



Joint generalized estimating equations for longitudinal binary data

DOI:

[10.1016/j.csda.2020.107110](https://doi.org/10.1016/j.csda.2020.107110)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Huang, Y., & Pan, J. (2021). Joint generalized estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis*, 155, 107110. <https://doi.org/10.1016/j.csda.2020.107110>

Published in:

Computational Statistics & Data Analysis

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Joint generalized estimating equations for longitudinal binary data

Youjun Huang^a, Jianxin Pan^{b,*}

^a*Mathematical College, Sichuan University, Chengdu 610065, China*

^b*Department of Mathematics, The University of Manchester, Manchester M13 9PL, U.K.*

Abstract

Modelling longitudinal binary data is challenging but common in practice. Existing methods on modelling of binary responses take no account of the fact that the correlation coefficient of binary responses must have an upper bound which is smaller than one. Ignoring this fact can lead to incorrect statistical inferences for longitudinal binary data. A novel method is proposed to model the mean and within-subject correlation coefficients for longitudinal binary data, simultaneously, by taking into account the constraints of the upper bounds. By introducing latent normally distributed random variables, the correlation coefficients of binary responses are connected to those for the latent variables, of which the correlation coefficients are modeled accordingly. A joint generalized estimating equation (GEE) method is developed for this purpose and the resulting correlation coefficients are shown to satisfy the constraints. Asymptotic normality of the parameter estimators is derived and simulation studies are made under various scenarios, showing that the proposed joint GEE method works very well even if the working covariance structures are misspecified. For illustration, the proposed method is applied to two real data practices to assess the effects of covariates on the mean and within-subject correlation coefficients.

Keywords: Correlation coefficients, Generalized estimating equations, Joint mean and correlation parameter estimation, Longitudinally correlated binary data

*Corresponding author

Email address: jianxin.pan@manchester.ac.uk (Jianxin Pan)

1. Introduction

Binary responses with repeat measurements are very common in many fields such as medical and biological sciences. Longitudinal normally distributed data have been studied well, while it is very challenging to model longitudinal binary data. A major issue is that the joint distribution of correlated binary data does not have an analytically tractable form. In the literature, various methods were proposed to model the conditional mean and marginal mean of correlated binary responses. The former was mainly studied within the framework of generalized linear mixed model (GLMM). For example, Heagerty (1999) considered a random effect model for longitudinal binary data. Wang and Louis (2003) developed an intercept random effect model with bridge distribution, and Parzen et al. (2011) extended the single random intercept model for longitudinal binary data. Although the GLMM-based approach produces good estimators of the fixed effects and has good predictions of the random effects, the within-subject correlations of binary responses remain unclear as they are analytically intractable. It becomes more difficult to understand how the within-subject correlations are associated with covariates of interest.

The latter was investigated by using generalized estimation equation (GEE) methods, see, e.g., Liang and Zeger (1986). With unnecessary assumption of distribution, the GEE methods directly solve estimating equations in order to obtain the estimators of the marginal mean parameters, which involves the specification of a working covariance matrix for binary responses. It was shown that the GEE estimators of the mean parameters are consistent even if the working covariance is misspecified. However, misspecification of the covariance can lead to a great loss of efficiency for the GEE estimators of the mean parameters (Sutradhar and Das, 1999, 2000). In addition, Crowder (1995) showed that the asymptotic properties of the parameter estimators may be broken down due to the uncertainty of the parameters involved in the misspecified working covariance. Daniels and Zhao (2003) demonstrated that when longitudinal data contain missing values the mean parameters estimators are biased if the working covariance is misspecified. On the other hand, the covariance itself may be of scientific interest under certain circumstances (Diggle and Verbyla, 1998). Many authors extended the GEE method to model the mean and covariance for longitudinal data, simultaneously. For example, using a modified Cholesky decomposition Ye and Pan (2006) proposed a joint GEE approach

to model the mean and covariance. However, their method only applies to longitudinal continuous data and is not appropriate for longitudinal binary data. Prentice (1988) considered regression methods by introducing additional estimating equations for correlation parameters. Prentice and Zhao (1991) further extended the methods to general multivariate discrete and continuous responses. Carey et al. (1993) considered an alternating logistic regression method for dealing with computational infeasibility when the cluster size is large.

When applying to longitudinal binary data, however, the aforementioned methods for modeling conditional mean and marginal mean are not appropriate, because they ignore an important fact, that is, the correlation coefficient of correlated binary responses has an upper bound that is smaller than one. As a result, invalid correlation estimators are produced and statistical inferences are not reliable. In spirit of Oman and Zucker (2001) who connected the correlation coefficients of binary responses to those of certain latent normally distributed responses, in this paper we propose a joint GEE method to estimate the mean and correlation parameters for longitudinal binary data. An appealing property of the proposed method is that the upper bound constraints of the estimated correlation coefficients are satisfied. The structure of this paper is as follows. In Section 2, we consider the models for both mean and correlation parameters and then develop the joint GEE method. The asymptotic normality of the parameter estimators is derived in Section 3. In Section 4, simulation studies are conducted to assess the performance of the proposed approach under various scenarios. For illustration, the approach is applied to the analysis of two real data examples in Section 5. Some concluding remarks are provided in Section 6.

2. Joint GEE for Mean and Correlation Parameters

2.1. Generalized Estimating Equation

For the j th measurement of the i th subject, let Y_{ij} be the binary outcome variable and x_{ij} be the $p \times 1$ vector of covariates ($j = 1, \dots, m_i$ and $i = 1, \dots, n$). The measurements from different subjects are assumed to be independent. Let $Y_i = (Y_{i1}, \dots, Y_{im_i})^T$ be the $m_i \times 1$ vector of responses and $X_i = (x_{i1}, \dots, x_{im_i})^T$ be the $m_i \times p$ matrix of covariates for the i th subject, $i = 1, \dots, n$. Denote the expectation vector of Y_i by

$\mu_i = (\mu_{i1}, \dots, \mu_{im_i})^T$ and the covariance matrix by $\text{var}(Y_i) = \Sigma_i = (\sigma_{ijk})$ with diagonal elements $\sigma_{ijj}^2 = \psi^2 \mu_{ij}(1 - \mu_{ij})$. Without loss of generality, the dispersion parameter ψ^2 is assumed to be 1 in this paper. In spirit of generalized linear models (GLM), the expectation $\mu_{ij} = \text{E}(Y_{ij}) = \text{Pr}\{Y_{ij} = 1\}$ is assumed to be associated with explanatory variables x_{ij} through logit or probit link function $g(\cdot)$, such that $g(\mu_{ij}) = x_{ij}^T \beta$ where β is a $p \times 1$ parameter vector. According to Liang and Zeger (1986), the GEE for mean parameter β is given by

$$U_1(\beta, \gamma) = \sum_{i=1}^n D_i^T \Sigma_i^{-1} (Y_i - \mu_i) = 0, \quad (1)$$

where $D_i = \partial \mu_i / \partial \beta$ is the $m_i \times p$ derivative matrix. Since the covariance matrix Σ_i is in general unknown, it was suggested to take a sandwich form like $\Sigma_i = A_i^{1/2} R_i(\gamma) A_i^{1/2}$, where $A_i = \text{diag}\{\text{var}(Y_i)\}$ and $R_i(\gamma)$ is the $m_i \times m_i$ correlation matrix of Y_i which may be related to new parameters γ .

Liang and Zeger (1986) proposed to use a working matrix to replace the unknown correlation matrix $R_i(\gamma)$ of Y_i and showed that the GEE method produces consistent estimators of the mean parameters β even if the working correlation matrix is misspecified. Chaganty and Joe (2004) indicated that the working correlation matrix should be viewed as a weight correlation matrix but cannot be treated as the true correlation matrix. By using a matrix Cauchy-Schwarz inequality, they further showed that the optimal choice of the working correlation matrix for correlated binary responses is the true correlation matrix, which is unfortunately unknown in practice. Simulation results by Dziak (2006); Wang et al. (2012) showed that the misspecification of correlation structure leads to poor performances in variable selection and causes estimation efficiency problems for correlated binary data. To avoid the misspecification problem, we propose to model the correlations for binary data using a simple parametric form, which in the meantime guarantees their upper bounds satisfied.

