Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach

Philip M. Westgate*

Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, USA

Received 31 May 2013; revised 30 September 2013; accepted 27 October 2013

Generalized estimating equations (GEE) are commonly used for the marginal analysis of correlated data, although the quadratic inference function (QIF) approach is an alternative that is increasing in popularity. This method optimally combines distinct sets of unbiased estimating equations that are based upon a working correlation structure, therefore asymptotically increasing or maintaining estimation efficiency relative to GEE. However, in finite samples, additional estimation variability arises when combining these sets of estimating equations, and therefore the QIF approach is not guaranteed to work as well as GEE. Furthermore, estimation efficiency can be improved for both analysis methods by accurate modeling of the correlation structure. Our goal is to improve parameter estimation, relative to existing methods, by simultaneously selecting a working correlation structure and choosing between GEE and two versions of the QIF approach. To do this, we propose the use of a criterion based upon the trace of the empirical covariance matrix (TECM). To make GEE and both QIF versions directly comparable for any given working correlation structure, the proposed TECM utilizes a penalty to account for the finite-sample variance inflation that can occur with either version of the QIF approach. Via a simulation study and in application to a longitudinal study, we show that penalizing the variance inflation that occurs with the QIF approach is necessary and that the proposed criterion works very well.

Keywords: Correlation selection; Efficiency; Empirical covariance matrix; Generalized estimating equations; Quadratic inference functions.

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

1 Introduction

Correlated data arise frequently in practice, such as in longitudinal studies in which independent subjects contribute multiple observations over time. For instance, we later give focus to a dataset from a study in which 59 epileptic subjects each contributed four outcomes, which were the number of seizures experienced throughout each of four consecutive two-week periods (Thall and Vail, 1990; Song, 2007). Each subject was randomized to serve as a control or to receive a treatment drug, in addition to being given standard chemotherapy. As the goal is to estimate the impact of the drug and to test whether or not it reduces seizure rates, while adjusting for other potential predictors, efficient parameter estimates from the corresponding regression model are desired.

*Corresponding author: e-mail: philip.westgate@uky.edu, Phone: +01-859-218-2082

Generalized estimating equations (Liang and Zeger, 1986) are commonly employed for the marginal analysis of correlated data. Assuming the mean structure is correctly specified, consistent parameter estimates are given whether or not the working correlation structure is correctly specified. However, accurate modeling of this structure can improve estimation efficiency (Liang and Zeger, 1986; Wang and Carey, 2003), and therefore two currently popular strategies have been proposed that do not require additional assumptions and can be used simultaneously. The first is based on the idea of selecting the best working correlation structure, for which multiple criteria have been proposed. For instance, see Hin et al. (2007), Hin and Wang (2009), Shults et al. (2009), Carey and Wang (2011), Gosho et al. (2011), and Westgate (submitted). The second is to optimally combine distinct sets of estimating equations, which are created from a selected working correlation structure, via the quadratic inference function (QIF) (Qu et al., 2000) approach. As with GEE, selecting the best working correlation structure for use with the QIF approach can also improve this method's efficiency (Song et al., 2009).

Originally, Qu et al. (2000) rewrote the inverse of the selected working correlation matrix inside the GEE as the weighted sum of m basis matrices. Using the generalized method of moments (GMM) (Hansen, 1982), the corresponding m distinct GEE components are then utilized inside the QIF and are optimally weighted in this method's estimating equations. Therefore, the QIF approach is more efficient than GEE when the working correlation structure is misspecified, whereas the two methods are equally efficient in the case of correct specification.

Unfortunately, the QIF's efficiency advantage is not necessarily observed in finite samples. Specifically, the estimating equations employed with this approach utilize a weighting matrix that is empirically estimated. Westgate and Braun (2012) and Westgate (2013c) discussed that this matrix's variability and usage of working parameter estimates increases the variability within these equations, and therefore the finite-sample variances of final parameter estimates are larger than theoretically expected. Westgate (2012) derived a formula that approximates this variance inflation and proposed a bias-corrected empirical covariance matrix to estimate the true covariance matrix of parameter estimates. Westgate (2013a) further extended this formula, and thus bias-corrected matrix, for the QIF approach that incorporates an unstructured working correlation for which the magnitude of the variance inflation increases. Westgate (2013c) showed via simulation study that although the choice of initial parameter estimates does not appear to influence the QIF's estimation performance, holding initial estimates constant within the weighting matrix throughout the iterative estimation procedure can improve parameter estimation in some instances. However, even when doing this, the finite-sample variance inflation can still exist, and therefore the QIF approach is not guaranteed to perform as well as the corresponding GEE with the same structure. Therefore, which combination of analysis method, either GEE or the QIF approach, and working correlation structure will lead to the least variable parameter estimates in any given setting in practice will be unknown.

Our goal is to improve parameter estimation, relative to existing methods, by simultaneously choosing between GEE and the QIF approach and selecting a working correlation structure. As mentioned previously, existing methods have been developed for correlation selection when using GEE or the QIF approach. However, these methods do not allow the direct comparison of these two analysis methods. Similarly, approaches have been proposed to select a set of basis matrices to utilize (Qu et al., 2008; Zhou and Qu, 2012), which is related to this idea of correlation selection. Additionally, Westgate and Braun (2013) recently proposed a modified QIF method that is essentially a combination of GEE and the QIF, or typical QIF, approach. Although this method was shown to improve overall estimation performance relative to the sole use of either GEE or the typical QIF approach for any given working correlation structure, it also has limitations. Particularly, this modified QIF is only applicable with the exchangeable or AR-1 working correlation structures, and it assumes that the data analyst already knows what correlation structure should be utilized. Therefore, our idea of simultaneously choosing an analysis method and working correlation structure is advantageous relative to these existing methods. Specifically, if both can be selected, the analyst is not limited with respect to the method and correlation combinations that can be considered, and therefore a combination may be selected that leads to improved parameter estimation relative to existing methods that do not consider this combination.

Furthermore, the modified QIF may still be advantageous relative to GEE and the typical QIF for any given analysis, and we therefore allow all three analysis methods to be considered for selection.

To simultaneously select both an analysis method and a working correlation structure, we propose the use of a criterion that utilizes the trace of the empirical covariance matrix (TECM), as motivated by Song et al. (2009) and Westgate (submitted). The idea behind this criterion is that the best combination of analysis method and correlation structure should give the least variable parameter estimates. Therefore, we propose selecting the combination that yields the smallest TECM. To make GEE and both versions of the QIF approach comparable, this criterion must be appropriately penalized for the increase in finite-sample variances in parameter estimates that can arise with either version of the QIF approach; otherwise, the QIF can be overselected. Therefore, we utilize the bias-corrected empirical covariance matrices proposed by Westgate (2012) and Westgate and Braun (2013) for the typical and modified QIF approaches, respectively, as they appropriately inflate the theoretical asymptotic empirical covariance matrix of parameter estimates. The use of this correction, or penalty, with the TECM allows the direct comparison of the realized estimation performances of distinct GEEs and QIF versions that utilize different correlation structures.

Section 2 introduces GEE and the QIF approach, and makes theoretical efficiency distinctions between these methods. In Section 3, we give details on the finite-sample estimation performance of the QIF approach, and discuss how to account for the finite-sample variance inflation that occurs with either version of this method. Furthermore, we propose the TECM criterion that imposes a penalty upon the QIF approach due to this variance increase. A simulation study is carried out in Section 4 to illustrate the finite-sample estimation performance of the QIF approach and the necessity and utility of the proposed TECM criterion. An applied example is then given in Section 5, and concluding remarks are made in Section 6.

