'+1}^{tt'})$, is an extension of the previous structure that additionally allows a time dependency. It is the special case of (6) in which the score parameters are the same for rows and columns for any given pair (t, t'). Having many more parameters than the other structures, we suggest using it only when the true association structure seems to be extremely complicated.

The time exchangeability and the RC structures are applicable to nominal and ordinal response scales, but the uniform and the category exchangeability structures are meaningful only for ordinal response categories.

The simulation findings in Section 4 suggest that parsimonious local odds ratios structures should be used when a nearly exchangeable underlying correlation pattern holds. Therefore we recommend the following guidelines: For ordinal response categories, apply the heterogeneity model to check the range of the estimated intrinsic parameters. If their range is small, say less than about two units, then regard the association pattern as nearly exchangeable and adopt the uniform structure. Likewise, for nominal response categories, a comparison of the L estimated intrinsic parameters is helpful to select between the time exchangeability and the more general RC structure. However, we recommend using the time exchangeable structure with small sample sizes, because convergence problems in the simulations did not allow us to explore confidently the performance of the more general RC structure.

3.2. *Software Implementation of GEE Local Odds Ratio Approach*

An R language software package *multgee* has been developed to implement the proposed GEE approach. It includes two routines for fitting the marginal models discussed in Section 2. The routine *ordLORgee* is designed for ordinal response categories with category excheangeability as the default local odds ratio structure, while the routine *nomLORgee* is for nominal response categories using the time exchangeability structure as default. More details about the package *multgee* can be found in Touloumis (2012).

4. Simulation Study

We conducted a simulation study to investigate the behavior of the proposed GEE approach under misspecification of the association structure. The marginal probabilities were specified by the cumulative probit model

$$\Phi^{-1}\left[\Pr(Y_{it} \leq j|\mathbf{x}_i)\right] = \beta_{0j} + \beta x_{it} \tag{7}$$

for $i = 1, \ldots, N$, $t = 1, \ldots, 4$, $j = 1, \ldots, 4$, where Φ is the cumulative distribution function of the standard normal distribution, $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$, $\boldsymbol{\beta} = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta)' = (-1.5, -0.5, 0.5, 1.5, 1)'$ and $N = 100, 500$. As in Wang and Carey (2003), a variety of distributional scenarios were used for the covariates:

Scenario 1: A time-stationary design with $x_{it} = x_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$ for all t.

Scenario 2: A time-varying design, in which $\mathbf{x}_i$ are i.i.d. from a tetra-variate normal distribution with mean vector $\mathbf{0}$, unit variances and a 4×4 correlation matrix with off-diagonal elements equal to 0.8.

Scenario 3: Scenario 1 with a monotone missing pattern, in which each subject had probability 0.25 of dropping out permanently after the first occasion.

Scenario 4: Scenario 2 with the missingness mechanism described in Scenario 3.

Conditional on $\mathbf{x}_i$ and $\boldsymbol{\beta}$, Y_{it} were determined by

$$Y_{it} = j \Leftrightarrow \beta_{0(j-1)} < \epsilon_{it} - \beta x_{it} \leq \beta_{0j} \tag{8}$$

for $j = 1, \ldots, 5$, $\beta_{00} = -\infty$, $\beta_{05} = \infty$ and i.i.d. latent vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4})'$ from a tetra-variate normal distribution

with mean vector $\mathbf{0}$, unit variances, and correlation matrix $\mathbf{R}_\epsilon$. For the latent correlation matrix $\mathbf{R}_\epsilon$ we considered exchangeable correlation forms, in which the off-diagonal elements were equal to ρ with $\rho = 0, 0.15, 0.50$ and 0.85, and also the correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.85 & 0.50 & 0.15 \\ 0.85 & 1.00 & 0.85 & 0.50 \\ 0.50 & 0.85 & 1.00 & 0.85 \\ 0.15 & 0.50 & 0.85 & 1.00 \end{bmatrix}.$$

We focused on the performance of the GEE estimators of β under the independence $(\hat{\beta}_I)$, the uniform $(\hat{\beta}_U)$, the category exchangeability $(\hat{\beta}_{CE})$, the time exchangeability $(\hat{\beta}_{TE})$ and the RC $(\hat{\beta}_{RC})$ structure. We originally planned to compare our GEE approach with existing ones, specifically the approach of Lipsitz et al. (1994) and that of Heagerty and Zeger (1996), as implemented by the routine *ordgee* in the R package *geepack* (Højsgaard, Halekoh, and Yan, 2006). Unfortunately, low convergence rates in every sampling scheme ($\leq 38\%$ for the GEE approach of Lipsitz et al. (1994) and $\leq 60\%$ for that of Heagerty and Zeger (1996)) did not permit a fair comparison with the proposed GEE approach. So, we focused only on comparing the various association model structures in our formulation, especially relative to the independence structure that is commonly used in practice (e.g., it is the only option with the SAS PROC GENMOD routine for multinomial GEE).

In each of the 40 sampling schemes we drew $10,000$ replications. For every estimating procedure in a sampling scheme we calculated the simulated bias and standard error of the GEE estimator of β, the percentage of times that the 95% Wald confidence interval for β using the estimated "robust" covariance matrix in (3) contained the true parameter value and the simulated relative efficiency (SRE) criterion treating $\hat{\beta}_I$ as a baseline estimator. The SRE criterion was defined as the ratio of the simulated mean squared errors of $\hat{\beta}_I$ to $\hat{\beta}$ measured over replications in which both estimators were obtained. Simulation results in sampling schemes associated with a latent exchangeable correlation pattern are available on Web Tables 1–4 and with the latent correlation matrix $\mathbf{R}$ in Table 1.

For our proposed GEE approach, we noticed that simple "working" local odds ratios structures exhibited fewer convergence problems, similar to those reported in Miller et al. (1993) for multinomial responses or in Lipsitz, Laird, and Harrington (1991) for binary responses, and clearly less than those of alternative multinomial GEE approaches. Convergence problems were encountered for the RC structure when $N = 100$, because the estimated local odds ratios were often close to the boundary values due to sparse marginalized contingency tables. For this reason, the performance of $\hat{\beta}_{RC}$ was evaluated only when $N = 500$, and we recommend use of this general structure only for large sample sizes.