2.2. Modelling Correlation Matrix

Since $\text{Pr}\{Y_{ij} \times Y_{ik} = 1\} \leq \min(\mu_{ij}, \mu_{ik})$, it can be shown that the correlation matrix $R_i = (r_{ijk})$ for correlated binary responses must satisfy

$$r_{ijk} \leq \min\{\psi_{ij}/\psi_{ik}, \psi_{ik}/\psi_{ij}\} = \bar{r}_{ijk},$$

where $\psi_{ij} = (\mu_{ij}/(1 - \mu_{ij}))^{1/2}$. Obviously, \bar{r}_{ijk} is a quantity related to the odds ratio. By studying the relationship between r_{ijk} and \bar{r}_{ijk} , Oman and Zucker (2001) conjectured that

$$r_{ijk} = \bar{r}_{ijk} c_{ijk}(\gamma), \quad (2)$$

are valid correlation coefficients for certain correlated binary data, where $C_i(\gamma) = (c_{ijk}(\gamma))$ may be any normal-model correlation matrix involving parameter γ . When the matrix $C_i(\gamma)$ takes compound symmetry (CS), order-one moving average (MA(1)) or order-one autoregressive (AR(1)) structures, which are all commonly encountered correlation structures for longitudinal continuous data, Oman and Zucker (2001) gave a theoretical justification that equation (2) is really a valid correlation coefficient for correlated binary data. Equation (2) implies that by introducing latent normally distributed random variables with the covariance structure MA(1), CS or AR(1), the correlations of correlated binary variables can be estimated through estimating the correlations of the latent normally distributed variables, provided the mean is given. In addition, the resulting correlations for binary data guarantee to satisfy the constraints of upper bounds. However, the estimation of parameter γ remains unclear.

2.3. GEE for Correlation Parameter

For each subject, we now define new variables $W_{ijk} = (Y_{ij} - Y_{ik})^2$, which are binary responses too and are actually the coordinates of the variogram clouds in spatial statistics. If $\mu_{ij} \leq \mu_{ik}$, the expectation and variance of W_{ijk} can be expressed straightforwardly as

$$E(W_{ijk}) = \mu_{ij}(1 - \mu_{ik}) + \mu_{ik}(1 - \mu_{ij}) + 2\mu_{ij}(\mu_{ik} - 1)c_{ijk}(\gamma) \equiv f_{ijk} + h_{ijk}c_{ijk}(\gamma),$$

and

$$\text{var}(W_{ijk}) = \{f_{ijk} + h_{ijk}c_{ijk}(\gamma)\}[1 - \{f_{ijk} + h_{ijk}c_{ijk}(\gamma)\}].$$

Let W_i denote the $M_i = m_i(m_i - 1)/2$ -dimensional vector of W_{ijk} without repetition ($j \neq k$). We then construct the second GEE in order to estimate the correlation parameter γ ,

$$U_2(\gamma, \beta) = \sum_{i=1}^n E_i^T G_i^{-1}(W_i - \xi_i) = 0, \quad (3)$$

where ξ_i is the expectation of W_i , $E_i = \partial \xi_i / \partial \gamma$ is the $M_i \times 1$ derivative matrix, and G_i is the $M_i \times M_i$ covariance matrix of W_i . Note G_i is in general unknown but can be

approximated by a sandwich matrix

$$G_i = B_i^{1/2} V_i(\alpha) B_i^{1/2},$$

where B_i is a diagonal matrix with entries $\text{var}(W_{ij})$ and $V_i(\alpha)$ is a working correlation matrix of W_i involving parameter α . Obviously, G_i becomes the true covariance matrix of W_i if V_i takes the true correlation matrix of W_i . In spirit of the GEE for the mean parameters, we propose to use certain structures such as CS and AR(1) structures to be the working correlation structure of $V_i(\alpha)$. An appealing property of using the CS or AR(1) matrix is that its inverse has an analytical form, so that the calculation of the inverse matrix V_i^{-1} becomes straightforward even if m_i is large.

We then iteratively use the two GEEs in (1) and (3) to calculate the estimators of the mean parameters β and the correlation parameter γ . We summarize the calculations in the following algorithm.

Algorithm 1: Joint GEE for longitudinally correlated binary data

- 1: Input an initial $\hat{\beta}^{(0)}$ and a fixed parameter α ;
 - 2: Compute all the W_i 's and the estimators of f_{ijk} 's and h_{ijk} 's;
 - 3: Solve the second GEE $U_2(\gamma, \beta) = 0$ in (3) to obtain correlation parameter $\hat{\gamma}^{(0)}$;
 - 4: Compute correlation matrix $R_i^{(0)}$ by (2) and form covariance matrix $\Sigma_i^{(0)}$;
 - 5: Input $\Sigma_i^{(0)}$ to the first GEE $U_1(\beta, \gamma) = 0$ in (1). Solve the equation to obtain new estimator $\hat{\beta}^{(1)}$. Update f_{ijk} 's and h_{ijk} 's;
 - 6: Repeat steps 3 to 5 until convergence.
-

For the initial $\hat{\beta}^{(0)}$, we suggest to use the consistent estimator obtained by ordinary GEE method (Liang and Zeger, 1986) with independent working structure. When iteratively solving the equations, a modified Fisher-scoring algorithm may be applied. For example, to solve the equation $U_1(\beta, \gamma) = 0$, with a starting value β_0 we can update the value of β at the $(s + 1)$ th iteration through

$$\beta_{s+1} = \beta_s + \left(\sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i \right)^{-1} \sum_{i=1}^n D_i^T \Sigma_i^{-1} (Y_i - \mu_i),$$

where μ_i and D_i are evaluated at $\beta = \beta_s$ and Σ_i at γ_s . To solve the equation $U_2(\gamma, \beta) = 0$,

the value of γ can be updated by

$$\gamma_{s+1} = \gamma_s + \left(\sum_{i=1}^n E_i^T G_i^{-1} E_i \right)^{-1} \sum_{i=1}^n E_i^T G_i^{-1} (W_i - \xi_i),$$

where ξ_i , E_i and G_i are all evaluated at $\beta = \beta_s$ and $\gamma = \gamma_s$. Due to those analytical forms, the algorithm converges very quickly and the solutions become stable.

In the above algorithm, the parameters β and γ are estimated iteratively, meaning that the estimation processes for estimating β and γ are taken independently, though theoretically the estimation of γ may be associated with the estimator of β and vice versa. As an alternative, we also consider a single but large GEE for the full parameters $\theta = (\beta^T, \gamma)^T$. Denote $\eta_i = (\mu_i^T, \xi_i^T)^T$ ($i = 1, \dots, n$). The full parameter vector θ then may be estimated by solving the single GEE

$$\sum_{i=1}^n \frac{\partial \eta_i^T}{\partial \theta} \Omega_i^{-1} \begin{pmatrix} Y_i - \mu_i \\ W_i - \xi_i \end{pmatrix} = 0, \quad (4)$$

where Ω_i is an $(m_i + M_i) \times (m_i + M_i)$ working covariance matrix of $(Y_i^T, W_i^T)^T$, which has a sandwich form as

$$\Omega_i = \begin{pmatrix} A_i & 0 \\ 0 & B_i \end{pmatrix}^{1/2} \Omega_i^* \begin{pmatrix} A_i & 0 \\ 0 & B_i \end{pmatrix}^{1/2},$$

where Ω_i^* is a working correlation matrix of $(Y_i^T, W_i^T)^T$. Note that the quasi-Fisher information matrix for the augmented responses $(Y_i^T, W_i^T)^T$ is not necessarily block-diagonal because its off-diagonal block sub-matrix in general is not exactly zero. The quasi Fisher-scoring algorithm can then be applied to solving the single GEE through, e.g., at the $(s + 1)$ th-step,

$$\theta_{s+1} = \theta_s + \left(\sum_{i=1}^n \frac{\partial \eta_i^T}{\partial \theta} \Omega_i^{-1} \frac{\partial \eta_i}{\partial \theta} \right)_{\theta=\theta_s}^{-1} \left[\sum_{i=1}^n \frac{\partial \eta_i^T}{\partial \theta} \Omega_i^{-1} \begin{pmatrix} Y_i - \mu_i \\ W_i - \xi_i \end{pmatrix} \right]_{\theta=\theta_s}.$$

