Received: 22 February 2019 | Accepted: 26 February 2019
DOI: 10.1111/biom.13050

ORIGINAL ARTICLE

WILEY Biometrics Since 1947 A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Marginal analysis of ordinal clustered longitudinal data with informative cluster size

Aya A. Mitani[1] | **Elizabeth K. Kaye**[2] | **Kerrie P. Nelson**[1]

[1]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118

[2]Department of Health Policy and Health Services Research, Boston University Henry M. Goldman School of Dental Medicine, Boston, Massachusetts 02118

Correspondence
Aya A. Mitani, Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118
Email: amitani@bu.edu

Abstract

The issue of informative cluster size (ICS) often arises in the analysis of dental data. ICS describes a situation where the outcome of interest is related to cluster size. Much of the work on modeling marginal inference in longitudinal studies with potential ICS has focused on continuous outcomes. However, periodontal disease outcomes, including clinical attachment loss, are often assessed using ordinal scoring systems. In addition, participants may lose teeth over the course of the study due to advancing disease status. Here we develop longitudinal cluster-weighted generalized estimating equations (CWGEE) to model the association of ordinal clustered longitudinal outcomes with participant-level health-related covariates, including metabolic syndrome and smoking status, and potentially decreasing cluster size due to tooth-loss, by fitting a proportional odds logistic regression model. The within-teeth correlation coefficient over time is estimated using the two-stage quasi-least squares method. The motivation for our work stems from the Department of Veterans Affairs Dental Longitudinal Study in which participants regularly received general and oral health examinations. In an extensive simulation study, we compare results obtained from CWGEE with various working correlation structures to those obtained from conventional GEE which does not account for ICS. Our proposed method yields results with very low bias and excellent coverage probability in contrast to a conventional generalized estimating equations approach.

KEYWORDS

clustered data, generalized estimating equations, informative cluster size, longitudinal data, ordinal outcome, quasi-least squares

1 | INTRODUCTION

The use of generalized estimating equations (GEE) to estimate parameters with a marginal interpretation is common in longitudinal studies (Liang and Zeger, 1986). The repeated measurements over time within a unit of observation are correlated and GEE accounts for the temporal correlation by use of a working correlation matrix and sandwich variance estimates. In some longitudinal studies, a unit can belong to a cluster of units and multiple clusters can exist. In such cases, two distinct correlations exist: the correlation between units within the same cluster and the correlation between the temporal observations on the same unit.

When fitting a marginal model with GEE, one assumption is the independence between cluster size and the outcome of interest. This assumption is often violated in data arising from periodontal disease studies. One consequence of periodontal disease is tooth loss. The probability of losing

a tooth increases with the severity of the disease, resulting in fewer teeth (smaller cluster size) among participants who are more prone to periodontal disease (outcome). This phenomenon, where the outcome of interest is related to the cluster size, is called informative cluster size (ICS). If the goal of the study is to describe the population-average effect of covariates on the outcome measured on a typical tooth from a randomly selected participant, then conventional GEE will over-weight healthy teeth and produce biased parameter estimates (Williamson et al., 2003).

Our motivation stems from the Department of Veterans Affairs Dental Longitudinal Study (VADLS) (Kapur et al., 1972). The study investigators collected participant-level demographics and health outcomes as well as each participant's tooth-level periodontal disease outcomes repeatedly over time. Clinical attachment loss (CAL) was one of the periodontal disease outcomes. CAL is the distance between the cementoenamel junction of the tooth and the point on the root where the gum begins to separate, and was recorded using an ordinal scoring system with four categories (0: <2mm, 1: 2-2.9mm, 2: 3-4.9mm, 3: ≥5mm). A higher score, indicating greater separation between the gum and the root of the tooth, is considered a worse prognosis of periodontal disease. We are interested in examining the population-average effect of participant-level covariates, such as metabolic syndrome and smoking, on *ordinal* CAL scores recorded at the tooth-level over time (Kaye et al., 2016). Here, each tooth (unit) belongs to a participant (cluster) and the temporal measurements are obtained for each tooth. The data likely has ICS because participants with high overall CAL scores tend to have fewer teeth at baseline.

ICS is frequently observed in other medical settings. For example, in a repeated pregnancy study, women who have experienced an adverse pregnancy outcome may have fewer subsequent pregnancies (Chaurasia et al., 2018). Hospitals or surgeons with good reputation may take on higher-risk patients or more difficult cases and thus experience more unfavorable outcomes such as post-operative complications (Panageas et al., 2007). Finally, in a psychological study, the frequency of depressive episodes by a participant may also be related to the severity level of each event (Iosif and Sampson, 2014).

Methods to account for ICS have been developed by several authors. Within-cluster resampling (WCR) was proposed by Hoffman et al. (2001) as a method to obtain parameter estimates with marginal interpretations in a cross-sectional study in which cluster size is informative. WCR involves sampling one unit (tooth) from each cluster (participant) Q times with replacement (Q is large), producing Q data sets where each data set contains one randomly selected observation per cluster. Because observations are now independent within each data set, we can fit a generalized linear model (GLM) to each data set to describe the relationship between the predictors and the outcome. The final WCR estimator is computed by taking the average of the Q GLM parameter estimates obtained from each of the Q data sets. WCR is simple and intuitive but is computationally intensive and unsuitable for ordinal outcomes. This is because there is no guarantee that all outcome categories will be represented in each of the Q sampled data sets.

For cross-sectional data with ICS, Williamson et al. (2003) proposed cluster-weighted GEE (CWGEE) as an alternate method to WCR. CWGEE involves taking the weighted average, where the weight is the inverse of cluster size, of the GEE score function during the estimation process while using an independence working correlation matrix. CWGEE is asymptotically equivalent to WCR but much more computationally efficient. Furthermore, unlike WCR, CWGEE can be applied to ordinal outcomes without any foreseeable issues.

Here, we are interested in modeling the association between covariates and the *longitudinal experience* of a typical tooth from a randomly selected participant. Wang et al. (2011) extended Williamson et al.'s CWGEE to the longitudinal setting for continuous outcomes, assuming constant cluster size over time and same number of visits across all participants. This longitudinal CWGEE approach also estimates the within-tooth correlation over time using the two-stage quasi-least squares (QLS) method (Chaganty, 1997). However, the assumption of constant cluster size over time is not realistic in a longitudinal study of periodontal disease because participants (especially those that are prone to periodontal disease) will likely lose one or more of their teeth over time. Indeed, in VADLS, almost half of the participants lose at least one tooth during follow-up. Bible et al. (2016) further extended CWGEE to address this issue by including another set of weights in the CWGEE score function for continuous outcomes. The additional set of weights is the inverse of the number of temporal observations made on each unit from each cluster. Note that in the periodontal disease setting, cluster size can only decrease over time because an adult tooth loss is permanent.

Although our article focuses on marginal models, mixed effects models are also affected by ICS and several authors have developed remedies. Among them is jointly modeling the outcome and the distribution of cluster size given the random effects and the covariates (Dunson et al., 2003). Seaman et al. (2014) has a comprehensive review article on methods for handling clustered data with ICS, and an article comparing the performance between the joint model approach and WCR for data with ICS is available by Zhang et al. (2017).

Additional challenges arise when modeling the marginal association between longitudinal clustered ordinal outcomes and the covariances in the presence of ICS. When applying the GEE method to ordinal data, the ordinal responses are transformed into a vector of binary indicators (Lipsitz et al., 1994). The marginal distribution of the vector of binary indicators

is multinomial and the mean and the variance-covariance matrix among binary indicators need to be modeled. In addition, the association parameter within each binary indicator over the repeated measures need to be estimated in order to improve the efficiency of GEE. Nooraee et al. (2014) have a comprehensive review article on GEE for longitudinal ordinal data and compare the relative performance of existing software packages. None of which, however, is designed to accommodate data with ICS.

