



Published in final edited form as:

Am Stat. 2017 ; 71(4): 344–353. doi:10.1080/00031305.2016.1200490.

A Comparison of Correlation Structure Selection Penalties for Generalized Estimating Equations

Philip M. Westgate^{*} and

Department of Biostatistics, College of Public Health, University of Kentucky

Woodrow W. Burchett

Department of Statistics, College of Arts and Sciences, University of Kentucky

Abstract

Correlated data are commonly analyzed using models constructed using population-averaged generalized estimating equations (GEEs). The specification of a population-averaged GEE model includes selection of a structure describing the correlation of repeated measures. Accurate specification of this structure can improve efficiency, whereas the finite-sample estimation of nuisance correlation parameters can inflate the variances of regression parameter estimates. Therefore, correlation structure selection criteria should penalize, or account for, correlation parameter estimation. In this manuscript, we compare recently proposed penalties in terms of their impacts on correlation structure selection and regression parameter estimation, and give practical considerations for data analysts.

Keywords

Bias-correction; Efficiency; Empirical covariance matrix; Longitudinal data

1 Introduction

The need to analyze correlated data arises in a variety of research settings. In general, we have N independent clusters, with the i th cluster contributing n_i , $i = 1, \dots, N$, correlated outcomes denoted by $Y_i = [Y_{i1}, \dots, Y_{in_i}]^T$. A common example is correlated repeated measures collected in longitudinal studies in which independent subjects contribute outcomes over time. If a marginal model is desired, generalized estimating equations (GEE) (Liang and Zeger, 1986) are commonly used for the data analysis. Although working marginal variances and a correlation structure must be selected, this approach often yields valid statistical inference as long as the mean structure is correct.

Examples of popular working correlation structures include independence, exchangeable, AR-1, the least parsimonious Toeplitz, and unstructured. For simplicity of example, suppose each subject in a balanced longitudinal study contributes n repeated measures. We note,

^{*}The authors gratefully acknowledge *the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR000117.*

however, that this is not a requirement and the results of this manuscript are generalizable for an unequal number of measurements per subject. With these working correlation matrices, the element in the j th row and k th column is equal to the working form for $Corr(Y_{ij}, Y_{ik})$. If $j = k$, $Corr(Y_{ij}, Y_{ik}) = 1$. Otherwise, exchangeable and AR-1 each use one parameter, α , such that $Corr(Y_{ij}, Y_{ik}) = \alpha$ and $Corr(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}$, respectively, whereas $Corr(Y_{ij}, Y_{ik}) = 0$ with independence. For the least parsimonious Toeplitz version, denote the $n - 1$ correlation parameters by α_l , $l = 1, \dots, n - 1$, such that $Corr(Y_{ij}, Y_{ik}) = \alpha_l$ for $l = |j - k|$. For unstructured, each distinct correlation element is allowed to differ such that $Corr(Y_{ij}, Y_{ik}) = \alpha_{jk}$, resulting in the need to estimate $n(n-1)/2$ different correlation parameters.

To improve efficiency for the estimation of regression parameters, accurate modeling of the working correlation structure is desired (Liang and Zeger, 1986; Wang and Carey, 2003). Therefore, multiple correlation selection criteria have been proposed. As a quick reference, we refer the reader to Westgate (2014), who summarized and studied the performances of many criteria. In short, the ‘correlation information criterion’ (CIC) (Hin and Wang, 2009), gaussian pseudolikelihood (GP) (Carey and Wang, 2011), and the ‘trace of the empirical covariance matrix’ (TECM) (Westgate, 2014) work well.