2 Generalized estimating equations and quadratic inference functions

For the ith independent cluster, or ith subject in a longitudinal study, $i = 1, \ldots, N$, denote $\boldsymbol{Y}_i = [Y_{i1}, \ldots, Y_{in_i}]^T$ as the corresponding vector of correlated outcomes, with corresponding marginal means, $\boldsymbol{\mu}_i = [\mu_{i1}, \ldots, \mu_{in_i}]^T$, modeled by $f(\mu_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta}$. Here, f is a known link function, $\mathbf{x}_{ij} = [1, x_{ij1}, \ldots, x_{ij(p-1)}]$ is the corresponding vector of covariate values, and $\beta = [\beta_0, \beta_1, \ldots, \beta_{p-1}]^T$ is a $p \times 1$ vector of regression parameters. Furthermore, $\mathrm{Var}(Y_{ij}) = \phi\nu(\mu_{ij})$, $j = 1, \ldots, n_i$, where ϕ is an assumed common dispersion parameter and ν is a known function.

Parameter estimates, $\hat{\boldsymbol{\beta}}$, are obtained from GEE (Liang and Zeger, 1986) by solving

$$\sum_{i=1}^{N} D_i^T V_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{1}$$

Here, $D_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^T$, $V_i = A_i^{1/2} R_i A_i^{1/2}$, $A_i = \mathrm{diag}[\phi\nu(\mu_{i1}), \ldots, \phi\nu(\mu_{in_i})]$, and R_i is the working correlation matrix for the ith cluster, $i = 1, \ldots, N$. When the working covariance structure for the data is misspecified, the model-based covariance matrix of $\hat{\boldsymbol{\beta}}$ is biased. Therefore, the empirical covariance matrix, given by

$$\widehat{\Sigma}_{E,GEE} = \left(\sum_{i=1}^{N} D_i^T V_i^{-1} D_i\right)^{-1} \left[\sum_{i=1}^{N} D_i^T V_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T V_i^{-1} D_i\right] \left(\sum_{i=1}^{N} D_i^T V_i^{-1} D_i\right)^{-1},$$

is commonly utilized with this approach since it is a consistent estimate for the true covariance matrix even when the working covariance structure is misspecified (Liang and Zeger, 1986). However, for small N, such as less than 40 or 50 (Mancl and DeRouen, 2001; Lu et al., 2007), $\widehat{\Sigma}_{E,GEE}$ can be negatively biased due to the use of estimated empirical covariances, $(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T, i = 1, \ldots, N$. Therefore, as studied in Lu et al. (2007), Mancl and DeRouen (2001), and Kauermann and Carroll (2001) derived distinct corrections that replace $(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$ with either $(I_n - H_i)^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T(I_n - H_i^T)^{-1}$ or $(I_n - H_i)^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T, i = 1, \ldots, N$, respectively. Here, I_{n_i} is an $n_i \times n_i$ identity matrix and $H_i = D_i(\sum_{i=1}^N D_i^T V_i^{-1} D_i)^{-1} D_i^T V_i^{-1}$. We denote $\widehat{\Sigma}_{BC,GEE}$ as the estimated empirical covariance matrix that incorporates one of these two corrections.

When the working correlation structure is misspecified, GEE can be inefficient. Therefore, Qu et al. (2000) proposed the QIF approach to improve efficiency, in which they use $R_i^{-1} \approx \sum_{r=1}^m \gamma_{ri} M_{ri}$, where $M_{ri}, r = 1, \ldots, m$, are basis matrices composed of 0's and 1's and $\gamma_{ri}, r = 1, \ldots, m$, are functions consisting of a correlation parameter(s) and possibly cluster size. Specific examples include the working independence, exchangeable, and AR-1 structures, which have inverses that can be exactly rewritten as the sum of one, two, and three weighted basis matrices, respectively. For all three structures, M_{1i} is an identity matrix. For exchangeable, M_{2i} is a matrix of 0's on the diagonal and 1's elsewhere, whereas M_{2i} uses 1's on the two main off-diagonals and 0's elsewhere for AR-1. Furthermore for AR-1, all elements of M_{3i} are 0 except for the two corner components, which are 1, on the main diagonal. Due to its negligible impact on estimation, this third basis matrix is sometimes ignored (Qu et al., 2000; Song et al., 2009). Replacing R_i^{-1} with $\sum_{r=1}^m \gamma_{ri} M_{ri}$, the GEE in Eq. (1) becomes

$$\sum_{r=1}^{m} I \sum_{i=1}^{N} D_i^T A_i^{-1/2} \gamma_{ri} M_{ri} A_i^{-1/2} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \sum_{r=1}^{m} I \sum_{i=1}^{N} g_{ri}. \tag{2}$$

When $\gamma_{r1} = \ldots = \gamma_{rN} = \gamma_r, r = 1, \ldots, m$, Eq. (2) can be simplified as

$$\sum_{r=1}^{m} \gamma_r I \sum_{i=1}^{N} D_i^T A_i^{-1/2} M_{ri} A_i^{-1/2} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \sum_{r=1}^{m} \gamma_r I \sum_{i=1}^{N} g_{ri}. \tag{3}$$

This simplification can be made when data are balanced, or even in settings in which cluster sizes vary and the AR-1 structure is applicable, such as when some subjects do not contribute data at one or more final time points and the temporal spacing is the same for all subjects. The claimed advantage of this formulation given in Eq. (3) is that the m sets of equations do not need to incorporate nuisance correlation parameters, which therefore do not need to be estimated (Qu et al., 2000). Alternatively, as discussed by Westgate and Braun (2012), this simplification is not always possible when clusters vary in size, as is the case when incorporating the exchangeable structure. In this case, $\gamma_{ri}, i = 1, \ldots, N; r = 1, \ldots, m$, depends on cluster size. We note that, as with GEE, if cluster sizes vary due to missing data, the QIF approach must assume data are missing completely at random (Song et al., 2009). Furthermore, even when Eq. (2) can be simplified to the form given in Eq. (3), such simplification is not necessary. Additionally, although the formulations for $g_{ri}, i = 1, \ldots, N; r = 1, \ldots, m$, are different for Eqs. (2) and (3), this equivalent notation is utilized in order to easily demonstrate their common usage with the QIF approach. We note that for the unstructured working correlation, the m sets of estimating equations are not exactly the same as given in Eqs. (2) and (3), and we refer the reader to Qu and Lindsay (2003) and Westgate (2013a) for technical details.

In short, Eqs. (2) and (3) show that GEE can be rewritten as a weighted linear combination of the given m sets of unbiased estimating equations. Specifically, the weight for the rth set is given by either I in Eq. (2) or $\gamma_r I$ in Eq. (3), where I is the identity matrix, and $\gamma_r, r = 1, \ldots, m$, are determined solely by the working correlation structure. Therefore, I and $\gamma_r I, r = 1, \ldots, m$, are optimal weights in Eqs. (2) and (3), respectively, only if the covariance structure is correctly specified (Qu et al., 2000).

