Statistics in Medicine

Research Article

Received 18 January 2013, Accepted 16 January 2014 Published online 6 February 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6106

Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data

Philip M. Westgate[*†]

Generalized estimating equations are commonly used to analyze correlated data. Choosing an appropriate working correlation structure for the data is important, as the efficiency of generalized estimating equations depends on how closely this structure approximates the true structure. Therefore, most studies have proposed multiple criteria to select the working correlation structure, although some of these criteria have neither been compared nor extensively studied. To ease the correlation selection process, we propose a criterion that utilizes the trace of the empirical covariance matrix. Furthermore, use of the unstructured working correlation can potentially improve estimation precision and therefore should be considered when data arise from a balanced longitudinal study. However, most previous studies have not allowed the unstructured working correlation to be selected as it estimates more nuisance correlation parameters than other structures such as AR-1 or exchangeable. Therefore, we propose appropriate penalties for the selection criteria that can be imposed upon the unstructured working correlation. Via simulation in multiple scenarios and in application to a longitudinal study, we show that the trace of the empirical covariance matrix works very well relative to existing criteria. We further show that allowing criteria to select the unstructured working correlation when utilizing the penalties can substantially improve parameter estimation. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: correlation structure; efficiency; empirical covariance matrix; generalized estimating equations; unstructured

1. Introduction

Longitudinal studies are very common and result in correlated data. Specifically, subjects are followed over time, and outcomes from the same subject are assumed to be correlated. For instance, we later utilize data from the Prevention of Alzheimer's Disease by Vitamin E and Selenium (PREADViSE) clinical trial [1]. From this study, we have data on 50 men who were yearly assessed. Our outcome of interest is cognitive status as measured using the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) T-score [2, 3].

Generalized estimating equations (GEE) [4] are commonly used for the marginal analysis of correlated data, as they produce consistent parameter estimates when the mean structure for the data is correctly specified, while choosing a working covariance structure is the only additional requirement. This structure does not need to be the true structure, although accurate correlation selection is ideal in order to estimate parameters efficiently [4, 5]. Therefore, Ziegler and Vens [6] have recently presented an informative discussion on this topic, and multiple criteria have been proposed to select the working correlation structure for the data. These include the 'quasi-likelihood under the independence model criterion' (QIC) [7], resampling-based approaches [8], the 'correlation information criterion' (CIC) [9], and multiple criteria based upon the work of Rotnitzky and Jewell [10–13]. Recently, Carey and Wang [13], Gosho *et al.* [14], and Zhou *et al.* [15] each proposed criteria that performed very well in their respective

List of ICs, look into this

Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, U.S.A.
Correspondence to: Philip M. Westgate, Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, U.S.A.
†*E-mail: philip.westgate@uky.edu*

Paper discussing the need for penalizing when including unstructured.

simulation studies, although these criteria have never been compared or extensively studied. Furthermore, most prior studies did not allow the unstructured working correlation to be selected, as it involves the estimation of multiple nuisance correlation parameters for which no penalty for any criterion had been derived [9]. Additionally, Barnett *et al.* [16] demonstrated the need for this type of penalty, as they showed that failing to penalize complex covariance structures with the QIC often led to incorrectly selecting more complex models.

Proposed IC: TECM

Although some criteria can be obtained from statistical packages, they typically are not part of standard output. Therefore, we propose a criterion that selects the working correlation structure that yields the smallest trace of the empirical covariance matrix (TECM). The TECM can be easily obtained from statistical packages via estimated standard errors (SEs) and is based upon the idea that the best working correlation structure under consideration is the one that produces the least variable parameter estimates.

Another paper on unstructured, finite samples

Utilizing the unstructured working correlation matrix can be advantageous when the temporal spacing between repeated measurements is the same for each subject. Assuming a correctly modeled covariance structure [17], use of the unstructured working correlation theoretically is equally or more efficient than the use of any other structure and should therefore be considered for selection. However, Westgate [18] formally showed that the estimation of multiple correlation parameters in the unstructured matrix potentially increases the variances of the regression parameter estimates in finite samples. Therefore, simpler correlation structures may perform better in some settings, especially when one of these is the true structure. To keep selection criteria from over selecting the unstructured working correlation in these settings, while still allowing this structure to be chosen in settings in which it does improve parameter estimation, we propose criterion-specific penalties based on the work of Westgate [18], which take into account the estimated increase in variances.

Big idea of penalities

Section 2 introduces notation and discusses GEE, common estimators of the covariance matrix of the regression parameter estimates, and a covariance correction for estimating numerous nuisance parameters with the unstructured working correlation matrix. In Section 3, we briefly discuss different correlation structure criteria that have been proposed, along with how they performed in past simulation studies. We also propose the TECM criterion and penalties that can be imposed upon the unstructured working correlation matrix for criteria that utilize an estimate for the covariance matrix of parameter estimates. In Section 4, we carry out a multi-scenario simulation study that compares the TECM and other selection criteria, in addition to demonstrating the utility of the proposed penalties. We also make comparisons in application to the PREADViSE study in Section 5. We give concluding remarks in Section 6.

2. Generalized estimating equations

We assume that N independent subjects each contribute up to n observations having the same temporal spacing. Therefore, subject i, $i = 1, \ldots, N$, contributes a vector of correlated outcomes, $\boldsymbol{Y}_i = \left[Y_{i1}, \ldots, Y_{in_i}\right]^T$, $n_i \leqslant n$, with marginal means denoted by $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i)$ and the diagonal matrix of marginal variances given by $\phi \boldsymbol{A}_i$, in which the jth diagonal element is given by $\phi\nu(\mu_{ij})$, $j = 1, \ldots, n_i$, ν is a known function, and ϕ is an assumed common dispersion parameter. The marginal mean for the jth observation, $j = 1, \ldots, n_i$, from the ith subject, $i = 1, \ldots, N$, uses a known link function, f, a vector of covariate values, $\mathbf{x}_{ij} = \left[1, x_{1ij}, \ldots, x_{(p-1)ij}\right]^T$, and a $p \times 1$ vector of regression parameters, $\boldsymbol{\beta} = \left[\beta_0, \beta_1, \ldots, \beta_{p-1}\right]^T$, giving $f(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta}$. The true correlation structure is unknown, and a working correlation matrix is chosen, denoted by $\boldsymbol{R}_i(\boldsymbol{\alpha})$, in which $\boldsymbol{\alpha}$ is the nuisance correlation parameter(s) to be estimated. The working covariance matrix for subject i is then given by $\boldsymbol{V}_i = \phi \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}$. We note that if the working structure is misspecified, estimates for $\boldsymbol{\alpha}$, and therefore $\boldsymbol{R}_i(\boldsymbol{\alpha})$, may not be consistent [19, 20]. Examples of estimates for ϕ and the exchangeable and AR-1 correlation parameters, based on estimates proposed by Liang and Zeger [4] or implemented in SAS GENMOD [21], which we later use in our simulation study of Section 4 are

$$\hat{\phi} = \frac{1}{\left(\sum_{i=1}^{N} n_i\right) - p}\sum_{i=1}^{N}\sum_{j=1}^{n_i} e_{ij}^2, \quad \hat{\alpha}_{exch} = \left[\left(0.5\sum_{i=1}^{N} n_i(n_i-1) - p\right)\hat{\phi}\right]^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i} e_{ij}e_{ik}$$

and

$$\hat{\alpha}_{AR-1} = \left[\left(\sum_{i=1}^{N}(n_i - 1) - p\right)\hat{\phi}\right]^{-1} \sum_{i=1}^{N}\sum_{j=1}^{n_i-1} e_{ij}e_{i(j+1)} \quad (1)$$

respectively, where $e_{ij} = (Y_{ij} - \mu_{ij})/\sqrt{v(\mu_{ij})}$.

