Review Article

Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments

Ming Wang

Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA 17033, USA

Correspondence should be addressed to Ming Wang; mwang@phs.psu.edu

Generalized Estimating Equation (GEE) is a marginal model popularly applied for longitudinal/clustered data analysis in clinical trials or biomedical studies. We provide a systematic review on GEE including basic concepts as well as several recent developments due to practical challenges in real applications. The topics including the selection of "working" correlation structure, sample size and power calculation, and the issue of informative cluster size are covered because these aspects play important roles in GEE utilization and its statistical inference. A brief summary and discussion of potential research interests regarding GEE are provided in the end.

1. Introduction

Generalized Estimating Equation (GEE) is a general statistical approach to fit a marginal model for longitudinal/clustered data analysis, and it has been popularly applied into clinical trials and biomedical studies [1–3]. One longitudinal data example can be taken from a study of orthodontic measurements on children including 11 girls and 16 boys. The response is the measurement of the distance (in millimeters) from the center of the pituitary to the pterygomaxillary fissure, which is repeatedly measured at ages 8, 10, 12, and 14 years. The primary goal is to investigate whether there exists significant gender difference in dental growth measures and the temporal trend as age increases [4]. For such data analysis, it is obvious that the responses from the same individual tend to be "more alike"; thus incorporating within-subject and between-subject variations into model fitting is necessary to improve efficiency of the estimation and the power [5].

There are several simple methods existing for repeated data analysis, that is, ANOVA/MANOVA for repeated measures, but the limitation is the incapability of incorporating covariates. There are two types of approaches, mixed-effect models and GEE [6, 7], which are traditional and are widely used in practice now. Of note is that these two methods have different tendencies in model fitting depending on the study objectives. In particular, the mixed-effect model is an individual-level approach by adopting random effects to capture the correlation between the observations of the same subject [7]. On the other hand, GEE is a population-level approach based on a quasilikelihood function and provides the population-averaged estimates of the parameters [8]. In this paper, we focus on the latter to provide a review and recent developments of GEE. As is well known, GEE has several defining features [9–11]. (1) The variance-covariance matrix of responses is treated as nuisance parameters in GEE and thus this model fitting turns out to be easier than mixed-effect models [12]. In particular, if the overall treatment effect is of primary interest, GEE is preferred. (2) Under mild regularity conditions, the parameter estimates are consistent and asymptotically normally distributed even when the "working" correlation structure of responses is misspecified, and the variance-covariance matrix can be estimated by robust "sandwich" variance estimator. (3) GEE relaxes the distribution assumption and only requires the correct specification of marginal mean and variance as well as the link function which connects the covariates of interest and marginal means.

However, several aspects of GEE are still in controversy since Liang and Zeger [6]. Crowder addressed some issues on inconsistent estimation of within-subject correlation

coefficient under a misspecified "working" correlation structure based on asymptotic theory [7]. In addition, the estimation of the correlation coefficients using the moment-based approach is not efficient; thus the correlation matrix may not be a positive definite matrix in certain cases. Also, Liang and Zeger did not incorporate the constraints on the range of correlation which was restricted by the marginal means because the estimation of the correlation coefficients was simply based on Pearson residuals [6]. Chaganty and Joe discussed this issue for dependent Bernoulli random variables [13], and later Sabo and Chaganty made future explanation [14]. For example, Sutradhar and Das pointed out under misspecification the correlation coefficient estimates did not converge to the true values [15]. Furthermore, for discrete random vectors, the correlation matrix was usually complicated, and it was not easy to attain multivariate distributions with specified correlation structures. These limitations lead researchers to actively work on this area to develop novel methodologies. Several alternative approaches for estimating the correlation coefficients have been proposed; for example, one method was based on "Gaussian" estimation [16, 17], and the basic idea was to estimate the correlation coefficients based on multivariate normal estimating equations, and the feature was that this estimation can ensure the estimated correlation matrix was positive-definite. Wang and Carey proposed to estimate the correlation coefficients by differentiating the Cholesky decomposition of the working correlation matrix [18]. Also, Qu and Lindsay (2003) proposed similar Gaussian or quadratic estimating equations [19]. In particular, for binary longitudinal data, the estimation of the correlation coefficients was proposed based on conditional residuals [20–22]. Nevertheless, in this paper, the above issues are not discussed in great depth, and the assumption that, under the regular mild conditions, the consistency of parameter estimates as well as within-subject correlation coefficient estimate holds is satisfied. Thus, three specific topics including model selection, power analysis, and the issue of informative cluster size are mainly focused on and the recent developments are reviewed in the following sections.

2. Method

2.1. Notation and GEE. Suppose that longitudinal/clustered data consists of K subjects/clusters. For subject/cluster i ($i = 1, 2, \ldots, K$), suppose that there are n_i observations and Y_{ij} denotes the jth response ($j = 1, \ldots, n_i$), and let X_{ij} denote a $p \times 1$ vector of covariates. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$ denote the response vector for the ith subject with the mean vector noted by $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i})'$ where μ_{ij} is the corresponding jth mean. The responses are assumed to be independent across subjects/clusters but correlated within each subject/cluster. The marginal model specifies that a relationship between μ_{ij} and the covariates X_{ij} is written as follows:

$$g\left(\mu_{ij}\right) = X'_{ij}\boldsymbol{\beta}, \tag{1}$$

where g is a known link function and $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of regression coefficients with the true value

as $\boldsymbol{\beta}_0$. The conditional variance of Y_{ij} given X_{ij} is specified as $\text{Var}(Y_{ij} \mid X_{ij}) = v(\mu_{ij})\phi$, where v is a known variance function of μ_{ij} and ϕ is a scale parameter which may need to be estimated. Mostly, v and ϕ depend on the distributions of outcomes. For instance, if Y_{ij} is continuous, $v(\mu_{ij})$ is specified as 1, and ϕ represents the error variance; if Y_{ij} is count, $v(\mu_{ij}) = \mu_{ij}$, and ϕ is equal to 1. Also, the variance-covariance matrix for Y_i is noted by $V_i = \phi A_i^{1/2} R_i(\boldsymbol{\alpha}) A_i^{1/2}$, where $A_i = \text{Diag}\{v(\mu_{i1}), \ldots, v(\mu_{in_i})\}$ and the so-called "working" correlation structure $R_i(\boldsymbol{\alpha})$ describes the pattern of measures within subject, which is of size $n_i \times n_i$ and depends on a vector of association parameters denoted by $\boldsymbol{\alpha}$. Table 1 provides summary of commonly used "working" correlation structures with the moment-based estimates for $\boldsymbol{\alpha}$ (more details in http://www.okstate.edu/sas/). Note that the iterative algorithm is applied for estimating $\boldsymbol{\alpha}$ using the Pearson residuals $e_{ij} = (y_{ij} - \mu_{ij})/\sqrt{v(\mu_{ij})}$ calculated from the current value of $\boldsymbol{\beta}$. Also, the scale parameter ϕ can be estimated by

$$\widehat{\phi} = \frac{1}{N - p} \sum_{i=1}^{K} \sum_{j=1}^{n_i} e_{ij}^2, \tag{2}$$

where $N = \sum_{i=1}^{K} n_i$ is the total number of observations and p is covariates dimensionality.