If the correlation structure is misspecified, I and $\gamma_r I, r = 1, \ldots, m$, are not optimal. Therefore, to asymptotically obtain optimal weighting matrices, Qu et al. (2000) utilized the GMM (Hansen, 1982) and composed the extended score vector that is comprised of the m sets of equations:

$$\bar{g}_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} g_i(\boldsymbol{\beta}) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} g_{1i} \\ \vdots \\ \sum_{i=1}^{N} g_{mi} \end{bmatrix}.$$

These are used to create the QIF, given by $Q_N(\boldsymbol{\beta}) = N\bar{g}_N^T(\boldsymbol{\beta})C_N^{-1}(\boldsymbol{\beta})\bar{g}_N(\boldsymbol{\beta})$, in which $C_N(\boldsymbol{\beta}) = (1/N)\sum_{i=1}^{N} g_i(\boldsymbol{\beta})g_i^T(\boldsymbol{\beta})$ is an empirical covariance matrix that is used to estimate the optimal weighting covariance matrix, $\mathrm{E}[C_N(\boldsymbol{\beta})] = (1/N)\sum_{i=1}^{N}\mathrm{Cov}[g_i(\boldsymbol{\beta})]$. The estimate for $\boldsymbol{\beta}$ is obtained by minimizing the QIF, which is asymptotically equivalent to solving

$$N\dot{g}_N^T(\boldsymbol{\beta})C_N^{-1}(\boldsymbol{\beta})\bar{g}_N(\boldsymbol{\beta}) = \sum_{r=1}^{m} B_r \sum_{i=1}^{N} g_{ri} = \mathbf{0}, \tag{4}$$

in which $\dot{g}_N(\boldsymbol{\beta}) = \mathrm{E}[\partial\bar{g}_N(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T]$ and B_r, $r = 1, \ldots, m$, is a matrix comprised of columns $p(r-1)+1$ to pr from $\dot{g}_N^T(\boldsymbol{\beta})C_N^{-1}(\boldsymbol{\beta})$. The estimating equations in Equation (4) are asymptotically equivalent to $N\dot{g}_N^T(\boldsymbol{\beta})\mathrm{E}[C_N(\boldsymbol{\beta})]^{-1}\bar{g}_N(\boldsymbol{\beta})$ since $C_N(\boldsymbol{\beta}) - \mathrm{E}[C_N(\boldsymbol{\beta})] \xrightarrow{p} 0$ (Qu et al., 2000; Han and Song, 2011). Therefore, B_r, $r = 1, \ldots, m$, are asymptotically optimal weighting matrices (Lindsay, 1982; Small and MCLeish, 1994; Qu et al., 2000). Specifically, Eq. (4) gives the optimal weighted linear combination of the m sets of estimating equations such that the minimum asymptotic covariance matrix, $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (1/N)(\dot{g}_N^T\ \mathrm{E}[C_N(\boldsymbol{\beta})]^{-1}\dot{g}_N)^{-1}$, in the Löwner ordering is attained (Hansen, 1982; Lindsay, 1982; Small and McLeish, 1994; Qu et al., 2000).

GEE, as given in Eq. (2), or Eq. (3) when applicable, and the QIF method's estimating equations, as given in Eq. (4), have the same form since both can be expressed as weighted linear combinations of the same m sets of equations. Therefore, they are said to be in the same class of estimating equations (Qu et al., 2000). However, because B_r, $r = 1, \ldots, m$, are asymptotically optimal weighting matrices, efficiency is improved upon GEE when the working covariance structure is misspecified, whereas efficiency is maintained when this structure is correct (Qu et al., 2000).

The modified QIF approach revises the estimating equations of the typical QIF approach given in Eq. (4) by replacing C_N^{-1} with C_N^{*-1}, where $C_N^* = \rho_N M_N + (1-\rho_N)C_N$. Here, M_N replaces the empirical covariances, $(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T$, within C_N with model-based covariances, V_i, $i = 1, \ldots, N$, and is thus the model-based estimate for $\mathrm{E}[C_N(\boldsymbol{\beta})]$. Furthermore, ρ_N minimizes the expected quadratic loss of C_N^*, which, like C_N, is also an asymptotically optimal weighting matrix that allows the modified QIF approach to maintain the theoretical efficiency advantage over GEE.

3 Finite-sample properties and the proposed selection criterion

Either version of the QIF approach incorporates $C_N(\boldsymbol{\beta})$ in some manner. Specifically, the typical QIF soley utilizes $C_N(\boldsymbol{\beta})$, whereas $\rho_N M_N + (1-\rho_N)C_N(\boldsymbol{\beta})$ is used by the modified version. The empirical covariance matrix C_N is our focus with respect to the variance inflation that can occur with either version of the QIF approach. Therefore, we start off by describing the impacts of C_N with the typical QIF approach, and then naturally extend these ideas for the modified QIF.

Theoretical efficiency results for the QIF approach rely upon $N\dot{g}_N^T(\boldsymbol{\beta})\mathrm{E}[C_N(\boldsymbol{\beta})]^{-1}\bar{g}_N(\boldsymbol{\beta})$ as estimating equations. However, $\mathrm{E}[C_N(\boldsymbol{\beta})]$ is unknown and replaced by its asymptotically equivalent estimator, $C_N(\boldsymbol{\beta})$, with the typical QIF approach. Realistically, we must utilize $C_N(\tilde{\boldsymbol{\beta}})$ rather than $C_N(\boldsymbol{\beta})$, where $\tilde{\boldsymbol{\beta}}$ is a consistent working estimate for $\boldsymbol{\beta}$. Due to the finite-sample variability within C_N, replacing $\boldsymbol{\beta}$ with $\tilde{\boldsymbol{\beta}}$ inside this matrix causes the estimation variability within Eq. (4) to increase (Westgate and Braun, 2012; Westgate, 2012, 2013c). Therefore, the finite-sample variances of final parameter estimates, $\hat{\boldsymbol{\beta}}$, are larger than theoretically expected (Westgate, 2012, 2013c). The asymptotic empirical covariance matrix of $\hat{\boldsymbol{\beta}}$, given by $\widehat{\Sigma}_{E,QIF} = (1/N)(\dot{g}_N^T C_N^{-1} \dot{g}_N)^{-1} = (1/N)J_N^{-1}$, does not take into account this finite-sample variance inflation, and is therefore biased in finite samples (Westgate, 2012).

The magnitude of this finite-sample variance inflation, and therefore the amount of bias within $\widehat{\Sigma}_{E,QIF}$, depends upon the sample size (Qu et al., 2008; Westgate and Braun, 2012; Westgate, 2012, 2013c). Specifically, the smaller the sample size, the greater the variance increase. Therefore, in the case of a correctly specified correlation structure and when N is not large, the QIF approach will not perform as well as GEE and may even be detrimental (Westgate and Braun, 2012; Westgate, 2012, 2013c). Alternatively, which method is best when the correlation structure is misspecified will depend upon the extent of the finite-sample variance inflation occurring with the QIF approach and the degree of efficiency gain this method theoretically has over GEE (Westgate, 2012).

Due to these finite-sample issues, it is clear that the finite-sample covariance of $\hat{\boldsymbol{\beta}}$, rather than the theoretical asymptotic covariance, must be considered when selecting an analysis method and working correlation structure. Therefore, the finite-sample inflation of $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$, for any working correlation structure with the QIF approach, must be known. Utilizing the following expansion that is conditioned on the use of $\tilde{\boldsymbol{\beta}}$ within C_N,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx -\boldsymbol{J}_N^{-1} \dot{\boldsymbol{g}}_N^T \boldsymbol{C}_N^{-1}(\tilde{\boldsymbol{\beta}}) \bar{\boldsymbol{g}}_N(\boldsymbol{\beta}), \tag{5}$$

and a first-order Taylor series expansion of Eq. (5), Westgate (2012) showed that the finite-sample covariance inflation can be approximated. The resulting bias-corrected empirical covariance matrix is given by $\widehat{\Sigma}_{BC,QIF} = (I_p + \hat{G})\widehat{\Sigma}_{E,QIF}(I_p + \hat{G})^T$, in which $\hat{G}$ is the estimate for

$$G = -\frac{\partial}{\partial \boldsymbol{\beta}^{*T}} \left[J_N^{-1} \dot{g}_N^T C_N^{-1}(\boldsymbol{\beta}^*) \bar{g}_N(\boldsymbol{\beta}) \right] |_{\boldsymbol{\beta}^* = \boldsymbol{\beta}}.$$