For a particular latent correlation pattern, similar trends occurred for the SRE criterion of each estimator. In particular, SRE increased in time-varying designs compared to time-stationary designs, and it changed little as N increased. The results suggest that the proposed GEE method tends to

be strikingly more efficient than the independence "working" model in designs with time-varying covariates, regardless of whether data are missing. Under a latent exchangeable correlation pattern, all the proposed GEE estimators were nearly equally efficient. Conditional on the covariates configuration, the Web Tables show that their efficiency appeared to be a strongly increasing function of ρ. The estimator $\hat{\beta}_I$ was adequate in Scenario 1 or if $\rho = 0.15$. However, the proposed GEE estimators offered substantial efficiency gains as ρ increased, for example up to 60% for $\rho = 0.85$. Under the latent correlation matrix $\mathbf{R}$ (for which results are summarized in Table 1), the choice of the "working" local odds ratios structure mattered. Comparing values for the SRE criterion, we see that the estimators $\hat{\beta}_{CE}$ and $\hat{\beta}_{RC}$ based on more complex local odds ratios structures performed slightly better than $\hat{\beta}_U$ and $\hat{\beta}_{TE}$, perhaps because the category exchangeability and the RC structure allow the local odds ratios to vary across time pairs and better approximate complicated association patterns such as the one implied by $\mathbf{R}$.

5. Data Analysis: Rheumatoid Arthritis Data

A randomized clinical trial (Lipsitz et al., 1994) was designed to evaluate the effectiveness of the drug auranofin versus the placebo therapy for the treatment of rheumatoid arthritis. The repeated response variable was the self-assessment of the rheumatoid arthritis classified on a five-level ordinal scale $(1 = \text{poor}, \ldots, 5 = \text{very good})$ and measured at three follow-up times (t), coded as (1) at one, (2) at three and (3) at 5 months. Dropout or intermittent missing responses occurred for 12 out of the 301 subjects, but Lipsitz et al. (1994) argued that MCAR was a plausible assumption for the missingness mechanism.

First we consider the association structure. The full marginalized contingency table had 11 out of 75 cells empty. An examination of the 5×5 marginalized contingency tables at each time pair indicates strong pairwise association, since the majority of observed frequencies fall on the diagonal, subdiagonal, and superdiagonal, with less than 11% falling in the remaining cells. The estimated intrinsic parameters using the heterogeneous linear-by-linear association model do not differ much, varying from 0.65 to 0.91, and thus the uniform local odds ratio structure seems a reasonable approximation for the association pattern. The local odds ratio estimate is 2.26 after fitting the homogeneous linear-by-linear association model.

Let γ_{itj} denote the cumulative probability of response level j $(j = 1, 2, 3, 4)$ at time t $(t = 1, 2, 3)$ for subject i $(i = 1, \ldots, 301)$, and let b_i and trt_i denote the subject-specific baseline self-assessment of rheumatoid arthritis and the binary treatment indicator for receiving the drug auranofin. To model the marginal probabilities, we used the cumulative logit marginal model

$$\text{logit}(\gamma_{itj}) = \beta_{0j} + \beta_1 I(t = 2) + \beta_2 I(t = 3) + \beta_3 trt_i + \beta_4 I(b_i = 2)$$
$$+ \beta_5 I(b_i = 3) + \beta_6 I(b_i = 4) + \beta_7 I(b_i = 5). \quad (9)$$

Table 2 contains the GEE estimates and the "robust" standard error estimates of the marginal regression vector $\boldsymbol{\beta}$ in (9) under the independence "working" model and under the uniform and the category exchangeability structures.

Table 1
Simulation results for the GEE estimators of $\beta = 1$ in $\Pr(Y_{it} \leq j|\mathbf{x}_i) = \Phi(\beta_{0j} + \beta x_{it})$ under the latent correlation matrix $\mathbf{R}$ as defined in Section 4.

			$N = 100$					$N = 500$		
Estimator	Bias	SE	SRE	EL	CR	Bias	SE	SRE	EL	CR
Scenario 1										
$\hat{\beta}_I$	0.0145	0.1087	1.00	93.95	100.00	0.0029	0.0472	1.00	95.09	100.00
$\hat{\beta}_U$	0.0140	0.1088	0.98	93.87	97.78	0.0027	0.0479	0.97	94.78	97.82
$\hat{\beta}_{CE}$	0.0092	0.1022	1.12	94.00	97.42	0.0018	0.0448	1.11	94.95	97.65
$\hat{\beta}_{TE}$	0.0044	0.1086	1.00	93.89	98.09	0.0010	0.0476	0.98	94.80	97.90
$\hat{\beta}_{RC}$						−0.0014	0.0447	1.12	94.93	97.20
Scenario 2										
$\hat{\beta}_I$	0.0153	0.1019	1.00	94.28	100.00	0.0033	0.0451	1.00	94.86	100.00
$\hat{\beta}_U$	0.0107	0.0891	1.30	94.48	98.54	0.0026	0.0397	1.28	95.28	98.88
$\hat{\beta}_{CE}$	0.0024	0.0813	1.59	93.92	98.91	0.0009	0.0359	1.58	95.32	98.84
$\hat{\beta}_{TE}$	0.0036	0.0895	1.31	94.22	98.58	0.0012	0.0397	1.29	95.15	98.71
$\hat{\beta}_{RC}$						−0.0045	0.0359	1.56	94.69	98.69
Scenario 3										
$\hat{\beta}_I$	0.0196	0.1270	1.00	93.89	100.00	0.0044	0.0551	1.00	94.92	100.00
$\hat{\beta}_U$	0.0171	0.1224	1.06	94.25	98.46	0.0037	0.0536	1.06	94.80	98.89
$\hat{\beta}_{CE}$	0.0104	0.1198	1.12	94.04	89.06	0.0024	0.0520	1.13	94.96	98.80
$\hat{\beta}_{TE}$	0.0082	0.1221	1.08	94.20	97.79	0.0020	0.0535	1.06	94.77	98.80
$\hat{\beta}_{RC}$						−0.0023	0.0522	1.11	94.72	94.45
Scenario 4										
$\hat{\beta}_I$	0.0183	0.1202	1.00	93.93	100.00	0.0038	0.0531	1.00	94.90	100.00
$\hat{\beta}_U$	0.0122	0.1058	1.28	94.29	98.82	0.0027	0.0472	1.26	94.60	99.27
$\hat{\beta}_{CE}$	-0.0035	0.1012	1.41	92.98	98.40	0.0002	0.0445	1.42	94.44	99.15
$\hat{\beta}_{TE}$	0.0031	0.1062	1.29	93.67	98.97	0.0010	0.0472	1.27	94.43	99.37
$\hat{\beta}_{RC}$						−0.0082	0.0447	1.36	93.16	98.80

The columns display the simulated bias (Bias), the simulated standard errors (SE), the simulated relative efficiency (SRE), the empirical level of the 95% Wald confidence intervals for β based on the robust covariance matrix (EL), and the convergence rate (CR) for each estimating procedure.