To obtain parameter estimates, $\hat{\boldsymbol{\beta}}$, using GEE, we solve $\sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}$, in which $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^T$ [4]. Assuming that the working covariance structure, and thus correlation structure, is correctly specified, the covariance of $\hat{\boldsymbol{\beta}}$ can be consistently estimated using the model-based matrix, $\widehat{\boldsymbol{\Sigma}}_{MB} = \left(\sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right)^{-1}$ [22]. This estimator will be biased if the covariance structure is misspecified, and therefore, the empirical covariance matrix, $\widehat{\boldsymbol{\Sigma}}_E = \widehat{\boldsymbol{\Sigma}}_{MB}\left[\sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right]\widehat{\boldsymbol{\Sigma}}_{MB}$, is routinely utilized with the GEE approach as it is a consistent estimate for the true covariance of $\hat{\boldsymbol{\beta}}$ even when the working structure is misspecified [4]. Therefore, assuming the true and working covariance structures are equivalent, these two matrices are asymptotically equivalent. However, as parameters must be estimated, there may be finite-sample differences between these matrices even if the true covariance structure is utilized. Furthermore, when the number of subjects is less than 40 or 50 [22, 23], there may be notable negative biases in variances obtained from $\widehat{\boldsymbol{\Sigma}}_E$ that must be corrected. For instance, see Mancl and DeRouen [22] and Kauermann and Carroll [24], who proposed replacing $(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$ with $(\boldsymbol{I}_n - \boldsymbol{H}_i)^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T\left(\boldsymbol{I}_n - \boldsymbol{H}_i^T\right)^{-1}$ and $(\boldsymbol{I}_n - \boldsymbol{H}_i)^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$, respectively, for $i = 1, \ldots, N$, where $\boldsymbol{H}_i = \boldsymbol{D}_i\widehat{\boldsymbol{\Sigma}}_{MB}\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}$.

Σ(MB): Model-based
Σ(E): Robust or empirical

Liang and Zeger [4] proposed estimating the unstructured correlation matrix with

$$\boldsymbol{R}_{U1} = \frac{1}{\hat{\phi}N}\sum_{i=1}^{N} \boldsymbol{A}_i^{-1/2}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{A}_i^{-1/2} \quad (2)$$

An alternative matrix could replace these diagonal elements with 1, and Equation (2) could be divided by $N - p$ rather than N. These modifications yield the unstructured correlation matrix, $\boldsymbol{R}_{U2}$, utilized by SAS GENMOD [21], in which the estimate for $Corr(Y_{ij}, Y_{ik})$ is given by

$$\hat{\alpha}_{jk} = \left[(N - p)\hat{\phi}\right]^{-1}\sum_{i=1}^{N} e_{ij}e_{ik}$$

Bias of traditional covariance estimators

Although $\widehat{\boldsymbol{\Sigma}}_{MB}$ and $\widehat{\boldsymbol{\Sigma}}_E$ are appropriate to use with working correlation structures such as independence, exchangeable, and AR-1, these covariance matrices can be biased when utilizing an unstructured working correlation matrix because they do not take into account the additional finite-sample variance in $\hat{\boldsymbol{\beta}}$ that arises from the estimation of the $n(n-1)/2$ nuisance correlation parameters [18]. Specifically, Westgate [18] explained that this variance increase occurs because of the use of working parameter estimates to estimate $\boldsymbol{\mu}_i$, $i = 1, \ldots, N$, in Equation (2) or $\hat{\alpha}_{jk}$, $1 \leqslant j \neq k \leqslant n$, therefore increasing the estimation variability of GEE. Westgate [18] derived an estimate for this variance inflation that can be applied to either $\widehat{\boldsymbol{\Sigma}}_{MB}$ or $\widehat{\boldsymbol{\Sigma}}_E$, yielding $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_{MB}\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ and $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_E\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ as bias-corrected model-based and empirical covariance matrices, respectively. Here, $\boldsymbol{I}_p$ is a $p \times p$ identity matrix,

How to adjust for bias when using unstructured

$$\boldsymbol{G} = \frac{\partial}{\partial\boldsymbol{\beta}^{*T}}\left[\left(\phi\sum_{i=1}^{N}\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right)^{-1}\sum_{i=1}^{N}\boldsymbol{D}_i^T\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_U^{-1}(\boldsymbol{\beta}^*)\boldsymbol{A}_i^{-1/2}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)\right]\Big|_{\boldsymbol{\beta}^*=\boldsymbol{\beta}} \quad (3)$$

and $\widehat{\boldsymbol{G}}$ replaces unknown parameters with their estimated values after the derivative is taken in Equation (3). Within Equation (3), $\boldsymbol{R}_U$ represents the unstructured working correlation matrix that is utilized, either $\boldsymbol{R}_{U1}$ or $\boldsymbol{R}_{U2}$. For more detail on the derivative that is taken in Equation (3), see [18]. We note that for fixed n, the impact and necessity of this covariance correction diminishes as N increases, as the correlation parameters will be estimated more precisely.

3. Selection criteria and penalties for the unstructured working correlation

What we saw!!!

Multiple correlation selection criteria utilize estimated covariance matrices and can therefore over select the unstructured matrix when not appropriately taking into account the increase in finite-sample variance in $\widehat{\boldsymbol{\beta}}$ that results from the estimation of multiple correlation parameters. We therefore propose penalizing the unstructured working correlation using $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ in place of $\widehat{\boldsymbol{\Sigma}}$. Here, $\widehat{\boldsymbol{\Sigma}}$ denotes an appropriate covariance matrix estimate that is criterion dependent, and $\widehat{\boldsymbol{G}}$ is the estimate of $\boldsymbol{G}$, which is defined in Equation (3) and depends on how the unstructured correlation matrix is estimated. We now discuss popular and relatively new selection criteria and how to incorporate this penalty with these criteria.

PENALTY!!!

3.1. Quasi-likelihood under the independence model criterion

Pan [7] proposed the QIC, which is also meant to aid in the model selection process. The QIC for a given working correlation structure, R, is given by

$$QIC(R) = -2Q(\boldsymbol{\beta}, \phi) + 2tr\left(\widehat{\boldsymbol{\Sigma}}_I^{-1}\widehat{\boldsymbol{\Sigma}}_E\right)$$

in which $Q(\boldsymbol{\beta}, \phi)$ is the quasi-likelihood under the assumption that all observations are independent and $\widehat{\boldsymbol{\Sigma}}_I = \phi\left(\sum_{i=1}^N \boldsymbol{D}_i^T \boldsymbol{A}_i^{-1}\boldsymbol{D}_i\right)^{-1}$. Here, we can replace unknown parameters with corresponding estimates obtained from GEE implementing the working correlation structure, which is the approach we take, although estimates from the use of the independence structure could alternatively be used [9, 25]. In the selection process, we choose the structure that gives the smallest value for the QIC. To penalize the unstructured working correlation, we propose replacing $\widehat{\boldsymbol{\Sigma}}_E$ with $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_E\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ inside $QIC\ (\boldsymbol{R}_U)$. This appropriately inflates the value for the QIC, as the corrected matrix takes into account the estimated increase in variances in the final parameter estimates resulting from the need to estimate numerous correlation parameters.

Penalty applied to robust only.

3.2. Rotnitzky–Jewell criterion

Rotnitzky and Jewell [10] suggested the use of $C_1 = tr(\boldsymbol{Q})/p$ and $C_2 = tr\left(\boldsymbol{Q}^2\right)/p$ to select the working correlation structure, where $\boldsymbol{Q} = \widehat{\boldsymbol{\Sigma}}_{MB}^{-1}\widehat{\boldsymbol{\Sigma}}_E$ or $\boldsymbol{Q} = \widehat{\boldsymbol{\Sigma}}_E\widehat{\boldsymbol{\Sigma}}_{MB}^{-1}$. As $\widehat{\boldsymbol{\Sigma}}_E$ and $\widehat{\boldsymbol{\Sigma}}_{MB}$ are asymptotically equivalent when the working and true covariance structures are equivalent, we desire values for C_1 and C_2 that are close to 1. Hin *et al.* [11] proposed a Rotnitzky–Jewell (RJ) criterion for a given working correlation structure, R, that is given by

$$RJ(R) = \sqrt{(1 - C_1)^2 + (1 - C_2)^2}$$

Shults *et al.* [12] and Carey and Wang [13] have alternatively proposed similar criteria. These include $|C_1 - 1|$, $|C_2 - 1|$, $DBAR = C_2 - 2C_1 + 1$, and $\Delta = \sum_{j=1}^p \left[log(c_j)\right]^2$, where $c_j,\ j = 1, \ldots, p$, are the eigenvalues of $\boldsymbol{Q}$. For any one of these given criteria, they selected the structure that gives the smallest value.