Unfortunately, the estimation of any nuisance correlation parameters can increase the finite-sample variances of regression parameter estimates (Westgate, 2013, 2015). This increase in variances was shown with the unstructured working correlation matrix in Westgate (2013), and an approximation for the covariance inflation was derived. Furthermore, Westgate (2015) demonstrated variance inflation and the utility of the covariance inflation approximation when GEE incorporates structured working correlation matrices. Due to this variance inflation the estimation of correlation parameters should be accounted for, or penalized, when selecting a working structure. Besides comparing the performances of many correlation structure selection criteria, Westgate (2014) showed that utilizing Westgate’s (2013) covariance inflation approximation as a penalty to be applied toward the unstructured working correlation matrix can improve the performances of criteria. Furthermore, Westgate (2015) showed that use of this particular penalty can be applied toward any working correlation structure, thus improving selection accuracy even further. We note that Westgate’s penalty can be applied with any criterion that incorporates an empirical estimate for the covariance matrix of the regression parameter estimates. Without such a penalty, selection criteria can overselect less parsimonious correlation structures such as the unstructured and least parsimonious Toeplitz matrices, potentially resulting in less precise regression parameter estimation (Barnett et al., 2010; Hardin and Hilbe, 2012; Shults and Hilbe, 2014; Westgate, 2014, 2015). This position of Westgate (2014, 2015) on the need to utilize correlation structure selection penalties is the position we take in this manuscript.

Other correlation structure selection penalties have recently been proposed that were not studied by Westgate (2014, 2015), and that have not been directly compared in terms of their advantages, disadvantages, and ultimately their impact on selection accuracy. Specifically, Hardin and Hilbe (2012) and Shults and Hilbe (2014) proposed penalties for use with the

well-known ‘quasi-likelihood under the independence model criterion’ (QIC) (Pan, 2001), which we later extend for use with the CIC. Also, penalties based on the AIC (Akaike, 1974) and Bayesian information criterion (BIC) (Schwarz, 1978) have been proposed for use with both the GP criterion (Xiaolu and Zhongyi, 2013) and a method proposed by Chen and Lazar (2012) that utilizes empirical likelihood (EL) (Owen, 1988; Qin and Lawless, 1994).

Our focus in this manuscript is therefore to compare the previously mentioned penalties in terms of their practical implications on correlation structure selection and ultimately regression parameter estimation. In Section 2, we briefly describe GEE and relevant correlation selection criteria. In Section 3, we discuss properties of different penalties for data analysts to consider. Penalties and corresponding selection criteria are then contrasted via a simulation study in Section 4 and an application example in Section 5. Concluding remarks, including practical recommendations, are given in Section 6.

2 GEE and Correlation Selection Criteria

2.1 Generalized Estimating Equations

Established in (Liang and Zeger, 1986), a consistent estimate of the regression parameters, $\beta = [\beta_0, \beta_1, \dots, \beta_{p-1}]^T$, is obtained by solving $\sum_{i=1}^N D_i^T V_i^{-1} (Y_i - \mu_i) = \sum_{i=1}^N D_i^T A_i^{-1/2} R_i^{-1}(\alpha) A_i^{-1/2} (Y_i - \mu_i) = \mathbf{0}$. Here, for the i th cluster, $D_i = \partial \mu_i / \partial \beta^T$, $V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$ is the working covariance structure for Y_i , $R_i(\alpha)$ is the working correlation matrix for Y_i composed of parameters given by α , A_i is a diagonal matrix of working marginal variances for Y_i and $E(Y_i) = \mu_i$ with link function f such that $f(\mu_{ij}) = X_{ij}^T \beta$ for $x_{ij} = [1, x_{1ij}, \dots, x_{(p-1)ij}]^T$. Assuming the nuisance correlation parameters are known, $Cov(\hat{\beta}) \approx$

$$\Sigma = \left(\sum_{i=1}^N D_i^T V_i^{-1} D_i \right)^{-1} \left(\sum_{i=1}^N D_i^T V_i^{-1} Cov(Y_i) V_i^{-1} D_i \right) \left(\sum_{i=1}^N D_i^T V_i^{-1} D_i \right)^{-1}. \quad (1)$$