As expected with time-stationary covariate designs under a nearly exchangeable latent correlation pattern, the uniform and the category exchangeability structure provided similar GEE estimates and standard error estimates, while based on the estimated standard errors they offer some efficiency gains relative to the independence "working" model. In every estimating procedure, the effects of the 5 months follow-up indicator, of the drug auranofin, and of the indicators of the

Table 2
Results for the proposed GEE method with a cumulative logit marginal model, using local odds ratios to characterize association structure, for a rheumatoid arthritis clinical trial.

	Local odds ratios structure					
	Independence		Uniform		Category Exchangeability	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
β_{01}	−1.8500	0.4068	−1.8432	0.3893	−1.8400	0.3874
β_{02}	0.2308	0.3690	0.2669	0.3501	0.2771	0.3484
β_{03}	2.1889	0.3857	2.2313	0.3663	2.2478	0.3651
β_{04}	4.4899	0.4431	4.5254	0.4212	4.5482	0.4199
β_1	−0.0127	0.1211	0.0014	0.1218	−0.0008	0.1218
β_2	−0.3852	0.1160	−0.3617	0.1139	−0.3605	0.1141
β_3	−0.5643	0.1679	−0.5121	0.1680	−0.5046	0.1673
β_4	−0.6270	0.4027	−0.6696	0.3804	−0.7029	0.3786
β_5	−1.1861	0.3754	−1.2607	0.3525	−1.2756	0.3507
β_6	−2.5281	0.4375	−2.6437	0.4128	−2.6558	0.4104
β_7	−3.9533	0.5351	−3.9661	0.5316	−3.9956	0.5325

baseline response level 3 or higher are statistically significant at $\alpha = 0.05$. The indicator of the baseline response level 2 has p-value=0.08 under the uniform structure, 0.06 under the category exchangeability structure, and only 0.12 under the independence "working" structure.

We interpret the main effects estimates using the uniform local odds ratio structure. Subjects at the 5 months follow-up time have estimated odds of self-assessment of rheumatoid arthritis in response level k or better that are 1.43 times the corresponding odds at the 1-month follow-up. The drug auranofin improved the self-assessment during the trial, as subjects in the treatment group have estimated cumulative odds 1.67 times those of subjects in the placebo group. The magnitude of these estimates is similar to the ones in Lipsitz et al. (1994), where a three-level ordinal response scale was used. Finally, a one-level increase in the baseline response 1, 2, 3 and 4 increases the estimated cumulative odds multiplicatively by 1.95, 1.81, 3.99, and 3.75, respectively.

6. Summary and Discussion

We proposed a GEE approach based on a local odds ratios parameterization that can handle both ordinal and nominal correlated multinomial responses. Simulations show that this approach is considerably more efficient than the independence "working" model in time-varying covariate designs when the multinomial responses are strongly correlated. The intrinsic parameters in the association models describe the overall strength of pairwise associations, and we can use their estimates to evaluate whether the correlation pattern changes dramatically across time. For ordinal responses, we recommend using the uniform structure for nearly exchangeable correlation patterns across times and the category exchangeability structure for more complicated association patterns. For nominal responses, we recommend using the time exchangeability structure, with the more general time-dependent RC structure being possible with large samples when the simpler structure seems inadequate. Note that the use of the proposed GEE approach is not restricted to longitudinal studies but it can be applied whenever a marginal model is sensible for correlated multinomial responses, such as for a wide variety of forms of clustered data.

Although we developed the proposed GEE approach to account for multinomial responses, it can be applied with binary responses. In this case, the proposed GEE approach does not directly relate to ordinary GEE approaches that use an odds ratio parameterization because it identifies $\boldsymbol{\alpha}$ as a "nuisance" vector and not as a "working" association vector. Hence, we believe that even in the binary case our approach is more robust to GEE pitfalls pointed out by Crowder (1995) and Chaganty and Joe (2004, 2006) for such responses.

In future research, we hope to extend the proposed GEE method to handle unsynchronized longitudinal studies in which the data for each subject are collected at irregular time points. One could also investigate the estimation procedure for $\boldsymbol{\alpha}$ for sparse marginalized contingency tables, evaluate data-based strategies for selecting the working structure for the local odds ratios, and consider alternative estimators for the robust covariance matrix $\boldsymbol{\Sigma}_G$.

7. Supplementary Materials

Web Appendix A and Web Tables 1-4 referenced in Sections 2.2 and 4 are available with this article at the Biometrics website on Wiley Online Library.

References

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley and Sons, Inc., Hoboken, New Jersey, 3rd edition.

Anderson, C. and Vermunt, J. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology* **30**, 81–121.

Becker, M. (1989). On the bivariate normal distribution and association models for ordinal categorical data. *Statistics & Probability Letters* **8**, 435–440.

Becker, M. and Clogg, C. (1989). Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association* **84**, 142–151.

Bergsma, W. and Rudas, T. (2002). Marginal models for categorical data. *The Annals of Statistics* **30**, 140–159.

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

Chaganty, N. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93**, 197–206.

Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82**, 407–410.

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

Fahrmeir, L. and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer, New York, 2nd edition.

Fréchet, M. (1940). *Les probabilités associées à un système d'événements compatibles et dépendants*. Hermann & Cie, Paris.

Goodman, L. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association* **74**, 537–552.

Goodman, L. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics* **13**, 10–69.

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

Højsgaard, S., Halekoh, U., and Yan, J. (2006). The R package geepack for Generalized Estimating Equations. *Journal of Statistical Software* **15**, 1–11.

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

Lipsitz, S., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* **13**, 1149–1163.

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

Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on the working correlation structures. *Biometrics* **52**, 354–361.

Miller, M., Davis, C., and Landis, J. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics* **49**, 1033–1044.

Parsons, N., Edmondson, R., and Gilmour, S. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society C* **55**, 507–524.