In this article, we focus on clustered longitudinal data, with potentially decreasing cluster size, that is subject to informative cluster size when the outcome of interest is an ordinal categorical variable. In particular, we further develop the method proposed by Bible et al. (2016) under a proportional odds logistic regression model framework to accommodate ordinal outcomes, and use the approach by Parsons et al. (2006) to construct the correlation matrix among the repeated responses. In section 2, we describe our proposed CWGEE approach in more detail. Extensive simulation studies with results are presented in section 3. In section 4, we apply our approach and conventional GEE to data from the VADLS. Finally, the article concludes with a discussion in section 5.

2 | METHOD

2.1 | Construction of QLS

Consider a longitudinal clustered data set in which units are grouped into clusters and each unit contributes repeated observations of unique length. Let Y_{ijk} be the ordinal outcome measurement of the kth visit on the jth unit from the ith cluster, where $k = 1, ..., t_{ij}$, $j = 1, ..., n_i$ and $i = 1, ..., N$. Let $\boldsymbol{x}_{ijk}$ be the $p \times 1$ vector of covariates for Y_{ijk}. The response Y_{ijk} is an ordinal score of $C > 2$ categories. Our goal is to fit a proportional odds logistic regression model to describe the relationship between the covariates and the ordinal response:

$$\text{logit}\{\Pr(Y_{ijk} \le c)\} = \eta_c + \boldsymbol{x}_{ijk}\boldsymbol{\beta}, \qquad c = 1, ..., C-1. \quad (1)$$

The dependence between the ordinal responses will be incorporated in the working correlation matrix. For a proportional odds model, the ordinal response of C categories can be transformed into $C-1$ binary responses such that $U_{ijkc} = 1$ if $Y_{ijk} \le c$ and $U_{ijkc} = 0$ if $Y_{ijk} > c$, for $c = 1, ..., C-1$. For each kth visit of the jth unit from the ith cluster, we have a response vector $\boldsymbol{U}'_{ijk} = (U_{ijk1}, ..., U_{ijk(C-1)})$ of length $C-1$. Then, equation (1) can be re-expressed as a logistic regression model for each of the $C-1$ binary responses (Kenward et al., 1994):

$$\text{logit}\{\Pr(U_{ijkc} = 1)\} = \eta_c + \boldsymbol{x}_{ijk}\boldsymbol{\beta}, \qquad c = 1, ..., C-1. \quad (2)$$

Let $\mu_{ijkc} = \text{E}(U_{ijkc}) = \Pr(Y_{ijk} \le c)$. Using matrix notation, we group each of the response vectors of the jth unit from the ith cluster as $\boldsymbol{U}'_{ij} = (\boldsymbol{U}'_{ij1}, ..., \boldsymbol{U}'_{ijt_{ij}})$. Similarly, let $\boldsymbol{\mu_{ij}} = \text{E}(\boldsymbol{U}_{ij})$, where $\boldsymbol{\mu'_{ij}} = (\boldsymbol{\mu}'_{ij1}, ..., \boldsymbol{\mu}'_{ijt_{ij}})$ and $\boldsymbol{\mu}'_{ijk} = (\mu_{ijk1}, ..., \mu_{ijk(C-1)})$. The covariates for the jth unit from the ith cluster are represented as $\boldsymbol{X}_{0ij} = (\boldsymbol{x}_{ij1}, ..., \boldsymbol{x}_{ijt_{ij}})'$. The complete data matrix needs to include the cut points of the response vector and the covariates. Let $\mathbf{1}_{t_{ij}}$ and $\mathbf{1}_{C-1}$ be vectors of 1's with length t_{ij} and $C-1$ respectively, and let $\boldsymbol{I}_{C-1}$ be the identity matrix of dimension $C-1$. The complete data matrix for the jth unit from the ith cluster is $\boldsymbol{X}_{ij} = (\mathbf{1}_{T_i} \otimes \boldsymbol{I}_{C-1}, \boldsymbol{X}_{0ij} \otimes \mathbf{1}_{C-1})$, where $\otimes$ represents the Kronecker product (Parsons et al., 2006). Finally, let $\boldsymbol{\beta} = (\eta_1, ..., \eta_{C-1}, \beta_1, ..., \beta_p)$ be the $(C-1+p) \times 1$ vector of coefficients. Let $\boldsymbol{V}_{ij}$ be the diagonal matrix containing the variance of elements of $\boldsymbol{U}_{ij}$, where $\text{var}(U_{ijkc}) = \mu_{ijkc}(1-\mu_{ijkc})$ and let $\boldsymbol{R}_{ij}$ be the matrix of correlations, which is assumed to be a function of α, between the elements of $\boldsymbol{U}_{ij}$. Finally, let $\boldsymbol{Z}'_{ij} = (\boldsymbol{U}_{ij} - \boldsymbol{\mu}_{ij})'\boldsymbol{V}_{ij}^{-1/2}$. We now construct the generalized sum of squares for error with two types of weights; cluster weights $(1/n_{ij})$ which is the inverse of cluster size at baseline for each participant i, and temporal weights $(1/t_{ij})$ which is the inverse of number of visits made by the jth tooth of the ith participant (Wang et al., 2011; Bible et al., 2016):

$$Q_W(\alpha, \boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \boldsymbol{Z}'_{ij}\boldsymbol{R}_{ij}^{-1}\boldsymbol{Z}_{ij} = \mathbf{0}. \quad (3)$$

2.2 | Specification of $\boldsymbol{R_{ij}}$

We specify $\boldsymbol{R}_{ij}$ using the same approach as Parsons et al. (2006). The matrix $\boldsymbol{R}_{ij}$ contains the correlation between every pair of elements in $\boldsymbol{U}_{ij}$. We let $\boldsymbol{R}_{ij} = \boldsymbol{C}(\alpha) \otimes \boldsymbol{S}$, where the $(t_{ij}-1) \times (t_{ij}-1)$ matrix $\boldsymbol{C}(\alpha)$ contains the temporal correlations between visits and the $(C-1) \times (C-1)$ matrix $\boldsymbol{S}$ contains the correlations between the binary responses within each visit, such that

$$\boldsymbol{C}(\alpha) = \begin{pmatrix} 1 & \alpha^{d_{12}} & \cdots & \alpha^{d_{1t_{ij}}} \\ \alpha^{d_{21}} & 1 & & \vdots \\ \vdots & & \ddots & \alpha^{d_{t_{ij}-1,t_{ij}}} \\ \alpha^{d_{t_{ij}1}} & \cdots & \alpha^{d_{t_{ij},t_{ij}-1}} & 1 \end{pmatrix},$$

$$\boldsymbol{S} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1(C-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(C-1)1} & \cdots & \rho_{(C-1)(C-1)} \end{pmatrix}.$$

In proportional odds logistic regression models, the correlation between U_{ijtc_1} and U_{ijtc_2} for $c_1 < c_2$ is given by $\rho_{c_1c_2} = \rho_{c_2c_1} = \exp(\hat{\eta}_{c_1} - \hat{\eta}_{c_2})^{1/2}$ (Kenward et al., 1994). The first-order autoregressive (AR1) structure is a popular choice for modeling the correlation between visits for longitudinal data.