As with $\widehat{\Sigma}_{E,GEE}$ for GEE, $\widehat{\Sigma}_{BC,QIF}$ can still be negatively biased for small N due to the use of $(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^T$, $i = 1, \ldots, N$. Therefore, based on the corrections of Mancl and DeRouen (2001) and Kauermann and Carroll (2001) for GEE, Westgate (2012) proposed utilizing $\widehat{\Sigma}_{BC,QIF} = (1/N)(I_p + \hat{G})J_N^{-1} \dot{g}_N^T C_N^{-1}(\tilde{\boldsymbol{\beta}}) \tilde{C}_N(\hat{\boldsymbol{\beta}}) C_N^{-1}(\tilde{\boldsymbol{\beta}}) \dot{g}_N J_N^{-1}(I_p + \hat{G})^T$, in which $\tilde{C}_N(\hat{\boldsymbol{\beta}})$ utilizes similar corrections for the biases within the estimated empirical covariances. Westgate (2012) showed that use of $\widehat{\Sigma}_{BC,QIF}$ works well for estimating variances of the final parameter estimates, as it led to appropriate SE estimates and thus inference, whereas use of $\widehat{\Sigma}_{E,QIF}$ can lead to notable negative bias in SE estimates and hence liberal test sizes and undercoverage of confidence intervals. Furthermore, when incorporating the unstructured working correlation, the formula for G must be modified, and details are given in Westgate (2013a).

The modified QIF approach is essentially a combination of GEE and the typical QIF approach that is meant to optimally take their finite-sample advantages and disadvantages into account. Specifically, within $C_N^* = \rho_N M_N + (1 - \rho_N)C_N$, M_N utilizes the model-based covariances on which GEE is based, whereas C_N corresponds to the typical QIF approach. Details on how to estimate ρ_N can be found in Westgate and Braun (2013). When $\hat{\rho}_N < 1$, weight is given to C_N, and therefore variance inflation also occurs with the modified QIF approach, although to a lesser extent relative to the typical QIF when $\hat{\rho}_N > 0$. For instance, results from the simulation study of Westgate and Braun (2013) showed that $\hat{\rho}_N$ can be very close to 1, on average, when the working and true structures are equivalent, implying that only negligible or no variance inflation occurs with the modified QIF approach in these situations. Extending the bias correction of Westgate (2012), Westgate and Braun (2013) derived a bias-corrected empirical covariance matrix, which we will denote as $\widehat{\Sigma}_{BC,MQIF}$, to estimate the covariance of $\hat{\boldsymbol{\beta}}$. We refer the reader to Westgate and Braun (2013) for the exact formula for $\widehat{\Sigma}_{BC,MQIF}$. However, we do point out that as with $\widehat{\Sigma}_{BC,QIF}$ for the typical QIF approach, $\widehat{\Sigma}_{BC,MQIF}$ takes into account any variance inflation that may occur with the modified QIF approach.

In practice, to select the best combination of analysis method and working correlation structure, estimates of the covariance matrices of resulting parameter estimates from the different combinations of interest must be compared. Therefore, a natural choice for a selection criterion is the TECM, given by $tr(\widehat{\Sigma}_{BC,GEE})$ for GEE, and $tr(\widehat{\Sigma}_{BC,QIF})$ and $tr(\widehat{\Sigma}_{BC,MQIF})$ for the typical and modified versions of the QIF approach, respectively. As the goal is to obtain the least variable parameter estimates, this criterion selects the combination that yields the smallest observed TECM. Alternatively, we could ignore the small-sample corrections given by Mancl and DeRouen (2001) and Kauermann and Carroll (2001), as these have nothing to do with the covariance inflation that occurs with the QIF method, and they are similarly done for each method and correlation structure. For example, the TECM value would then be $tr(\widehat{\Sigma}_{E,GEE})$ for GEE. We further point out that the proposed TECM criterion addresses the two limitations we discussed with respect to the modified QIF approach. Specifically, the proposed TECM criterion can be used to select either exchangeable or AR-1 for use with the modified QIF. However, the utility of the proposed TECM goes beyond this, as it also allows the modified QIF to be compared with the typical QIF approach and GEE, and these latter methods are not limited to only exchangeable and AR-1 structures.

We note that the TECM is also motivated by the work of Westgate (submitted) and Song et al. (2009). Westgate (submitted) proposed the TECM, as given by $tr(\widehat{\Sigma}_{E,GEE})$, for GEE since it is simple to obtain and was shown to work well relative to existing correlation selection criteria. Song et al. (2009) proposed the use of the trace of the Godambe Information (TGI) matrix, which is the inverse of $\widehat{\Sigma}_{E,QIF}$, to select a single working correlation structure with the QIF approach, although they noted that $tr(\widehat{\Sigma}_{E,QIF})$ could similarly be utilized. However, since $\widehat{\Sigma}_{E,QIF}$ does not use the finite-sample covariance inflation correction, which we view as a penalty that $tr(\widehat{\Sigma}_{BC,QIF})$ imposes upon the QIF approach, the corresponding unpenalized TGI and TECM criteria can be detrimental. Specifically, the typical QIF approach will be over selected relative to GEE and possibly the modified QIF approach. Furthermore, with respect to selecting a correlation structure for use with the typical QIF approach, Westgate (2012) stated the need for future research to determine if utilizing the variance inflation formula within the TGI criterion has any additional utility relative to using the uncorrected TGI.

4 Simulation study

4.1 Study description

We now carry out a simulation study to exhibit realistic estimation performances of the typical and modified QIF approaches, relative to GEE, in addition to demonstrating the necessity and utility of the proposed TECM criterion that penalizes the finite-sample variance inflation that results with these QIF versions. For simplicity, we consider scenarios in which we are interested in six specific data analysis options. These include combinations of either a working exchangeable or AR-1 correlation structure with either GEE or the typical or modified QIF approaches. We use these six combinations because they have been commonly used in the QIF literature and allow us to easily compare theoretical estimation efficiencies and realistic performances. We note that, due to the negligible impact of the third basis matrix, we use the working AR-1 version of the QIF method that uses only two basis matrices. Furthermore, based on the work of Westgate (2013c), and as used by Westgate and Braun (2013), initial parameter estimates for both QIF versions were obtained from GEE incorporating independence, and initial estimates were held fixed within the weighting matrices C_N and C_N^*.

We simulate data from the model given by

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \epsilon_{ij}; \quad \epsilon_{ij} \sim N(0, 1); \quad j = 1, 2, 3, 4,$$

in which $\boldsymbol{\beta} = [0, 0.5, 1]^T$, $x_{1ij} = j/4$ and $x_{2ij} \sim N(j/4, 1)$, $i = 1, \ldots, N$, similar to a model used by Qu et al. (2000). The number of subjects utilized in any given setting is either 25, 50, 100, or 500, which allows us to demonstrate the impacts sample size has on the estimation performance of the

Table 1 Empirical frequencies of selecting each of the six data analysis options out of 1000 replications, with corresponding ratios (Ratio) that, for each setting (N), compare the empirical mean squared error (MSE) from the sole use of the best analysis option(s) ($MR = 1.00$) to the MSE from the use of the given trace of the empirical covariance matrix (TECM) criterion. MSE ratios (MR) that compare the best analysis option(s) ($MR = 1.00$) to the other options are italicized. The true correlation structure is exchangeable (Exch).