We denote a consistent estimator for Σ as $\hat{\Sigma}$. For instance, this could be the popular Liang and Zeger (1986) empirical, or robust, estimator which replaces $Cov(Y_i)$ with $(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T, i = 1, \dots, N$, in Equation (1). In small-sample settings, this estimator can be biased because the residual matrices incorporated within these empirical covariances tend to be too small (Mancl and DeRouen, 2001). Therefore, multiple corrections for this bias have been proposed. For instance, see Kauermann and Carroll (2001); Mancl and DeRouen (2001); Fay and Graubard (2001); Morel et al. (2003); McCaffrey and Bell (2006); Fan et al. (2013). We also note that the correlation matrices may fail to be consistent if they do not model the true structure, in which case these empirical covariance estimators may also fail to be consistent (Crowder, 1995; Sutradhar and Das, 1999).

2.2 Correlation Selection Criteria

We give focus to the CIC, TECM, and GP for correlation structure selection because these criteria were found by Westgate (2014) to notably outperform the well-known QIC (Pan, 2001) and criteria motivated by the work of Rotnitzky and Jewell (Rotnitzky and Jewell, 1990; Hin et al., 2007; Shults et al., 2009; Carey and Wang, 2011). We also give focus to criteria based upon empirical likelihood versions of the AIC and BIC that were proposed by Chen and Lazar (2012). The CIC is taken from the QIC. The unpenalized QIC for a given working correlation structure, R , is given by

$QIC(R) = -2Q(\beta, \phi) + 2tr(\widehat{\Sigma}_I^{-1} \widehat{\Sigma}) = -2Q(\beta, \phi) + 2CIC(R)$. Here, $Q(\beta, \phi)$ is the quasi-likelihood under the assumption that all outcomes are independent, and

$\widehat{\Sigma}_I = (\sum_{i=1}^N \mathbf{D}_i^T \mathbf{A}_i^{-1} \mathbf{D}_i)^{-1}$ is the model-based covariance matrix assuming independence. To improve upon the performance of the QIC, Hin and Wang (2009) proposed using the CIC, as only this last part of the QIC contains information about the correlation structure. The unpenalized TECM and GP are given by $TECM(R) = tr(\widehat{\Sigma})$ and

$GP(R) = \sum_{i=1}^N [(Y_i - \mu_i)^T V_i^{-1} (Y_i - \mu_i) + \log(|V_i|)]$, respectively. The structure that yields the smallest value for the given criterion is selected. We note that this form for the GP is based on the setup of Xiaolu and Zhongyi (2013), and $\widehat{\Sigma}$ is utilized within each of these criteria except for the GP. As the Chen and Lazar (2012) criteria require a penalty, we discuss them in detail in the following section.

3 Correlation Selection Penalties

3.1 Penalties of Hardin & Hilbe and Shults & Hilbe

Hardin and Hilbe (2012) and Shults and Hilbe (2014) proposed penalized versions of the QIC given by

$$-2Q(\beta, \phi) + 2CIC(R) + 2 \frac{[p+r+m][p+r+m+1]}{N-p-r-m-1} = QIC(R) + 2P(p, r, m, N) \quad (2)$$

in which $m=0$ and $m=1$ for the Hardin and Hilbe (2012) and Shults and Hilbe (2014) penalties, respectively, and r is the number of estimated correlation parameters. These penalties are motivated by the adjusted Akaike information criterion (AIC) of Hurvich and Tsai (1989), who proposed adding a similar penalty to the AIC as a small-sample correction factor. The CIC has been shown to perform better than the QIC with respect to correlation selection, and therefore we note that these penalties can also be incorporated within the CIC. Specifically, $-2Q(\beta, \phi)$ in Equation (2) can be ignored, and multiplying by 2 has no impact. Therefore, the penalized CIC is given by $CIC(R) + P(p, r, m, N)$.

As N increases, any inflation of the variances of the regression parameter estimates, and therefore the needed penalty value, reduces in magnitude because correlation parameters are estimated more precisely. The use of these penalties is therefore intuitive because $P(p, r, m, N) \rightarrow 0$ as $N \rightarrow \infty$, given $p, r \leq c < \infty$ for some constant c . However, this penalty could still use greater theoretical justification (Shults and Hilbe, 2014). Furthermore, Hardin

and Hilbe (2012) and Shults and Hilbe (2014) found their penalties, when used with the QIC, tended to favor simpler structures and over-select independence, particularly for smaller N in the Hardin and Hilbe (2012) study. In short, $P(p, r, m, N)$ can over-penalize, although this overpenalization diminishes as N increases.