If $d_{mn} = |m - n|$, then $\boldsymbol{C}(\alpha)$ has an AR1 structure and if all $d_{mn} = 1$, then $\boldsymbol{C}(\alpha)$ has an exchangeable structure.

2.3 | Estimation of β and α

Taking the partial derivative of $Q_W(\alpha, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and setting it equal to $\mathbf{0}$, we obtain the CWGEE score function:

$$\sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \boldsymbol{D}_{ij}' \boldsymbol{W}_{ij}^{-1} (\boldsymbol{U}_{ij} - \boldsymbol{\mu}_{ij}) = \mathbf{0} \tag{4}$$

where $\boldsymbol{D}_{ij} = \partial \boldsymbol{\mu}_{ij} / \partial \boldsymbol{\beta}$ and $\boldsymbol{W}_{ij} = \boldsymbol{V}_{ij}^{1/2} \boldsymbol{R}_{ij} \boldsymbol{V}_{ij}^{1/2}$. With the correct specification of $\boldsymbol{R}_{ij}$, this marginalization has the interpretation of describing a typical longitudinal experience of a typical unit from a typical cluster (Bible et al., 2016). The correlation between units within a cluster is accounted for by the cluster-level weights in equation (4). We still need to model the correlation structure between visits within a unit to increase the precision of $\boldsymbol{\beta}$ estimates.

The estimating procedure for α has two parts. Similar to the estimating equation for $\boldsymbol{\beta}$, the stage one estimating equation for α is given by taking the partial derivative of $Q_W(\alpha, \boldsymbol{\beta})$ with respect to α and setting it equal to 0:

$$\sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \boldsymbol{Z}_{ij}' \frac{\partial \boldsymbol{R}_{ij}^{-1}}{\partial \alpha} \boldsymbol{Z}_{ij} = 0. \tag{5}$$

The stage one estimator for α, however, is asymptotically biased (Chaganty and Shults, 1999). Therefore, a consistent stage two estimator for α is obtained by solving the following equation for α:

$$\sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} trace \left(\frac{\partial \boldsymbol{R}_{ij}^{-1}(\hat{\alpha}_0)}{\partial \hat{\alpha}_0} \boldsymbol{R}_{ij}(\alpha) \right) = \mathbf{0} \tag{6}$$

where $\hat{\alpha}_0$ is the solution to equation (5). The estimators for $\boldsymbol{\beta}$ and α are obtained by choosing a starting value for $\boldsymbol{\beta}$ (typically from fitting a GLM assuming independence between observations) and iterating through equations (4), (5) and (6) until convergence is reached.

2.4 | AR1 working correlation structure

If we choose the AR1 structure for $\boldsymbol{C}(\alpha)$, then equation (5) has a closed-form solution. The stage one estimator for α can be solved by the following formula:

$$\hat{\alpha}_0 = \frac{F_a + \sqrt{(F_a + F_b)(F_a - F_b)}}{F_b} \tag{7}$$

where

$$F_a = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \left(\sum_{k=1}^{t_{ij}} \boldsymbol{Z}_{ijk}' \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk} + \sum_{k=2}^{t_{ij}-1} \boldsymbol{Z}_{ijk}' \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk} \right)$$

and $F_b = 2 \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \sum_{k=1}^{t_{ij}-1} \boldsymbol{Z}_{ijk}' \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk+1}$. The stage two estimator for α is given by solving equation (6) which reduces to:

$$\hat{\alpha} = \frac{2\hat{\alpha}_0}{1 + \hat{\alpha}_0^2}. \tag{8}$$

Details of the derivations for equations (7) and (8) are shown in the Appendix and Supplementary Material respectively.

2.5 | Exchangeable working correlation structure

If we choose the exchangeable structure, the stage one estimating equation for α reduces to

$$\sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \left(g_{ai} \sum_{k=1}^{t_{ij}} \boldsymbol{Z}_{ijk}' \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk} - 2 g_{bi} \sum_{k=1}^{t_{ij}-1} \sum_{k'=k+1}^{t_{ij}} \boldsymbol{Z}_{ijk}' \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk'} \right) = 0$$

where $g_{ai} = \frac{\alpha_0^2(t_{ij}-1)(t_{ij}-2)+2\alpha_0(t_{ij}-1)}{\left[1+\alpha_0(t_{ij}-1)\right]^2}$ and $g_{bi} = \frac{1+\alpha_0^2(t_{ij}-1)}{\left[1+\alpha_0(t_{ij}-1)\right]^2}$. The solution for α_0 can be obtained through a root finding algorithm. The stage two estimator for α is given by

$$\hat{\alpha} = \left[\sum_{i=1}^{N} \left(\frac{1}{n_i} \right) \frac{(1 - t_{ij}) \left\{ \hat{\alpha_0}^2 (t_{ij} - 1) + 1 \right\}}{\left\{ 1 + \hat{\alpha}_0 (t_{ij} - 1) \right\}^2} \right]^{-1} \sum_{i=1}^{N} \left(\frac{1}{n_i} \right) \frac{(1 - t_{ij}) \hat{\alpha}_0 \left\{ \hat{\alpha}_0 (t_{ij} - 2) + 2 \right\}}{\left\{ 1 + \hat{\alpha}_0 (t_{ij} - 1) \right\}^2}.$$

2.6 | Variance-covariance matrix for $\hat{\boldsymbol{\beta}}$

The robust variance-covariance matrix, $\boldsymbol{\Psi}$, for $\hat{\boldsymbol{\beta}}$ is constructed in a similar way as done by Wang et al. (2011) using the familiar sandwich estimator:

$$\hat{\boldsymbol{\Psi}} = \hat{\boldsymbol{B}}^{-1} \hat{\boldsymbol{M}} \hat{\boldsymbol{B}}^{-1} \tag{9}$$

where

$$\hat{\boldsymbol{B}} = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \hat{\boldsymbol{D}}' \hat{\boldsymbol{W}}_{ij}^{-1} \hat{\boldsymbol{D}} \tag{10}$$

and

$$\hat{\boldsymbol{M}} = \sum_{i=1}^{N} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \hat{\boldsymbol{D}}' \hat{\boldsymbol{W}}_{ij}^{-1} (\boldsymbol{Y}_{ij} - \hat{\boldsymbol{\mu}}_{ij}) \right\} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \hat{\boldsymbol{D}}' \hat{\boldsymbol{W}}_{ij}^{-1} (\boldsymbol{Y}_{ij} - \hat{\boldsymbol{\mu}}_{ij}) \right\}' \quad (11)$$

The proof for the asymptotic normality of $\hat{\boldsymbol{\beta}}$ estimated using CWGEE is provided in the appendix of Williamson et al. (2003).

An R package to compute our proposed CWGEE is available for download from the first author's personal website.

3 | SIMULATION STUDY

3.1 | Simulating longitudinal clustered ordinal data with informative cluster size

To simulate longitudinal clustered data with continuous outcomes, Bible et al. (2016) used a linear mixed effects model with a random participant-level effect and a random tooth-level effect. To induce ICS, they let the cluster size per participant be a function of the random participant-level effect. However, unlike the linear setting, with an ordinal outcome, the cluster-specific parameter estimates estimated by fitting a generalized linear mixed model and the marginal (or the population-average) parameter estimates estimated using GEE are not the same (Fitzmaurice et al., 2011). To overcome this issue, the bridge distribution (Wang and Louis, 2003) was used to obtain the marginal probability of success when fitting a random intercept logistic regression model of the form:

$$p_{ijkc} = \Pr(U_{ijkc} = 1 | b_i, x_{ijk}, \boldsymbol{\beta}) = \frac{\exp(b_i + \phi^{-1} x'_{ijk} \boldsymbol{\beta})}{1 + \exp(b_i + \phi^{-1} x'_{ijk} \boldsymbol{\beta})} \quad (12)$$

where b_i follows a bridge distribution with density $f_b(b_i|\phi) = \frac{1}{2\pi} \frac{\sin(\phi\pi)}{\cosh(\phi b_i) + \cos(\phi\pi)}, -\infty < b_i < \infty, 0 < \phi < 1$. In equation (12), $\boldsymbol{\beta}$ has a marginal interpretation, as desired.