Our simulation studies presented in Section 4 show that the iterative GEE algorithm yields solutions very close to those by the single GEE for full parameters θ . It implies that the use of the single but large GEE method is not absolutely necessary. Besides, the iterative GEEs successfully separate the big estimating equation into two relatively small equations, largely reducing computational loads with little compromise of the accuracy of parameter estimators.

3. Asymptotic Property

Similar to Prentice (1988), the joint asymptotic distribution of the estimators $\hat{\beta}$ and $\hat{\gamma}$ through the GEEs (1) and (3) can be obtained. A sketch of the proof is provided in the Appendix.

Theorem 1. *Under regularity conditions, the joint GEE estimators $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically normally distributed as*

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \rightarrow N \left\{ 0, n \begin{pmatrix} Q_{11} & 0 \\ Q_{21} & Q_{22} \end{pmatrix}^{-1} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{21}^T \\ 0 & Q_{22} \end{pmatrix}^{-1} \right\}, \quad (5)$$

where

$$\begin{aligned} Q_{11} &= \sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i, \\ Q_{21} &= \sum_{i=1}^n E_i^T G_i^{-1} \frac{\partial \xi_i}{\partial \beta}, \\ Q_{22} &= \sum_{i=1}^n E_i^T G_i^{-1} E_i, \\ \Lambda_{11} &= Q_{11} = \sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i, \\ \Lambda_{12} &= \Lambda_{21}^T = \sum_{i=1}^n D_i^T \Sigma_i^{-1} \text{cov}(Y_i, W_i) G_i^{-1} E_i \quad \text{and} \\ \Lambda_{22} &= \sum_{i=1}^n E_i^T G_i^{-1} \text{cov}(W_i) G_i^{-1} E_i. \end{aligned}$$

Note that if only the variance of $\hat{\beta}$ is of interest, we may simply calculate Q_{11} with the parameters replaced by their estimators. When $\hat{\beta}$ and $\hat{\gamma}$ are both of interest, their variances involve the calculation of the whole asymptotic covariance matrix. In this case, the covariance matrices $\text{cov}(Y_i, W_i)$ and $\text{cov}(W_i)$ may be simply estimated by $(Y_i - \hat{\mu}_i)(W_i - \hat{\xi}_i)^T$ and $(W_i - \hat{\xi}_i)(W_i - \hat{\xi}_i)^T$, respectively, where $\hat{\mu}_i$ and $\hat{\xi}_i$ are evaluated at $(\hat{\beta}, \hat{\gamma})$, see Liang and Zeger (1986).

4. Simulation Study

In this section, simulation studies for longitudinally correlated binary data, generated under the three scenarios AR(1), CS and MA(1) for the normal-model correlation matrix $C(\gamma)$, are conducted. For each scenario of the simulations below, 1000 random samples are generated by the using the models of Oman and Zucker (2001). The generation mechanisms are described as follows.

- Scenario 1: AR(1) for $C_i(\gamma)$. Denote $q_{ij} = \Phi^{-1}(\mu_{ij})$ ($i = 1, \dots, n; j = 1, \dots, m_i$) where $\mu_i = (\mu_{i1}, \dots, \mu_{im_i})^T$ is the marginal mean of Y_i and $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard normal distribution. For each i , we generate m_i independent variables ε_{ij} from standard normal distribution and m_i independent variables U_{ij} from Bernoulli distribution with parameter γ ($0 \leq \gamma \leq 1$) ($j = 1, \dots, m_i$). Let Z_{i1} be a standard normal variable independent with the U_{ij} and ε_{ij} and let $Z_{ij} = U_{ij}Z_{i(j-1)} + (1 - U_{ij})\varepsilon_{ij}$ ($j > 1$). Furthermore, let $Y_{ij} = \mathbf{1}_{(Z_{ij} \leq q_{ij})}$ where $\mathbf{1}_{(\cdot)}$ is the indicator function. Then the constructed binary variables Y_{ij} must have the desired within-subject correlations $r_{ijk} = \bar{r}_{ijk}c_{ijk}(\gamma)$ with $c_{ijk}(\gamma) = \gamma^{|j-k|}$ ($j \neq k$).
- Scenario 2: CS for $C_i(\gamma)$. The generation mechanism is the same as the Scenario 1 except that, a) the Bernoulli parameter changes to $\gamma^{1/2}$ ($0 \leq \gamma \leq 1$), and b) an extra standard normal variable ε_{i0} is generated and let $Z_{ij} = U_{ij}\varepsilon_{i0} + (1 - U_{ij})\varepsilon_{ij}$. Then Y_i with $Y_{ij} = \mathbf{1}_{(Z_{ij} \leq q_{ij})}$ have correlations $r_{ijk} = \bar{r}_{ijk}c_{ijk}(\gamma)$ with $c_{ijk}(\gamma) = \gamma$ ($j \neq k$).
- Scenario 3: MA(1) for $C_i(\gamma)$. The generation mechanism is the same as the Scenario 1 but the Bernoulli parameter is replaced by $[1 - (1 - 4\gamma)^{1/2}]/2$ ($0 \leq \gamma \leq 0.25$) and $Z_{ij} = U_{ij}\varepsilon_{i(j-1)} + (1 - U_{ij})\varepsilon_{ij}$. Then Y_i with $Y_{ij} = \mathbf{1}_{(Z_{ij} \leq q_{ij})}$ have correlations $r_{ijk} = \bar{r}_{ijk}c_{ijk}(\gamma)$ with $c_{ijk}(\gamma) = \gamma$ for $|j - k| = 1$ and $c_{ijk}(\gamma) = 0$ for $|j - k| > 1$.

We apply the proposed iterative joint GEE methods to fit each of these generated data sets respectively. We also consider different structures specified to the correlation matrix $V_i(\alpha)$ in the second GEE $U_2(\gamma, \beta)$. In what follows, we use abbreviated term “AR(1)-CS” structures to represent “AR(1) structure specified to $C_i(\gamma)$ in the first GEE $U_1(\beta, \gamma)$ ”

and CS working correlation structure specified to $V_i(\alpha)$ in the second GEE $U_2(\gamma, \beta)$ ". Similarly, Other types of structures are expressed in similar abbreviations.