Based on Liang and Zeger [6], GEE yields asymptotically consistent $\widehat{\boldsymbol{\beta}}$ even when the "working" correlation structure ($R_i(\boldsymbol{\alpha})$) is misspecified, and the estimate of $\boldsymbol{\beta}$ is obtained by solving the following estimating equation:

$$U\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{K} D'_i V_i^{-1} \left(Y_i - \mu_i\right) = 0, \tag{3}$$

where $D_i = \partial \mu_i / \partial \boldsymbol{\beta}'$. Under mildregularity conditions, $\widehat{\boldsymbol{\beta}}$ is asymptotically normally distributed with a mean $\boldsymbol{\beta}_0$ and a covariance matrix estimated based on the sandwich estimator

$$\widehat{V}_{LZ} = \left(\sum_{i=1}^{K} D'_i V_i^{-1} D_i\right)^{-1} \widehat{M}_{LZ} \left(\sum_{i=1}^{K} D'_i V_i^{-1} D_i\right)^{-1} \tag{4}$$

with

$$\widehat{M}_{LZ} = \sum_{i=1}^{K} D'_i V_i^{-1} \text{Cov}\left(Y_i\right) V_i^{-1} D_i \tag{5}$$

by replacing $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and ϕ with their consistent estimates, where $\text{Cov}(Y_i) = \widehat{r}_i \widehat{r}'_i$ with $\widehat{r}_i = Y_i - \widehat{\mu}_i$ is an estimator of the variance-covariance matrix of Y_i [6, 23]. This "sandwich" estimator is robust in that it is consistent even if the correlation structure (V_i) is misspecified. Note that if V_i is correctly specified, then $\widehat{V}_{LZ}$ reduces to $(\sum_{i=1}^{K} D'_i V_i^{-1} D_i)^{-1}$, which is often referred to as the model-based variance estimator [24]. Thus, a Wald Z-test can be performed based on asymptotic normal distribution of the test statistic. Next, we will overview model selection criteria and particularly "working" correlation structure selection criteria with regard to GEE.

TABLE 1: Summary of commonly used "working" correlation structures for GEE.

Correlation structure	$\mathrm{Corr}(Y_{ij}, Y_{ik})$	Sample matrix	Estimator
Independent	$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	NA
Exchangeable	$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{pmatrix}$	$\widehat{\alpha} = \dfrac{1}{(N' - p)\phi} \sum\limits_{i=1}^{K} \sum\limits_{j \neq k} e_{ij} e_{ik}$ $N' = \sum\limits_{i=1}^{K} n_i (n_i - 1)$
k-dependent	$\mathrm{Corr}(Y_{ij}, Y_{i,j+m}) = \begin{cases} 1 & m = 0 \\ \alpha_m & m = 1, 2, \ldots, k \\ 0 & m > k \end{cases}$	$\begin{pmatrix} 1 & \alpha_1 & 0 \\ \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_1 & 1 \end{pmatrix}$	$\widehat{\alpha}_m =$ $\dfrac{1}{(K_m - p)\phi} \sum\limits_{i=1}^{K} \sum\limits_{j \leq n_i - m} e_{ij} e_{i,j+t}$ $K_t = \sum\limits_{i=1}^{K} (n_i - m)$
Autoregressive AR(1)	$\mathrm{Corr}(Y_{ij}, Y_{i,j+m}) = \alpha^m,\ m = 0, 1, 2, \ldots, n_i - j$	$\begin{pmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{pmatrix}$	$\widehat{\alpha} = \dfrac{1}{(K_1 - p)\phi} \sum\limits_{i=1}^{K} \sum\limits_{j \leq n_i - 1} e_{ij} e_{i,j+1}$ $K_1 = \sum\limits_{i=1}^{K} (n_i - 1)$
Toeplitz	$\mathrm{Corr}(Y_{ij}, Y_{i,j+m}) = \begin{cases} 1 & m = 0 \\ \alpha_m & m = 1, 2, \ldots, n_i - j \end{cases}$	$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ \alpha_1 & 1 & \alpha_1 \\ \alpha_2 & \alpha_2 & 1 \end{pmatrix}$	$\widehat{\alpha} = \dfrac{1}{(N' - p)\phi} \sum\limits_{i=1}^{K} \sum\limits_{j \neq k} e_{ij} e_{ik}$ $N' = \sum\limits_{i=1}^{K} n_i (n_i - 1)$
Unstructured	$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 1 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 1 \end{pmatrix}$	$\widehat{\alpha}_{jk} = \dfrac{1}{(K - p)\phi} \sum\limits_{i=1}^{K} e_{ij} e_{ik}$

2.2. Model Selection of GEE. In this section, we will discuss the model selection criteria available of GEE. There are several reasons why model selection of GEE models is important and necessary: (1) GEE has gained increasing attention in biomedical studies which may include a large group of predictors [25–28]. Therefore, variable selection is necessary for determining which are included in the final regression model by identifying significant predictors; (2) it is already known that one feature of GEE is that the consistency of parameter estimates can still hold even when the "working" correlation structure is misspecified. But, correctly specifying "working" correlation structure can definitely enhance the efficiency of the parameter estimates in particular when the sample size is not large enough [16, 24, 25, 29]. Therefore, how to select intrasubject correlation matrix plays a vital role in GEE with improved finite-sample performance; (3) the variance function $v(\mu)$ is another potential factor affecting the goodness-of-fit of GEE [25, 30]. Correctly specified variance function can assist in the selection of covariates and an appropriate correlation structure [31, 32]. Different criteria might be needed due to the goal of model selection [24, 29, 33], and next I will particularly introduce the existing approaches on the selection of "working" correlation structure with its own merits and limitations [34].

According to Rotnitzky and Jewell, the adequacy of "working" correlation structure can be examined through

$\Gamma = (\sum_{i=1}^{K} D_i' V_i^{-1} D_i)^{-1} \widehat{M}_{LZ}$, where $\widehat{M}_{LZ}$ has been defined in Section 2.1 [35]. The statistic RJ(R) is defined by

$$\mathrm{RJ}(R) = \sqrt{(1 - \mathrm{RJ1})^2 + (1 - \mathrm{RJ2})^2}, \tag{6}$$

where $\mathrm{RJ1} = \mathrm{trace}(\Gamma)/p$ and $\mathrm{RJ2} = \mathrm{trace}(\Gamma^2)/p$, respectively. If the "working" correlation structure R is correctly specified, RJ1 and RJ2 will be thus close to 1, leading to RJ(R) approaching 0. Thus, RJ1, RJ2, and RJ(R) can all be used for correlation structure selection.