Multiple papers have carried out simulation studies involving at least one of these criteria and possibly the QIC. Neither Hin *et al.* [11] nor Gosho *et al.* [14] found that the RJ criterion or QIC is clearly better than the other. Shults *et al.* [12] compared $|C_1 - 1|$, $|C_2 - 1|$, and $DBAR$, finding that $|C_1 - 1|$ and $|C_2 - 1|$ had no notable differences and worked considerably better than $DBAR$.

To penalize the unstructured working correlation, we propose replacing $\widehat{\boldsymbol{\Sigma}}_E$ or $\widehat{\boldsymbol{\Sigma}}_{MB}$ with $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_E\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ or $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_{MB}\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$, respectively, inside $\boldsymbol{Q}$. Specifically, the replacement that is made is the one that results in a larger criterion value. The idea is that we are inflating the value of the criterion by an estimated amount that takes into account the difference between a theoretical covariance matrix that assumes correlation parameters are known and the corresponding matrix that appropriately accounts for the finite-sample estimation of correlation parameters.

3.3. Correlation information criterion

To improve upon the performance of the QIC, Hin and Wang [9] proposed the CIC, which selects the structure that minimizes the second part of the QIC that depends on the true and working correlation structures via the empirical covariance matrix. Specifically, the CIC for a given working correlation structure, R, is

$$CIC(R) = tr\left(\widehat{\boldsymbol{\Sigma}}_I^{-1}\widehat{\boldsymbol{\Sigma}}_E\right)$$

The reasoning that the CIC will improve the accuracy of correlation structure selection is that the first term of the QIC does not contain information about the correlation structure, and is therefore only useful for modeling the mean structure [9]. Via simulation, Hin and Wang [9], among others [14, 15], have found that the CIC notably improves correlation structure selection relative to the QIC. Furthermore, the CIC worked better than the RJ criterion in the study presented by Gosho *et al.* [14]. As with the QIC, we propose penalizing the unstructured working correlation by replacing $\widehat{\boldsymbol{\Sigma}}_E$ with $\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_E\left(\boldsymbol{I}_p + \widehat{\boldsymbol{G}}\right)^T$ inside $CIC\left(\boldsymbol{R}_U\right)$.

Penalty for CIC

3.4. Shults and Chaganty and Gaussian pseudolikelihood criteria

As discussed by Shults *et al.* [12], Shults and Chaganty [26] proposed the Shults–Chaganty (SC) criterion, which is given by

$$SC(R) = \sum_{i=1}^{N}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)^T \boldsymbol{V}_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)$$

for a given working correlation structure, R. The idea is to select the working correlation structure that yields the smallest weighted, via the working covariance structure, error sum of squares. Carey and Wang [13] suggested a similar criterion, the Gaussian pseudolikelihood (GP), and it is given by

ICs based on residuals that failed

$$GP(R) = -0.5\sum_{i=1}^{N}\left[\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)^T \boldsymbol{V}_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right) + log\left(|\boldsymbol{V}_i|\right)\right]$$

We select the structure yielding the largest GP value. In the simulation study of Carey and Wang [13], the GP criterion worked better than $DBAR$ and Δ in terms of selecting the correct covariance structure. Furthermore, Shults *et al.* [12] found that the SC criterion did not work as well as $|C_1 - 1|$, $|C_2 - 1|$, or $DBAR$, and the results of Gosho *et al.* [14] show that the SC criterion did not work as well as the QIC, CIC, and RJ criteria. As the SC and GP criteria do not utilize an estimated covariance matrix of the final parameter estimates, we cannot utilize our proposed method for penalizing the unstructured working correlation with these two criteria.

3.5. C statistic criterion

Gosho *et al.* [14] proposed selecting the working correlation structure, R, that gives the smallest value for

$$C(R) = tr\left(\left[\left(\sum_{i=1}^{N}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)^T\right)\left(\sum_{i=1}^{N}\boldsymbol{V}_i\right)^{-1} - \boldsymbol{I}\right]^2\right)$$

in which $\boldsymbol{I}$ is a $n \times n$ identity matrix. This C statistic essentially selects the structure that yields the smallest estimated difference between the sums of the empirical and model-based covariances. This criterion worked very well in their simulation study, in which it outperformed the QIC, RJ, CIC, and SC criteria in terms of selecting the true structure. However, in terms of relative efficiencies, the CIC appeared to do as well as the C statistic criterion, and all other criteria, except for the SC, worked almost as well. As with the SC and GP criteria, the C statistic is not composed of an estimated covariance matrix of the final parameter estimates and therefore cannot utilize the proposed method for penalizing the unstructured working correlation.

3.6. Information ratio p-value criterion

Zhou *et al.* [15] proposed a criterion based upon a p-value from a test that utilizes C_1, which they refer to as the information ratio (IR) statistic, as the test statistic. Because the null hypothesis of the test is that the true and working covariance structures are the same, this criterion selects the correlation structure that yields the largest p-value. Details on how to obtain p-values can be found in their paper. In their simulation study, they showed that their IR p-value criterion worked better than the CIC and QIC in terms of choosing the correct structure. We propose the same penalty for the unstructured working correlation that was imposed upon the RJ-based criteria, which will decrease the IR p-value.

3.7. Trace of the empirical covariance matrix criterion

TECM can be calculated easily!

Lastly, we propose selecting the working correlation structure that gives the smallest TECM, $tr\left(\widehat{\boldsymbol{\Sigma}}_E\right)$. The goal is to choose the structure that produces the least variable parameter estimates, and therefore, it is reasonable to select the structure that gives the smallest sum of the estimates of the variances of the parameter estimates. This criterion is advantageous over the other criteria in that it is very simple to obtain and calculate from statistical packages. Specifically, we can obtain the TECM simply by taking the sum of the squared estimated SEs. We further propose the obvious penalty for the unstructured working correlation, yielding $tr\left[\left(\boldsymbol{I}_p+\widehat{\boldsymbol{G}}\right)\widehat{\boldsymbol{\Sigma}}_E\left(\boldsymbol{I}_p+\widehat{\boldsymbol{G}}\right)^T\right]$ as the TECM.

Where TECM came from

The idea of using the TECM comes from Song *et al.* [27]. They proposed a correlation structure selection criterion for the quadratic inference function (QIF) method [28] that is based upon the trace of the Godambe information (TGI) matrices of the structures under comparison, where the Godambe information matrix is the inverse of the empirical covariance matrix. This idea is essentially the same as using the TECM, and Song *et al.* [27] suggested that the TGI could be applicable with GEE.

4. Simulation study

4.1. Study description

Why MSEs are used and why frequency selection may be flawed.

To assess differences in how well the selection criteria work, in addition to demonstrating the utility of the proposed penalties that are imposed upon the unstructured working correlation, we carry out a simulation study using three distinct scenarios, each of which is based on a previous study [9, 11, 14, 15]. Three scenarios are used because no single study has directly compared all of the discussed criteria, and, as shown in the results, the performance of a given criterion can be scenario dependent. To assess differences, we utilize empirical mean squared errors (MSEs) of parameter estimates resulting from the use of a given criterion, in addition to the empirical frequencies of how many times each correlation structure was selected by the given criterion. Most studies, except for [14], have only presented empirical frequencies and have concluded that the criterion that chose the true structure most frequently is best. Although this reasoning is correct, selection frequencies are not fully informative, because we are truly interested in the variances of parameter estimates. For instance, empirical MSEs resulting from the use of any of the working correlation structures were very similar in multiple settings of our study. Therefore, all criteria performed well in these settings, although distinct differences in empirical frequencies could be notable.