				GEE		Typical QIF		Modified QIF	
N	TECM	Ratio		Exch	AR-1	Exch	AR-1	Exch	AR-1
			MR:	*1.00*	*0.93*	*0.95*	*0.86*	*1.00*	*0.92*
25	Asymptotic	0.88		19	15	145	742	71	8
	Corrected	0.97		44	76	257	191	357	75
			MR:	*1.00*	*0.94*	*0.99*	*0.92*	*1.00*	*0.93*
50	Asymptotic	0.94		23	9	255	581	128	4
	Corrected	0.99		22	53	318	136	433	38
			MR:	*1.00*	*0.92*	*0.99*	*0.93*	*1.00*	*0.91*
100	Asymptotic	0.97		28	5	395	421	150	1
	Corrected	0.99		12	24	384	108	459	13
			MR:	*1.00*	*0.91*	*1.00*	*0.94*	*1.00*	*0.93*
500	Asymptotic	1.00		68	0	741	17	174	0
	Corrected	1.00		8	0	448	6	538	0

N - number of independent subjects; GEE - generalized estimating equations
QIF - quadratic inference function; Asymptotic - the unpenalized TECM criterion
Corrected - the proposed, penalized TECM criterion

QIF approach and on the utility of the TECM. Furthermore, results from settings in which the true structures are exchangeable, AR-1, and a specific Toeplitz, with correlation parameter values of 0.7, are presented in Tables 1, 2, and 3, respectively. AR-1 and exchangeable are used since they are the working structures under consideration, therefore giving settings in which the GEE, and possibly the modified QIF, that incorporates the true structure performs best. Specifically, this GEE and modified QIF approach will be more efficient than GEE and either QIF version incorporating the incorrect structure. They can also perform better than the corresponding typical QIF incorporating the true structure due to the finite-sample variance inflation that occurs with this method and that can be notably greater than with the modified QIF. The Toeplitz correlation is specifically given as $\rho_{ijk} = 0.7/|k - j|$; $k \neq j$; $k, j = 1, 2, 3, 4$, for the jth and kth outcomes from subject i, $i = 1, \ldots, N$, and is utilized to demonstrate settings in which the QIF approach can be advantageous over GEE.

To demonstrate differences in the estimation performances of the six specific data analysis combination options, each setting presents italicized empirical mean squared error (MSE) ratios, denoted by "MR" in Tables 1–3. Specifically, the MSE values are the sum of the empirical MSEs from all nonintercept parameters, as the intercept is usually not of interest. The numerator of a given combination's ratio is the smallest empirical MSE value that results from the use of what we therefore call the best of the six specific data analysis options for the given setting, whereas the denominator is the empirical MSE value that results from the use of the given combination in that setting. Therefore, the analysis option that works best has a ratio of *1.00*. Also, we can assess relative realistic estimation performances of the six combinations, as smaller ratios indicate larger MSEs. Furthermore, in some settings multiple combinations appear to work best and have equivalent ratios of *1.00* after rounding.

To demonstrate the necessity and utility, in terms of parameter estimation, of the proposed selection criterion that penalizes the typical and modified QIF approaches, in each setting we present

Table 2 Empirical frequencies of selecting each of the six data analysis options out of 1 000 replications, with corresponding ratios (Ratio) that, for each setting (N), compare the empirical mean squared error (MSE) from the sole use of the best analysis option ($MR = 1.00$) to the MSE from the use of the given trace of the empirical covariance matrix (TECM) criterion. MSE ratios (MR) that compare the best analysis option ($MR = 1.00$) to the other options are italicized. The true correlation structure is AR-1.

				GEE		Typical QIF		Modified QIF	
N	TECM	Ratio		Exch	AR-1	Exch	AR-1	Exch	AR-1
			MR:	*0.95*	*1.00*	*0.91*	*0.85*	*0.95*	*0.98*
25	Asymptotic	0.86		1	28	76	848	21	26
	Corrected	0.96		21	317	148	221	115	178
			MR:	*0.94*	*1.00*	*0.90*	*0.93*	*0.94*	*0.99*
50	Asymptotic	0.92		0	44	72	846	16	22
	Corrected	0.97		14	387	111	199	96	193
			MR:	*0.95*	*1.00*	*0.95*	*0.96*	*0.95*	*0.99*
100	Asymptotic	0.97		0	104	26	832	13	25
	Corrected	0.98		1	484	62	206	46	201
			MR:	*0.96*	*1.00*	*0.96*	*0.99*	*0.96*	*0.99*
500	Asymptotic	0.99		0	401	1	589	0	9
	Corrected	1.00		0	771	1	125	1	102

N - number of independent subjects; GEE - generalized estimating equations
QIF - quadratic inference function; Asymptotic - the unpenalized TECM criterion
Corrected - the proposed, penalized TECM criterion
Exch - Exchangeable

results from the use of two different versions of the TECM. The first, denoted by "Asymptotic" in Tables 1–3, utilizes asymptotic covariance formulas given by $\widehat{\Sigma}_{E,GEE}$, $\widehat{\Sigma}_{E,QIF}$, and $\widehat{\Sigma}_{E,MQIF}$ for GEE and the typical and modified QIF approaches, respectively, and demonstrates the necessity of penalizing the QIF approach. The second is the proposed TECM criterion and is denoted by "Corrected". It utilizes $\widehat{\Sigma}_{BC,GEE}$ for GEE, and $\widehat{\Sigma}_{BC,QIF}$ and $\widehat{\Sigma}_{BC,MQIF}$ are used to penalize, or correct for, the variance inflations occurring with the typical and modified QIF approaches, respectively. Results are in terms of empirical selection frequencies and MSE ratios. Specifically, the number of times each combination under consideration was selected out of 1000 replications is reported. The goal is to select the best analysis option as often as possible, thus decreasing MSEs. Therefore, we have additionally presented ratios that divide the smallest observed MSE value from the best analysis option, depending on the setting, by the empirical MSE values resulting from the use of the two TECM versions. This is done to demonstrate the relative estimation performances of the two TECM versions and the best analysis option that could have been selected. We note that in all settings presented in Tables 1 and 2, the GEE, along with the modified QIF in some instances, incorporating the true structure produced the smallest MSEs, in which case the empirical MSE value resulting from the use of this best GEE comprises the numerators of the ratios in the given setting.

With respect to the proposed TECM, estimates for $\widehat{\Sigma}_{BC,GEE}$, $\widehat{\Sigma}_{BC,QIF}$, and $\widehat{\Sigma}_{BC,MQIF}$ incorporate corrections based on the work of Mancl and DeRouen (2001). We do this since Westgate (2012) found that this correction worked slightly better than the one derived for the QIF based on the work of Kauermann and Carroll (2001). Furthermore, this type of correction is recommended for standard error (SE) estimation, particularly for $N < 40$ or $N < 50$ (Mancl and DeRouen, 2001; Lu et al., 2007). However, since this correction is similarly done for GEE and both QIF versions, and any type of

Table 3 Empirical frequencies of selecting each of the six data analysis options out of 1000 replications, with corresponding ratios (Ratio) that, for each setting (N), compare the empirical mean squared error (MSE) from the sole use of the best analysis option ($MR = 1.00$) to the MSE from the use of the given trace of the empirical covariance matrix (TECM) criterion. MSE ratios (MR) that compare the best analysis option ($MR = 1.00$) to the other options are italicized. The true correlation structure is Toeplitz.