We extended the method described by Parzen et al. (2011) that utilizes a Gaussian copula of the bridge distribution to simulate temporally correlated clustered ordinal data, Y_{ijk}. We used the exchangeable correlation structure with parameter τ to generate the correlation between teeth and used the AR1 correlation structure with parameter α to generate the correlation within a tooth over time. For each participant i, we computed the baseline hazard λ_i as a function of the participant-specific set of random effects $\boldsymbol{b}_i$, which followed the bridge distribution. The number of teeth for each participant (n_i) was generated from a binomial distribution with size 28 and probability λ_i. We also varied the number of temporal observations made on each tooth j of each participant i from 2 to 5. Following Bible et al. (2016), the probabilities of each tooth j's number of observations was determined by n_i. A detailed description of how the data was simulated is presented in Supplementary Material.

The ordinal outcome for the simulation study had $C = 4$ categories. The covariates included a participant-level binary exposure indicator (x_i) with the first half ($N/2$) of the participants having the exposure, visit number (visit_{ijk}), and the interaction between exposure and visit. For the jth tooth of the ith participant at the kth visit, our simulated model had the form:

$$\text{logit}\{\Pr(Y_{ijk} \leq c)\} = \eta_c + \beta_1 x_i + \beta_2 \text{visit}_{ijk} + \beta_3 x_i \times \text{visit}_{ijk} \quad c = 1, 2, 3. \quad (13)$$

The true values for the parameters were: $(\eta_1 - \eta_3, \beta_1 - \beta_3) = (1, 2, 3, -0.5, 0.1, 0.5)$.

We simulated 1,000 data sets for each scenario. We varied the number of participants from small to large ($N = 20, 100, 500$). The maximum number of teeth per participant (m) was set at 28 (maximum total number of teeth in an adult human excluding the third molars). We also varied the levels of correlation between teeth (τ) and between visits within a tooth (α) from none to high (none: $\tau = 0, \alpha = 0$; low: $\tau = 0.25, \alpha = 0.4$; medium: $\tau = 0.5, \alpha = 0.6$, high: $\tau = 0.75, \alpha = 0.8$). For each simulated data set, we applied CWGEE with independence (Ind), AR1, and exchangeable (Exch) working correlation structures. We also applied GEE functions from two existing R packages, the *ordgee* function in *geepack* (Hojsgaard et al., 2006) using the independence working correlation structure (ORDGEE Ind), and the *ordLORgee* function in *multgee* (Touloumis, 2015) also using the independence working correlation structure (MULTGEE Ind).

For each simulation scenario and method, we computed the mean estimates, the mean robust standard errors (SEs), the standard deviation (SD) of the 1,000 parameter estimates (a.k.a. empirical SEs), the mean relative biases, and the coverage probabilities from 95% confidence intervals of each parameter estimate in equation (13). The relative bias was obtained by calculating the relative difference between each of the 1,000 parameter estimates and the respective true value. The coverage probability was obtained by calculating the percentage of times the 95% confidence interval for each of the 1,000 parameter estimates include the respective true parameter value.

3.2 | Simulation results: ICS

Results from the simulation scenarios with informative cluster size with small and medium sample sizes ($N = 20, 100$)

and medium levels of correlation ($\tau = 0.5, \alpha = 0.6$) are shown in Table 1. Results from the large sample size ($N = 500$) are presented in Supplementary Material Table 1. The mean estimates, mean SEs, and SDs of the estimates across the three CWGEE methods were all similar. The difference between mean SEs and SDs within each method decreased as the sample size increased. In general, the mean SEs of the CWGEE methods were slightly larger than the mean SEs of MULTGEE Ind regardless of sample size. ORDGEE Ind encountered convergence issues. Only 444 simulations out of 1,000 converged when sample size was small ($N = 20$). The convergence rate did not improve when we relaxed the convergence criteria for the Fisher-scoring algorithm from the default of 0.0001 to 0.01. Convergence was not an issue with medium and large sample sizes nor with the other methods using any sample size. The mean SEs and the SDs of the parameter estimates from ORDGEE Ind were consistently larger compared to those from the other methods.

In Figure 1, we depict how sample size and levels of correlation impact coverage probabilities and absolute relative biases. The vertical axis indicates coverage probability with the dotted horizontal line representing 95%, and the horizontal axis indicates increasing levels of correlation. The size of the bubble is proportional to the absolute relative bias (%). An ideal bubble is small in size and located close to the 95% line. Because all CWGEE methods performed similarly, we only show results from CWGEE AR1. In general, the parameter estimates from CWGEE AR1 (black) were lowest in bias and had better coverage probability compared to those from MULTGEE Ind (white) and ORDGEE Ind (gray) across all sample sizes and correlation levels. The largest discrepancies in bias and coverage probability between the methods were observed in the three cut-off parameters ($\eta_1 - \eta_3$). For some of the parameter estimates (η_2, η_3, β_3), CWGEE AR1 suffered lower coverage probabilities, especially when the sample size was small, but still performed better than the other two methods.

3.3 | Simulation results: No ICS

To evaluate how CWGEE performs for ordinal outcomes when cluster size is not informative (i.e. outcome is unrelated to cluster size), we simulated ordinal data where the number of teeth for each participant was randomly generated from a binomial distribution with size 28 and probability 0.75. Results from the scenario with medium sample size ($N = 100$) and medium levels of correlation ($\tau = 0.5, \alpha = 0.6$) are shown in Table 2.

Simulation results from CWGEE were extremely similar across all three working correlation matrix structures (Ind, AR1, Exch). The results from CWGEE Ind were almost identical to the results from MULTGEE Ind. In general, all methods exhibited low relative biases and coverage probabilities close to 95%. Importantly, CWGEE methods performed comparably well to non-weighted GEE methods in the case of no ICS. We observed some discrepancy between the two non-weighted methods, especially in the mean SEs. In general, the SEs of ORDGEE Ind were larger than those of MULTGEE Ind for all parameters. Relative biases and coverage probabilities for the cut-off estimates were comparable between the two methods, but relative biases of β_2 and β_3 estimated by ORDGEE Ind were higher compared to those estimated by MULTGEE Ind. Coverage probabilities of all the predictors ($\beta_1 - \beta_3$) estimated by ORDGEE Ind were lower compared to those estimated by MULTGEE Ind.

4 | THE DEPARTMENT OF VETERANS AFFAIRS DENTAL LONGITUDINAL STUDY (VADLS)

The VADLS was initiated in 1969 as an extension of the Normative Aging Study (Kapur et al., 1972). Each participant's health status was measured approximately every three years. Baseline medical and dental examinations for these participants occurred between 1981 and 2011. We restricted the analysis to participants who had complete CAL records on all existing teeth. The total number of participants for our analysis was 456 with a total of 9622 teeth. Over the course of the study, 245 participants lost at least one tooth, a total of 965 teeth.