Shults and Chaganty [36] proposed a criterion for selecting "working" correlation structure based on the minimization of the generalized error sum of squares (ESS) given as follows:

$$\mathrm{ESS}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{K} (Y_i - u_i)' V_i^{-1} (Y_i - u_i)$$
$$= \sum_{i=1}^{K} Z_i'(\boldsymbol{\beta}) R_i^{-1}(\boldsymbol{\alpha}) Z_i(\boldsymbol{\beta}), \tag{7}$$

where $Z_i(\boldsymbol{\beta}) = A^{1/2}(Y_i - u_i)$. The criterion is defined by

$$\mathrm{SC} = \frac{\mathrm{ESS}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{(N - p - q)}, \tag{8}$$

where $N = \sum_{i=1}^{K} n_i$ is the total number of observations, p is the number of regression parameters, and q is the number of correlation coefficients within the "working" correlation structure. Another extended criterion from SC was proposed by Carey and Wang [37], where the Gaussian pseudolikelihood (GP) is adopted, and it is given by

$$\text{GP}(R) = -0.5 \times \sum_{i=1}^{K} (Y_i - u_i)' V_i^{-1} (Y_i - u_i) + \log(|V_i|), \quad (9)$$

where a better "working" correlation structure yields a larger GP. In their work, they also showed that GP criterion held better performance than RJ via simulation.

Another criterion is proposed by Pan [38], which modified Akaike information criterion (AIC) [39] in adaption to GEE. Due to the fact that GEE is not likelihood-based, thus it is called quasi-likelihood under the independence model criterion (QIC) [40]. The basic idea is to calculate the expected Kullback-Leibler discrepancy using the quasi-likelihood under the independence "working" correlation assumption due to the lack of a general and tractable quasi-likelihood for the correlated data under any other complex "working" correlation structures. QIC(R) is defined by

$$\text{QIC}(R) = -2\Psi\left(\widehat{\boldsymbol{\beta}}(R); I\right) + 2\text{trace}\left(\widehat{\Omega}_I \widehat{V}_{LZ}\right), \quad (10)$$

where the quasilikelihood $\Psi(\widehat{\boldsymbol{\beta}}(R); I) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} Q(\widehat{\boldsymbol{\beta}}(R), \widehat{\phi}; \{Y_{ij}, X_{ij}\})$ with $Q(\mu, \widehat{\phi}; y) = \int_y^u ((y-t)/\widehat{\phi}V(t))dt$ defined by [12], $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}$ are obtained under the hypothesized "working" correlation structure R, $\widehat{\Omega}_I = \sum_{i=1}^{K} D_i' V_i^{-1} D_i|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, R=I}$, and $\widehat{V}_{LZ}$ is defined above with replacement of $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}(R)$ [38]. Note that, in this work, Pan ignored the second term in Taylor's expansion of the discrepancy and showed its influence was not substantial among his simulation set-ups. Later on, Hardin and Hilbe (2003) made slight modification on QIC(R) by using $\widehat{\boldsymbol{\beta}}(I)$ and $\widehat{\phi}(I)$ for more stability, and QIC(R)$_{\text{HH}}$ is given by

$$\text{QIC}(R)_{\text{HH}} = -2\Psi\left(\widehat{\boldsymbol{\beta}}(I); I\right) + 2\text{trace}\left(\widehat{\Omega}_I \widehat{V}_{LZ}\right). \quad (11)$$

Note that QIC(R) and QIC(R)$_{\text{HH}}$ do not perform well in distinguishing the independence and exchangeable "working" correlation structures because, in certain cases, the same regression parameter estimates can be obtained under these two structures. Also, the attractive property of the QIC criterion is that it allows the selection of the covariates and "working" correlation structure simultaneously [41, 42], but this measure is more sensitive to the mean structure because QIC is particularly impacted by the first term and the second term which plays a role as a penalty. To better select "working" correlation structure, Hin and Wang proposed correlation information criterion (CIC) defined by

$$\text{CIC} = \text{trace}\left(\widehat{\Omega}_I \widehat{V}_{LZ}\right). \quad (12)$$

In their work, CIC was shown to outperform QIC when the outcomes were binary through simulation studies [43].

One limitation of this criterion is that it cannot penalize the overparameterization; thus the performance is not well in comparison with two correlation structures having quite different numbers of correlation parameters.

Another attractive criterion is the extended quasilikelihood information criterion (EQIC) proposed by Wang and Hin [25] by using the extended quasilikelihood (EQL) defined by Nelder and Pregibon based on the deviance function, which is shown below under the independent correlation structure [44]:

$$Q^*(\boldsymbol{\beta}, \phi; I) = -\frac{1}{2\phi} D(\boldsymbol{\beta}; I) - \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log\left(2\pi\phi A\left(\mu_{ij}\right)\right), \quad (13)$$

where the sum of deviances $D(\boldsymbol{\beta}; I) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} -2\phi\{Q(y_{ij}; \mu_{ij}) - Q(y_{ij}, y_{ij})\}$ with $Q(\cdot)$ being the quasilikelihood defined as above. Therefore, EQIC is defined by

$$\begin{aligned} \text{EQIC}(R) = \frac{1}{\phi} D(\boldsymbol{\beta}; I) + \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log\left(2\pi\phi A\left(\mu_{ij}\right)\right) \\ + 2\text{trace}\left(\widehat{\Omega}_I \widehat{V}_{LZ}\right), \end{aligned} \quad (14)$$

where some adjustments were applied to $A(\mu)$ by adding a small constant k with the optimal chosen value as $1/6$. The author indicated that the covariates were first selected based on QIC, and the variance function could be identified as the one minimizing EQIC given the selected covariates; then "working" correlation structure selection could be achieved based on CIC; in addition, they found out that the covariates selection by EQIC given different working variance functions was more consistent than that based on QIC [45].

Besides those criteria mentioned above, Cantoni et al. also discussed the covariate selection for longitudinal data analysis [46]; also, a variance function selection was mentioned by Pan and Mackenzie [30] as well as Wang and Lin [47]; in addition, more work on "working" correlation structure selection was addressed by Chaganty and Joe [48], Wang and Lin [47], Gosho et al. [49, 50], Jang [51], Chen [52], and Westgate [53–55], among others. Overall, the model selection of GEE is nontrivial, where the best selection criterion is still being pursued [56], and the recent work by Wang et al. can be followed up as the rule of thumb [45].