In practice, the data analyst should consider that $P(p, r, m, N)$ may favor structures with fewer, or no, correlation parameters. This property can be advantageous when simpler structures, such as exchangeable or AR-1, perform better than less parsimonious structures. However, this property can be disadvantageous if the true structure is not independence and ignoring the correlation leads to a loss in small sample efficiency, such as described in Fitzmaurice (1995), Mancl and Leroux (1996), and Wang and Carey (2003). Therefore, the analyst may not want to consider independence for selection in some instances. We also point out that $P(p, r, m, N)$ has a practical advantage in that it is simple to calculate and therefore does not need to be incorporated within existing software in order to be easily used.

3.2 Westgate Penalty

The need to penalize correlation parameter estimation arises because this estimation potentially increases the estimation variability of GEE, thus inflating $Cov(\hat{\beta})$ (Westgate, 2013, 2015). Specifically, Westgate (2013) showed that when accounting for correlation estimation, $Cov(\hat{\beta}) \approx (\mathbf{I}_p + \mathbf{G}) \boldsymbol{\Sigma} (\mathbf{I}_p + \mathbf{G})^T$, where \mathbf{I}_p is a $p \times p$ identity matrix and

$$\mathbf{G} = (\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_{p-1}),$$

$$\mathbf{G}_r = - \left(\sum_{i=1}^N \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \sum_{i=1}^N \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{R}_i^{-1} \frac{\partial \mathbf{R}_i(\hat{\alpha}(\beta))}{\partial \beta_r} \mathbf{R}_i^{-1} \mathbf{A}_i^{-1/2} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\beta)).$$

We note that although Westgate (2013) focused on the use of an unstructured working correlation matrix, Westgate (2015) demonstrated that $Cov(\hat{\beta}) \approx (\mathbf{I}_p + \mathbf{G}) \boldsymbol{\Sigma} (\mathbf{I}_p + \mathbf{G})^T$ for any working correlation structure given by \mathbf{R}_i , $i = 1, \dots, N$, within \mathbf{G} . Therefore, Westgate (2014, 2015) proposed penalizing the estimation of correlation parameters by using $(\mathbf{I}_p + \hat{\mathbf{G}}) \hat{\boldsymbol{\Sigma}} (\mathbf{I}_p + \hat{\mathbf{G}})^T$ in place of $\hat{\boldsymbol{\Sigma}}$, the estimator that had been historically utilized to estimate $Cov(\hat{\beta})$ within correlation selection criteria. As a result, versions of the CIC and TECM that incorporate this penalty are given by $CIC_w(R) = tr \left[\hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{I}_p + \hat{\mathbf{G}}) \hat{\boldsymbol{\Sigma}} (\mathbf{I}_p + \hat{\mathbf{G}})^T \right]$ and $TECM_w(R) = tr \left[(\mathbf{I}_p + \hat{\mathbf{G}}) \hat{\boldsymbol{\Sigma}} (\mathbf{I}_p + \hat{\mathbf{G}})^T \right]$ respectively. We note that $\hat{\mathbf{G}}$ is an estimate for \mathbf{G} in which $\boldsymbol{\beta}$ is replaced with $\hat{\boldsymbol{\beta}}$.