Table 1: Simulation results by the joint GEE with AR(1)-Independent, AR(1)-AR(1) structures and AR(1)-CS structures for Scenario 1

	independent	AR(1)			CS		
α	-	0.1	0.4	0.7	0.1	0.4	0.7
$\hat{\beta}$	0.794	0.794	0.795	0.795	0.795	0.796	0.796
sd1($\hat{\beta}$)	0.135	0.135	0.135	0.135	0.135	0.135	0.135
sd2($\hat{\beta}$)	0.134	0.134	0.134	0.134	0.135	0.135	0.135
cr($\hat{\beta}$)	0.955	0.955	0.955	0.954	0.954	0.954	0.954
$\hat{\gamma}$	0.300	0.300	0.300	0.301	0.301	0.301	0.302
sd1($\hat{\gamma}$)	0.042	0.041	0.042	0.055	0.046	0.052	0.054
sd2($\hat{\gamma}$)	0.041	0.040	0.041	0.055	0.046	0.052	0.054
cr($\hat{\gamma}$)	0.957	0.959	0.962	0.955	0.960	0.957	0.959

We set the number of subjects and the number of repeated measurements as $n = 100$ and $m_i = 10$ ($i = 1, \dots, n$), respectively. The true values of the mean parameter and the correlation parameter are taken as $\beta = 0.8$ and $\gamma = 0.3$, separately. The single covariate x_{ij} is generated independently from a uniform distribution $U(0, 1)$ and the mean is defined by $\mu_{ij} = \exp\{x_{ij}\beta\}/(1 + \exp\{x_{ij}\beta\})$. Table 1 summarizes the simulation results for Scenario 1 by using the models with AR(1)-Independent, AR(1)-AR(1) and AR(1)-CS structures, with the parameter $\alpha = 0.1, 0.4$ and 0.7 in $V_i(\alpha)$ that takes the AR(1) and CS structures, respectively. The estimated standard deviation (sd1) of the parameter estimator is calculated through the square root of the asymptotic variance presented in (5). The empirical standard deviation (sd2) is the standard deviation of 1000 parameter estimators for the simulated data. The coverage rate is the percentage of 1000 estimators $\hat{\beta}$ or $\hat{\gamma}$ that fall into 95% confidence interval. From Table 1, the resulting averages of $\hat{\beta}$ and $\hat{\gamma}$ are both very close to the true values. The estimated standard deviation differs little from the empirical standard deviation, showing the validity of the algorithm and the asymptotic property. The coverage rates are around 95.5% and very close to the nominal

level. Moreover, the estimators $\widehat{\beta}$ and $\widehat{\gamma}$ have little change when varying the structures of $V_i(\alpha)$ and changing the value of the parameter α , implying that the estimators of β and γ are robust against the misspecification of the working correlation structure of $V_i(\alpha)$ and its parameter α .

Table 2: Simulation results by the joint GEE with CS-Independent, CS-CS structures and CS-AR(1) structures for Scenario 2

	independent	CS			AR(1)		
α	-	0.1	0.4	0.7	0.1	0.4	0.7
$\widehat{\beta}$	0.799	0.800	0.803	0.810	0.799	0.799	0.801
sd1($\widehat{\beta}$)	0.168	0.168	0.167	0.166	0.168	0.168	0.168
sd2($\widehat{\beta}$)	0.164	0.165	0.166	0.168	0.164	0.164	0.165
cr($\widehat{\beta}$)	0.951	0.950	0.948	0.938	0.951	0.950	0.950
$\widehat{\gamma}$	0.298	0.298	0.296	0.294	0.298	0.298	0.297
sd1($\widehat{\gamma}$)	0.037	0.036	0.056	0.107	0.037	0.037	0.043
sd2($\widehat{\gamma}$)	0.037	0.040	0.058	0.113	0.037	0.036	0.043
cr($\widehat{\gamma}$)	0.952	0.948	0.939	0.945	0.952	0.951	0.947

The simulation results for correlated binary data generated in Scenario 2 by using the joint GEE with CS-Independent, CS-CS and CS-AR(1) structures are presented in Table 2. Again, from Table 2 we see that the average values of the parameter estimators are very close to their true values and the estimators are very stable when changing the working correlation structure for $V_i(\alpha)$ and the value of α as well. But the standard deviation of $\widehat{\gamma}$ increases from 0.04 to 0.1 when α changes from 0.1 to 0.7. It implies that misspecification of the correlation structure for $V_i(\alpha)$ in $U_2(\gamma, \beta)$ has little impact on the estimators of β and γ , but has some effects on the variance of the estimator $\widehat{\gamma}$.

Table 3 presents the simulation results for the binary data under Scenario 3 using the joint GEE with MA(1)-Independent, MA(1)-MA(1) and MA(1)-CS structures, that is, $C_i(\gamma)$ having MA(1) structure and $V_i(\alpha)$ having Independent, MA(1) and CS structures, respectively. We take the same true value $\beta = 0.8$ but choose $\gamma = 0.2$ this time. From Table 3, the conclusions made under Scenario 3 are the same as these under Scenarios 1-

Table 3: Simulation results by the joint GEE with MA(1)-Independent, MA(1)-MA(1) structures and MA(1)-CS structures for Scenario 2

	independent	MA(1)			CS		
α	-	0.1	0.4	0.7	0.1	0.4	0.7
$\widehat{\beta}$	0.795	0.795	0.795	0.795	0.795	0.796	0.797
sd1($\widehat{\beta}$)	0.126	0.126	0.126	0.126	0.126	0.126	0.126
sd2($\widehat{\beta}$)	0.123	0.123	0.123	0.123	0.123	0.123	0.124
cr($\widehat{\beta}$)	0.966	0.966	0.966	0.966	0.966	0.965	0.966
$\widehat{\gamma}$	0.200	0.200	0.200	0.199	0.200	0.199	0.196
sd1($\widehat{\gamma}$)	0.038	0.038	0.040	0.057	0.038	0.043	0.074
sd2($\widehat{\gamma}$)	0.038	0.038	0.040	0.056	0.038	0.043	0.075
cr($\widehat{\gamma}$)	0.943	0.944	0.940	0.959	0.949	0.951	0.949

2. In other words, the joint GEE method yields satisfactory estimators for both β and γ . Also, the misspecification of the correlation matrix $V_i(\alpha)$ does not affect the estimators of β and γ , though the standard deviation of $\widehat{\gamma}$ is somewhat affected.

We now turn to compare the joint GEE method with the single but big GEE approach in (4), aiming to observe the difference of the two GEE methods in terms of estimation performance. We choose three cases $\gamma = 0.1, 0.4$ and 0.7 , reflecting weak, medium and strong correlations specified to $C_i(\gamma)$ when generating correlated binary data under Scenario 1 and Scenario 2, respectively. Since the estimators of β and γ are both robust against misspecification of the working correlation matrix $V_i(\alpha)$ in (3), we only choose an AR(1) structure with $\alpha = 0.1$ for $V_i(\alpha)$ when applying the joint GEE. The two columns named “AR(1)-AR(1)” and “CS-AR(1)” in Table 4 present simulation results of the joint GEE for β and γ with their estimated standard deviations in parenthesis. When using the single but big GEE in (4), we use an AR(1) structure with $\alpha = 0.1$ to be the working correlation matrix for Ω_i^* , and use the method by Liang and Zeger (1986) to obtain the standard deviations of $\widehat{\theta} = (\widehat{\beta}^T, \widehat{\gamma}^T)^T$. The simulation results using single GEE for the binary data generated under two scenarios are provided in the columns with the name “Single GEE” in Table 4.

Table 4: Simulation results: Comparison of the joint GEE method to the single GEE approach, where simulated binary data are generated under Scenario 1 and Scenario 2

		Scenario 1		Scenario 2	
		Single GEE	AR(1)-AR(1)	Single GEE	CS-AR(1)
true $\gamma = 0.1$	$\hat{\beta}$	0.792(0.134)	0.795(0.122)	0.790(0.157)	0.796(0.141)
	$\hat{\gamma}$	0.100(0.040)	0.100(0.039)	0.099(0.029)	0.099(0.028)
true $\gamma = 0.4$	$\hat{\beta}$	0.800(0.162)	0.795(0.141)	0.798(0.202)	0.799(0.175)
	$\hat{\gamma}$	0.399(0.042)	0.400(0.042)	0.398(0.040)	0.398(0.040)
true $\gamma = 0.7$	$\hat{\beta}$	0.800(0.177)	0.802(0.156)	0.802(0.218)	0.807(0.183)
	$\hat{\gamma}$	0.698(0.037)	0.698(0.037)	0.699(0.036)	0.699(0.038)

From Table 4, it is clear that the joint GEE method performs as equally well as the single but big GEE approach. It confirms that it is feasible to separate the big GEE estimation into two relatively small GEE estimation procedures. This strategy largely reduces computational loads especially if m_i is large, because the single GEE involves the calculation of the inverse of Ω_i in (4), a very big matrix with size $(m_i + M_i) \times (m_i + M_i)$ where $M_i = m_i(m_i - 1)/2$. In contrast, the joint GEE calculates the inverses of two matrices with size $m_i \times m_i$ and $M_i \times M_i$, respectively. On the other hand, the joint GEE method produces accurate estimators for β and γ whenever the correlations of Y_i are weak, medium or strong. We notice that the standard deviation of $\hat{\beta}$ may increase slightly with the correlation parameter γ .