The ordinal CAL score has four categories (0–3), where a higher score indicates a worse prognosis of periodontal disease. The relationship between baseline number of teeth and mean CAL score per participant is shown in the left panel of Figure 2. Participants with a greater number of teeth at baseline are more likely to have lower mean baseline CAL and vice versa. The Pearson correlation coefficient (95% CI) was -0.374 (-0.450, -0.292). There exists a moderately strong indirect association between baseline number of teeth and mean CAL score per participant, indicating the presence of ICS in this data set. Because the outcome, CAL score, is related to cluster size, the use of CWGEE to account for ICS is appropriate. In the right panel in Figure 2, we see that as the baseline number of teeth increases, the maximum number of temporal observations made on each participant increases as well. Participants with fewer baseline number of teeth are quicker to lose their existing teeth compared to those more teeth. This relationship corroborates the use of the second set of weights based on the number of temporal observations made on each tooth per participant.

We modeled the association of CAL score as an ordinal variable with visit number, baseline age, smoking status (yes/no), metabolic syndrome (MetS) status (yes/no)

TABLE 1 Simulation results when cluster size is informative ($N = 20, 100$, $\tau = 0.5$, $\alpha = 0.6$)

Parameter	Truth	Results	MULTGEE Ind	ORDGEE Ind	CWGEE Ind	AR1	Exch
				$N = 20$			
η_1	1	Mean Est	1.688	1.700	1.069	1.067	1.059
		Mean SE	0.569	0.891	0.627	0.625	0.621
		SD Est	0.613	1.340	0.739	0.736	0.726
		Rel Bias (%)	68.8	70.0	6.9	6.7	5.9
		Cov Prob (%)	75.4	77.5	88.1	87.8	88.7
η_2	2	Mean Est	2.983	2.942	2.223	2.222	2.214
		Mean SE	0.610	0.915	0.683	0.681	0.679
		SD Est	0.666	1.370	0.834	0.832	0.826
		Rel Bias (%)	49.2	47.1	11.1	11.1	10.7
		Cov Prob (%)	62.1	73.4	85.9	85.2	86.8
η_3	3	Mean Est	4.406	4.199	3.608	3.608	3.604
		Mean SE	0.735	0.968	0.815	0.814	0.812
		SD Est	0.820	1.388	1.140	1.138	1.136
		Rel Bias (%)	46.9	40.0	20.3	20.3	20.1
		Cov Prob (%)	53.0	70.3	78.5	78.1	79.0
β_1	-0.5	Mean Est	-0.693	-0.492	-0.593	-0.585	-0.588
		Mean SE	0.863	1.443	0.950	0.950	0.944
		SD Est	0.952	5.162	1.113	1.113	1.100
		Rel Bias (%)	38.6	-1.6	18.6	17.0	17.6
		Cov Prob (%)	90.5	77.4	89.9	89.9	90.5
β_2	0.1	Mean Est	0.155	0.245	0.114	0.116	0.120
		Mean SE	0.196	0.377	0.225	0.226	0.221
		SD Est	0.218	0.611	0.293	0.291	0.284
		Rel Bias (%)	55.1	145.2	13.6	15.7	20.5
		Cov Prob (%)	89.5	76.4	84.6	85.1	84.8
β_3	0.5	Mean Est	0.710	0.892	0.633	0.629	0.630
		Mean SE	0.371	0.655	0.412	0.418	0.414
		SD Est	0.431	1.994	0.558	0.565	0.556
		Rel Bias (%)	42.0	78.4	26.6	25.9	26.0
		Cov Prob (%)	86.9	73.1	84.4	85.1	85.7
Convergence rate (%)			100.0	44.4	100.0	100.0	100.0
				$N = 100$			
η_1	1	Mean Est	1.742	1.662	1.044	1.035	1.033
		Mean SE	0.272	0.506	0.316	0.309	0.304
		SD Est	0.275	0.592	0.330	0.323	0.317
		Rel Bias (%)	74.2	66.2	4.4	3.5	3.3
		Cov Prob (%)	23.6	70.5	92.9	93.2	93.1
η_2	2	Mean Est	2.998	2.919	2.066	2.057	2.056
		Mean SE	0.293	0.515	0.3.05	0.345	0.341
		SD Est	0.296	0.600	0.369	0.363	0.360
		Rel Bias (%)	49.9	46.0	3.3	2.9	2.8
		Cov Prob (%)	8.5	56.1	92.3	92.8	92.7
η_3	3	Mean Est	4.318	4.245	3.127	3.121	3.121
		Mean SE	0.344	0.544	0.431	0.428	0.427
		SD Est	0.350	0.623	0.475	0.471	0.472
		Rel Bias (%)	43.9	41.5	4.2	4.0	4.0
		Cov Prob (%)	2.6	37.9	90.2	90.3	90.3
β_1	-0.5	Mean Est	-0.723	-0.844	-0.517	-0.513	-0.517
		Mean SE	0.412	0.905	0.476	0.465	0.460
		SD Est	0.418	1.034	0.502	0.483	0.478

(Continues)

TABLE 1 (Continued)

Parameter	Truth	Results	MULTGEE Ind	ORDGEE Ind	CWGEE Ind	AR1	Exch
		Rel Bias (%)	44.6	68.7	3.4	2.6	3.5
		Cov Prob (%)	91.0	89.1	93.6	93.9	94.1
β_2	0.1	Mean Est	0.136	0.200	0.082	0.088	0.09
		Mean SE	0.096	0.235	0.120	0.116	0.112
		SD Est	0.097	0.281	0.129	0.125	0.120
		Rel Bias (%)	35.7	100.2	-18.3	-12.5	-10.2
		Cov Prob (%)	92.8	90.5	92.0	92.9	92.5
β_3	0.5	Mean Est	0.689	0.819	0.525	0.523	0.525
		Mean SE	0.180	0.531	0.216	0.214	0.209
		SD Est	0.184	0.668	0.234	0.225	0.220
		Rel Bias (%)	37.9	63.7	5.0	4.6	5.1
		Cov Prob (%)	82.7	88.1	92.2	93.8	93.1
Convergence rate (%)			100.0	98.0	100.0	100.0	100.0