Westgate's penalty has a natural advantage over other penalties in that it directly accounts for the inflation of $Cov(\hat{\beta})$ that arises from the estimation of nuisance correlation parameters. If we were able to show $Cov(\hat{\beta}) = (\mathbf{I}_p + \mathbf{G}) \boldsymbol{\Sigma} (\mathbf{I}_p + \mathbf{G})^T$, and then utilize $(\mathbf{I}_p + \mathbf{G}) \boldsymbol{\Sigma} (\mathbf{I}_p + \mathbf{G})^T$ within correlation selection criteria, then this penalty would be ideal. Unfortunately, this is not the case. Specifically, $Cov(\hat{\beta}) \approx (\mathbf{I}_p + \mathbf{G}) \boldsymbol{\Sigma} (\mathbf{I}_p + \mathbf{G})^T$, and so we can

only approximate the covariance inflation. Furthermore, Westgate's penalty is an estimate, as \hat{G} must be used in place of G . As N decreases, the magnitude of the covariance inflation can increase, whereas the precision of \hat{G} decreases. Therefore, Westgate's penalty can be less reliable for smaller N , which is when the penalty is needed most. For instance, a resulting consequence can be that Westgate's penalty sometimes selects less parsimonious structures simply because the corresponding covariance inflation is underestimated. Although this selection may be advantageous when less parsimonious structures lead to the least variable parameter estimates, it will be disadvantageous when simpler structures work best.

3.3 Penalties for Empirical Likelihood and Gaussian Pseudolikelihood

The method proposed by Chen and Lazar (2012) uses EL forms of the AIC and BIC given by $EAIC(R) = -2\log ELR(\hat{\beta}, \hat{\alpha}) + 2(p+r)$ and $EBIC(R) = -2\log ELR(\hat{\beta}, \hat{\alpha}) + (p+r)\log(N)$, respectively. With this approach, a maximal correlation structure is assumed, and this structure and nested structures are possibly assumed to be under consideration for selection. In short, any given structure under consideration is compared with the maximal structure via the empirical likelihood ratio (ELR). Although Chen and Lazar (2012) assumed this maximal structure to be the stationary, or least parsimonious Toeplitz correlation matrix, we feel this is too restrictive relative to other criteria. In this manuscript, we use the unstructured working correlation as the maximal model. We refer the reader to Chen and Lazar (2012) for technical details. Xiaolu and Zhongyi (2013) took the same penalization approach and applied it with the GP criterion, resulting in penalized versions given by $AGP(R) = GP(R) + 2(p+r)$ and $BGP(R) = GP(R) + (p+r)\log(N)$. We note that Fu et al. (2015) also utilized this penalty when focusing on quantile regression.

Penalties are either $2(p+r)$ or $(p+r)\log(N)$. Although these seem intuitive with respect to the form of the AIC and BIC, they are not truly representative of the need for correlation penalization. Specifically, the magnitude of the penalty should decrease as N increases, whereas $2(p+r)$ does not depend on N and $(p+r)\log(N)$ gives a harsher penalty toward less parsimonious structures as N increases. However, because EL and GP do not incorporate an estimate for $Cov(\hat{\beta})$, the impacts of these penalties are not as clear as for the previously discussed penalties. We note that the simulation study of Xiaolu and Zhongyi (2013) potentially suggests that these penalties may overly favor simple structures. However, we will show in our simulation study that the impact of these penalties rely upon the utility of the given method, EL or GP.

4 Simulation Study

4.1 Study Description

We now demonstrate in a simulation study the impacts of the different penalties, and inherently the criteria that incorporate these penalties, on correlation structure selection frequencies and ultimately regression parameter estimation. Because penalties may overselect a certain type of structure, depending on parsimony, we gave focus to three scenarios. In each scenario, we assumed the analyst was interested in selecting one of multiple working structures. In one scenario, interest was in independence, exchangeable,

AR-1, Toeplitz, and unstructured. In another scenario, these same structures except for independence were considered, avoiding over-selection of independence. Finally, only independence, exchangeable, and AR-1 were considered, avoiding over-selection of the least parsimonious structures.