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

Touloumis, A. (2012). *multgee: GEE Solver for Correlated Nominal or Ordinal Multinomial Responses.* R package version 1.1.

Wang, Y. (1987). The probability integrals of bivariate normal distributions: A contingency table approach. *Biometrika* **74**, 185–190.

Wang, Y. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* **90**, 29–41.

Williamson, J., Kim, K., and Lipsitz, S. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association* **90**, 1432–1437.

Web-based Supplementary Materials for 'GEE for Multinomial Responses Using a Local Odds Ratios Parameterization' by Touloumis, A., Agresti, A., and Kateri, M.

Web Appendix A: Comparison of Association Structures for Multinomial Responses

First, suppose that bivariate multinomial responses are collected at times $t < t'$ for all subjects. Dobra and Fienberg (2000) proved the tightness of the Fréchet bounds in a 2-way probability table with prescribed marginals. Thus $\mathbf{V}_i$ is a 'working' covariance matrix if and only if the parameter vector $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ can determine 'working' probabilities $\{\pi_{itjt'j'} : j, j' = 1, \ldots, I\}$ with respect to (4) and (5). We examine the range of the correlation coefficients (Lipsitz, Kim, and Zhao, 1994; Miller, Davis, and Landis, 1993; Parsons, Edmondson, and Gilmour, 2006), the global odds ratios (Williamson, Kim, and Lipsitz, 1995; Heagerty and Zeger, 1996; Lumley, 1996) and the local odds ratios parameterization for $\boldsymbol{\alpha}$.

The correlation coefficients parameterization expresses the association structure via the correlation coefficients between the binary indicators Y_{itj} and $Y_{it'j'}$

$$\rho_{itjt'j'} = \frac{\pi_{itjt'j'} - \pi_{itj}\pi_{it'j'}}{\sqrt{\pi_{itj}(1 - \pi_{itj})\pi_{it'j'}(1 - \pi_{it'j'})}} \tag{1}$$

for $j, j' = 1, \ldots, I$. Theorem 1 provides the necessary and sufficient constraints imposed by the marginal model specification on the $(I - 1)^2$ non-redundant correlation coefficients.

Theorem 1. *Let $(Y_{it}, Y_{it'})'$ be the bivariate random vector of multinomial responses for subject i at times t and t', with given marginal probabilities $\{\pi_{itj} : j = 1, \ldots, I\}$ and $\{\pi_{it'j'} : j' = 1, \ldots, I\}$ respectively. The correlation coefficients $\{\rho_{itjt'j'} : j, j' = 1, \ldots, I - 1\}$ define a distribution for $(Y_{it}, Y_{it'})'$ if and only if:*

$$LB(\pi_{itj}, \pi_{it'j'}) < \rho_{itjt'j'} < UB(\pi_{itj}, \pi_{it'j'}) \tag{2}$$

for all $j, j' = 1, \ldots, I$ with $j \neq j'$ whenever $j = I$ or $j' = I$. Here

$$\rho_{itjt'I} = -\frac{\sum_{j'=1}^{I-1} \rho_{itjt'j'} \sqrt{\pi_{it'j'}(1 - \pi_{it'j'})}}{\sqrt{\pi_{it'I}(1 - \pi_{it'I})}}$$

for $j = 1, \ldots, I - 1$,

$$\rho_{itIt'j'} = -\frac{\sum_{j=1}^{I-1} \rho_{itjt'j'} \sqrt{\pi_{itj}(1 - \pi_{it'j'})}}{\sqrt{\pi_{itI}(1 - \pi_{itI})}}$$

for $j' = 1, \ldots, I - 1$,

$$LB(p, q) = -\min\left\{\frac{pq}{(1-p)(1-q)}, \frac{(1-p)(1-q)}{pq}\right\}^{1/2}$$

and

$$UB(p, q) = \left(\frac{\min\{p, q\}(1 - \max\{p, q\})}{(1 - \min\{p, q\})\max\{p, q\}}\right)^{1/2}.$$

Consequently, the only set of correlation coefficients that is compatible with any sets of marginal probabilities corresponds to independence of Y_{it} and $Y_{it'}$.

Proof. To prove inequalities (7) for $j, j' = 1, \ldots, I-1$, apply the Fréchet bounds (4) for $\{\pi_{itjt'j'}, j, j' = 1, \ldots, I - 1\}$ in (6). Simple algebraic manipulation gives the desired result. For $j \neq j' = I$, note that restrictions (5) imply that

$$\sum_{j'=1}^{I} \rho_{itjt'j'}\sqrt{\pi_{it'j'}(1 - \pi_{it'j'})} = 0$$

and hence

$$\rho_{itjt'I} = -\frac{\sum_{j=1}^{I-1} \rho_{itjt'j'}\sqrt{\pi_{it'j'}(1 - \pi_{it'j'})}}{\sqrt{\pi_{it'I}(1 - \pi_{it'I})}}.$$

Application of the Fréchet bounds (4) provides the claimed inequalities. By symmetry, the same procedure proves (7) for $j' = 1, \ldots, I - 1$ and $j = I$. Finally, the tightness of the Fréchet bounds completes the proof of the first statement.

For the second statement, it suffices to prove that if at least one $\rho_{itjt'j'} \neq 0$ then there exist π_{itj} and $\pi_{it'j'}$ such that the Fréchet bounds for $\pi_{itjt'j'}$ are not met. Set $\pi_{it'j'} = 0.5$. If $\rho_{itjt'j'} > 0$, then set $\pi_{itj} > 1/(1 + \rho_{itjt'j'}^2)$ to obtain $\pi_{itjt'j'} > 0.5 = \pi_{it'j'}$. If $\rho_{itjt'j'} < 0$, then set $\pi_{itj} > \rho_{itjt'j'}^2/(1 + \rho_{itjt'j'}^2)$ to obtain $\pi_{itjt'j'} < 0$. $\square$

Constraints (2) become more severe as the number of response categories increases. For example, one can show that a marginal model with time-stationary covariates always exists for $I = 2$ if the correlation parameter lies on $(-1, 1)$ but not if $I \geq 4$ and $-1 < \rho_{it1t'1}, \rho_{it2t'2}, \rho_{it3t'3} < -0.5$. Further it can be shown that there exist positive definite correlation matrices that do not correspond to any distribution for bivariate multinomial responses. The second statement in Theorem 1, adopting a common correlation structure across all subjects might not be more beneficial than the independence 'working' model, that is treating all responses as independent, when the linear predictor in (1) is unrestricted on the real line. For these reasons, we believe that the correlation coefficients parameterization is not a suitable metric for the association of correlated multinomial responses.