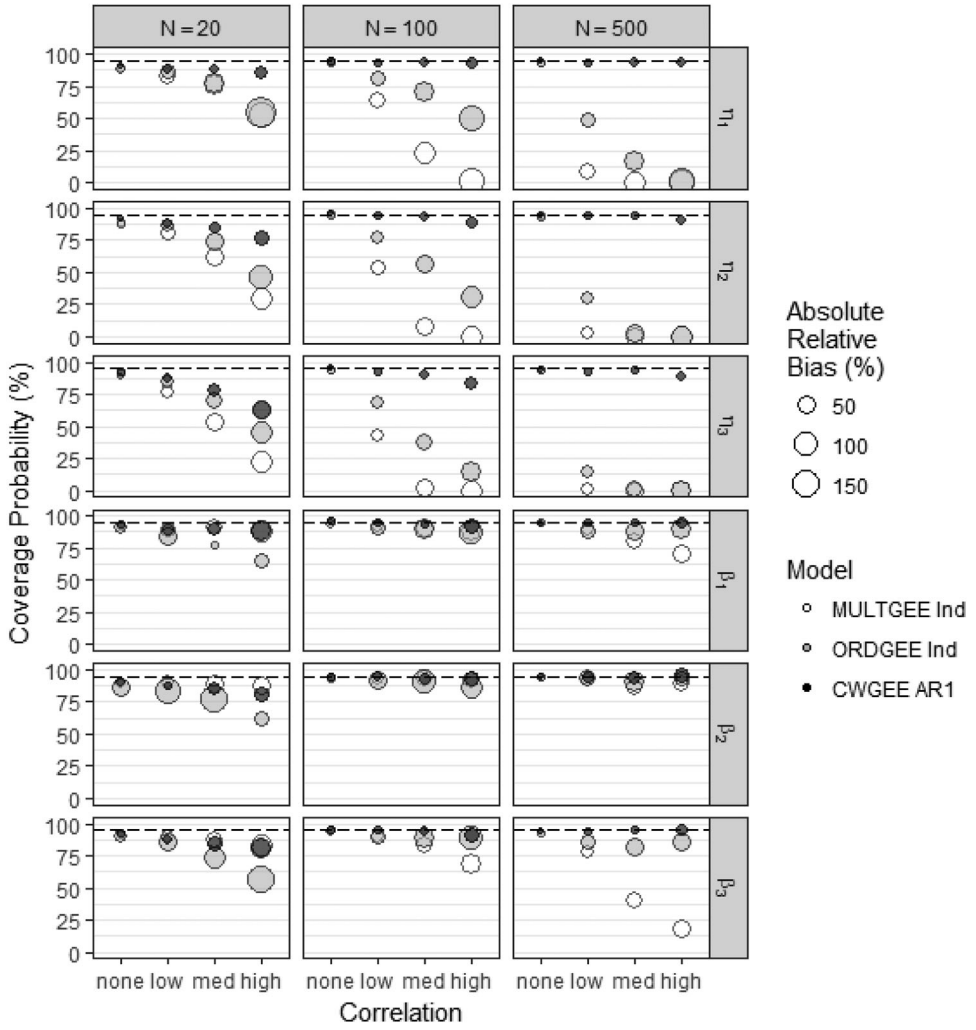

FIGURE 1 Coverage probability and absolute relative bias of all parameters by sample size and correlation for three models, MULTGEE Ind, ORDGEE Ind, CWGEE AR1 (proposed method) from simulation study. Correlation (τ is the correlation parameter between teeth and α is the correlation parameter over time within a tooth): none ($\tau = 0, \alpha = 0$); low ($\tau = 0.25, \alpha = 0.4$); med ($\tau = 0.5, \alpha = 0.6$); high ($\tau = 0.75, \alpha = 0.8$). Parameters: $\eta_1 - \eta_3, \beta_1 - \beta_3$.

TABLE 2 Simulation results when cluster size is not informative ($N = 100$, $\tau = 0.5$, $\alpha = 0.6$)

			MULTGEE	ORDGEE	CWGEE		
Parameter	**Truth**	**Results**	**Ind**	**Ind**	**Ind**	**AR1**	**Exch**
η_1	1	Mean Est	0.990	0.990	0.990	0.990	0.990
		Mean SE	0.240	0.310	0.240	0.240	0.240
		SD Est	0.240	0.370	0.240	0.250	0.240
		Rel Bias (%)	-1.1	-1.4	-0.9	-0.8	-0.9
		Cov Prob (%)	94.8	87.4	94.4	94.1	94.5
η_2	2	Mean Est	2.000	2.000	2.000	2.000	2.000
		Mean SE	0.260	0.320	0.260	0.260	0.260
		SD Est	0.270	0.390	0.270	0.270	0.270
		Rel Bias (%)	0.0	-0.1	0.1	0.1	0.1
		Cov Prob (%)	93.6	87.3	93.3	93.4	93.7
η_3	3	Mean Est	3.020	3.020	3.020	3.020	3.020
		Mean SE	0.310	0.350	0.310	0.310	0.310
		SD Est	0.320	0.430	0.320	0.330	0.320
		Rel Bias (%)	0.6	0.7	0.6	0.7	0.6
		Cov Prob (%)	93.0	88.2	92.7	92.4	92.5
β_1	-0.5	Mean Est	-0.490	-0.510	-0.490	-0.490	-0.490
		Mean SE	0.350	0.510	0.350	0.350	0.350
		SD Est	0.350	0.630	0.350	0.350	0.350
		Rel Bias (%)	-1.2	1.0	-1.2	-1.0	-1.0
		Cov Prob (%)	94.7	88.0	94.2	93.9	94.3
β_2	0.1	Mean Est	0.110	0.120	0.110	0.110	0.110
		Mean SE	0.070	0.110	0.070	0.070	0.070
		SD Est	0.080	0.140	0.080	0.080	0.080
		Rel Bias (%)	8.7	18.1	8.5	7.8	8.3
		Cov Prob (%)	92.7	88.3	92.6	92.5	92.7
β_3	0.5	Mean Est	0.500	0.520	0.500	0.500	0.500
		Mean SE	0.130	0.220	0.130	0.130	0.130
		SD Est	0.140	0.280	0.140	0.140	0.140
		Rel Bias (%)	0.2	4.9	0.1	0.3	0.3
		Cov Prob (%)	92.0	88.5	92.2	92.2	92.9
Convergence rate (%)			100.0	100.0	100.0	100.0	100.0

as defined by the National Cholesterol Education Program Adult Treatment Panel III criteria (Kaye et al., 2016), education-level (college degree or higher/none) and the interaction between visit and each of the aforementioned covariates:

$$\begin{aligned}\text{logit}\{\Pr(\text{CAL}_{ijk} \le c)\} &= \eta_c + \beta_1 \text{visit}_{ijk} + \beta_2 \text{age}_i + \beta_3 \text{smoke}_i \\ &+ \beta_4 \text{MetS}_i + \beta_5 \text{education}_i + \beta_6 \text{age}_i \times \text{visit}_{ijk} \\ &+ \beta_7 \text{smoke}_i \times \text{visit}_{ijk} + \beta_8 \text{MetS}_i \times \text{visit}_{ijk} \\ &+ \beta_9 \text{education}_i \times \text{visit}_{ijk}\end{aligned}$$

where $c = 0, 1, 2$ and the probabilities modeled are cumulated over the lower (healthier) scores. Results from the analysis are presented in Table 3. As anticipated from the simulation study in Section 3, the results from the three CWGEEs (Ind, AR1, Exch) were similar. The interaction between visit and smoking was statistically significant at the 0.05 level of significance in CWGEE Ind ($p = 0.034$), CWGEE AR1 ($p = 0.023$), and CWGEE Exch ($p = 0.008$) but not in MULTGEE Ind ($p = 0.123$) and ORDGEE Ind ($p = 0.279$). Based on the simulation study and the data structure, we feel most appropriate to interpret the results from CWGEE AR1. Holding other variables constant, the odds ratio of smokers and non-smokers having a healthier CAL score over each consecutive visit are 0.951 and 1.088 respectively, indicating that smokers are more likely to experience worse prognosis of periodontal disease over time compared to non-smokers.