Data were generated from a multivariate normal distribution such that $E(Y_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$; $j = 1, \dots, n$, where $\beta = [0, 0.3, 0.3]^T$, outcomes had unit marginal variance, and both covariates were independently generated across all observations from $Uniform(0, 1)$. We note that this model setup is similar to settings used in Hin et al. (2007), Hin and Wang (2009), and Westgate (2015). The true structures were either AR-1 with correlation value of 0.5 or unstructured with $Corr(Y_{i1}, Y_{i2}) = 0.8$, $Corr(Y_{i1}, Y_{i3}) = 0.3$, and $Corr(Y_{i2}, Y_{i3}) = 0.6$. Corresponding to true AR-1 settings, either 20 or 100 subjects each contributed 4 repeated measurements. Alternatively, to ensure stable empirical standard deviations (ESDs) were produced by the unstructured working correlation matrix, we only present results for settings in which 100 subjects each contributed 3 repeated measurements when unstructured was the true structure. Simulations were conducted in R version 2.13.1 (R Development Core Team, 2011), and outcomes were generated using

```
rmvnorm
```

of the

```
mvtnorm
```

package (Genz et al., 2013; Genz and Bretz, 2009)

In Tables 1–3, we present the number of times out of 1,000 replications that each working structure under consideration was selected. These frequencies are given for each scenario and each of the following selection criteria and penalties: the TECM criterion with Westgate penalty ($TECM_W$), the CIC with Westgate penalty (CIC_W), the CIC with Hardin and Hilbe penalty (CIC_{HH}), the CIC with Shults and Hilbe penalty (CIC_{SH}), the EAIC and EBIC with unstructured as the maximal structure, and the AGP and BGP. ESDs of $\hat{\beta}_2$ are also presented. We note that results with respect to $\hat{\beta}_1$ were very similar because values for both covariates were generated in the same manner.

Direct comparison of penalty values is only meaningful when these penalties are used with the same criterion. Three different penalties are available with the CIC. Therefore, the Westgate, Hardin and Hilbe, and Shults and Hilbe penalties are explicitly demonstrated in Table 4. Specifically, empirical means of the unpenalized CIC, CIC_W , CIC_{HH} , and CIC_{SH} are presented. Furthermore, the empirical mean of Westgate's penalty, along with the Hardin and Hilbe and Shults and Hilbe penalties, are given and are equal to the difference between the empirical means of the corresponding penalized and unpenalized CIC. We note that the mean of Westgate's penalty should be approximately equal to the optimal penalty. We utilized the empirical covariance matrix estimator that incorporates the Kauermann and Carroll (2001) correction for $\hat{\Sigma}$ within selection criteria, as it has been shown to work well

with respect to attaining valid inference (Westgate, 2015). However, other available corrections could alternatively be utilized. Fortunately, the performances of selection criteria are not sensitive to which of these corrections, if any, are utilized (Westgate, 2015).

4.2 Description of Results

As expected, using $P(p, r, m, N)$ as a penalty with the CIC led to a high selection frequency of the independence structure when $N = 20$. Furthermore, because the Shults and Hilbe penalty is greater than the Hardin and Hilbe penalty, it selected independence even more often. ESDs show that independence is a poor choice in this setting. Therefore, it is advantageous to not even consider independence for selection here, in which case the CIC with these penalties actually works best. Alternatively, when $N = 100$, use of the CIC with these penalties selected independence one time at most. Furthermore, use of these penalties never resulted in unstructured being selected when AR-1 was the true structure, and Toeplitz was selected only twice by the Hardin and Hilbe penalty. Alternatively, when unstructured was the true structure, Toeplitz and unstructured resulted in the smallest ESDs, and use of both penalties resulted in either of these two structures being selected most often. However, they both favored the simpler Toeplitz structure.

As found in Westgate (2014, 2015), his penalized versions of the TECM and CIC worked similarly, although selection frequencies were notably different when unstructured was the true structure. Use of these criteria and penalties, relative to other criteria and penalties with the exception of the EL methods, tended to select Toeplitz and unstructured more often when AR-1 was the true structure. This may be viewed as over-selection of less parsimonious structures. When $N = 20$, these structures did not perform as well, i.e. ESDs for $\hat{\beta}_2$ were larger, as AR-1 due to notable variance inflation from the estimation of additional correlation parameters. Therefore, not allowing the TECM and CIC with Westgate's penalty to select these two structures resulted in slightly smaller ESDs. However, when $N = 100$, the variance inflation is much smaller, and considering these structures was not detrimental.