2.3. Sample Size and Power of GEE. It is well known that the calculation of sample size and power is necessary and important for planning a clinical trial, which have been well studied for independent observations [1]. With the wide applications of GEE in clinical trials, this topic for correlated/clustered data has gained more attention than ever [5, 57]. The general method for sample size/power calculated was discussed by Liu and Liang [58], where the generalized score test was utilized to draw statistical inference and the resulting noncentral chi-square distribution of test statistic under the alternative hypothesis was derived; however, in some special cases, that is, correlated binary data with nonexchangeable correlation structure, there was no close

form available along the outline of that formula. Afterwards, Shih provided an alternative formula on sample size/power calculation, which relied on Wald tests using the estimates of regression parameters and robust variance estimators [59]. For example, in a study with one parameter of interest β, the hypothesis of interest can be formulated as

$$H_0: \beta = 0 \quad \text{versus} \quad H_a: \beta = b \neq 0, \tag{15}$$

where b is the expected value. Thus, based on a two-sided Z-test with type I error η, the power denoted by δ can be obtained by

$$\delta = 1 - \Phi\left(Z_{\eta/2} - \frac{b\sqrt{K}}{\sqrt{\nu_R}}\right), \tag{16}$$

where K is sample size and ν_R is the robust variance estimator corresponding to β in the estimate of $K\widehat{V}_{LZ}$. Accordingly, the sample size is given by

$$K = \frac{\nu_R\left(Z_{\eta/2} - Z_{1-\delta}\right)^2}{b^2}. \tag{17}$$

For correlated continuous data, the calculation is straightforward using (16); however, in particular, for correlated binary data, more work will be needed [60], and Pan provided explicit formulas for ν_R under various situations as follows [61]:

$$\nu_R = \Omega\left[\frac{1}{\pi p_0\left(1 - p_0\right)} + \frac{1}{\left(1 - \pi\right)p_1\left(1 - p_1\right)}\right], \tag{18}$$

where $\Omega = K(\sum_{i=1}^{K}\mathbf{1}'_{n_i}R_i^{-1}V_iR_i^{-1}\mathbf{1}_{n_i})/(\sum_{i=1}^{K}\mathbf{1}'_{n_i}R_i^{-1}\mathbf{1}_{n_i})^2$ with π as the proportion of subjects assigned to the control group and p_0 and p_1 as the mean for control and case groups [61]. The detailed calculations of ν_R under several important special cases are given by

If $R_i = V_I = \text{CS}: \Omega = \dfrac{K}{\sum_{i=1}^{K}\left(n_i/\left(1 + \left(n_i - 1\right)\alpha\right)\right)};$

If $R_i = I, \quad V_I = \text{CS}: \Omega = \dfrac{K\sum_{i=1}^{K}n_i\left[1 + \left(n_i - 1\right)\alpha\right]}{\left(\sum_{i=1}^{K}n_i\right)^2};$

If $R_i = V_I = \text{AR}(1): \Omega = \dfrac{K\left(1 + \alpha\right)}{\sum_{i=1}^{K}\left[n_i - \left(n_i - 2\right)\alpha\right]};$

If $R_i = I, \quad V_I = \text{AR}(1):$

$$\Omega = K\frac{\sum_{i=1}^{K}\left[n_i + 2\left(n_i - 1\right)\alpha + 2\left(n_i - 2\right)\alpha^2 + \cdots + 2\alpha^{n_i-1}\right]}{\left(\sum_{i=1}^{K}n_i\right)^2}. \tag{19}$$

These formulas can be directly used in practice, which has covered most situations encountered in clinical trials [61]. Note that when $R_i = V_I = \text{CS}$, Liu and Liang (1997) provided

a different formula of sample size compared with (17) with $n_i = n$, which is

$$\begin{aligned}K = \Big(&\left(z_{1-\eta/2} + z_{1-\delta}\right)^2 \\ &\times \left(\left(1 - \pi\right)p_0\left(1 - p_0\right) + \pi p_1\left(1 - p_1\right)\right) \\ &\times \left[1 + \left(n - 1\right)\alpha\right]\Big) \\ &\times \left(n\pi\left(1 - \pi\right)\left(p_1 - p_0\right)^2\right)^{-1}.\end{aligned} \tag{20}$$

Be aware that the difference is due to the test methods, the Wald Z-test used by Pan [61] and the score test applied by Liu and Liang [58]. Note that, in some cases, the score test may be preferred [62]. Although some other works exist for sample size/power calculation, they focused on the other alternative approaches rather than GEE [63, 64]; thus we do not discuss them here. For correlated Poisson data, the sample size/power calculation is more challenging due to the occurrence of overdispersion or sparsity, where negative binomial regression model may be explored [62, 65–67].

On the other hand, there are several concerns [68]. First, we here focus on the calculation of the sample size K assuming n_i is known; however, based on the power formula (16), ν_R depends on n_i and thus increasing n_i can also assist in power improvement but turns out to be less effective than K [69]. Second, the sample size/power calculation may be restricted to the limitation of clusters, for example, clustered randomized trials (CRTs), where the number of clusters could be relatively small. For example, by the literature review of published CRTs, the median number of clusters is shown as 21 [70]. In such situations, the power formula adjusted for the small samples in GEE is necessary, which has drawn attention from researchers recently [71–75].

2.4. Clustered Data with Informative Cluster Size. The application of GEE in clustered data with informative cluster size is another special topic [76]. Taking an example of a periodontal disease study, the number of teeth for each patient may be related to the overall oral health of the individual; in other words, the worse the oral health is, the less the number of teeth is and, thus, cluster size n_i may influence the distribution of the oral outcomes, which is called informative cluster size [45, 77]. Such issues commonly occur in biomedical studies (e.g., genetic disease studies), and rigorous statistical methods are needed for valid statistical inference [78]. Note that if the maximum of cluster size exists and is known, then this can be treated as (informative) missing data problem, which can be solved via the weighted estimating equations proposed by Robins et al. [79]; however, if the maximum is unknown or not accessible, the method of within-cluster resampling (WCR) proposed by Hoffman et al. could be applied [80]. The basic idea is that, for each of L resampled replicate data based on a Monte Carlo method (L is a large number, i.e., 10,000), one observation is randomly extracted from each cluster, where $\widehat{\beta}_l$ with variance estimator $\widehat{\Sigma}_l$ can be obtained from a regular score equation denoted by $S_l(\beta)$ for independent observations (i.e., linear regression for

continuous data; logistic regression for binary data; Poisson regression for count data), $l = 1, 2, \ldots, L$. The details are shown as follows:

$$
S_l(\boldsymbol{\beta}) = \sum_{i=1}^{K} S_{ij}(\boldsymbol{\beta}_l) I[j \in r_l] = 0;
$$

$$
\widehat{\boldsymbol{\beta}}_{\text{wcr}} = \frac{1}{L} \sum_{l=1}^{L} \widehat{\boldsymbol{\beta}}_l; \tag{21}
$$