5 | DISCUSSION

In this article, we developed a longitudinal CWGEE to model ordinal categorical outcomes, which extends a method proposed by Bible et al. (2016) for continuous outcomes. The study of ordinal outcomes raises a unique set of challenges beyond continuous outcomes. Our research is an important contribution to the current literature of ICS because many clinical outcomes that are potentially associated with cluster size are measured using an ordinal scoring system, including dental studies and patient-satisfactory surveys. Thus far, much

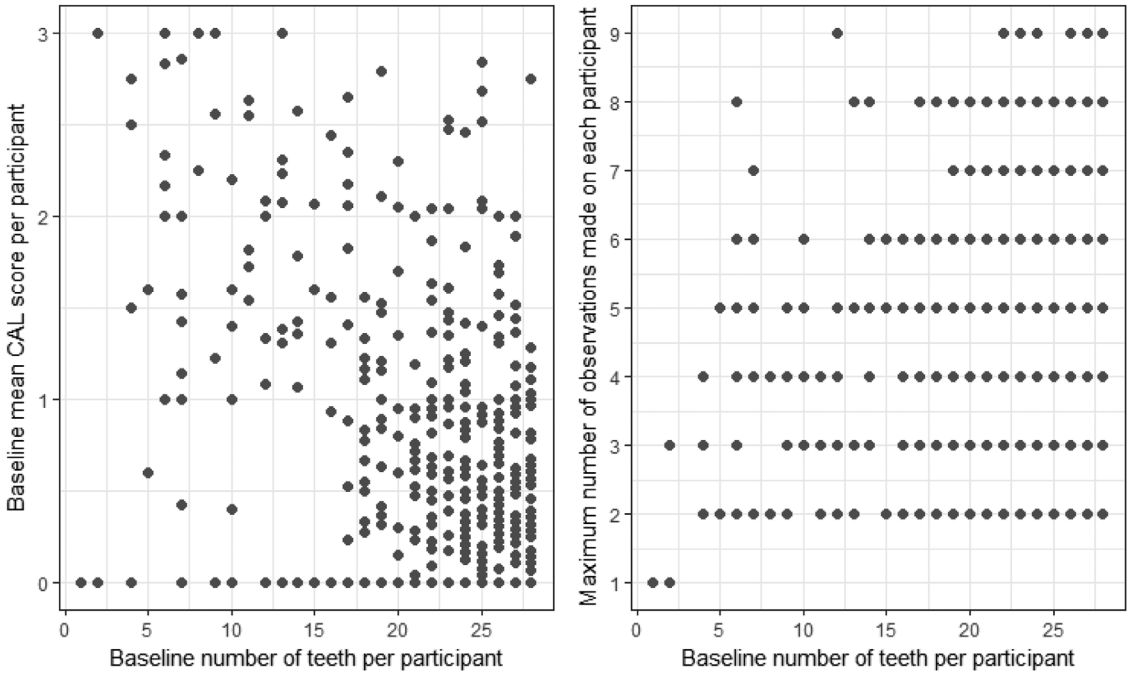

FIGURE 2 Left Panel: Relationship between number of teeth and mean clinical attachment loss (CAL) score (0: <2mm, 1: 2-2.9mm, 2: 3-4.9mm, 3: ≥5mm) at baseline per participant from Dapartment of Veterans Affairs Longitudinal Dental Study ($N = 456$). Right Panel: Relationship between number of teeth at baseline and maximum number of temporal observation made on each participant's tooth.

of the work on modeling marginal inference in clustered longitudinal studies with potential ICS has focused on continuous outcomes.

In our simulation study, we did not observe noticeable differences in relative biases and coverage probabilities between various choices of the working correlation structure within the proposed CWGEE approach. This is in agreement with the simulation results for continuous outcomes from Wang et al. (2011) and Bible et al. (2016). We observed a considerable improvement in bias and coverage probability when using CWGEE instead of conventional GEE to analyze data with ICS. When no ICS is present, the results from CWGEE methods were comparable to the results from conventional GEE methods. This was also observed in the results from Wang et al. (2011) and we feel comfortable analyzing data with any degree of ICS using the proposed CWGEE method.

We observed an underestimation of the SEs using the sandwich estimators for the small sample size scenario in our simulation study. This is a recognized problem in GEE and several authors have developed adjustments. We applied the degrees of freedom (DF) correction proposed by MacKinnon and White (1985) and observed a closer agreement between the DF-corrected SEs and the empirical SEs, and

TABLE 3 Results from the analysis of VA Dental Longitudinal Study showing coefficient estimates (SEs), $N = 456$

	MULTGEE	ORDGEE	CWGEE		
Variable	**Ind**	**Ind**	**Ind**	**AR1**	**Exch**
Int 1	1.949 (0.528)	1.908 (0.700)	2.038 (0.688)	2.022 (0.703)	1.922 (0.665)
Int 2	3.139 (0.530)	3.099 (0.701)	3.169 (0.692)	3.142 (0.708)	3.046 (0.670)
Int 3	4.082 (0.533)	4.053 (0.705)	4.052 (0.695)	4.014 (0.713)	3.918 (0.675)
Visit	0.109 (0.179)	0.073 (0.224)	0.080 (0.189)	0.084 (0.185)	0.106 (0.162)
Age	-0.030 (0.009)	-0.030 (0.011)	-0.035 (0.011)	-0.034 (0.011)	-0.032 (0.011)
Smoking	0.033 (0.221)	0.022 (0.293)	-0.012 (0.232)	0.023 (0.228)	0.085 (0.218)
MetS	-0.274 (0.122)	-0.251 (0.161)	-0.343 (0.150)	-0.314 (0.143)	-0.311 (0.138)
Education	0.420 (0.134)	0.505 (0.159)	0.389 (0.178)	0.373 (0.180)	0.366 (0.170)
Visit × Age	-0.004 (0.003)	-0.004 (0.004)	-0.004 (0.003)	-0.004 (0.003)	-0.005 (0.003)
Visit × Smoking	-0.147 (0.095)	-0.138 (0.128)	-0.152 (0.083)	-0.157 (0.079)	-0.175 (0.074)
Visit × MetS	0.053 (0.038)	0.047 (0.052)	0.069 (0.041)	0.063 (0.038)	0.062 (0.035)
Visit × Education	0.008 (0.042)	0.017 (0.057)	0.035 (0.046)	0.040 (0.045)	0.048 (0.039)

also an improvement in coverage probabilities for small number of clusters (20 and 50). When number of clusters was greater than 100, the DF-corrected SEs were comparable to the uncorrected SEs.

We set the maximum cluster size to be 28 in our simulations to closely resemble a dental study. However, maximum cluster size may be much smaller in other applications. To investigate the performance of the proposed CWGEE approach on smaller cluster size, we conducted additional simulations of maximum cluster sizes equal to 3, 5, and 10 for number of clusters of 100 and 500. The difference in performance between the proposed CWGEE and unweighted GEEs is more apparent when cluster sizes are larger because the degree of informative cluster size is stronger with larger cluster size but we did not observe convergence issues or reduction in performance on parameter estimation with our proposed model when maximum cluster size is small. The additional simulation results are presented in Supplementary Material.

Our approach to directly estimate the correlation structure between visits within a unit adheres closely to the approach described by Shults and Ardythe (2002) when implementing the method of QLS. An alternative approach to model the association between ordinal observations is to use the global odds ratio (Williamson et al., 1995) or the local odds ratio (Touloumis et al., 2013) parameterization. However, none of these approaches have yet been developed for use in QLS estimation. Although the odds ratio is a more natural measure of association and is subject to less constraints for categorical variables compared to the correlation coefficient (Shults and Hilbe, 2014), in our article, we decided to treat the association parameter as a "nuisance" because our focus lies in estimating the association between the ordinal outcome and its predictors and not in estimating the measure of temporal association between the ordinal observations. We obtained good results in bias and coverage probability from our simulation studies with a diverse range of correlations for both intra-teeth and inter-teeth association in situations when ICS is present and absent. Nonetheless, comparing the performance of CWGEE based on global or local odds ratios is a future topic of interest.

ACKNOWLEDGEMENTS

We thank the Associate Editor and the two anonymous referees for comments leading to an improved manuscript. This research was supported by NIH grants F31DE027589 (PI: Mitani), R03DE021730 (PI: Kaye), and R01CA226805 (PI: Nelson). We also acknowledge Raul Garcia, DMD, MMedSc who is the Principal Investigator and examiner for the Dental Longitudinal Study. The Dental Longitudinal Study and Normative Aging Study are components of the Massachusetts Veterans Epidemiology Research and Information Center which is supported by the VA Cooperative Studies Program. Views expressed in this article are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs.