Table 5: Simulation results: Comparison of the joint GEE method to the ordinary GEE with two working correlation structures for the binary data generated under Scenario 1

	independent model	CS model	AR(1)-AR(1)
$\hat{\beta}_1$	0.502(0.069)	0.503(0.066)	0.500(0.060)
$\hat{\beta}_2$	-1.720(0.256)	-1.716(0.246)	-1.708(0.221)
$\hat{\beta}_3$	0.020(0.022)	0.020(0.020)	0.019(0.017)
$\hat{\gamma}$	-	-	0.697(0.040)

We now consider the models with multiple covariates by increasing the number of covariates to three, i.e., $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})^T$, and carry out simulation studies in this case. Note the intercept is ignored in the simulation setup. The first covariate x_{ij1} is generated by resampling from $\{1, 2, 3\}$ with replacement. The second one x_{ij2} is taken as the squares of random samples from the uniform distribution on $(0, 1)$, and the third covariate x_{ij3} is a random sample from the standard normal distribution but multiplied by 3. The true values of the mean parameters and the correlation parameter in the correlation matrix $C_i(\gamma)$ are chosen as $\beta = (0.5, -1.7, 0.02)^T$ and $\gamma = 0.7$, respectively. The proposed joint GEE method is now used to analyze the simulated binary data under Scenario 1. In the second GEE, $V_i(\alpha)$ takes AR(1) structure with $\alpha = 0.1$. To compare with existing methods, we also apply the ordinary GEE method by Liang and Zeger (1986) to analyze the same data sets. Specifically, two working correlation structures, independent and CS structures, are used to approximate the correlation matrices $R_i(\gamma)$ in $U_1(\beta, \gamma)$. The parameter estimators are calculated directly by using the R package `geepack` (Højsgaard et al., 2006). These results, together with the ones obtained by the proposed joint GEE method, are presented in Table 5. From Table 5, we observe that the proposed joint GEE method is doing better than the ordinary GEE with the two working correlation structures, independent and CS structures, in terms of not only producing unbiased estimators of the mean parameters β but also yielding an unbiased estimator of the correlation parameter γ . On the other hand, the standard deviations of $\widehat{\beta}$ by the joint GEE method are notably smaller than those by the ordinary GEE. It implies that the joint GEE method improves the efficiency of the mean parameter estimators by the ordinary GEE. Importantly, the proposed GEE method provides the accurate estimator of the correlation parameter γ , which guarantees that the natural upper bounds of correlation coefficients of binary data are satisfied. In contrast, the ordinary GEE method does not address the issue of estimating the correlation coefficients of binary data. Hence, the proposed GEE method performs much better than the ordinary GEE approach when analyzing correlated binary data.

Lastly, we consider the scenarios of unbalanced longitudinal binary data. We generate such data sets by specifying AR(1), CS and MA(1) correlation structures for $C_i(\gamma)$ and using the models aforementioned Scenarios 1-3. For each individual, 10 measurement

Table 6: Simulation results by the joint GEE with AR(1) structure in $V_i(\alpha)$ for the unbalanced longitudinal binary data under Scenario 1-3, including the average, estimated standard deviation (sd1), empirical standard deviation (sd2), and coverage rates(cr)

	AR(1)-AR(1)	CS-AR(1)	MA(1)-AR(1)
true β	0.8	0.8	0.8
$\hat{\beta}$	0.806	0.801	0.803
sd1($\hat{\beta}$)	0.151	0.184	0.141
sd2($\hat{\beta}$)	0.153	0.181	0.142
cr($\hat{\beta}$)	0.954	0.952	0.953
true γ	0.3	0.3	0.2
$\hat{\gamma}$	0.300	0.297	0.200
sd1($\hat{\gamma}$)	0.047	0.045	0.043
sd2($\hat{\gamma}$)	0.047	0.045	0.045
cr($\hat{\beta}$)	0.948	0.944	0.937

times at which observations are taken are randomly sampled from $\{1, 2, \dots, 15\}$ with replacement. Let the true values $\beta = 0.8$ and $\gamma = 0.3$ for AR(1) and CS structures and $\gamma = 0.2$ for MA(1) structure. The single covariate x_{ij} is generated from a uniform distribution $U(0, 1)$. Again, the number of individuals are $n = 100$ and the simulation runs are 1000. When using the proposed joint GEE method, the AR(1) structure with $\alpha = 0.1$ is specified as the working correlation matrix for W_i . The simulation results are presented in Table 6, from which it is clear that the proposed joint GEE method performs also well for unbalanced longitudinal binary data. For example, it leads to the averaged estimators are very close to truths and the associated coverage rates are close to 0.95.

5. Real Data Analysis

5.1. Ohio Data Analysis

The Ohio data set is a subset of a six-city study, which was a longitudinal study of air pollution on health from a Harvard University technical report by Ware et al.

(1984). The data set was analyzed by Fitzmaurice and Laird (1993), which focused on the effect of maternal smoking on children’s respiratory illness. The data set contains binary responses representing the absence or presence of wheezing for 537 children from ages 7 to 10. Maternal smoking status is denoted by 1 meaning that the mother smoked regularly and 0 otherwise. Similar to Fitzmaurice and Laird (1993), maternal smoking is fixed at its first visiting value. The other two covariates are the age of the child and the interaction effect of the child’s age and the mother’s maternal smoking status. The model of the marginal mean of the binary response may be expressed as

$$g(\mu) = \beta_0 + \beta_1 age + \beta_2 smoke + \beta_3 age \times smoke,$$

where $g(\cdot)$ is the logistic link function and “smoke” is the maternal smoking status. When modelling the correlation coefficients of the longitudinal binary data, we propose to use three different types of correlation structures, i.e., AR(1), CS and MA(1), to approximate the normal-model structure in $C_i(\gamma)$ in (2) and also to be the working correlation structure $V_i(\alpha)$ in the second GEE in (3). The working correlation parameter α is taken as 0.1 and 0.7, measuring the possible low or high correlation for W_i . The parameter estimators are presented in Table 7-9, respectively.