Simulations allow unstructured.

Scenarios utilized settings in which the number of independent subjects was either 50 or 100, each of whom contributed four repeated measurements. Each setting was examined via 1000 simulations, which were carried out using R version 2.13.1 [29]. Similar to many other studies that have been used to compare various criteria, we consider independence, exchangeable, and AR-1 as the working structures that are under consideration to be selected and allow the true structure to be either exchangeable or AR-1 with correlation values of either 0.2 or 0.5. However, unlike other studies, we also allow the unstructured correlation matrix to be selected and to be the true structure, in which case it is given by

$$\begin{bmatrix} 1 & 0.8 & 0.3 & 0.2 \\ 0.8 & 1 & 0.6 & 0.4 \\ 0.3 & 0.6 & 1 & 0.8 \\ 0.2 & 0.4 & 0.8 & 1 \end{bmatrix}$$

for normal outcomes. For binary outcomes, there are additional constraints for the correlations [12, 30]. Therefore, to ensure correlated binary outcomes could be generated, and to present results in settings in which the unstructured working correlation did not necessarily perform best even though the true correlation matrix was indeed unstructured, we use correlation values that are only half as large as the values in the above matrix. We generated normal outcomes using `rmvnorm` of the `mvtnorm` package [31, 32], for which we specified the true covariance matrix of outcomes and used the default eigen-value matrix decomposition. We generated binary outcomes using `rmvbin` of the `bindata` package [33]. With `rmvbin`, we input the true marginal probabilities and correlation matrix, and the correlated binary outcomes are generated by thresholding a normal distribution [33, 34]. We note that, using our own function, the correlation parameter for working AR-1 was estimated using Equation (1), because we found that the use of this estimate was slightly more efficient overall than the estimate used by the `geeglm` function of `geepack` package [35–37] in terms of regression parameter estimation when the true structure was not AR-1. As is often carried out in practice, we set $\phi = 1$ when outcomes were binary. However, as is carried out in SAS GENMOD [21], we still estimate ϕ while obtaining correlation parameter estimates.

Simulation results involve the unstructured working correlation matrix given by $\boldsymbol{R}_{U2}$, as this matrix has diagonal elements that are equal to 1, an attribute that is typically utilized in practice. However, when unstructured is the true structure and outcomes are normal, we present results from the use of $\boldsymbol{R}_{U1}$ in order to avoid numerical instability that can occur with $\boldsymbol{R}_{U2}$ in these settings. Here, $\boldsymbol{R}_{U1}$ is an unstructured working covariance matrix divided by a scalar.

Scenario 1 is based on models used by Hin *et al.* [11] and Hin and Wang [9], although no intercept was included in their models. For subject i, we generated normal outcomes from $\mu_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$ with unit marginal variances, and we generated binary outcomes from $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$, where x_{1ij} and x_{2ij}, $j = 1, 2, 3, 4$ were independently drawn from $Uniform(0, 1)$. For normal outcomes, $\boldsymbol{\beta} = [0, 0.3, 0.3]^T$, whereas $\boldsymbol{\beta} = [0, 0.62, 0.62]^T$ for binary outcomes.

Scenario 2 is based on models used by Gosho *et al.* [14], although no intercept was included in their model for normal outcomes. For subject i, we generated normal outcomes from $\mu_{ij} = \beta_0 + \beta_1(j-1) + \beta_2 x_{2ij}$ with unit marginal variances, and we generated binary outcomes from $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1(j-1) + \beta_2 x_{2ij}$, where x_{2ij}, $j = 1, 2, 3, 4$ were independently drawn from $Bernoulli(0.5)$. For normal outcomes, $\boldsymbol{\beta} = [0, 1, 1]^T$, whereas $\boldsymbol{\beta} = [0.5, -0.2, -0.2]^T$ for binary outcomes.

Scenario 3 is based on models used by Zhou *et al.* [15]. For subject i, we generated normal outcomes from $\mu_{ij} = \beta_0 + \beta_1 x_{1ij}$ with unit marginal variances, and we generated binary outcomes from $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij}$, where x_{1ij}, $j = 1, 2, 3, 4$ were independently drawn from $Uniform(j, j+1)$. For normal outcomes, $\boldsymbol{\beta} = [3, 5]^T$, whereas $\boldsymbol{\beta} = [-1, 1/6]^T$ for binary outcomes.

Tables I–III present results for settings from the three scenarios when the true correlation matrix is unstructured, and Tables IV and V present results from scenario 1 settings in which the true structure is either exchangeable or AR-1. For space considerations, these five tables only contain results for $N = 100$, and results from scenarios 2 and 3 for settings in which the true correlation structure was either exchangeable or AR-1 are presented in eight tables in the Supporting information. However, results from these settings are discussed in the succeeding text.

We refer to the working correlation structure that yields the smallest empirical MSE quantity for a given setting as the best working correlation structure for that setting, or as the structure that performed best. We therefore view a correct choice as the selection of the best structure. We note that the best structure, and therefore the correct choice, is always the true structure except in some settings in which the true correlation matrix is unstructured and outcomes are binary. MSE quantities that are used take the sum of the empirical MSEs from all non-intercept parameters. Specifically, the tables report MSE ratios in which the numerator comprised the smallest observed empirical MSE value that results from the use of the best working correlation structure under consideration for the given setting and the denominators take the empirical MSEs resulting from each working correlation structure and each selection criterion. Furthermore, the empirical frequencies of selecting each structure are given for each criterion. As most previous studies have only allowed independence, exchangeable, and AR-1 to be selected, we present two sets of empirical MSE and frequency results for each criterion. The first allows only one of these three structures to be chosen, whereas the second also considers the unstructured working correlation for selection.

We note that results from the use of the QIC, $|C_1 - 1|$, $|C_2 - 1|$, $DBAR$, and the SC criteria are not presented for space and performance considerations. As with other studies, we found that the QIC did not perform as well as other criteria such as the CIC. Furthermore, $DBAR$ performed notably inferior to the other RJ-based criteria, and $|C_1 - 1|$ and $|C_2 - 1|$ performed very similarly to the IR p-value

Statistics in Medicine

Table I. Results from scenario 1 for settings in which the true correlation was unstructured and the number of subjects was 100.

		MSE ratios		Selection frequencies						
				Ind		Exch		AR-1		Un
Y	Criterion	1	(2)	1	(2)	1	(2)	1	(2)	(2)
Normal				MR=0.29		MR=0.46		MR=0.88		MR=1.00
	TECM	0.88	(0.98)	0	(0)	0	(0)	1000	(72)	(928)
	CIC	0.88	(0.96)	0	(0)	0	(0)	1000	(310)	(690)
	C statistic	0.88	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	GP	0.88	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	IR p-value	0.49	(0.52)	4	(1)	872	(780)	124	(109)	(110)
	RJ	0.51	(0.54)	0	(0)	795	(719)	205	(186)	(95)
	Δ	0.53	(0.58)	0	(0)	737	(569)	263	(160)	(271)
Binary				MR=0.83		MR=0.94		MR=1.00		MR=0.99
	TECM	1.00	(1.00)	1	(1)	173	(129)	826	(605)	(265)
	CIC	1.00	(1.00)	1	(1)	158	(124)	841	(690)	(185)
	C statistic	0.99	(0.99)	0	(0)	87	(0)	913	(4)	(996)
	GP	0.99	(0.99)	0	(0)	99	(0)	901	(1)	(999)
	IR p-value	0.97	(0.97)	24	(24)	554	(545)	423	(414)	(20)
	RJ	0.97	(0.97)	4	(4)	573	(565)	423	(413)	(18)
	Δ	0.96	(0.96)	0	(0)	649	(622)	351	(309)	(69)

Presented are empirical frequencies of selecting each working correlation structure out of 1000 replications, with corresponding ratios that compare empirical MSEs from the sole use of the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the MSEs from the use of the given criterion. MSE ratios (MR) that compare the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the other structures are also presented. Results for the situations in which the unstructured working correlation was not allowed to be selected are indicated by 1, whereas results from when this structure was a selection option are indicated by Equation (2) and are in parentheses. Unstructured is best for normal outcomes, whereas AR-1 is best for binary outcomes.
MSE, mean squared error; TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewell; Y, denotes whether outcomes were normal or binary; Ind, independence; Exch, exchangeable; Un, unstructured.

and RJ criteria, respectively, in terms of both empirical MSEs and selection frequencies. Finally, the SC criterion did not select the best structure as often as the GP criterion and was notably inferior in many settings.