$$
\widehat{V}_{\text{wcr}} = \frac{1}{L} \sum_{l=1}^{L} \widehat{\Sigma}_l - \frac{1}{L} \sum_{l=1}^{L} (\widehat{\boldsymbol{\beta}}_l - \widehat{\boldsymbol{\beta}}_{\text{wcr}})(\widehat{\boldsymbol{\beta}}_l - \widehat{\boldsymbol{\beta}}_{\text{wcr}})^T,
$$

where $S_{ij}(\boldsymbol{\beta}_l) = X'_{ij} V_{ij}^{-1}(Y_{ij} - X'_{ij}\beta_l)$ with r_l as the set of data index selected from the ith cluster in lth replicate data. Alternatively, the approach considered by Williamson et al. by adopting the weighted estimating equations performs asymptotically equivalently as WCR and also avoids intensive computing, and it is referred to as the cluster-weighted GEE (CWGEE) [81]. The estimating equation is

$$
S(\boldsymbol{\beta}) = \sum_{i=1}^{K} \frac{1}{K} \sum_{j=1}^{n_i} S_{ij}(\boldsymbol{\beta}) = 0, \tag{22}
$$

where S_{ij} is defined the same as above, but what is different is that the subscription j ranges from 1 to n_i, not restricted by the index r_l. Note that as $L \to \infty$, $(1/L) \sum_{i=1}^{L} S_l(\boldsymbol{\beta})$ converges to its expected estimating function and is asymptotically equivalent to $S(\boldsymbol{\beta})$.

This method was also explored or extended for the correlated data with nonignorable cluster size by Benhin et al. and Cong et al. [82, 83]. Furthermore, a more efficient method called modified WCR (MWCR) was proposed by Chiang and Lee, where minimum cluster size $n_i > 1$ subjects were randomly sampled from each cluster, and then GEE models for balanced data were applied for estimation by incorporating the intracluster correlation; thus MWCR might be a more efficient way for analysis [84]. But MWCR is not always satisfactory and Pavlou et al. recognized the sufficient conditions of the data structure and the choice of "working" correlation structure, which allowed the consistency of the estimates from MWCR [85]. In addition, Wang et al. extended the above work to the clustered longitudinal data, which are collected as repeated measures on subjects arising in clusters, with potential informative cluster size [45]. Examples include health studies of subjects from multiple hospitals or families. With the adoption and comparison of GEE, WCR, and CWGEE, the author claimed that CWGEE was recommended because of the comparable performance with WCR and the lack of intensive Monte Carlo computation in terms of well preserved coverage rates and desirable power properties, while GEE models led to invalid inference due to the biased parameter estimates via extensive simulation studies and real data application of a periodontal disease study [45]. In addition, for observed-cluster inference, Seaman et al. discussed the methods, including weighted and doubly weighted GEE and the shared random-effects models for comparison, and

showed the conditions under which the shared random-effects model described members with observed outcomes **Y** [86]. More work can be found in [87–90], among others.

3. Simulation

In this section, we focus on "working" correlation structure selection and compare the performances of the existing criteria through simulation studies. Two types of outcomes are considered, continuous and count responses. The models for data generation are as follows:

$$
\begin{aligned}
u_{ij} &= \beta_0 + \beta_1 \times x_{ij}; \\
\log(u_{ij}) &= \beta_0 + \beta_1 \times x_{ij},
\end{aligned} \tag{23}
$$

where $\beta_0 = \beta_1 = 0.5$, $i = 1, 2, \ldots, I$ with $I = 50, 100, 200, 500$ and $j = 1, 2, \ldots, J$ with $J = 4, 8$. The covariates x_{ij} are i.i.d. from a standard uniform distribution Unif$(0, 1)$. For each scenario, we generate the data based on the underlying true correlation structures as independent (IND), exchangeable (EXCH), and autoregressive (AR-1) with $\alpha = 0.3, 0.7$. 1,000 Monte Carlo data sets are generated for each scenario, where the estimates of regression parameters and within-subject correlation matrix and seven model selection criteria measures are calculated using the "working" correlation structure of IND, EXCH, and AR-1. The partial simulation results are provided in Tables 2, 3, and 4, where the results of CIC are not shown because they are the same as those of QIC.

Based on the results, RJ does not perform well for the scenarios with either continuous or binary outcomes, while RJ1 and RJ2 have comparable performances and can select the true underlying correlation structure in most scenarios with better performance under large sample size. QIC is not satisfactory when the true correlation structure is independent but has advantageous performance for the scenarios with the true correlation structure as exchangeable or AR-1. On the other hand, SC and GP do not perform well for longitudinal data with normal responses, but the performance is slightly improved for longitudinal data with binary outcomes. The results may vary due to variety of factors including the types of "working" correlation structure considered for model fitting, the sample size, and/or the magnitude of correlation coefficient. For the future work, there is a necessity to find out a robust criterion for "working" correlation structure selection of GEE, and more advanced approaches are emerging currently.

4. Future Direction and Discussion

In this paper, we provide a review of several specific topics such as model selection with emphasis on the selection of "working" correlation structure, sample size and power calculation, and clustered data analysis with informative cluster size related to GEE for longitudinal/correlated data. The simulation studies are conducted for providing numerical comparisons among five types of model selection criteria [91, 92]. Until now, novel methodologies are still needed and being developed due to the increasing usage and potential

TABLE 2: Simulation for longitudinal data with independent correlation matrix.

n	K	Criterion	IND	EXCH	AR-1	IND	EXCH	AR-1
				Normal			Binary	
4	50	QIC	198	393	409	202	374	424
		RJ	327	423	250	312	421	267
		RJ1	388	322	290	399	316	285
		RJ2	384	327	289	388	320	292
		SC	488	1	512	351	310	339
		GP	547	0	453	368	306	326
	100	QIC	209	377	414	185	407	408
		RJ	338	415	247	340	410	250
		RJ1	389	349	262	381	358	261
		RJ2	389	353	258	372	357	271
		SC	482	1	517	352	346	302
		GP	520	0	480	360	348	292
8	50	QIC	200	411	389	203	363	434
		RJ	282	497	221	292	476	232
		RJ1	402	354	244	386	340	274
		RJ2	402	357	241	373	347	280
		SC	465	1	535	351	325	324
		GP	558	0	442	382	311	307
	100	QIC	188	393	419	201	398	401
		RJ	321	442	237	287	466	247
		RJ1	347	385	268	385	367	248
		RJ2	347	382	271	377	369	254
		SC	492	0	508	355	343	302
		GP	541	0	459	370	341	289

TABLE 3: Simulation for longitudinal data with exchangeable correlation matrix with $\alpha = 0.3$.