Now consider the global odds ratios parameterization. First, we note that such a parameterization is valid only for ordinal response categories and is not sensible with nominal response categories. Let $F_{itjt'j'} = \Pr(Y_{it} \leq j, Y_{it'} \leq j'|\mathbf{x}_i)$ be the probability of subject i to have response outcome at category j or less at time t and at category j' or less at time t'. Let $\psi_{itjt'j'}$ be the global odds ratio for subject i at the category cutpoint (j, j')

$$\psi_{itjt'j'} = \frac{\Pr(Y_{it} \leq j, Y_{it'} \leq j'|\mathbf{x}_i)\Pr(Y_{it} > j, Y_{it'} > j'|\mathbf{x}_i)}{\Pr(Y_{it} > j, Y_{it'} \leq j'|\mathbf{x}_i)\Pr(Y_{it} \leq j, Y_{it'} > j'|\mathbf{x}_i)} = \frac{F_{itjt'j'}(1 - \gamma_{itj} - \gamma_{it'j'} + F_{itjt'j'})}{(\gamma_{itj} - F_{itjt'j'})(\gamma_{it'j'} - F_{itjt'j'})}$$

for $j, j' = 1, \ldots, I-1$, where $\gamma_{itj} = \sum_{k=1}^{j} \pi_{itk}$ is the cumulative probability of the response category j at time t for subject i. The probabilities $\{\pi_{itjt'j'} : j, j' = 1, \ldots, I\}$ depend on $\boldsymbol{\beta}$ and $\{\psi_{itjt'j'} : j, j' = 1, \ldots, I-1\}$ since

$$\pi_{itjt'j'} = F_{itjt'j'} - F_{itjt'(j-1)} - F_{it(j-1)t'j'} + F_{it(j-1)t'(j'-1)}$$

where

$$F_{itjt'j'} = \begin{cases} \frac{1 + (\gamma_{itj} + \gamma_{it'j'})(\psi_{itjt'j'} - 1) - S_{itjt'j'}}{2(\psi_{itjt'j'} - 1)} & \text{if } \psi_{itjt'j'} \neq 1 \\ \gamma_{itj}\gamma_{it'j'} & \text{if } \psi_{itjt'j'} = 1 \end{cases}$$

and

$$S^2_{itjt'j'} = [1 + (\gamma_{itj} + \gamma_{it'j'})(\psi_{itjt'j'} - 1)]^2 + 4\psi_{itjt'j'}\left(1 - \psi_{itjt'j'}\right)\gamma_{itj}\gamma_{it'j'}.$$

Theorem 2 proves that the set of global odds ratios $\{\psi_{itjt'j'} : j, j' = 1, \ldots, I-1\}$ defines a distribution for subject i compatible with any sets of marginal probabilities if and only if a uniform global odds ratio structure is adopted, that is $\psi_{itjt'j'} = \psi_{itt'}$ for all $j, j' = 1, \ldots, I-1$.

Theorem 2. *Let $(Y_{it}, Y_{it'})'$ be the bivariate random vector of multinomial responses for subject i at times t and t', with given marginal probabilities $\{\pi_{itj} : j = 1, \ldots, I\}$ and $\{\pi_{it'j'} : j' = 1, \ldots, I\}$ respectively. The set of global odds ratios $\{\psi_{itjt'j'} : j, j' = 1, \ldots, I-1\}$ defines a joint distribution for $(Y_{it}, Y_{it'})'$ compatible with any sets of marginal probabilities if and only if $\psi_{itjt'j'} = \psi_{itt'}$ for all $j, j' = 1, \ldots, I-1$.*

Proof. First, we prove that $\psi_{itjt'j'}$ must be nondecreasing in j and j'. Assume otherwise that $\psi_{itjt'j'} < \psi_{itjt'(j'-1)}$ and let $\gamma_{itj} = \gamma_{it'j'} = 0.5$. The function

$$\Delta_{jj'} = F_{itjt'j'} - F_{itjt'(j'-1)} = \frac{1}{2}\frac{\psi_{itjt'j'}^{1/2}}{1 + \psi_{itjt'j'}^{1/2}} - F_{itjt'(j'-1)}$$

is a continuous function of $\gamma_{it'(j'-1)}$ with $\Delta_{jj'} > 0$, if $\gamma_{it'(j'-1)} = 0$, and $\Delta_{jj'} < 0$, if $\gamma_{it'(j'-1)} = 0.5$. According to the intermediate value theorem there exists $0 < \gamma_{it'(j'-1)} < 0.5$ such that $F_{itjt'j'} < F_{itjt'(j'-1)}$. However, $F_{itjt'j'}$ is nondecreasing in j' and thus $\psi_{itjt'j'} \geq \psi_{itjt'(j'-1)}$. By symmetry, $\psi_{itjt'j'}$ is nondecreasing in j. Next suppose that $\psi_{it(I-1)t'j'} > \psi_{it(I-1)t'(j'-1)}$ and let $\gamma_{it(I-1)} = \gamma_{it'j'} = 0.5$. Then

$$\pi_{itIt'j'} = 0.5 - \gamma_{it(I-1)} - \frac{1}{2}\frac{\psi_{it(I-1)t'j'}^{1/2}}{1 + \psi_{it(I-1)t'j'}^{1/2}} + F_{it(I-1)t'(j'-1)}$$

is a continuous function of $\gamma_{it'(j'-1)}$ with $\pi_{itIt'j'} > 0$, if $\gamma_{it'(j'-1)} = 0$, and $\pi_{itIt'j'} < 0$, if $\gamma_{it'(j'-1)} = 0.5$. Similarly to before, there exists $0 < \gamma_{it'(j'-1)} < 0.5$ such that $\pi_{itIt'j'} < 0$ and thus $\psi_{it(I-1)t'j'} = \psi_{it(I-1)t'(j'-1)}$ for all j'. By symmetry, $\psi_{itjt'(I-1)} = \psi_{it(j-1)t'(I-1)}$ for all j.
Now suppose that $\psi_{it(I-1)t'(I-2)} > \psi_{it(I-2)t'(I-2)}$ and let $\gamma_{it(I-1)} = \gamma_{it'(I-1)} = 0.5$ and $\gamma_{it(I-2)} = \gamma_{it'(I-2)}$. Then