Table 7-9 show that the children are less likely to get wheezing as they grow whereas the maternal smoking status increases the possibility of wheezing. This conclusion is in agreement with Fitzmaurice and Laird (1993)’s. On the other hand, we notice that the parameter estimators of $\hat{\beta}$ and $\hat{\gamma}$, particularly their standard deviations, more or less may vary with the specification of the correlation structure specified to $C_i(\gamma)$ and $V_i(\alpha)$, and the parameter values therein, due to the finite sample issue. In order to select appropriate correlation structures for both $C_i(\gamma)$ and $V_i(\alpha)$, the quasi-likelihood under independence model criterion (QIC) introduced by Pan (2001) is applied. From the simulation studies, see Table 4, it is concluded that the mean parameter β and the correlation parameter γ can be estimated well by using the proposed GEEs, so that their model selection criteria can be considered separately as briefed below. Let I denote the independent model where the working correlation matrix takes the identity matrix, and the QIC for $\hat{\beta}$ is defined as follows

$$\text{QIC}_\beta = -2Q\left(\hat{\beta}(R); I\right) + 2\text{trace}\left(\hat{\Omega}_I^{-1}\widehat{\text{cov}}(\hat{\beta}(R))\right), \quad (6)$$

where the quasi-likelihood function $Q(\hat{\beta}(R); I)$ is calculated using the independent covariance structure but the mean parameter β is replaced by $\hat{\beta}(R)$, which is obtained from a general working correlation structure R . The matrices $\hat{\Omega}_I$ and $\widehat{cov}(\hat{\beta}(R))$ are the estimators of covariance matrix of $\hat{\beta}$ under the working correlation matrices I and R , respectively. With the logistic link, the quasi-likelihood function can be written as

$$Q(\hat{\beta}(R); I) = \sum_{i=1}^n \sum_{j=1}^{m_i} \left[Y_{ij} \log \frac{\hat{\mu}_{ij}}{1 - \hat{\mu}_{ij}} + \log(1 - \hat{\mu}_{ij}) \right],$$

where $\hat{\mu}_{ij}$ is the fitted value of μ_{ij} using the estimated coefficients in the GEE model with the correlation matrix R . Similarly, the QIC for $\hat{\gamma}$ takes the form

$$\text{QIC}_\gamma = -2 \sum_{i=1}^n \sum_{j=1}^{M_i} \left[W_{ij} \log \frac{\hat{\xi}_{ij}}{1 - \hat{\xi}_{ij}} + \log(1 - \hat{\xi}_{ij}) \right] + 2 \text{trace} \left(\tilde{\Omega}_I^{-1} \widehat{cov}(\hat{\gamma}) \right),$$

where $\tilde{\Omega}_I$ and $\widehat{cov}(\hat{\gamma})$ are the estimators of the covariance matrices of $\hat{\gamma}$ with working correlation matrices I and $V_i(\alpha)$, respectively. As a general rule, the best correlation structures for $C_i(\gamma)$ and $V_i(\alpha)$ are the ones corresponding to the smallest values of QIC_β and QIC_γ .

Table 7: Ohio Data Analysis: Parameter estimates (standard deviations in parentheses) when $C_i(\gamma)$ takes an AR(1) structure and $V_i(\alpha)$ takes the structures AR(1), CS and MA(1)

$V_i(\alpha)$	AR(1)		CS		MA(1)	
α	0.1	0.7	0.1	0.7	0.1	0.7
$\hat{\beta}_0$	-1.92(0.12)	-1.94(0.14)	-1.92(0.11)	-1.94(0.14)	-1.92(0.12)	-1.92(0.11)
$\hat{\beta}_1$	-0.15(0.07)	-0.15(0.06)	-0.15(0.07)	-0.15(0.06)	-0.15(0.07)	-0.15(0.07)
$\hat{\beta}_2$	0.29(0.18)	0.26(0.22)	0.29(0.18)	0.25(0.23)	0.29(0.18)	0.30(0.17)
$\hat{\beta}_3$	0.08(0.11)	0.09(0.09)	0.08(0.11)	0.089(0.09)	0.08(0.11)	0.08(0.11)
$\hat{\gamma}$	0.46(0.06)	0.80(0.06)	0.44(0.06)	0.83(0.05)	0.45(0.06)	0.35(0.08)
QIC_β	1831.13	1834.71	1831.00	1835.11	1831.11	1830.20
QIC_γ	2926.24	3174.13	2926.57	3227.92	2926.24	2941.11

From Table 7-9, it is clear that the joint GEE method with AR(1) structure specified to $C_i(\gamma)$ has large values in both QIC_β and particularly QIC_γ when compared to the method with $C_i(\gamma)$ having CS or MA(1) structure. See the values of QIC_β and QIC_γ across Tables 7-9 under each setting of $V_i(\alpha)$. In addition, we notice that the estimator

Table 8: Ohio Data Analysis: Parameter estimates (standard deviations in parentheses) when $C_i(\gamma)$ takes a CS structure and $V_i(\alpha)$ takes the structures AR(1), CS and MA(1)

$V_i(\alpha)$	AR(1)		CS		MA(1)	
	0.1	0.7	0.1	0.7	0.1	0.7
$\widehat{\beta}_0$	-1.90(0.12)	-1.90(0.12)	-1.90(0.12)	-1.90(0.12)	-1.90(0.12)	-1.90(0.12)
$\widehat{\beta}_1$	-0.14(0.06)	-0.14(0.06)	-0.14(0.06)	-0.14(0.06)	-0.14(0.06)	-0.14(0.06)
$\widehat{\beta}_2$	0.31(0.19)	0.31(0.19)	0.31(0.19)	0.31(0.19)	0.31(0.19)	0.31(0.19)
$\widehat{\beta}_3$	0.07(0.09)	0.07(0.09)	0.07(0.09)	0.07(0.09)	0.07(0.09)	0.07(0.09)
$\widehat{\gamma}$	0.39(0.06)	0.39(0.06)	0.39(0.06)	0.38(0.06)	0.39(0.06)	0.42(0.06)
QIC_β	1830.45	1830.47	1830.45	1830.42	1830.46	1830.68
QIC_γ	2901.52	2902.47	2901.52	2902.45	2901.52	2905.07

Table 9: Ohio Data Analysis: Parameter estimates (standard deviations in parentheses) when $C_i(\gamma)$ takes an MA(1) structure and $V_i(\alpha)$ takes the structures AR(1), CS and MA(1)

$V_i(\alpha)$	AR(1)		CS		MA(1)	
	0.1	0.7	0.1	0.7	0.1	0.7
$\widehat{\beta}_0$	-1.92(0.11)	-1.92(0.11)	-1.92(0.11)	-1.91(0.10)	-1.92(0.11)	-1.91(0.10)
$\widehat{\beta}_1$	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)	-0.15(0.07)
$\widehat{\beta}_2$	0.30(0.17)	0.30(0.17)	0.30(0.17)	0.31(0.16)	0.30(0.17)	0.31(0.16)
$\widehat{\beta}_3$	0.09(0.12)	0.09(0.12)	0.09(0.12)	0.08(0.12)	0.09(0.12)	0.08(0.12)
$\widehat{\gamma}$	0.39(0.06)	0.38(0.06)	0.39(0.06)	0.29(0.05)	0.39(0.06)	0.29(0.15)
QIC_β	1830.18	1830.13	1830.18	1829.54	1830.18	1829.54
QIC_γ	1417.74	1417.82	1417.76	1423.32	1417.74	1447.51

$\hat{\gamma}$ with $C_i(\gamma)$ taking an AR(1) structure is not as stable as the one with CS or MA(1) structure when varying α in $V_i(\alpha)$. We thus conclude that AR(1) structure is not a good working structure for $C_i(\gamma)$ in this case. In contrast, the joint GEE method with $C_i(\gamma)$ having CS or MA(1) structure is doing better in the sense of relatively small values of QIC_β and QIC_γ . In these two cases, the resulting mean parameter estimators are very close to these by Fitzmaurice and Laird (1993). We also observe that in these cases $\hat{\gamma}$ remains stable when α increases from 0.1 to 0.7, with an exception that a MA(1) structure with $\alpha = 0.7$ is specified to $V_i(\alpha)$, see Tables 8-9. The reason is that $\alpha = 0.7$ is almost identical to the boundary value 0.7071 of MA(1), and beyond this value the MA(1) covariance matrix is no longer positive definite. Our simulations show that the estimator $\hat{\gamma}$ changes substantially when α in MA(1) is more than 0.65. Note that the QIC_β 's from the model with MA(1) specified to $C_i(\gamma)$ are very close to those by the model with $C_i(\gamma)$ having a CS structure, implying that MA(1) is also a good structure for $C_i(\gamma)$. When estimating the parameter γ , as indicated by QIC_γ either AR(1), CS or MA(1) structure can be specified to $V_i(\alpha)$ as long as a small value of α is chosen (e.g., $\alpha = 0.1$), where $C_i(\gamma)$ takes either CS or MA(1) structure. In this case, both the parameter estimators $\hat{\beta}$ and $\hat{\gamma}$, as well as their standard deviations, are robust against the working correlation structure in the second GEE. The resulting estimator $\hat{\gamma}$ of the correlation parameter in the normal-model for this data set is around 0.4 with standard deviation 0.06.