4.2. *Description of results*

Results show that the C statistic and GP criteria work very similarly in terms of MSEs and structure selection. The TECM and CIC criteria also worked similarly because of their related usage of the empirical covariance matrix. Although simulation results do not necessarily favor one criterion over the other in general, the TECM more frequently selected the unstructured working correlation in settings in which it was the best structure and was allowed to be selected, whereas the CIC typically chose the best structure more often than the TECM in other settings. In these latter settings, empirical MSEs resulting from the use of these two criteria were usually similar because in many settings, the correct selection frequencies were not notably different or multiple structures under consideration produced similar empirical MSEs. However, the TECM resulted in smaller empirical MSEs in the settings in which unstructured was best. Finally, the IR p-value, RJ, and Δ criteria worked similarly in many settings, at least in terms of empirical MSEs, although Δ performed notably better in multiple settings and was more likely to select the unstructured working correlation.

When the unstructured working correlation was not allowed to be selected, the C statistic and GP criteria tended to select the best structure more frequently than the other criteria, resulting in smaller empirical MSEs than with other criteria in multiple settings. However, they appeared to have only a negligible advantage over the TECM and CIC. For instance, excluding settings in which the true correlation was unstructured, MSE ratios ranged from 0.99 to 1.00 for the C statistic and GP criteria, whereas MSE

Table II. Results from scenario 2 for settings in which the true correlation was unstructured and the number of subjects was 100.

		MSE ratios		Selection frequencies						
				Ind		Exch		AR-1		Un
Y	Criterion	1	(2)	1	(2)	1	(2)	1	(2)	(2)
Normal				MR=0.34		MR=0.52		MR=0.81		MR=1.00
	TECM	0.81	(1.00)	0	(0)	0	(0)	1000	(23)	(977)
	CIC	0.81	(0.98)	0	(0)	0	(0)	1000	(72)	(928)
	C statistic	0.81	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	GP	0.81	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	IR p-value	0.81	(0.83)	0	(0)	1	(0)	999	(830)	(171)
	RJ	0.81	(0.86)	0	(0)	0	(0)	1000	(685)	(315)
	Δ	0.81	(0.97)	0	(0)	0	(0)	1000	(96)	(904)
Binary				MR=0.85		MR=0.97		MR=1.00		MR=0.98
	TECM	1.01	(1.00)	0	(0)	172	(134)	828	(576)	(290)
	CIC	1.01	(1.00)	0	(0)	125	(93)	875	(618)	(289)
	C statistic	1.00	(0.98)	0	(0)	89	(0)	911	(0)	(1000)
	GP	0.99	(0.98)	0	(0)	94	(0)	906	(0)	(1000)
	IR p-value	0.99	(0.99)	0	(0)	113	(109)	887	(821)	(72)
	RJ	0.99	(0.99)	0	(0)	60	(57)	940	(864)	(79)
	Δ	1.00	(0.99)	0	(0)	127	(43)	873	(511)	(446)

Presented are empirical frequencies of selecting each working correlation structure out of 1000 replications, with corresponding ratios that compare empirical MSEs from the sole use of the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the MSEs from the use of the given criterion. MSE ratios (MR) that compare the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the other structures are also presented. Results for the situations in which the unstructured working correlation was not allowed to be selected are indicated by 1, whereas results from when this structure was a selection option are indicated by Equation (2) and are in parentheses. Unstructured is best for normal outcomes, whereas AR-1 is best for binary outcomes.

MSE, mean squared error; TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewell; Y, denotes whether outcomes were normal or binary; Ind, independence; Exch, exchangeable; Un, unstructured.

ratios ranged from 0.97 to 1.00 for the TECM and CIC across all except for two settings, in which the ratios were 0.96 or 1.01 for the CIC. These ratios imply that all four of these criteria worked very well. However, the IR p-value, RJ, and Δ criteria were less reliable and even worked very poorly in some settings, although they did work well in many other settings. For instance, these three criteria had the tendency to over select exchangeable in settings of scenario 1 in which the true correlation structure was AR-1, which was also seen for the RJ criterion in the simulation study of Hin *et al.* [11].

When the unstructured working correlation was allowed to be selected, the TECM and CIC worked best overall. They notably outperformed the C statistic and GP in multiple settings in which the unstructured working correlation was not the best structure, as the C statistic and GP criteria often selected the unstructured working correlation because of not imposing a penalty. The IR p-value and RJ criteria tended to have lower frequencies of selecting the unstructured working correlation relative to the CIC and TECM, whereas these selection frequencies were not consistent for Δ relative to the CIC and TECM. Furthermore, the IR p-value, RJ, and Δ worked well in many settings but were also detrimental in some situations.

In settings in which the exchangeable or AR-1 structures produced the smallest empirical MSEs, allowing the unstructured working correlation to be selected typically only reduced the MSE ratios by a negligible or small amount for criteria that imposed a penalty. However, allowing the unstructured working correlation to be selected in settings in which it worked best notably improved empirical MSEs, especially when using the TECM or CIC, and was therefore advantageous. Furthermore, imposing penalties is necessary. For instance, because the C statistic and GP criteria often chose the unstructured working correlation, they were detrimental in multiple settings. Similarly, we found (results not shown) that although the other criteria did not always select the unstructured working correlation when

Table III. Results from scenario 3 for settings in which the true correlation was unstructured and the number of subjects was 100.

		MSE ratios		Selection frequencies						
				Ind		Exch		AR-1		Un
Y	Criterion	1	(2)	1	(2)	1	(2)	1	(2)	(2)
Normal				MR=0.60		MR=0.61		MR=0.75		MR=1.00
	TECM	0.75	(0.99)	0	(0)	0	(0)	1000	(17)	(983)
	CIC	0.75	(0.90)	0	(0)	0	(0)	1000	(141)	(859)
	C statistic	0.75	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	GP	0.75	(1.00)	0	(0)	0	(0)	1000	(0)	(1000)
	IR p-value	0.75	(0.81)	0	(0)	0	(0)	1000	(482)	(518)
	RJ	0.75	(0.85)	0	(0)	0	(0)	1000	(359)	(641)
	Δ	0.75	(0.94)	0	(0)	0	(0)	1000	(73)	(927)
Binary				MR=0.95		MR=0.96		MR=1.00		MR=1.00
	TECM	1.00	(1.01)	112	(85)	174	(125)	714	(437)	(353)
	CIC	1.00	(1.00)	66	(50)	109	(69)	825	(560)	(321)
	C statistic	1.00	(1.00)	0	(0)	93	(0)	907	(0)	(1000)
	GP	1.00	(1.00)	0	(0)	99	(0)	901	(0)	(1000)
	IR p-value	1.00	(1.00)	0	(0)	32	(23)	968	(789)	(190)
	RJ	1.00	(1.00)	0	(0)	17	(17)	983	(808)	(175)
	Δ	0.99	(1.01)	0	(0)	152	(24)	848	(296)	(680)

Presented are empirical frequencies of selecting each working correlation structure out of 1000 replications, with corresponding ratios that compare empirical MSEs from the sole use of the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the MSEs from the use of the given criterion. MSE ratios (MR) that compare the best structure (the best structure for a given setting is the structure that produced the smallest empirical MSE value) with the other structures are also presented. Results for the situations in which the unstructured working correlation was not allowed to be selected are indicated by 1, whereas results from when this structure was a selection option are indicated by Equation (2) and are in parentheses. Unstructured is best for normal outcomes, whereas either Unstructured or AR-1 is best for binary outcomes.