$$\pi_{it(I-1)t'(I-1)} = \frac{1}{2}\frac{\psi_{it(I-1)t'(I-1)}^{1/2}}{1 + \psi_{it(I-1)t'(I-1)}^{1/2}} - F_{it(I-1)t'(I-2)} + F_{it(I-2)t'(I-1)} - F_{it(I-2)t'(I-2)}$$

is a continuous function of $\gamma_{it(I-2)}$ with $\pi_{it(I-1)t'(I-1)} > 0$, if $\gamma_{it(I-2)} = 0$, and $\pi_{it(I-1)t'(I-1)} < 0$, if $\gamma_{it(I-2)} = 0.5$. Thus there exists $0 < \gamma_{it(I-2)} < 0.5$ such that $\pi_{it(I-1)t'(I-1)} < 0$. Applying this

argument successively, we can conclude that $\psi_{itjt'j} = \psi_{itt'} > 0$ for all $j, j' = 1, \ldots, I-1$. Finally, if $\psi_{itjt'j'} = \psi_{itt'} > 0$ for all $j, j' = 1, \ldots, I-1$, then $F_{itjt'j'}$ is a two increasing copula (Nelsen, 2006) and consequently $\pi_{itjt'j'} > 0$ for all $j, j' = 1, \ldots, I$. Since the bivariate probabilities satisfy the fixed marginals, the tightness of the Fréchet bounds completes the proof. $\square$

Heagerty and Zeger (1996) and Williamson et al. (1995) correctly speculated that the global odds ratios are less restrictive than the correlation coefficients parameterization. Theorem 2 clarifies the extent to which the global odds ratios parameterization does not depend on the marginal model specification. With a uniform global odds ratios structure, the association pattern can be modeled as a single global odds ratio for all subjects ($\psi_{itt'} = \psi$) or as a function of the covariates (Williamson et al., 1995). However, Dale (1984) and Goodman (1981) provided examples where a uniform global odds ratio does not approximate well underlying association patterns. Although uniform global odds ratios structures are consistent with the marginal model specification in (1), they might lead to some efficiency loss in estimating $\boldsymbol{\beta}$. Perhaps more importantly, they are not sensible for use with nominal-scale response variables.

Finally, let $\theta_{itjt'j'}$ be the local odds ratio for subject i at the category cutpoint (j, j')

$$\theta_{itjt'j'} = \frac{\pi_{itjt'j'} \pi_{it(j+1)t'(j'+1)}}{\pi_{it(j+1)t'j'} \pi_{itjt'(j'+1)}}$$

for $j, j' = 1, \ldots, I-1$. The distribution of $(Y_{it}, Y_{it'})$ can be viewed as the unique solution of the iterative proportional fitting procedure proposed by Deming and Stephan (1940), with local odds ratios structure $\{\theta_{itjt'j'} : j, j' = 1, \ldots, I-1\}$, row marginals $\{\pi_{itj} : j = 1, \ldots, I\}$ and column marginals $\{\pi_{it'j'} : j' = 1, \ldots, I\}$. The local odds ratios parameterization has certain advantages over the global odds ratios and correlation coefficient alternatives for modeling the association of bivariate multinomial responses. Rudas (1998) argued that in a 2-way probability table with prescribed marginals and positive cells, the local odds ratios are unconstrained and they describe the totality of 2-way probability tables. Hence the local odds ratios are variation independent to $\boldsymbol{\beta}$ and this permits parsimonious, flexible and efficient 'working' forms for $\boldsymbol{\alpha}$ that may depend on the category cutpoints and/or the covariates. Essentially the local odds ratios parameterization maps all the correlation coefficients in (1) that satisfy Theorem 1 and unlike the global odds ratios, local odds ratios can be meaningful both for ordinal and nominal responses.

Next, suppose that $T_i = 3$ and note that for given univariate marginals the three sets of $I \times I$ contingency tables that satisfy a uniform odds ratios structure are included to the three sets of $I \times I$ contingency tables that satisfy a local odds ratio structure. Now, by dichotomizing the correlated multinomial responses at the each response category when $T_i = 3$, the results in Chaganty and Joe (2006) give necessary bounds for a uniform global odds ratios structure and consequently, for a local odds ratios structure. The finer implication of this result is that the set of admissible values for $\widehat{\boldsymbol{\alpha}}$ depends on the unknown marginal regression vector $\boldsymbol{\beta}$ when $T_i \geq 3$ for all the parameterizations of $\boldsymbol{\alpha}$ considered here. This proves that the conditions (4) and (5) are not sufficient when $T_i \geq 3$. Therefore we cannot guarantee the existence of $(\widehat{\boldsymbol{\beta}}_G, \widehat{\boldsymbol{\alpha}})$ when a 'working' distribution for $\mathbf{Y}_i$ cannot be obtained for at least one i and hence treating $\widehat{\boldsymbol{\alpha}}$ as an estimate of the 'working' association structure does not seem appropriate. In the relevant GEE literature, these potential pitfalls have not been addressed, but instead the relevant estimating procedures are based on introducing another set of estimating equations which implicitly assumes that $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are variation independent. However, as has been noted (Crowder, 1995), it is crucial for the estimating mechanism of $\boldsymbol{\alpha}$ to ensure the existence and the convergence of $\widehat{\boldsymbol{\alpha}}$ to a well-defined parameter vector $\boldsymbol{\alpha}$, since $\boldsymbol{\alpha}$ is involved in the asymptotic properties of the GEE estimator $\widehat{\boldsymbol{\beta}}_G$.

References

Chaganty, N. and Joe, H. (2006). Range of correlation matrices for dependent bernoulli random variables. *Biometrika* **93,** 197–206.

Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82,** 407–410.

Dale, J. (1984). Local versus global association for bivariate ordered responses. *Biometrika* **71,** 507–514.

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

Dobra, A. and Fienberg, S. (2000). Bounds for cell entries in contingency tables given marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences of the United States of America* **22,** 11885–11892.

Goodman, L. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika* **68,** 347–355.

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

Lipsitz, S., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* **13,** 1149–1163.

Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on the working correlation structures. *Biometrics* **52,** 354–361.

Miller, M., Davis, C., and Landis, J. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics* **49,** 1033–1044.

Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag.

Parsons, N., Edmondson, R., and Gilmour, S. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society C* **55,** 507–524.

Rudas, T. (1998). *Odds Ratios in the Analysis of Contingency Tables*. Sage.

Williamson, J., Kim, K., and Lipsitz, S. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association* **90,** 1432–1437.

Web Table 1: Simulation results for the GEE estimators of $\beta = 1$ in $\Pr(Y_{it} \leq j | \mathbf{x}_i) = \Phi(\beta_{0j} + \beta x_{it})$ under a latent exchangeable correlation matrix with parameter ρ in Scenario 1. The columns display the simulated bias (Bias), the simulated standard errors (SE), the simulated relative efficiency (SRE), the empirical level of the nominal 95% Wald confidence intervals for β based on the robust covariance matrix (EL), and the convergence rate (CR) for each estimating procedure.

Sample Size	Estimator	Bias	SE	SRE	EL	CR	Bias	SE	SRE	EL	CR
$N = 100$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0063	0.0683	1.00	94.59	100.00	0.0083	0.0786	1.00	94.11	100.00
	$\hat{\beta}_U$	0.0065	0.0689	0.98	94.10	99.56	0.0086	0.0791	0.97	93.73	99.21
	$\hat{\beta}_{CE}$	0.0018	0.0691	0.98	94.05	99.40	0.0044	0.0793	0.98	93.87	99.31
	$\hat{\beta}_{TE}$	-0.0030	0.0692	0.97	93.75	99.42	-0.0012	0.0792	0.99	93.57	99.04
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0128	0.1011	1.00	93.71	100.00	0.0165	0.1229	1.00	93.87	100.00
	$\hat{\beta}_U$	0.0126	0.1016	0.98	93.22	98.40	0.0144	0.1219	0.99	94.08	97.40
	$\hat{\beta}_{CE}$	0.0090	0.1012	0.99	93.28	98.04	0.0110	0.1213	0.99	94.23	95.84
	$\hat{\beta}_{TE}$	0.0034	0.1013	0.99	93.24	98.34	0.0083	0.1211	1.00	94.03	96.69
$N = 500$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0018	0.0303	1.00	94.94	100.00	0.0019	0.0348	1.00	94.50	100.00
	$\hat{\beta}_U$	0.0018	0.0307	0.98	94.71	99.80	0.0019	0.0351	0.98	94.38	99.54
	$\hat{\beta}_{CE}$	0.0009	0.0307	0.98	94.81	99.87	0.0012	0.0351	0.98	94.46	99.75
	$\hat{\beta}_{TE}$	-0.0001	0.0306	0.98	94.69	99.74	0.0000	0.0350	0.99	94.41	99.39
	$\hat{\beta}_{RC}$	-0.0038	0.0307	0.97	94.58	99.47	-0.0033	0.0351	0.98	94.12	99.33
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0017	0.0443	1.00	94.37	100.00	0.0038	0.0541	1.00	94.59	100.00
	$\hat{\beta}_U$	0.0017	0.0445	0.98	94.40	98.31	0.0035	0.0543	0.99	94.40	97.81
	$\hat{\beta}_{CE}$	0.0011	0.0445	0.99	94.37	98.16	0.0029	0.0543	0.99	94.42	96.76
	$\hat{\beta}_{TE}$	0.0000	0.0443	1.00	94.38	98.09	0.0024	0.0541	1.00	94.46	97.44
	$\hat{\beta}_{RC}$	-0.0026	0.0442	0.99	94.26	97.90	0.0011	0.0541	1.00	94.39	96.78

Web Table 2: Simulation results for the GEE estimators of $\beta = 1$ in $\Pr(Y_{it} \leq j | \mathbf{x}_i) = \Phi(\beta_{0j} + \beta x_{it})$ under a latent exchangeable correlation matrix with parameter ρ in Scenario 2.

Sample Size	Estimator	Bias	SE	SRE	EL	CR	Bias	SE	SRE	EL	CR
$N = 100$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0054	0.0686	1.00	94.37	100.00	0.0079	0.0769	1.00	93.67	100.00
	$\hat{\beta}_U$	0.0032	0.0711	0.93	94.53	99.25	0.0057	0.0769	1.00	94.03	99.30
	$\hat{\beta}_{CE}$	-0.0043	0.0715	0.92	94.03	99.36	-0.0023	0.0773	0.99	93.74	99.23
	$\hat{\beta}_{TE}$	-0.0043	0.0715	0.92	94.04	99.13	-0.0020	0.0772	0.99	93.66	99.13
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0114	0.0946	1.00	93.83	100.00	0.0176	0.1165	1.00	93.59	100.00
	$\hat{\beta}_U$	0.0083	0.0870	1.18	93.86	98.98	0.0123	0.0922	1.58	93.81	97.72
	$\hat{\beta}_{CE}$	-0.0012	0.0877	1.17	93.09	99.13	0.0014	0.0927	1.59	93.07	97.72
	$\hat{\beta}_{TE}$	0.0009	0.0868	1.19	93.57	99.01	0.0074	0.0916	1.60	93.86	97.56
$N = 500$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0010	0.0308	1.00	94.63	100.00	0.0015	0.0341	1.00	94.68	100.00
	$\hat{\beta}_U$	0.0006	0.0318	0.94	94.56	99.57	0.0010	0.0339	1.01	94.61	99.57
	$\hat{\beta}_{CE}$	-0.0009	0.0318	0.94	94.52	99.51	-0.0006	0.0339	1.01	94.57	99.53
	$\hat{\beta}_{TE}$	-0.0010	0.0318	0.93	94.56	99.18	-0.0006	0.0339	1.01	94.58	99.25
	$\hat{\beta}_{RC}$	-0.0060	0.0319	0.90	93.95	99.24	-0.0058	0.0340	0.98	93.99	99.33
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0023	0.0422	1.00	94.59	100.00	0.0028	0.0503	1.00	94.89	100.00
	$\hat{\beta}_U$	0.0014	0.0386	1.19	94.45	98.74	0.0021	0.0402	1.56	94.88	98.41
	$\hat{\beta}_{CE}$	-0.0004	0.0387	1.19	94.34	98.94	-0.0001	0.0403	1.56	94.69	98.19
	$\hat{\beta}_{TE}$	0.0000	0.0386	1.20	94.33	98.73	0.0011	0.0401	1.56	94.78	97.98
	$\hat{\beta}_{RC}$	-0.0060	0.0386	1.16	93.76	99.10	-0.0057	0.0404	1.53	93.89	98.11

Web Table 3: Simulation results for the GEE estimators of $\beta = 1$ in $\Pr(Y_{it} \leq j|\mathbf{x}_i) = \Phi(\beta_{0j} + \beta x_{it})$ under a latent exchangeable correlation matrix with parameter ρ in Scenario 3.