5.2. CGI Data Analysis

The Clinical Global Impression (CGI) is widely accepted measurement of illness severity in a variety of psychiatric disorders. The CGI data set we analyse contains longitudinal binary responses representing if the CGI improvement (CGI-I) is responded (1 for yes and 0 otherwise) in a follow-up study after certain treatments. In total, there were 454 patients whose number of repeated measurements varies from 1 to 7, leading to 2805 unbalanced longitudinal observations.

The explanatory variables include: treatment (TRX), Race, Sex, Age, Body Mass Index (BMI), Severity of Illness (SEV) and Time. The TRX contains three types of treatment, which are Dose 1, Dose 2 and Placebo denoted by TRX1, TRX2, TRX3 respectively. The Race contains White, Black, Oriental and Other denoted by Race1,

Race2, Race3 and Race4 respectively. As the TRX and Race each have more than two categories, so that a dummy variable is introduced to each category to represent the category's appearance/disappearance. Sex is set a binary variable, 0 for female and 1 for male. The SEV is a variable taking the numbers from 0 to 7, where 0 refers to "Not assessed" and 1 to 7 refer to "Not at ill" to "Extremely ill". The aim is to assess the treatment effects on the CGI-I responses, by taking into account of other explanatory variables including Race, Sex, Age, BMI, SEV and Time. A logistic model

$$\begin{aligned} \text{logit}(\mu) = & \beta_0 + \beta_1 TRX1 + \beta_2 TRX2 + \beta_3 Race2 + \beta_4 Race3 + \beta_5 Race4 + \beta_6 Sex \\ & + \beta_7 Age + \beta_8 BMI + \beta_9 SEV + \beta_{10} Time + \beta_{11} TRX1 \times Sex + \beta_{12} TRX2 \times Sex, \end{aligned}$$

is use to model the mean for the CGI data. The proposed joint GEE methods with AR(1) and CS specified to $C_i(\gamma)$ and AR(1), CS and MA(1) to $V_i(\alpha)$ are applied. Table 10 presents the numerical results for the CGI data analysis with $\alpha = 0.1$. Other values of α were also considered and the results are very similar and so omitted here.

From Table 10, the estimators of the mean parameters β , as well as their standard deviations, are close whatever structures are specified to matrices $C_i(\gamma)$ and $V_i(\alpha)$, implying that they are robust against misspecification of the structures of $C_i(\gamma)$ and $V_i(\alpha)$. In contrast, the estimator of the correlation parameter γ seems to have some changes with different structures specified to $C_i(\gamma)$ but is robust against to the choice of the structure for $V_i(\alpha)$. This is reasonable because γ in $C_i(\gamma)$ represents a different degree of correlation in different correlation structures.

On the other hand, in terms of the QIC values the model with AR(1)-CS structure yields the smallest value of QIC_β (1390.85) for the mean parameter β while the model with CS-CS structure achieves the lowest value of QIC_γ (4314.03) for the correlation parameter γ . For the overall performance of both the estimators of β and γ , we suggest to consider $QIC_{\beta,\gamma} = \omega QIC_\beta + (1 - \omega)QIC_\gamma$, where ω is a weight taking value in $[0, 1]$, measuring either the mean or correlation is more important. Since the mean and correlation are treated equally important in this paper, it is reasonable to set $\omega = 0.5$. As a result, the model with AR(1)-CS structure, meaning $C_i(\gamma)$ and $V_i(\alpha)$ specified with AR(1) and CS structures respectively, turns to be the best one, since it has the smallest $QIC_{\beta,\gamma}$ (2853.13). Based on the model with AR(1)-CS structure, the results presented in Table 10 show that, for the females Dose 2 seems likely to be more effective

Table 10: CGI data. Parameter estimates(standard deviances) based on the models with AR(1) and CS specified to $C_i(\gamma)$ and AR(1), CS and MA(1) to $V_i(\alpha)$ where $\alpha = 0.1$

structure	AR(1)-AR(1)	AR(1)-CS	AR(1)-MA(1)	CS-AR(1)	CS-CS	CS-MA(1)
Intercept	9.88(0.86)	9.92(0.86)	9.88(0.86)	9.88(0.85)	9.96(0.85)	9.90(0.86)
TRX1	-0.25(0.32)	-0.25(0.31)	-0.25(0.32)	-0.16(0.31)	-0.17(0.31)	-0.16(0.32)
TRX2	0.22(0.32)	0.22(0.31)	0.22(0.32)	0.25(0.31)	0.25(0.30)	0.25(0.31)
Race2	0.03(0.31)	0.02(0.30)	0.03(0.31)	-0.02(0.31)	-0.02(0.30)	-0.02(0.31)
Race3	-0.94(0.66)	-0.97(0.65)	-0.93(0.66)	-0.48(0.65)	-0.54(0.64)	-0.52(0.65)
Race4	-0.09(0.32)	-0.08(0.32)	-0.10(0.32)	0.27(0.32)	0.25(0.32)	0.27(0.32)
Sex	-0.07(0.35)	-0.07(0.34)	-0.07(0.35)	-0.06(0.35)	-0.07(0.35)	-0.06(0.35)
Age	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)
BMI	-0.01(0.01)	-0.01(0.01)	-0.01(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)
SEV	-3.11(0.18)	-3.11(0.18)	-3.11(0.18)	-3.03(0.18)	-3.05(0.18)	-3.04(0.18)
Time	0.15(0.03)	0.15(0.03)	0.15(0.03)	0.13(0.03)	0.13(0.03)	0.13(0.03)
TRX1×Sex	-0.25(0.48)	-0.25(0.47)	-0.25(0.48)	-0.30(0.48)	-0.29(0.47)	-0.31(0.48)
TRX2×Sex	-1.07(0.49)	-1.08(0.48)	-1.06(0.49)	-1.10(0.49)	-1.10(0.49)	-1.10(0.49)
$\hat{\gamma}$	0.78(0.09)	0.74(0.07)	0.78(0.09)	0.65(0.11)	0.62(0.08)	0.65(0.11)
QIC_{β}	1392.48	1390.85	1392.55	1397.58	1395.63	1397.06
QIC_{γ}	4325.96	4315.41	4326.37	4326.21	4314.03	4325.61
$QIC_{\beta,\gamma}$	2859.22	2853.13	2859.46	2861.895	2854.83	2861.335

than Placebo because the estimated coefficient of TRX2 is positive although it is not statistically significant, while for the males Dose 2 is not as effective as Placebo. An interesting phenomenon is that both the female and male patients in Dose 1 (TRX1) are less likely to respond their CGI improvements than those in Placebo group. It seems that the black patients have little difference with the white patients in responding the CGI improvements, while the oriental patients and others are less likely to respond than white patients. Females are more likely to respond than males. Increasing Age, BMI or SEV can reduce the probability to respond the treatment, whereas the responding becomes more likely with time increasing. Among those, SEV and Time are significant covariates. Note that the parameter estimator corresponding to Age being -0.00 is actually -1.8×10^{-3} here. Furthermore, $\hat{\gamma} = 0.74$ with standard deviation 0.07, leading the maximum within-subject correlation coefficient between two adjacent binary measurements to 0.69, implies that the correlation coefficient for the longitudinal binary data is strong and should not be simply ignored.