MSE, mean squared error; TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewell; Y, denotes whether outcomes were normal or binary; Ind, independence; Exch, exchangeable; Un, unstructured.

not imposing a penalty, they still selected it many more times than they would have if imposing a penalty, therefore increasing empirical MSEs in settings in which this structure did not perform best.

In general, how well the criteria worked depended upon the number of subjects, the correlation value, and the differences in the performances of each structure that could be selected. Specifically, as the number of subjects increased from 50 to 100, estimates that were involved within the criteria were more precise, and therefore the proportion of correct selections typically increased. Furthermore, the unstructured working correlation was more precisely estimated for larger N, and therefore selecting this structure when either AR-1 or exchangeable was the true structure was not as detrimental in settings consisting of 100, relative to 50, subjects. Also, larger correlation values typically caused greater differences between the empirical MSEs resulting from the independence, exchangeable, and AR-1 structures, making it more likely for any given criterion to select the best structure. Furthermore, there were many settings, especially in the results for scenario 3, in which only negligible differences existed between the empirical MSEs resulting from the different structures. Therefore, only negligible differences in empirical MSEs for the different criteria were evident, although notable differences could exist in terms of the empirical selection frequencies.

5. Application

To demonstrate differences between the criteria that were compared in the simulation study, in addition to showing the impacts of the proposed penalties, we now give focus to data from the PREADViSE study [1]. Specifically, we have data from 50 men who came in for annual assessments for three or four consecutive years. The reason some men were not assessed at the fourth occasion was simply because rolling enrollment was utilized and the study period ended. The outcome we give focus to is the CERAD

Table IV. Results from scenario 1 for settings in which the true correlation was exchangeable and the number of subjects was 100.

			MSE ratios		Selection frequencies						
					Ind		Exch		AR-1		Un
Y	ρ	Criterion	1	(2)	1	(2)	1	(2)	1	(2)	(2)
Normal	0.2				MR=0.93		MR=1.00		MR=0.96		MR=0.96
		TECM	0.99	(0.98)	20	(17)	803	(685)	177	(162)	(136)
		CIC	0.99	(0.98)	17	(15)	800	(680)	183	(169)	(136)
		C statistic	1.00	(0.96)	0	(0)	917	(25)	83	(3)	(972)
		GP	1.00	(0.96)	0	(0)	924	(14)	76	(1)	(985)
		IR p-value	0.99	(0.99)	43	(43)	636	(632)	321	(318)	(7)
		RJ	0.99	(0.99)	17	(17)	691	(681)	292	(290)	(12)
		Δ	1.00	(0.99)	3	(2)	835	(743)	162	(139)	(116)
Normal	0.5				MR=0.63		MR=1.00		MR=0.87		MR=0.94
		TECM	0.99	(0.99)	0	(0)	961	(839)	39	(34)	(127)
		CIC	0.99	(0.99)	0	(0)	959	(852)	41	(38)	(110)
		C statistic	1.00	(0.97)	0	(0)	995	(380)	5	(0)	(620)
		GP	1.00	(0.96)	0	(0)	997	(306)	3	(0)	(694)
		IR p-value	0.98	(0.98)	0	(0)	801	(774)	199	(194)	(32)
		RJ	0.99	(0.99)	0	(0)	868	(832)	132	(127)	(41)
		Δ	0.98	(0.98)	0	(0)	851	(755)	149	(128)	(117)
Binary	0.2				MR=0.91		MR=1.00		MR=0.97		MR=0.98
		TECM	1.00	(0.99)	10	(9)	802	(662)	188	(163)	(166)
		CIC	0.99	(0.99)	13	(13)	812	(697)	175	(158)	(132)
		C statistic	1.00	(0.98)	0	(0)	918	(0)	82	(0)	(1000)
		GP	1.00	(0.98)	0	(0)	918	(1)	82	(0)	(999)
		IR p-value	0.99	(0.99)	29	(29)	646	(642)	328	(326)	(8)
		RJ	0.99	(0.99)	11	(11)	722	(713)	267	(266)	(10)
		Δ	1.00	(1.00)	1	(1)	883	(817)	116	(105)	(77)
Binary	0.5				MR=0.62		MR=1.00		MR=0.86		MR=0.97
		TECM	0.99	(1.00)	0	(0)	911	(658)	89	(58)	(284)
		CIC	0.99	(0.99)	0	(0)	942	(764)	58	(42)	(194)
		C statistic	1.00	(0.97)	0	(0)	989	(30)	11	(0)	(970)
		GP	1.00	(0.97)	0	(0)	985	(23)	15	(0)	(977)
		IR p-value	0.97	(0.97)	0	(0)	765	(742)	235	(226)	(33)
		RJ	0.98	(0.98)	0	(0)	826	(795)	174	(167)	(38)
		Δ	0.97	(0.98)	0	(0)	846	(760)	154	(132)	(108)

Presented are empirical frequencies of selecting each working correlation structure out of 1000 replications, with corresponding ratios that compare empirical MSEs from the sole use of the exchangeable structure with the MSEs from the use of the given criterion. MSE ratios (MR) that compare the exchangeable with the other structures are also presented. Results for the situations in which the unstructured working correlation was not allowed to be selected are indicated by 1, whereas results from when this structure was a selection option are indicated by (2) and are in parentheses.
MSE, mean squared error; TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewell; Y, denotes whether outcomes were normal or binary; ρ, exchangeable correlation value; Ind, independence; Exch, exchangeable; Un, unstructured.

T-score [2, 3], which is a summary of global cognitive status [1]. T-scores ranged in value from 24 to 77, with higher scores indicating greater cognition. We also have information on each subject's age and estimated full-scale IQ [38] when they entered the trial. Age in years and IQ ranged from 62 to 87 and 89.6 to 120.8, respectively.

We use the model given by

$$\mu_{ij} = \beta_0 + \beta_1 time_{ij} + \beta_2 (IQ_i - 110) + \beta_3 (Age_i - 70) + \beta_4 (Age_i - 70)^2 ; \quad j = 1, \ldots, n_i$$

where age and IQ are centered at rounded values that are near their respective sample means and medians, and n_i is either 3 or 4. In this model, μ_{ij} is the mean T-score for the ith subject in the jth year, and $time = j - 1$ is the number of years since baseline. We fit the model using GEE that incorporate either independence or exchangeable or AR-1 or unstructured ($\boldsymbol{R}_{U2}$) working correlation matrices. For each

Table V. Results from scenario 1 for settings in which the true correlation was AR-1 and the number of subjects was 100.