n	K	Criterion	IND	EXCH	AR-1	IND	EXCH	AR-1
				Normal			Binary	
4	50	QIC	106	699	195	53	758	189
		RJ	419	139	442	869	5	126
		RJ1	0	963	37	12	898	90
		RJ2	0	959	41	22	876	102
		SC	0	593	407	282	650	68
		GP	1	593	406	412	524	64
	100	QIC	31	879	90	7	867	126
		RJ	350	88	562	911	2	87
		RJ1	0	995	5	2	946	52
		RJ2	0	996	4	10	933	57
		SC	0	598	402	339	635	26
		GP	0	501	499	445	531	24
8	50	QIC	80	828	92	50	876	74
		RJ	10	395	595	813	6	181
		RJ1	0	1000	0	0	987	13
		RJ2	0	1000	0	0	966	25
		SC	0	488	513	302	696	2
		GP	0	511	489	497	500	3
	100	QIC	17	953	30	8	973	19
		RJ	0	408	592	861	0	139
		RJ1	0	1000	0	0	997	3
		RJ2	0	1000	0	0	993	7
		SC	0	470	530	328	672	0
		GP	0	526	474	486	514	0

TABLE 4: Simulation for longitudinal data with AR-1 correlation matrix with $\alpha = 0.3$.

n	K	Criterion	IND	EXCH	AR-1	IND	EXCH	AR-1
				Normal			Binary	
4	50	QIC	91	166	743	66	170	764
		RJ	712	142	146	925	12	63
		RJ1	0	478	522	7	505	488
		RJ2	0	466	534	20	499	481
		SC	0	480	520	220	350	430
		GP	0	543	457	303	332	365
	100	QIC	25	116	859	7	122	871
		RJ	770	95	135	972	4	24
		RJ1	0	475	525	1	569	430
		RJ2	0	481	519	5	571	424
		SC	0	491	509	237	371	392
		GP	0	540	460	290	353	357
8	50	QIC	50	88	862	44	77	879
		RJ	646	148	206	934	5	61
		RJ1	0	445	555	0	535	465
		RJ2	0	443	557	10	535	455
		SC	0	467	533	168	397	435
		GP	0	549	451	269	406	325
	100	QIC	16	39	945	7	33	960
		RJ	648	154	198	972	0	28
		RJ1	0	455	545	1	603	396
		RJ2	0	455	545	1	609	390
		SC	0	480	520	177	458	365
		GP	0	532	468	247	457	296

theoretical constraints of GEE as well as new challenges emerging from practical applications in clinical trials or biomedical studies.

In addition, current research of interest related to GEE also includes a robust and optimal model selection criterion of GEE under missing at random (MAR) or missing not at random (MNAR) [93, 94], sample size/power calculation for correlated sparse or overdispersion count data or longitudinal data with small sample [57–60], GEE with improved performance under the situations with informative cluster size and/or MAR and/or small sample size [95–98], and GEE for high-dimensional longitudinal data [99]. Although GEE has attractive features, flexible application, and easy implementation in software, the application in practice should be cautious depending on the context of study design or data structure and the goals of research interest.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] Z. Feng, P. Diehr, A. Peterson, and D. McLerran, "Selected statistical issues in group randomized trials," *Annual Review of Public Health*, vol. 22, pp. 167–187, 2001.

[2] G. Fitzmaurice, N. M. Larid, and J. H. Ware, *Applied Longitudinal Data*, John Wiley & Sons, 2004.

[3] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, Chapman and Hall/CRC Press, Boca Raton, Fla, USA, 2003.

[4] R. F. Potthoff and S. N. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, pp. 313–326, 1964.

[5] L. M. Friedman, C. D. Furberg, and D. L. DeMets, *Fundamentals of Clinical Trials*, Springer, New York, NY, USA, 3nd edition, 1989.

[6] K. Y. Liang and S. L. Zeger, "A comparison of two bias-corrected covariance estimators for generalized estimating equations," *Biometrika*, vol. 73, pp. 13–22, 1986.

[7] M. Crowder, "On the use of a working correlation matrix in using generalised linear models for repeated measures," *Biometrika*, vol. 82, no. 2, pp. 407–410, 1995.

[8] R. W. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika*, vol. 61, pp. 439–447, 1974.

[9] P. Diggle, P. Heagerty, K. Y. Liang, and S. L. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, Oxford, UK, 2002.

[10] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, *Longitudinal Data Anlaysis*, Chapman & Hall/CRC Press, 2008.

[11] D. Hedeker and R. D. Gibbons, *Analysis of Longitudinal Data*, John Wiley & Sons, 2006.

[12] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, UK, 1989.

[13] N. R. Chaganty and H. Joe, "Range of correlation matrices for dependent Bernoulli random variables," *Biometrika*, vol. 93, no. 1, pp. 197–206, 2006.

[14] R. T. Sabo and N. R. Chaganty, "What can go wrong when ignoring correlation bounds in the use of generalized estimating equations," *Statistics in Medicine*, vol. 29, no. 24, pp. 2501–2507, 2010.

[15] B. C. Sutradhar and K. Das, "On the efficiency of regression estimators in generalised linear models for longitudinal data," *Biometrika*, vol. 86, no. 2, pp. 459–465, 1999.

[16] Y.-G. Wang and V. Carey, "Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance," *Biometrika*, vol. 90, no. 1, pp. 29–41, 2003.

[17] S. R. Lipsitz, G. Molenberghs, G. M. Fitzmaurice, and J. Ibrahim, "GEE with Gaussian estimation of the correlations when data are incomplete," *Biometrics*, vol. 56, no. 2, pp. 528–536, 2000.

[18] Y.-G. Wang and V. J. Carey, "Unbiased estimating equations from working correlation models for irregularly timed repeated measures," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 845–853, 2004.

[19] A. Qu and B. G. Lindsay, "Building adaptive estimating equations when inverse of covariance estimation is difficult," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 65, no. 1, pp. 127–142, 2003.

[20] S. R. Lipsitz and G. M. Fitzmaurice, "Estimating equations for measures of association between repeated binary responses," *Biometrics*, vol. 52, no. 3, pp. 903–912, 1996.

[21] Y. Lee and J. A. Nelder, "Conditional and marginal models: another view," *Statistical Science*, vol. 19, no. 2, pp. 219–238, 2004.

[22] Y. Lee and J. A. Nelder, "Likelihood inference for models with unobservables: another view," *Statistical Science*, vol. 24, no. 3, pp. 255–269, 2009.

[23] A. Qu, B. G. Lindsay, and B. Li, "Improving generalised estimating equations using quadratic inference functions," *Biometrika*, vol. 87, no. 4, pp. 823–836, 2000.

[24] G. Kauermann and R. J. Carroll, "A note on the efficiency of sandwich covariance matrix estimation," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1387–1396, 2001.

[25] Y. G. Wang and L. Y. Hin, "Modeling strategies in longitudinal data analysis: covariate, variance function and correlation structure selection," *Computational Statistics and Data Analysis*, vol. 54, no. 12, pp. 3359–3370, 2010.