				GEE		Typical QIF		Modified QIF	
N	TECM	Ratio		Exch	AR-1	Exch	AR-1	Exch	AR-1
			MR:	*0.88*	*1.00*	*0.84*	*0.95*	*0.88*	*0.98*
25	Asymptotic	0.96		1	15	16	924	3	41
	Corrected	0.98		4	344	53	348	16	235
			MR:	*0.87*	*0.98*	*0.85*	*1.00*	*0.87*	*0.97*
50	Asymptotic	1.00		0	19	2	961	1	17
	Corrected	0.99		0	340	13	479	1	167
			MR:	*0.84*	*0.94*	*0.83*	*1.00*	*0.83*	*0.94*
100	Asymptotic	1.00		0	7	1	986	0	6
	Corrected	0.97		0	246	2	661	0	91
			MR:	*0.80*	*0.90*	*0.81*	*1.00*	*0.80*	*0.91*
500	Asymptotic	1.00		0	0	0	1000	0	0
	Corrected	1.00		0	5	0	989	0	6

N - number of independent subjects; GEE - generalized estimating equations
QIF - quadratic inference function; Asymptotic - the unpenalized TECM criterion
Corrected - the proposed, penalized TECM criterion
Exch - Exchangeable

working correlation structure, we found (results not shown) that its impact on the TECM criterion's performance was negligible in all settings.

4.2 Study results

For settings in which the true correlation structure was either AR-1 or exchangeable, the corresponding GEE and modified QIF incorporating the true structure worked best, and very similarly, among all six analysis options. Although theoretically as efficient, the corresponding typical QIF versions that also incorporated the true structure did not work as well when $N \leq 100$ due to notable finite-sample variance inflations. Specifically, MSE ratios corresponding to these versions showed that variance inflations were more notable for smaller N.

In the settings of Table 3 in which neither of the working structures modeled the true Toeplitz structure, the typical QIF incorporating AR-1 outperformed both GEEs and modified QIFs when $N \neq 25$. This occurred because the theoretical efficiency gains with the QIF approach were greater than the corresponding finite-sample variance inflations. Additionally, the larger N became, the better this QIF that incorporated AR-1 performed relative to the other combinations.

Due to the notable differences in theoretical and finite-sample estimation performances of the typical QIF approach, particularly for settings in which $N \leq 100$, use of the proposed correction penalty with the TECM was necessary. This was the case since the unpenalized TECM tended to overselect the typical QIF that, although theoretically being equally or more efficient, could realistically not perform as well as GEE or the modified QIF. Alternatively, the proposed penalty appropriately accounted for the finite-sample variance inflations occurring with the QIF approach. Therefore, although the unpenalized TECM was detrimental in multiple settings, the proposed TECM worked very well. For instance, MSE ratios resulting from the use of the proposed TECM were only as low as 0.96 and were typically between 0.97 and 1.00. Alternatively, these ratios were as small as 0.86 for the unpenalized TECM and were often lower than for the proposed TECM, especially in settings in which the true structure was AR-1 or exchangeable and $N \leq 50$.

In general, the smaller N was, the better the proposed penalized TECM worked relative to the uncorrected TECM. Although the penalized TECM performed much better overall, the uncorrected version worked slightly better when $N \neq 25$ in Table 3. This occurred since the unpenalized TECM overselects the QIF approach, and use of AR-1 with the typical QIF performed best here.

For settings in which $N = 500$, no notable differences in terms of empirical MSEs were apparent between these two selection criteria and the best analysis option(s) under consideration. This was the case since finite-sample variance inflations with the different QIF versions were negligible, as was the impact of the variability in the TECM criterion that arises from the use of estimated variances. Although the empirical distributions of selection frequencies resulting from the use of these two criteria could still be notably different, the distinct combinations that were most frequently chosen by these criteria resulted in similar empirical MSEs.

In some settings in which the true structure was exchangeable or Toeplitz, the typical QIF approach incorporating AR-1 outperformed the corresponding GEE with AR-1. This type of comparison has commonly been used in the QIF literature to demonstrate the efficiency advantage the QIF approach has over GEE in instances of correlation structure misspecification. However, results suggest that this specific improvement in estimation performance for the working AR-1 structure can be less impactful when utilizing the proposed TECM for the selection of both analysis method and working structure, as this criterion allows the selection of a better working correlation structure if one is under consideration. We also note that when the true structure was not exchangeable, neither the typical nor modified QIF incorporating exchangeable ever performed better than the corresponding GEE, although theoretically they should have had some efficiency gain.

An interesting finding in settings in which $N \leq 100$ and the true structure was exchangeable is that the unpenalized TECM tended to select AR-1 more often than exchangeable with the typical QIF approach, although use of exchangeable is theoretically more efficient in these settings. The reason for this is that $\widehat{\Sigma}_{E,QIF}$ assumes the use of $C_N(\boldsymbol{\beta})$, rather than $C_N(\tilde{\boldsymbol{\beta}})$, in Eq. (4) as the weighting covariance matrix. As explained in detail in Westgate (2013c), the unrealistic use of $C_N(\boldsymbol{\beta})$ would work better in finite samples than the use of $\mathrm{E}[C_N(\boldsymbol{\beta})]$, which is the matrix that the QIF approach's theoretical efficiency is based upon. Furthermore, the impact of C_N on parameter estimation distinctly depends on the working correlation structure via utilized basis matrices, and is more influential for AR-1 relative to exchangeable. Specifically, we found that the use of $C_N(\boldsymbol{\beta})$ allowed the AR-1 structure to actually perform better than the exchangeable structure in these settings (results not shown), whereas the opposite occurred when realistically using $C_N(\tilde{\boldsymbol{\beta}})$. Therefore, since $\widehat{\Sigma}_{E,QIF}$ is based upon the use of $C_N(\boldsymbol{\beta})$ as the weighting matrix, the unpenalized TECM tended to select AR-1 more often than exchangeable with the typical QIF in these settings.

5 Application

To illustrate the use of the proposed TECM criterion, we utilize the dataset from the seizures study in which 28 and 31 subjects were randomized to the control and treatment arms of the trial, respectively (Thall and Vail, 1990; Song, 2007). One subject had large numbers of seizures, and therefore Song (2007) and Song et al. (2009) fit the marginal model given by

$$log(\mu_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij}; \quad j = 1, 2, 3, 4$$

to demonstrate the robustness of the typical QIF approach, relative to GEE, in the presence of outliers (Qu and Song, 2004). Furthermore, Westgate (2012) used this example to demonstrate the use of

Table 4 Parameter estimates, asymptotic standard error estimates (SE_A), bias-corrected standard error estimates (SE_{BC}), and values for the unpenalized (Asymptotic) and penalized (Corrected) traces of empirical covariance matrices (TECMs) resulting from six analyses, each using a distinct analysis method and correlation structure combination, of the epileptic seizures dataset.