Sample Size	Estimator	Bias	SE	SRE	EL	CR	Bias	SE	SRE	EL	CR
$N = 100$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0099	0.0871	1.00	94.55	100.00	0.0117	0.0974	1.00	93.71	100.00
	$\hat{\beta}_U$	0.0097	0.0883	0.97	94.47	99.74	0.0113	0.0979	0.99	93.76	99.52
	$\hat{\beta}_{CE}$	0.0021	0.0887	0.97	94.46	99.49	0.0045	0.0981	0.99	93.81	99.52
	$\hat{\beta}_{TE}$	-0.0021	0.0889	0.97	94.15	99.57	-0.0006	0.0981	1.00	93.54	99.45
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0162	0.1157	1.00	93.68	100.00	0.0227	0.1381	1.00	93.35	100.00
	$\hat{\beta}_U$	0.0148	0.1127	1.05	93.67	99.00	0.0184	0.1284	1.14	94.30	98.32
	$\hat{\beta}_{CE}$	0.0099	0.1125	1.06	93.89	98.36	0.0130	0.1282	1.14	94.30	87.50
	$\hat{\beta}_{TE}$	0.0042	0.1122	1.06	93.91	98.97	0.0127	0.1284	1.15	94.16	92.61
$N = 500$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0024	0.0386	1.00	94.55	100.00	0.0030	0.0418	1.00	94.64	100.00
	$\hat{\beta}_U$	0.0023	0.0394	0.96	94.52	99.79	0.0029	0.0422	0.98	94.57	99.78
	$\hat{\beta}_{CE}$	0.0009	0.0394	0.96	94.53	99.84	0.0017	0.0421	0.99	94.44	99.79
	$\hat{\beta}_{TE}$	-0.0000	0.0394	0.96	94.37	99.79	0.0006	0.0421	0.99	94.49	99.62
	$\hat{\beta}_{RC}$	-0.0057	0.0394	0.94	94.02	99.46	-0.0047	0.0421	0.98	94.14	99.51
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0040	0.0500	1.00	94.63	100.00	0.0034	0.0596	1.00	94.50	100.00
	$\hat{\beta}_U$	0.0035	0.0486	1.06	94.79	98.83	0.0029	0.0561	1.12	94.79	98.74
	$\hat{\beta}_{CE}$	0.0024	0.0485	1.06	94.75	99.00	0.0025	0.0561	1.12	94.77	98.38
	$\hat{\beta}_{TE}$	0.0014	0.0484	1.07	94.93	98.90	0.0020	0.0559	1.13	94.85	98.71
	$\hat{\beta}_{RC}$	-0.0024	0.0485	1.07	94.72	98.97	-0.0009	0.0565	1.11	94.68	86.44

Web Table 4:
Simulation results for the GEE estimators of $\beta = 1$ in $\Pr(Y_{it} \leq j|\mathbf{x}_i) = \Phi(\beta_{0j} + \beta x_{it})$ under a latent exchangeable correlation matrix with parameter ρ in Scenario 4.

Sample Size	Estimator	Bias	SE	SRE	EL	CR	Bias	SE	SRE	EL	CR
$N = 100$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0093	0.0879	1.00	94.21	100.00	0.0121	0.0946	1.00	94.00	100.00
	$\hat{\beta}_U$	0.0065	0.0902	0.95	94.12	99.59	0.0089	0.0950	1.00	94.12	99.53
	$\hat{\beta}_{CE}$	-0.0066	0.0909	0.94	93.44	99.56	-0.0046	0.0959	0.98	93.45	99.50
	$\hat{\beta}_{TE}$	-0.0039	0.0904	0.95	93.68	99.49	-0.0021	0.0949	1.00	93.75	99.39
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0153	0.1109	1.00	93.76	100.00	0.0235	0.1317	1.00	92.95	100.00
	$\hat{\beta}_U$	0.0108	0.1032	1.15	93.74	99.39	0.0155	0.1075	1.49	93.20	98.65
	$\hat{\beta}_{CE}$	-0.0054	0.1054	1.12	92.67	99.31	-0.0037	0.1088	1.47	91.83	96.80
	$\hat{\beta}_{TE}$	0.0003	0.1034	1.16	93.42	99.20	0.0075	0.1074	1.51	93.08	98.40
$N = 500$		$\rho = 0$					$\rho = 0.15$				
	$\hat{\beta}_I$	0.0024	0.0383	1.00	95.27	100.00	0.0028	0.0414	1.00	94.69	100.00
	$\hat{\beta}_U$	0.0018	0.0394	0.95	95.24	99.78	0.0021	0.0414	1.00	94.76	99.69
	$\hat{\beta}_{CE}$	-0.0004	0.0394	0.95	95.00	99.66	-0.0002	0.0415	1.00	94.67	99.62
	$\hat{\beta}_{TE}$	-0.0003	0.0394	0.95	95.02	99.51	0.0001	0.0415	1.00	94.65	99.61
	$\hat{\beta}_{RC}$	-0.0083	0.0396	0.90	93.69	99.25	-0.0082	0.0415	0.96	93.63	99.30
		$\rho = 0.5$					$\rho = 0.85$				
	$\hat{\beta}_I$	0.0027	0.0480	1.00	95.08	100.00	0.0036	0.0560	1.00	95.08	100.00
	$\hat{\beta}_U$	0.0020	0.0449	1.14	95.12	99.40	0.0027	0.0462	1.47	95.06	99.12
	$\hat{\beta}_{CE}$	-0.0008	0.0449	1.14	94.93	99.41	-0.0006	0.0462	1.47	94.98	98.91
	$\hat{\beta}_{TE}$	-0.0001	0.0447	1.15	95.05	99.39	0.0013	0.0461	1.48	95.00	98.99
	$\hat{\beta}_{RC}$	-0.0092	0.0450	1.09	93.78	99.37	-0.0092	0.0465	1.40	93.55	97.99