ORCID

Aya A. Mitani http://orcid.org/0000-0002-0373-5032

REFERENCES

Bible, J., Beck, J. D., and Datta, S. (2016). Cluster adjusted regression for displaced subject data (cards): Marginal inference under potentially informative temporal cluster size profiles. *Biometrics* 72, 441–451.

Chaganty, N. R. (1997). An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference* 63, 39–54.

Chaganty, N. R. and Shults, J. (1999). On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *Journal of Statistical Planning and Inference* 76, 145–161.

Chaurasia, A., D., L., and Albert, P. S. (2018). Pattern-mixture models with incomplete informative cluster size: application to a repeated pregnancy study. *Journal of the Royal Statistical Society – Series C* 67, 255–273.

Dunson, D. B., Chen, Z., and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* 59, 521–530.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*. Wiley, 2 edition.

Hoffman, E. B., Sen, P. K., and Weinberg, C. R. (2001). Within-cluster resampling. *Biometrika* 88, 1121–1134.

Hojsgaard, S., Halekoh, U., and Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software* 15, 1–11.

Iosif, A. and Sampson, A. R. (2014). A model for repeated clustered data with informative cluster size. *Statistics in Medicine* 33, 738–759.

Kapur, K. K., Glass, R. L., Loftus, E. R., Alman, J. E., and Feller, R. P. (1972). The veterans administration longitudinal study of oral health and disease. *Aging and Human Development* 3, 125–137.

Kaye, E. K., Chen, N., Cabral, H. J., Vokonas, P., and Garcia, R. I. (2016). Metabolic syndrome and periodontal disease progression in men. *Journal of Dental Research* 95, 822–828.

Kenward, M. G., Lesaffre, E., and Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* 50, 945–953.

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

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

MacKinnon, J. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Economics* 29, 305–325.

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

Panageas, K. S., Schrag, D., Localio, R. A., Venkatraman, E. S., and Begg, C. B. (2007). Properties of analysis methods that account for

clustering in volume–outcome studies when the primary predictor is cluster size. *Statistics in Medicine* 26, 2017–2035.

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

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

Seaman, S., Pavlou, M., and Copas, A. (2014). Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Statistics in Medicine* 33, 5371–5387.

Shults, J. and Ardythe, M. L. (2002). Use of quasi-least squares to adjust for two levels of correlation. *Biometrics* 58, 521–530.

Shults, J. and Hilbe, J. M. (2014). *Quasi-Least Squares Regression*. Chapman & Hall/CRC.

Touloumis, A. (2015). R package multgee: A generalized estimating equations solver for multinomial response. *Journal of Statistical Software* 64, 1–14.

Touloumis, A., Agresti, A., and Kateri, M. (2013). Gee for multinomial responses using a local odds ratios parameterization. *Biometrics* 69, 633–640.

Wang, M., Kong, M. Y., and Datta, S. (2011). Inference for marginal linear models for clustered longitudinal data with potentially informative cluster sizes. *Statistical Methods in Medical Research* 20, 347–367.

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

Williamson, J. M., Datta, S., and Satten, G. A. (2003). Marginal analyses of clustered data when cluster size is informative. *Biometrics* 59, 36–42.

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

Zhang, B., Liu, W., Zhang, Z., Qu, Y., Chen, Z., and Albert, P. S. (2017). Modeling of correlated data with informative cluster sizes: An evaluation of joint modeling and within-cluster resampling approaches. *Statistical Methods in Medical Research* 26, 1881–1895.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2, 3, and 5 are available with this article at the Biometrics website on Wiley Online Library. The URL for the R package mentioned at the end of Section 2 is https://github.com/AyaMitani/CWGEE.

How to cite this article: Mitani AA, Kaye EK, Nelson KP. Marginal analysis of ordinal clustered longitudinal data with informative cluster size. *Biometrics.* 2019;1–12. https://doi.org/10.1111/biom.13050

Appendix

Derivation of equation (7) in Section 2.4

If $\boldsymbol{C}_{ij}(\alpha)$ is an first-order autoregressive (AR1) structure,

$$\boldsymbol{C}_{ij}(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{t_{ij}-1} \\ \alpha & 1 & \alpha & \ddots & \alpha^{t_{ij}-2} \\ \alpha^2 & \alpha & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \alpha \\ \alpha^{t_{ij}-1} & \cdots & \alpha^2 & \alpha & 1 \end{pmatrix} \text{and}$$

$$\boldsymbol{C}(\alpha)^{-1} = \frac{1}{1-\alpha^2} \begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \ddots & \vdots \\ 0 & -\alpha & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\alpha^2 & -\alpha \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix}.$$

Then,

$$\boldsymbol{R}_{ij}(\alpha)^{-1} = \boldsymbol{C}_{ij}(\alpha)^{-1} \otimes \boldsymbol{S}^{-1} = \frac{1}{1-\alpha^2} \times \begin{pmatrix} \boldsymbol{S}^{-1} & -\alpha\boldsymbol{S}^{-1} & 0 & \cdots & 0 \\ -\alpha\boldsymbol{S}^{-1} & (1+\alpha^2)\boldsymbol{S}^{-1} & -\alpha\boldsymbol{S}^{-1} & \ddots & \vdots \\ 0 & -\alpha\boldsymbol{S}^{-1} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & (1+\alpha^2)\boldsymbol{S}^{-1} & -\alpha\boldsymbol{S}^{-1} \\ 0 & \cdots & 0 & -\alpha\boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} \end{pmatrix}.$$

Because $\boldsymbol{S}^{-1}$ is free of α,

$$\frac{\partial \boldsymbol{R}_{ij}(\alpha)^{-1}}{\partial \alpha} = \frac{\partial \boldsymbol{C}_{ij}(\alpha)^{-1}}{\partial \alpha} \otimes \boldsymbol{S}^{-1} = \frac{1}{(1-\alpha^2)^2} \begin{pmatrix} 2\alpha\boldsymbol{S}^{-1} & -(1+\alpha^2)\boldsymbol{S}^{-1} & 0 & \cdots & 0 \\ -(1+\alpha^2)\boldsymbol{S}^{-1} & 4\alpha\boldsymbol{S}^{-1} & -(1+\alpha^2)\boldsymbol{S}^{-1} & \ddots & \vdots \\ 0 & -(1+\alpha^2)\boldsymbol{S}^{-1} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 4\alpha\boldsymbol{S}^{-1} & -(1+\alpha^2)\boldsymbol{S}^{-1} \\ 0 & \cdots & 0 & -(1+\alpha^2)\boldsymbol{S}^{-1} & 2\alpha\boldsymbol{S}^{-1} \end{pmatrix}.$$

Then, equation (5) is equivalent to

$$\sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} \left[\alpha S_1 - (1+\alpha^2) S_2\right]$$

where

$$S_1 = \sum_{k=1}^{t_{ij}} \boldsymbol{Z}'_{ijk} \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk} + \sum_{k=2}^{t_{ij}-1} \boldsymbol{Z}'_{ijk} \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk},$$

$$S_2 = \sum_{k=1}^{t_{ij}-1} \boldsymbol{Z}'_{ijk} \boldsymbol{S}^{-1} \boldsymbol{Z}_{ijk+1}.$$

We can rearrange the above equation to

$$\alpha^2 \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} S_2 - \alpha \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} S_1 + \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{t_{ij}} S_2 = 0$$

and solve for α using the quadratic formula, yielding equation (7).