[26] W. Pan, "Goodness-of-fit tests for GEE with correlated binary data," *Scandinavian Journal of Statistics*, vol. 29, no. 1, pp. 101–110, 2002.

[27] A. M. Wood, I. R. White, and P. Royston, "How should variable selection be performed with multiply imputed data?" *Statistics in Medicine*, vol. 27, no. 17, pp. 3227–3246, 2008.

[28] M. D. Begg and M. K. Parides, "Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data," *Statistics in Medicine*, vol. 22, no. 16, pp. 2591–2602, 2003.

[29] L. Y. Hin, V. J. Carey, and Y. G. Wang, "Criteria for working-correlation-structure selection in GEE: assessment via simulation," *The American Statistician*, vol. 61, no. 4, pp. 360–364, 2007.

[30] J. X. Pan and G. Mackenzie, "On modelling mean-covariance structures in longitudinal studies," *Biometrika*, vol. 90, no. 1, pp. 239–244, 2003.

[31] M. Davidian and R. J. Carroll, "Variance function estimation," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1079–1091, 1987.

[32] M. Pourahmadi, "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999.

[33] S. Konishi and G. Kitagawa, "Generalised information criteria in model selection," *Biometrika*, vol. 83, no. 4, pp. 875–890, 1996.

[34] B. Zhang, "Summarizing the goodness of fit o f generalized linear models for longitudinal data," *Statistics in Medicine*, vol. 19, pp. 1265–1275, 2000.

[35] A. Rotnitzky and N. P. Jewell, "Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data," *Biometrika*, vol. 77, no. 3, pp. 485–497, 1990.

[36] J. Shults and N. R. Chaganty, "Analysis of serially correlated data using quasi-least squares," *Biometrics*, vol. 54, no. 4, pp. 1622–1630, 1998.

[37] V. J. Carey and Y.-G. Wang, "Working covariance model selection for generalized estimating equations," *Statistics in Medicine*, vol. 30, no. 26, pp. 3117–3124, 2011.

[38] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.

[39] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the 2nd International Symposium on Information Theory*, vol. 15, pp. 267–281, 1973.

[40] J. A. Nelder and Y. Lee, "Likelihood, quasi-likelihood and pseudolikelihood: some comparisons," *Journal of the Royal Statistical Society B*, vol. 54, no. 1, pp. 273–284, 1992.

[41] J. Cui, "QIC program and model selection in GEE analyses," *The Stata Journal*, vol. 7, no. 2, pp. 209–220, 2007.

[42] J. Cui and G. Qian, "Selection of working correlation structure and best model in GEE analyses of longitudinal data," *Communications in Statistics—Simulation and Computation*, vol. 36, no. 4–6, pp. 987–996, 2007.

[43] L. Y. Hin and Y. G. Wang, "Working-correlation-structure identification in generalized estimating equations," *Statistics in Medicine*, vol. 28, no. 4, pp. 642–658, 2009.

[44] J. A. Nelder and D. Pregibon, "An extended quasi-likelihood function," *Biometrika*, vol. 74, no. 2, pp. 221–232, 1987.

[45] M. Wang, M. Kong, and S. Datta, "Inference for marginal linear models for clustered longitudinal data with potentially informative cluster sizes," *Statistical Methods in Medical Research*, vol. 20, no. 4, pp. 347–367, 2011.

[46] E. Cantoni, J. M. Flemming, and E. Ronchetti, "Variable selection for marginal longitudinal generalized linear models," *Biometrics. Journal of the International Biometric Society*, vol. 61, no. 2, pp. 507–514, 2005.

[47] Y.-G. Wang and X. Lin, "Effects of variance-function misspecification in analysis of longitudinal data," *Biometrics*, vol. 61, no. 2, pp. 413–421, 2005.

[48] N. R. Chaganty and H. Joe, "Efficiency of generalized estimating equations for binary responses," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 66, no. 4, pp. 851–860, 2004.

[49] M. Gosho, C. Hamada, and I. Yoshimura, "Criterion for the selection of a working correlation structure in the generalized estimating equation approach for longitudinal balanced data," *Communications in Statistics*, vol. 40, no. 21, pp. 3839–3856, 2011.

[50] M. Gosho, C. Hamada, and I. Yoshimura, "Selection of working correlation structure in weighted generalized estimating equation method for incomplete longitudinal data," *Communications in Statistics*, vol. 43, no. 1, pp. 62–81, 2014.

[51] M. J. Jang, *Working correlation selection in generalized estimating equations [Dissertation]*, University of Iowa, 2011.

[52] J. Chen and N. A. Lazar, "Selection of working correlation structure in generalized estimating equations via empirical likelihood," *Journal of Computational and Graphical Statistics*, vol. 21, no. 1, pp. 18–41, 2012.

[53] P. M. Westgate, "A bias-corrected covariance estimator for improved inference when using an unstructured correlation with quadratic inference functions," *Statistics and Probability Letters*, vol. 83, no. 6, pp. 1553–1558, 2013.

[54] P. M. Westgate, "Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach," *Biometrical Journal*, vol. 56, no. 3, pp. 461–476, 2014.

[55] P. M. Westgate, "Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data," *Statistics in Medicine*, vol. 33, no. 13, pp. 2222–2237, 2014.

[56] J. Ye, "On measuring and correcting the effects of data mining and model selection," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 120–131, 1998.

[57] J. J. Shuster, *Practical Handbook of Sample Size Guidelines for Clinical Trials*, CRC Press, Boca Raton, Fla, USA, 1993.

[58] G. Liu and K.-Y. Liang, "Sample size calculations for studies with correlated observations," *Biometrics*, vol. 53, no. 3, pp. 937–947, 1997.

[59] W. J. Shih, "Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations," *Biometrical Journal*, vol. 39, no. 8, pp. 899–908, 1997.

[60] S. R. Lipsitz and G. M. Fitzmaurice, "Sample size for repeated measures studies with binary responses," *Statistics in Medicine*, vol. 13, no. 12, pp. 1233–1239, 1994.

[61] W. Pan, "Sample size and power calculations with correlated binary data," *Controlled Clinical Trials*, vol. 22, no. 3, pp. 211–227, 2001.

[62] N. Breslow, "Tests of hypotheses in overdispersed Poisson regression and other quasi likelihood models," *Journal of the American Statistical Association*, vol. 85, pp. 565–571, 1990.

[63] E. W. Lee and N. Dubin, "Estimation and sample size considerations for clustered binary responses," *Statistics in Medicine*, vol. 13, no. 12, pp. 1241–1252, 1994.

[64] D. J. Sargent, J. A. Sloan, and S. S. Cha, "Sample size and design considerations for phase II clinical trials with correlated observations," *Controlled Clinical Trials*, vol. 20, no. 3, pp. 242–252, 1999.