		GEE		Typical QIF		Modified QIF	
Covariate	Estimate	Exch	AR-1	Exch	AR-1	Exch	AR-1
Intercept	$\hat{\beta}_0$	–2.32	–2.52	–1.94	–2.32	–2.33	–2.49
	SE_A	1.04	1.03	0.99	1.02	1.04	1.01
	SE_{BC}	1.20	1.20	1.24	1.25	1.20	1.18
Treatment	$\hat{\beta}_1$	–0.01	–0.02	–0.02	–0.04	–0.01	–0.03
Indicator	SE_A	0.19	0.19	0.14	0.14	0.19	0.18
	SE_{BC}	0.27	0.28	0.19	0.16	0.27	0.27
Baseline	$\hat{\beta}_2$	1.23	1.25	1.17	1.20	1.23	1.24
Seizures	SE_A	0.16	0.16	0.11	0.09	0.16	0.15
	SE_{BC}	0.26	0.28	0.16	0.14	0.26	0.26
Natural Log	$\hat{\beta}_3$	0.60	0.65	0.52	0.60	0.60	0.65
of Age	SE_A	0.29	0.29	0.28	0.27	0.29	0.28
	SE_{BC}	0.34	0.34	0.36	0.34	0.34	0.33
Two-Week	$\hat{\beta}_4$	–0.06	–0.06	–0.06	–0.05	–0.06	–0.06
Interval	SE_A	0.03	0.03	0.03	0.02	0.03	0.03
	SE_{BC}	0.04	0.04	0.03	0.03	0.04	0.03
	Asymptotic TECM	1.24	1.22	1.09	1.13	1.23	1.15
	Corrected TECM	1.70	1.70	1.72	1.74	1.69	1.64

GEE - generalized estimating equations;
QIF - quadratic inference function;
Exch - Exchangeable;
Baseline Seizures - the natural log of one fourth of the total number of seizures experienced during the eight weeks prior to randomization

$\widehat{\Sigma}_{BC,QIF}$ relative to $\widehat{\Sigma}_{E,QIF}$ in terms of standard error estimation, and Westgate and Braun (2013) further utilized this example by demonstrating the modified QIF approach relative to GEE and the typical QIF method. In the model, μ_{ij} is the ith subject's marginal mean number of seizures during the jth two week interval, x_{1i} is an indicator for subject i receiving the treatment drug rather than serving as a control, the natural log of one fourth of the total number of seizures experienced by the ith subject during the eight weeks prior to randomization is given by x_{2i}, the natural log of the ith subject's age is denoted by x_{3i}, and $x_{4ij} = j$ is the jth two-week interval.

We fit this model using the six specific data analysis options that were considered in the simulation study. As in that study, we incorporate the work of Westgate (2013c) by obtaining initial parameter estimates for both QIF versions from GEE incorporating independence, and initial estimates were held fixed within the weighting matrices C_N and C_N^*. Table 4 presents the corresponding parameter estimates, asymptotic SE estimates from $\widehat{\Sigma}_{E,GEE}$, $\widehat{\Sigma}_{E,QIF}$, or $\widehat{\Sigma}_{E,MQIF}$, bias-corrected SE estimates from versions of $\widehat{\Sigma}_{BC,GEE}$, $\widehat{\Sigma}_{BC,QIF}$, or $\widehat{\Sigma}_{BC,MQIF}$ that utilize corrections based on the work of Mancl and DeRouen (2001), and values for both TECM versions. Parameter estimates and corresponding estimated SEs varied across the **six** distinct analyses. Furthermore, due to negative bias, asymptotic SE estimates were notably smaller than corresponding bias-corrected estimates. Therefore, for any given combination, penalized and unpenalized TECM values were notably different.

As in the simulation study, the unpenalized TECM favored the typical QIF approach. However, using the corrected estimated covariance matrices, the proposed TECM criterion indicates that the modified QIF incorporating the AR-1 structure should be selected over the other five combinations. Additionally, GEE with either correlation structure and the modified QIF with exchangeable all yielded similar values for the proposed TECM criterion, and the typical QIF with either correlation structure produced the largest values. We note that these results extend the results given in Westgate and Braun (2013) for the analysis of this dataset. Specifically, the proposed criterion not only indicates that the AR-1 structure is preferred over the exchangeable when utilizing the modified QIF, but it also supports that the modified QIF should be used rather than GEE or the typical QIF for the analysis of this particular dataset.

As noted by Song et al. (2009), outliers may influence SE estimates obtained from GEE, and their impact on the modified QIF requires further study. Therefore, caution may be necessary when making comparisons in this example. In Table 4, it is apparent that the penalized TECM values for the typical QIF analyses are largest due to notably larger bias-corrected SE estimates for their corresponding intercept estimates. Furthermore, the typical QIF analyses appeared to perform notably better, based on comparing bias-corrected standard error estimates, than either GEE or modified QIF version with respect to estimating β_1 and β_2.

6 Concluding remarks

Asymptotically, the QIF approach should be utilized rather than GEE for parameter estimation. However, this theoretical result will not necessarily hold in practice due to the finite-sample inflation of the variances of resulting parameter estimates from the QIF approach. For instance, results from our simulation study demonstrate that GEE can perform better than the QIF approach when the true structure is incorporated. Therefore, we proposed a selection criterion that takes the trace of the empirical covariance matrix that has corrected for, or penalized, this variance inflation that occurs with either version of the QIF approach. This penalty was shown to be necessary in the simulation study and application example, and the resulting TECM criterion was shown to work well, in terms of parameter estimation, for selecting both an analysis method, GEE or the typical or modified QIF, and a working correlation structure.

In practice, neither the true correlation structure nor the best analysis method will be known, although the plausibility of different correlation structures can be considered. Therefore, the ability to simultaneously choose an analysis method and plausible working correlation structure that results from our proposed approach is very important. Specifically, our proposed approach of simultaneous selection, for which we used the penalized TECM, can improve parameter estimation relative to the sole use of the modified QIF approach and existing approaches that focus on correlation selection for only GEE or only the typcial QIF. This is the case because the combination of correlation structure and analysis method that will result in the least variable parameter estimates may be a selection option with our proposed approach but not even considered by each of these existing methods. For instance, in some settings of the simulation study, the typical QIF approach resulted in the smallest empirical MSEs. Therefore, using an existing method to select either AR-1 or exchangeable for GEE only would be disadvantageous. The proposed TECM also addresses the limitations of the modified QIF by not only allowing for correlation structure selection within this method, but by also allowing for structures other than AR-1 and exchangeable to be considered for selection when used in combination with GEE or the typical QIF approach. We also point out that the typical and modified QIF approaches, and therefore our proposed method of comparing correlation structures simultaneously with these methods and GEE, can be used whether or not data are balanced. However, the correlation and data structure determine whether Eq. (2) or Eq. (3) is used with either QIF approach.

Although for simplicity of presentation we only utilized exchangeable and AR-1 in our simulation study and application, an example of another structure that can be selected when the temporal spacing of repeated measurements is the same for all subjects in a longitudinal study is the unstructured working correlation (Liang and Zeger, 1986; Qu and Lindsay, 2003; Song, 2007; Song et al., 2009). This structure potentially involves the estimation of multiple nuisance correlation parameters, and Westgate (2013b) showed that this can increase the finite-sample variances of corresponding regression parameter estimates from GEE, similar to the variance inflation that occurs with the QIF approach. As previously discussed, Westgate (2013a) further showed that this can also increase the degree of variance inflation for the typical QIF approach. Westgate (2013a) and Westgate (2013b) derived bias-corrected covariance formulas that take these variance inflations into account when utilizing an unstructured working correlation. Therefore, these formulas can be used with our proposed usage of the TECM, as already mentioned with the QIF approach, because they take into account, or penalize, the resulting variance inflation (Westgate, 2013a, submitted). We note that the use of the TECM, in addition to penalizing the unstructured working correlation, was proposed by Westgate (submitted) specifically for correlation selection with GEE.

Westgate (submitted) compared the TECM criterion with multiple other GEE correlation selection criteria. For instance, some well known criteria include the "quasi-likelihood under the independence model criterion" (QIC) (Pan, 2001), the "correlation information criterion" (CIC) (Hin and Wang, 2009), and criteria based on the work of Rotnitzky and Jewell (1990), such as is given in Hin et al. (2007). The results presented in Westgate (submitted) showed that the TECM and CIC perform similarly and are preferable relative to other criteria. In this manuscript, we extend the work of Westgate (submitted) by proposing the use of the TECM for selecting not just a working correlation structure, but also an analysis method. Although no other criteria have been proposed to select both the structure and analysis method, the results of Westgate (submitted) justify future work that extends the applicability of the CIC for this purpose. This CIC could then be compared with the TECM for simultaneous selection.