			MSE ratios		Selection frequencies						
					Ind		Exch		AR-1		Un
Y	ρ	Criterion	1	(2)	1	(2)	1	(2)	1	(2)	(2)
Normal	0.2				MR=0.95		MR=0.98		MR=1.00		MR=0.98
		TECM	0.99	(0.99)	48	(43)	169	(147)	783	(690)	(120)
		CIC	1.00	(0.99)	41	(39)	152	(136)	807	(704)	(121)
		C statistic	1.00	(0.98)	0	(0)	103	(1)	897	(16)	(983)
		GP	1.00	(0.98)	0	(0)	107	(1)	893	(10)	(989)
		IR p-value	0.97	(0.97)	238	(237)	420	(417)	343	(342)	(5)
		RJ	0.98	(0.98)	172	(171)	440	(437)	388	(386)	(6)
		Δ	0.98	(0.98)	21	(20)	580	(539)	399	(334)	(107)
Normal	0.5				MR=0.69		MR=0.89		MR=1.00		MR=0.93
		TECM	1.00	(0.99)	0	(0)	31	(28)	969	(840)	(132)
		CIC	1.00	(0.99)	0	(0)	18	(15)	982	(880)	(105)
		C statistic	1.00	(0.97)	0	(0)	4	(0)	996	(297)	(703)
		GP	1.00	(0.96)	0	(0)	4	(1)	996	(249)	(751)
		IR p-value	0.93	(0.93)	23	(15)	540	(518)	438	(414)	(54)
		RJ	0.94	(0.94)	0	(0)	549	(523)	451	(425)	(52)
		Δ	0.94	(0.93)	0	(0)	523	(475)	477	(379)	(146)
Binary	0.2				MR=0.94		MR=0.97		MR=1.00		MR=0.97
		TECM	1.00	(1.00)	19	(17)	169	(146)	812	(688)	(149)
		CIC	1.00	(0.99)	20	(19)	137	(125)	843	(729)	(127)
		C statistic	1.00	(0.96)	0	(0)	108	(0)	892	(2)	(998)
		GP	0.99	(0.96)	0	(0)	104	(0)	896	(2)	(998)
		IR p-value	0.98	(0.98)	188	(184)	415	(413)	397	(397)	(7)
		RJ	0.98	(0.98)	125	(124)	434	(433)	441	(438)	(5)
		Δ	0.97	(0.98)	13	(11)	660	(629)	327	(287)	(73)
Binary	0.5				MR=0.68		MR=0.86		MR=1.00		MR=0.98
		TECM	1.00	(1.00)	0	(0)	25	(18)	975	(744)	(238)
		CIC	1.00	(0.99)	0	(0)	13	(10)	987	(826)	(164)
		C statistic	1.00	(0.98)	0	(0)	7	(0)	993	(46)	(954)
		GP	1.00	(0.98)	0	(0)	6	(0)	994	(42)	(958)
		IR p-value	0.92	(0.93)	2	(1)	546	(528)	456	(433)	(42)
		RJ	0.92	(0.92)	0	(0)	550	(537)	450	(433)	(30)
		Δ	0.90	(0.90)	0	(0)	628	(609)	372	(318)	(73)

Presented are empirical frequencies of selecting each working correlation structure out of 1000 replications, with corresponding ratios that compare empirical MSEs from the sole use of the AR-1 structure with the MSEs from the use of the given criterion. MSE ratios (MR) that compare the AR-1 with the other structures are also presented. Results for the situations in which the unstructured working correlation was not allowed to be selected are indicated by 1, whereas results from when this structure was a selection option are indicated by Equation (2) and are in parentheses.

MSE, mean squared error; TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewell; Y, denotes whether outcomes were normal or binary; ρ, AR-1 correlation value; Ind, independence; Exch, exchangeable; Un, unstructured.

working structure, parameter estimates, their corresponding estimated SEs, and values for the selection criteria are presented in Table VI. We note that the C statistic requires an equal number of repeated measurements for all subjects, and therefore it is not applicable for use with this data. We also present two sets of results for the unstructured working correlation, one that incorporates the inflated covariance formula for obtaining SE estimates and penalized values for the criteria and another that does not correct for the estimation of the six nuisance correlation parameters.

Results show that SE estimates are too small when using $\boldsymbol{R}_{U2}$ and not incorporating the covariance inflation formula. Additionally, not using this formula for penalizing $\boldsymbol{R}_{U2}$ causes the TECM, CIC, and RJ criteria to select the unstructured working correlation, although the TECM could also select AR-1. Alternatively, when correctly utilizing the covariance inflation formula, we see that $\boldsymbol{R}_{U2}$ actually results in the largest estimated SEs and is not selected by any criteria. Specifically, the GP, IR p-value, RJ,

Table VI. Parameter estimates, standard error (SE) estimates, and values for correlation structure selection criteria from analyses of the longitudinal study that incorporates either independence (Ind) or exchangeable (Exch) or AR-1 or an unstructured (Un) working correlation with the generalized estimating equations.

	Ind $\hat{\beta}$ (SE)	Exch $\hat{\beta}$ (SE)	AR-1 $\hat{\beta}$ (SE)	Un[1] $\hat{\beta}$ (SE)	Un[2] $\hat{\beta}$ (SE)
Parameter					
β_0	48.5 (1.36)	48.8 (1.35)	48.7 (1.30)	48.0 (1.44)	48.0 (1.29)
β_1	2.10 (0.37)	1.79 (0.36)	1.76 (0.39)	2.11 (0.49)	2.11 (0.41)
β_2	0.24 (0.15)	0.24 (0.15)	0.20 (0.15)	0.17 (0.15)	0.17 (0.13)
β_3	−0.57 (0.22)	−0.55 (0.22)	−0.54 (0.22)	−0.44(0.26)	−0.44(0.22)
β_4	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
Criterion					
TECM	2.06	2.02	1.91	2.41	1.91
CIC	9.87	9.68	9.47	11.32	8.57
GP	−496	−448	−453	−462	−462
IR p-value	0.02	0.66	0.33	0.09	0.47
RJ	4.19	0.15	0.26	1.06	0.14
Δ	4.02	0.84	0.86	2.27	1.53

TECM, trace of the empirical covariance matrix; CIC, correlation information criterion; GP, Gaussian pseudolikelihood; IR, information ratio; RJ, Rotnitzky–Jewel.
[1]Standard error estimates come from the corrected empirical covariance matrix, and selection criteria, with the exception of GP, impose a penalty.
[2]Standard error estimates are not corrected, and selection criteria do not impose a penalty.

and Δ criteria all select the working exchangeable structure, and the TECM and CIC both select AR-1. Taking into consideration that our simulation results show that the TECM and CIC are favorable in general, in addition to the fact that this is a longitudinal study, AR-1 may be preferred over exchangeable. However, results from either structure are similar.

6. Concluding Remarks

Selecting an accurate working correlation structure to use with GEE is important in terms of estimation precision. To ease and improve upon the selection process, we proposed both a criterion that selects the structure that gives the smallest value for the trace of its corresponding empirical covariance matrix and penalties that can be utilized with this TECM and other criteria that allow the unstructured working correlation to be selected. The TECM can simply be obtained by summing the squared estimated SEs of resulting parameter estimates from standard statistical output and is similar in nature to the CIC. Therefore, only negligible differences in performances between these criteria were notable in our simulation study. When we did not allow the unstructured working correlation to be selected, the C statistic and GP criteria worked very similarly and even had a slight advantage over the TECM and CIC. However, allowing the unstructured working correlation to be selected by applying penalties to the TECM and CIC was shown to substantially improve estimation precision in settings in which this structure resulted in smaller empirical MSEs than the other structures under consideration, while only allowing a slight loss in estimation precision when either the exchangeable or AR-1 resulted in the smallest empirical MSEs. Additionally, multiple times the use of the C statistic or GP was detrimental when allowed to select the unstructured working correlation in settings in which it did not perform best, as no penalty was applied to these criteria. Finally, we found that RJ-based criteria worked well in many settings but could also be detrimental relative to other studied criteria.

In our simulation study, we gave focus toward balanced longitudinal data settings in which all subjects contribute the same number of repeated measurements. However, as was seen in our application example, when the number of measurements contributed by each subject varies, the unstructured working correlation may still be reasonably applied if temporal spacing is the same for all subjects. In such situations, the covariance inflation with the working unstructured correlation matrix can still be derived, and therefore the proposed penalties can also be implemented.

Derivation of a penalty that can be incorporated with the C statistic or GP criteria requires future work. As these criteria do not utilize at least one estimated covariance matrix of the parameter estimates, a straightforward penalty could not be applied. We suspect, however, that any such penalty for these criteria will only have slight utility, as our simulation study showed that the CIC and TECM worked almost as well as these criteria in a practical sense when the unstructured working correlation was not considered.