[65] C. S. Li, "Semiparametric negative binomial regression models," *Communications in Statistics: Simulation and Computation*, vol. 39, no. 3, pp. 475–486, 2010.

[66] W. H. Greene, "Accounting for excess zeros and sample selection in Poisson and negative binomial regression models," Tech. Rep., New York University, 1994.

[67] P. Lambert, "Modeling of repeated series of count data measured at unequally spaced times," *Applied Statistics*, vol. 45, pp. 31–38, 1996.

[68] M. S. Pepe and G. L. Anderson, "A cautionary note on in ference for marginal regression models with longitudinal data and general correlated response data," *Communications in Statistics, Series B*, vol. 23, pp. 939–951, 1994.

[69] M. Wang and Q. Long, "Modified robust variance estimator for generalized estimating equations with improved small-sample performance," *Statistics in Medicine*, vol. 30, no. 11, pp. 1278–1291, 2011.

[70] M. Taljaard, A. D. McRae, C. Weijer et al., "Inadequate reporting of research ethics review and informed consent in cluster randomised trials: Review of random sample of published trials," *British Medical Journal*, vol. 342, Article ID d2496, 2011.

[71] L. A. Mancl and T. A. DeRouen, "A covariance estimator for GEE with improved small-sample properties," *Biometrics*, vol. 57, no. 1, pp. 126–134, 2001.

[72] M. P. Fay and B. I. Graubard, "Small-sample adjustments for Wald-type tests using sandwich estimators," *Biometrics*, vol. 57, no. 4, pp. 1198–1206, 2001.

[73] W. Pan, "On the robust variance estimator in generalised estimating equations," *Biometrika*, vol. 88, no. 3, pp. 901–906, 2001.

[74] W. Pan and M. M. Wall, "Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations," *Statistics in Medicine*, vol. 21, no. 10, pp. 1429–1441, 2002.

[75] X. Guo, W. Pan, J. E. Connett, P. J. Hannan, and S. A. French, "Small-sample performance of the robust score test and its modifications in generalized estimating equations," *Statistics in Medicine*, vol. 24, no. 22, pp. 3479–3495, 2005.

[76] D. M. Farewell, "Marginal analyses of longitudinal data with an informative pattern of observations," *Biometrika*, vol. 97, no. 1, pp. 65–78, 2010.

[77] J. D. Beck, T. Sharp, G. G. Koch, and S. Offenbacher, "A 5-year study of attachment loss and tooth loss in community-dwelling older adults," *Journal of Periodontal Research*, vol. 32, no. 6, pp. 516–523, 1997.

[78] S. J. Arbes Jr., H. Ágústsdóttir, and G. D. Slade, "Environmental tobacco smoke and periodontal disease in the United States," *American Journal of Public Health*, vol. 91, no. 2, pp. 253–257, 2001.

[79] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Analysis of semi-parametric regression models for repeated outcomes in the

presence of missing data," *Journal of the American Statistical Association*, vol. 90, pp. 106–121, 1995.

[80] E. B. Hoffman, P. K. Sen, and C. R. Weinberg, "Within-cluster resampling," *Biometrika*, vol. 88, no. 4, pp. 1121–1134, 2001.

[81] J. M. Williamson, S. Datta, and G. A. Satten, "Marginal analyses of clustered data when cluster size is informative," *Biometrics*, vol. 59, no. 1, pp. 36–42, 2003.

[82] E. Benhin, J. N. Rao, and A. J. Scott, "Mean estimating equation approach to analysing cluster-correlated data with nonignorable cluster sizes," *Biometrika*, vol. 92, no. 2, pp. 435–450, 2005.

[83] X. J. Cong, G. Yin, and Y. Shen, "Marginal analysis of correlated failure time data with informative cluster sizes," *Biometrics*, vol. 63, no. 3, pp. 663–672, 2007.

[84] T. C. Chiang and K. Y. Lee, "Efcient estimation methods for informative cluster size data," *Statistical Sinica*, vol. 80, pp. 121–123, 2008.

[85] M. Pavlou, S. R. Seaman, and A. J. Copas, "An examination of a method for marginal inference when the cluster size is informative," *Statistica Sinica*, vol. 23, no. 2, pp. 791–801, 2013.

[86] S. R. Seaman, M. Pavlou, and A. J. Copas, "Methods for observed-cluster inference when cluster size is informative: a review and clarifications," *Biometrics*, vol. 70, no. 2, pp. 449–456, 2014.

[87] Z. Chen, B. Zhang, and P. S. Albert, "A joint modeling approach to data with informative cluster size: robustness to the cluster size model," *Statistics in Medicine*, vol. 30, no. 15, pp. 1825–1836, 2011.

[88] Y. Huang and B. Leroux, "Informative cluster sizes for subcluster-level covariates and weighted generalized estimating equations," *Biometrics*, vol. 67, no. 3, pp. 843–851, 2011.

[89] B. F. Kurland, L. L. Johnson, B. L. Egleston, and P. H. Diehr, "Longitudinal data with follow-up truncated by death: match the analysis method to research aims," *Statistical Science*, vol. 24, no. 2, pp. 211–222, 2009.

[90] J. M. Neuhaus and C. E. McCulloch, "Estimation of covariate effects in generalized linear mixed models with informative cluster sizes," *Biometrika*, vol. 98, no. 1, pp. 147–162, 2011.

[91] S. R. Lipsitz, G. M. Fitzmaurice, E. J. Orav, and N. M. Laird, "Performance of generalized estimating equations in practical situations," *Biometrics*, vol. 50, no. 1, pp. 270–278, 1994.

[92] D. B. Hall and T. A. Severini, "Extended generalized estimating equations for clustered data," *Journal of the American Statistical Association*, vol. 93, no. 444, pp. 1365–1375, 1998.

[93] C.-W. Shen and Y.-H. Chen, "Model selection for generalized estimating equations accommodating dropout missingness," *Biometrics*, vol. 68, no. 4, pp. 1046–1054, 2012.

[94] C.-W. Shen and Y.-H. Chen, "Model selection of generalized estimating equations with multiply imputed longitudinal data," *Biometrical Journal*, vol. 55, no. 6, pp. 899–911, 2013.

[95] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[96] R. J. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, New York, NY, USA.

[97] P. Diggle, D. Farewell, and R. Henderson, "Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal," *Journal of the Royal Statistical Society C*, vol. 56, no. 5, pp. 499–550, 2007.

[98] A. J. Copas and S. R. Seaman, "Bias from the use of generalized estimating equations to analyze incomplete longitudinal binary data," *Journal of Applied Statistics*, vol. 37, no. 6, pp. 911–922, 2010.

[99] L. Wang, J. Zhou, and A. Qu, "Penalized generalized estimating equations for high-dimensional longitudinal data analysis," *Biometrics*, vol. 68, no. 2, pp. 353–360, 2012.