6. Discussions

In this paper we proposed a GEE-based approach to estimate the mean parameters and within-subject correlation parameters, simultaneously, for longitudinal binary data. We also considered the asymptotic normality of the parameter estimators under certain mild conditions. A parametric form of the within-subject correlation coefficients is considered so that the resulting correlation coefficients for correlated binary data satisfy their natural upper bounds. Such correlation coefficients are connected to $C(\gamma)$, the correlation matrix from a latent normal-model so that the modelling reduces to modelling of the normal-model correlation coefficients. The choice of the structure for $C(\gamma)$ is important because as we already see from the real data analysis in the previous section it may affect the estimator of γ and particularly its standard deviation. We proposed to use the QIC to select the most appropriate correlation structure for $C(\gamma)$, but alternative criteria may be good too. For the correlation matrix $V(\alpha)$ of the augmented random variables W , we suggested to use independent, CS or AR(1) structure as its working correlation structure. The reason is that the inverse of $V(\alpha)$ has an explicit form in these cases, so that its calculation becomes straightforward, which is helpful especially

when the dimension of $V(\alpha)$ is high. Also, the estimators of mean parameters β and correlation coefficient γ are robust against the misspecification of the working structure of $V(\alpha)$, indicating that the proposed method works efficiently.

Acknowledgements

We would like to thank the editors and referees for their very constructive suggestions.

Appendix A. Proof of asymptotic normality

PROOF OF THEOREM 1. According to Taylor expansion, we can have

$$\begin{bmatrix} n^{1/2}(\hat{\beta} - \beta) \\ n^{1/2}(\hat{\gamma} - \gamma) \end{bmatrix} = \begin{bmatrix} -n^{-1} \frac{\partial U_1(\beta, \gamma)}{\partial \beta^T} & -n^{-1} \frac{\partial U_1(\beta, \gamma)}{\partial \gamma} \\ -n^{-1} \frac{\partial U_2(\gamma, \beta)}{\partial \beta^T} & -n^{-1} \frac{\partial U_2(\gamma, \beta)}{\partial \gamma} \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2} U_1(\beta, \gamma) \\ n^{-1/2} U_2(\gamma, \beta) \end{bmatrix} + o_p(n). \quad (\text{A.1})$$

The first element of the inverse matrix in (A.1) can be expressed as

$$\begin{aligned} -n^{-1} \frac{\partial U_1(\beta, \gamma)}{\partial \beta^T} &= -n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \beta^T} (D_i^T \Sigma_i^{-1} (Y_i - \mu_i)) \\ &\quad - n^{-1} \sum_{i=1}^n D_i^T \frac{\partial}{\partial \beta^T} (\Sigma_i^{-1}) (Y_i - \mu_i) \\ &\quad + n^{-1} \sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i. \end{aligned} \quad (\text{A.2})$$

The first two terms of right hand in (A.2) are $o_p(1)$ if we assume their variances are finite, because they involve the sum of n independent variables with zero mean. In consequence, only the third term dominates and

$$-n^{-1} \frac{\partial U_1(\beta, \gamma)}{\partial \beta^T} = n^{-1} \sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i + o_p(1).$$

In a similar manner, we have

$$\begin{aligned} -n^{-1} \frac{\partial U_1(\beta, \gamma)}{\partial \gamma} &= o_p(1), \\ -n^{-1} \frac{\partial U_2(\gamma, \beta)}{\partial \beta^T} &= n^{-1} \sum_{i=1}^n E_i^T G_i^{-1} \frac{\partial \xi_i}{\partial \beta^T} + o_p(1), \\ -n^{-1} \frac{\partial U_2(\gamma, \beta)}{\partial \gamma} &= n^{-1} \sum_{i=1}^n E_i^T G_i^{-1} E_i + o_p(1). \end{aligned}$$

So the inverse matrix in (A.1) converges to the matrix

$$n \begin{pmatrix} \sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i & 0 \\ \sum_{i=1}^n E_i^T G_i^{-1} \frac{\partial \xi_i}{\partial \beta^T} & \sum_{i=1}^n E_i^T G_i^{-1} E_i \end{pmatrix}^{-1},$$

as $n \rightarrow \infty$. On the other hand, the score functions $[n^{-1/2}U_1(\beta, \gamma)^T, n^{-1/2}U_2(\gamma, \beta)^T]^T$ have an asymptotic normal distribution with zero mean and covariance matrix

$$\lim_{n \rightarrow \infty} n^{-1} \begin{pmatrix} \sum_{i=1}^n D_i^T \Sigma_i^{-1} \text{cov}(Y_i) \Sigma_i^{-1} D_i & \sum_{i=1}^n D_i^T \Sigma_i^{-1} \text{cov}(Y_i, W_i) G_i^{-1} E_i \\ \sum_{i=1}^n E_i^T G_i^{-1} \text{cov}(W_i, Y_i) \Sigma_i^{-1} D_i & \sum_{i=1}^n E_i^T G_i^{-1} \text{cov}(W_i) G_i^{-1} E_i \end{pmatrix}.$$

It is assumed that we model the covariance of Y_i correctly. Thus, the first element of the asymptotic covariance matrix reduces to $\sum_{i=1}^n D_i^T \Sigma_i^{-1} D_i$. Denote the Q s and A s as introduced in the Theorem and then the proof is completed.

References

- Carey, V., Zeger, S., Diggle, P., 1993. Modelling multivariate binary data with alternating logistic regression. *Biometrika* 80, 517–526.
- Chaganty, N., Joe, H., 2004. Efficiency of generalized estimating equations for binary responses. *J.R.Statist.Soc. B* 66, 851 – 860.
- Crowder, M., 1995. On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* 82, 407–410.
- Daniels, M., Zhao, Y., 2003. Modelling the random effects covariance matrix in longitudinal data. *Statistics in medicine* 22(10), 16311647.
- Diggle, P., Verbyla, A., 1998. Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* 54, 401–415.
- Dziak, J., 2006. Variable selection for longitudinal data by penalized quadratic inference functions. Ph.d. dissertation. Department of statistics, The Pennsylvania State University.
- Fitzmaurice, G., Laird, N., 1993. A likelihood-based method for analysing longitudinal binary responses. *Biometrika* 80, 141–151.
- Heagerty, P., 1999. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* 55, 688–698.
- Højsgaard, S., Halekoh, U., Yan, J., 2006. The *r* package *geepack* for generalized estimating equations. *J. Stat. Softw.* 15, 1–11.
- Liang, K., Zeger, S., 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.
- Oman, S., Zucker, D., 2001. Modelling and generating correlated binary data. *Biometrika* 88, 287–290.
- Pan, W., 2001. Akaike’s information criterion in generalized estimating equations. *Biometrics* 57, 120–125.

- Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G., Mallick, B., Ibrahim, J., 2011. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *Ann.Appl.Stat.* 5(1), 449–467.
- Prentice, R., 1988. Correlated binary regression with covariates specific to each binary observation. *Biometrics* 44, 1033–1048.
- Prentice, R., Zhao, L., 1991. Estimating equations for generalised linear models for multivariate normal covariance matrix. *Biometrics* 47, 825–839.
- Sutradhar, B., Das, K., 1999. On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* 86, 459–465.
- Sutradhar, B., Das, K., 2000. On the accuracy of efficiency of estimating equation approach. *Biometrics* 56(2), 622–625.
- Wang, L., Zhou, J., Qu, A., 2012. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* 68, 353 – 360.
- Wang, Z., Louis, T., 2003. Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika* 90, 765–775.
- Ware, J., Dockery, D., Spiro, A., Speizer, F., Ferris, B.J., 1984. Passive smoking, gas cooking and respiratory health in children living in six cities. *Am. Rev. Respir. Dis.* 129, 366–374.
- Ye, H., Pan, J., 2006. Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* 93, 927–941.