Extensions of the QIF approach that have been proposed involve the optimal combination of estimating equations derived from more than one working correlation structure, therefore resulting in more potential analysis options to select from using the proposed TECM criterion. For the interested reader, see Qu et al. (2008) and Leung et al. (2009). In short, these approaches have drawn theoretical interest because they asymptotically increase or maintain efficiency relative to the use of a single correlation structure. However, the use of additional structures will increase the dimension of C_N, which will have the opposite impact and increase the variances of finite-sample parameter estimates (Qu et al., 2008; Westgate and Braun, 2012). As expected, we found in simulation results (not shown) that our proposed TECM criterion also works when these QIF extensions are considered as analysis options. However, we also found that allowing such analysis options to be selected can be detrimental when they are not the best option under consideration. Therefore, future work is needed in terms of deciding the specific analysis method and correlation structure combinations that will be considered as options for the analysis of a particular dataset. Possible options to look into are the methods of Qu et al. (2008) and Zhou and Qu (2012), each of which take a different approach to select a working set of basis matrices. Ziegler and Vens (2010) also give useful comments on selecting an appropriate working correlation structure, and, as previously mentioned, it is reasonable to only consider plausible structures. We also note that the proposed criterion can be used with the approaches of Stoner and Leroux (2002) and Stoner et al. (2010), who optimally combine multiple sets of estimating equations that are derived using data contrasts rather than basis matrices.

As mentioned in the application example, Song et al. (2009) warned that the TGI, and therefore the TECM, may be a difficult criterion to utilize with GEE in the presence of outliers. This is the case since the empirical covariance matrix from GEE, which is used to obtain the TECM criterion value, may be biased in some form due to these outliers. Although previous work by Qu and Song (2004) has demonstrated the robustness of the typical QIF approach to outliers relative to GEE, future work is required to determine any possible impact outliers have on the TECM criterion when comparing different GEEs and typical and modified QIFs.

An alternative to the quadratic inference function approach to optimally combine multiple sets of equations is the use of empirical likelihood (EL) (Owen, 1988; Qin and Lawless, 1994). Leung et al. (2009) and Li and Pan (2013) both utilized EL to optimally combine multiple sets of estimating equations. Asymptotically, the resulting optimal estimating equations have the same form as those utilized by the QIF approach as given in Equation (4). However, to our knowledge, work is needed to derive the magnitude of any potential finite-sample variance inflation that may arise with the EL approach.

We have proposed a criterion that works well in terms of selecting an analysis method, GEE or either QIF approach, and working correlation structure. Although this criterion was shown to work well in terms of parameter estimation, it does not take into account the other theoretical advantages the QIF approach has over GEE. Specifically, the QIF can be used as a statistic in likelihood ratio score-type tests and to test for the correctness of the working mean structure (Qu et al., 2000; Song et al., 2009), and, as previously mentioned, this method is more robust to outliers than GEE.

Functions and code used for our simulation study and application are given in Supporting Information and can alternatively be obtained by contacting the author at philip.westgate@uky.edu. Included in the code are functions to carry out both versions of the QIF approach. The output includes, among other necessities, the bias-corrected SE estimates and the value for the proposed TECM criterion. The functions are based on the function given in Westgate (2012), which modified the QIF function that can be found at http://www-personal.umich.edu/p∼xsong/, Dr. Peter X.K. Song's website, and comes from the work of the first two authors of Song et al. (2009).

Acknowledgments We thank two anonymous reviewers for their constructive comments that helped improve this paper.

Conflict of interest

The author has declared no conflict of interest.

References

Carey, V. J. and Wang, Y.-G. (2011). Working covariance model selection for generalized estimating equations. *Statistics in Medicine* **30**, 3117–3124.

Gosho, M., Hamada, C. and Yoshimura, I. (2011). 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* **40**, 3839–3856.

Han, P. and Song, P. X.-K. (2011). A note on improving quadratic inference functions using a linear shrinkage approach. *Statistics and Probability Letters* **81**, 438–445.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

Hin, L.-Y., Carey, V. J. and Wang, Y.-G. (2007). Criteria for working-correlation-structure selection in GEE. *The American Statistician* **61**, 360–364.

Hin, L.-Y. and Wang, Y.-G. (2009). Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine* **28**, 642–658.

Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* **96**, 1387–1396.

Leung, D. H. Y., Wang, Y.-G. and Zhu, M. (2009). Efficient parameter estimation in longitudinal data analysis using a hybrid GEE method. *Biostatistics* **10**, 436–445.

Li, D. and Pan, J. (2013). Empirical likelihood for generalized linear models with longitudinal data. *Journal of Multivariate Analysis* **114**, 63–73.

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

Lindsay, B. G. (1982). Conditional score functions: some optimality results. *Biometrika* **69**, 503–512.
Lu, B., Preisser, J. S., Qaqish, B. F., Suchindran, C., Bangdiwala, S. I. and Wolfson, M. (2007). A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* **63**, 935–941.
Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57**, 126–134.
Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.
Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.
Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Annals of Statistics* **22**, 300–325.
Qu, A. and Lindsay, B. G. (2003). Building adaptive estimating equations when inverse of covariance estimation is difficult. *Journal of the Royal Statistical Society, Series B* **65**, 127–142.
Qu, A., Lee, J. J. and Lindsay, B. G. (2008). Model diagnostic tests for selecting informative correlation structure in correlated data. *Biometrika* **95**, 891–905.
Qu, A., Lindsay, B. G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.
Qu, A. and Song, P. X.-K. (2004). Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika* **91**, 447–459.
Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.
Shults, J., Sun, W., Tu, X., Kim, H., Amsterdam, J., Hilbe, J. M. and Ten-Have, T. (2009). A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine* **28**, 2338–2355.
Small, C. G., McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. Wiley, New York, NY.
Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer: New York, NY.
Song, P. X.-K, Jiang, Z., Park, E. and Qu, A. (2009). Quadratic inference functions in marginal models for longitudinal data. *Statistics in Medicine* **28**, 3683–3696.
Stoner, J. A. and Leroux, B. G. (2002). Analysis of clustered data: a combined estimating equations approach. *Biometrika* **89**, 567–578.
Stoner, J. A., Leroux, B. G. and Puumala, M. (2010). Optimal combination of estimating equations in the analysis of multilevel nested correlated data. *Statistics in Medicine* **29**, 464–473.
Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.
Wang, Y.-G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika* **90**, 29–41.
Westgate, P. M. (2012). A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine* **31**, 4003–4022.
Westgate, P. M. (2013a). A bias-corrected covariance estimator for improved inference when using an unstructured correlation with quadratic inference functions. *Statistics and Probability Letters* **83**, 1553–1558.
Westgate, P. M. (2013b). A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. *Statistics in Medicine* **32**, 2850–2858.
Westgate, P. M. (2013c). A comparison of utilized and theoretical covariance weighting matrices on the estimation performance of quadratic inference functions. *Communications in Statistics - Simulation and Computation*. Doi: 10.1080/03610918.2012.752839
Westgate, P. M. (submitted). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine*.
Westgate, P. M. and Braun, T. M. (2012). The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions. *Statistics in Medicine* **31**, 2209–2222.
Westgate, P. M. and Braun, T. M. (2013). An improved quadratic inference function for parameter estimation in the analysis of correlated data. *Statistics in Medicine* **32**, 3260–3273.
Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistical Association* **107**, 701–710.
Ziegler, A. and Vens, M. (2010). Generalized estimating equations: Notes on the choice of the working correlation matrix. *Methods of Information in Medicine* **49**, 421–425.