When there are multiple competing correlation structures to choose from, it is reasonable to suspect that the use of criteria such as the TECM or CIC will lead to negatively biased standard error estimates because these criteria favor smaller estimated variances. In turn, this would result in sub-nominal confidence interval coverage probabilities. We note, however, that if one structure is consistently chosen, this will not be an issue because valid inference can be obtained via the sole use of any working correlation structure we considered. Furthermore, we studied this issue when conducting our simulation study and did not find any notable evidence of undercoverage by confidence intervals or negative bias in standard error estimates (results not shown).

Future work will give focus to situations in which we are confident that working marginal variances are correctly specified and there is a common correlation structure across subjects. In this case, we can more efficiently estimate $cov(\boldsymbol{Y}_i)$ within the empirical covariance matrix with $\phi \boldsymbol{A}_i^{1/2} \boldsymbol{R}_U \boldsymbol{A}_i^{1/2}$ rather than $(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T$ [17], $i = 1, \ldots, N$, thus possibly enhancing the selection performances of criteria. For instance, Chen and Lazar [39] proposed criteria based upon empirical likelihood versions of the AIC and BIC, which performed better than the CIC and QIC in their simulation study. However, their criteria assume correctly specified working marginal variances and a common correlation structure, which was only as general as the stationary structure in their simulation study. We note that future work can also consider the selection of the working covariance structure, as in Carey and Wang [13].

The QIF method [28] is an alternative to GEE that is increasing in popularity. Two advantages it has over GEE are that it is more robust to outliers and is asymptotically more efficient when the working covariance structure is misspecified [27, 28, 40], although its finite-sample estimation performance will not always be as good as GEE's [41, 42]. Song *et al.* [27] warned that their TGI criterion may be difficult to apply when outliers in the data are present, as they may bias the empirical covariance matrix and thus the selection process. We note that outliers can potentially bias the selection process in this manner for any of the criteria we examined, and not just for the TECM. Similarly, as mentioned by Hin and Wang [9], utilizing a misspecified mean structure can also mislead the selection process. Although work has been performed to improve estimation performance for a given working correlation structure [43], further work is needed with respect to the TECM or TGI criteria in order to make GEE and QIF directly comparable in terms of which approach and working correlation structure to select.

Acknowledgements

We would like to thank the anonymous associate editor and two reviewers for their constructive comments that helped improve this paper. We would also like to thank Dr. Richard J. Kryscio, Dr. Frederick A. Schmitt, and Dr. Erin Abner for providing us the data from the PREADViSE trial, which was supported by a grant from the National Institute on Aging (R01 AG019241). We would further like to thank Dr. Abner for her direct assistance in providing us the data set and relevant information on the trial.

References

1. Caban-Holt A, Abner E, Kryscio RJ, Crowley JJ, Schmitt FA. Age-expanded normative data for the Ruff 2&7 Selective Attention Test: evaluating cognition in older males. *The Clinical Neuropsychologist* 2012; **26**:751–768.
2. Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Weiner MF, Cullum CM. A total score for the CERAD neuropsychological battery. *Neurology* 2005; **65**:102–106.
3. Mathews M, Abner E, Caban-Holt A, Kryscio R, Schmitt F. CERAD practice effects and attrition bias in a dementia prevention trial. *International Psychogeriatrics* 2013; **25**:1115–1123.
4. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
5. Wang Y-G, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika* 2003; **90**:29–41.
6. Ziegler A, Vens M. Generalized estimating equations: notes on the choice of the working correlation matrix. *Methods of Information in Medicine* 2010; **49**:421–425.
7. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics* 2001; **57**:120–125.
8. Pan W, Connett JE. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. *Statistica Sinica* 2002; **12**:475–490.

9. Hin L-Y, Wang Y-G. Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine* 2009; **28**:642–658.
10. Rotnitzky A, Jewell NP. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 1990; **77**:485–497.
11. Hin L-Y, Carey VJ, Wang Y-G. Criteria for working-correlation-structure selection in GEE. *The American Statistician* 2007; **61**:360–364.
12. Shults J, Sun W, Tu X, Kim H, Amsterdam J, Hilbe JM, Ten-Have T. A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine* 2009; **28**:2338–2355.
13. Carey VJ, Wang Y-G. Working covariance model selection for generalized estimating equations. *Statistics in Medicine* 2011; **30**:3117–3124.
14. Gosho M, Hamada C, Yoshimura I. Criterion for the selection of a working correlation structure in the generalized estimating equation approach for longitudinal balanced data. *Communications in Statistics - Theory and Methods* 2011; **40**:3839–3856.
15. Zhou QM, Song P-K, Thompson ME. Information ratio test for model misspecification in quasi-likelihood inference. *Journal of the American Statistical Association* 2012; **107**:205–213.
16. Barnett AG, Koper N, Dobson AJ, Schmiegelow F, Manseau M. Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution* 2010; **1**:15–24.
17. Pan W. On the robust variance estimator in generalised estimating equations. *Biometrika* 2001; **88**:901–906.
18. Westgate PM. A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. *Statistics in Medicine* 2013; **32**:2850–2858.
19. Crowder M. On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* 1995; **82**:407–410.
20. Sutradhar BC, Das K. On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika* 1999; **86**:459–465.
21. SAS Institute Inc. *SAS/STAT 9.3 User's Guide*. SAS Institute Inc.: Cary, NC, 2011.
22. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**:126–134.
23. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala SI, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 2007; **63**:935–941.
24. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; **96**:1387–1396.
25. Hardin J, Hilbe J. *Generalized Estimating Equations*. Chapman & Hall/CRC: Boca Raton, Florida, 2003.
26. Shults J, Chaganty NR. Analysis of serially correlated data using quasi-least squares. *Biometrics* 1998; **54**:1622–1630.
27. Song P-K, Jiang Z, Park E, Qu A. Quadratic inference functions in marginal models for longitudinal data. *Statistics in Medicine* 2009; **28**:3683–3696.
28. Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika* 2000; **87**:823–836.
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0, 2011. Available from: http://www.R-project.org.
30. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics* 1988; **44**:1033–1048.
31. Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T. *mvtnorm: Multivariate Normal and t Distributions*, 2013. R package version 0.9-9995. Available from: http://CRAN.R-project.org/package=mvtnorm.
32. Genz A, Bretz F. *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, Vol. 195. Springer-Verlage: Heidelberg, 2009. ISBN 978-3-642-01688-2.
33. Leisch F, Weingessel A, Hornik K. *bindata: Generation of Artificial Binary Data*, 2011. R package version 0.9-18. Available from: http://CRAN.R-project.org/package=bindata.
34. Leisch F, Weingessel A, Hornik K. *On the generation of correlated artificial binary data*, *Working Paper Series*, SFB "Adaptive Information Systems and Modeling in Economics and Management Science". Vienna University of Economics, 1998. http://www.wu-wien.ac.at/am.
35. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. *Journal of Statistical Software* 2006; **15**:1–11.
36. Yan J, Fine J. Estimating equations for association structures. *Statistics in Medicine* 2004; **23**:859–874.
37. Yan J. geepack: yet another package for generalized estimating equations. *R-News* 2002; **2/3**:12–14.
38. Boekamp JR, Strauss ME, Adams N. Estimating premorbid intelligence in African-American and white elderly veterans using the American version of the national adult reading test. *Journal of Clinical and Experimental Neuropsychology* 1995; **17**:645–653.
39. Chen J, Lazar NA. Selection of working correlation structure in generalized estimating equations via empirical likelihood. *Journal of Computational and Graphical Statistics* 2012; **21**:18–41.
40. Qu A, Song PX-K. Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika* 2004; **91**:447–459.
41. Westgate PM, Braun TM. The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions. *Statistics in Medicine* 2012; **31**:2209–2222.
42. Westgate PM. A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine* 2012; **31**:4003–4022.
43. Westgate PM, Braun TM. An improved quadratic inference function for parameter estimation in the analysis of correlated data. *Statistics in Medicine* 2013; **32**:3260–3273.

Statistics
in Medicine

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.