Table 4 shows that, when incorporated with the CIC, the Hardin and Hilbe and Shults and Hilbe penalties are larger than the average Westgate penalty. We point out that the former penalties actually have a penalty value for independence, whereas the Westgate penalty is 0 because no covariance inflation occurs with this structure. Therefore, penalty comparisons must take this into account. For instance, Westgate's average penalty toward the exchangeable structure is 0.06 when $N = 20$. Alternatively, the Hardin and Hilbe (Shults and Hilbe) penalties toward independence and exchangeable are 0.75 (1.33) and 1.33 (2.14), respectively, for a difference of 0.58 (0.81), which is notably larger than the 0.06 difference with the Westgate penalty. Therefore, the Hardin and Hilbe and Shults and Hilbe penalties result in the over-selection of independence, and simpler structures in general.

In short, relative to the Hardin and Hilbe and Shults and Hilbe penalties, Westgate's penalty resulted in smaller ESDs when $N = 20$ and independence was allowed to be selected. However, when independence was not considered, but Toeplitz and unstructured were options, these former penalties resulted in slightly smaller ESDs. Finally, when $N = 100$, no notable differences were observed in ESDs resulting from the use of these three penalties.

The EAIC and EBIC over-selected Toeplitz and unstructured when $N = 20$, and corresponding ESDs were unacceptably higher than ESDs produced from the use of any other criterion and penalty. Alternatively, not allowing Toeplitz and unstructured to be selected by the EAIC and EBIC reduced ESDs to acceptable levels. Furthermore, when $N = 100$, selecting less parsimonious structures was not detrimental, and no notable differences in ESDs were notable across criteria and penalties.

The AGP and BGP worked very well and similarly to the TECM and CIC with Westgate's penalty in terms of resulting ESDs. When AR-1 was the true structure, the AGP and BGP did not over-select independence, Toeplitz, or unstructured, and they had high frequencies of correctly selecting AR-1. Alternatively, when unstructured was the true structure, the AGP and BGP selected Toeplitz and unstructured most often, as did the other criteria and penalties.

5 Application

We now demonstrate the penalties in the context of the Prevention of Alzheimer's Disease by Vitamin E and Selenium (PREADViSE) clinical trial (Caban-Holt et al., 2012). We have yearly global cognitive status data, measured using the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) T-score (Chandler et al., 2005; Mathews et al., 2013; Caban-Holt et al., 2012), from fifty men. Larger scores indicate higher cognitive status. Each subject came in for an annual assessment for either three or four sequential years. The number of assessments simply depended on when the subject was recruited into the study. Available predictors of the continuous T-score include baseline age and estimated full-scale IQ (Boekamp et al., 1995).

Our working marginal model, based on the application of Westgate (2014), centers age and IQ at values near their corresponding sample medians and means. Specifically,

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

where $n_i = 3$ or 4 , $time_{ij} = j - 1$ is years since baseline, and Y_{ij} is the i th subject's T-score in their j th year.

Table 5 presents results from using independence, exchangeable, AR-1, Toeplitz, and unstructured working correlation matrices with GEE to fit the above model. Specifically, for each working structure, parameter estimates, their corresponding estimated standard errors, and penalized correlation selection criterion values are presented. We also give unpenalized values for the TECM and CIC to illustrate the degree of penalizations. We note that because some subjects contribute only three observations we are unable to use the EL criteria.

The TECM and CIC, when not penalizing for correlation estimation, select unstructured. We note that Toeplitz could also have been chosen by the unpenalized TECM. For these criteria Westgate's penalty is small for exchangeable and AR-1 and increases in magnitude for Toeplitz and especially for unstructured. The reason for this is because the estimated covariance inflation was negligible (notable) when estimating one (multiple) correlation parameter(s). We note that Westgate's penalty is more apparent with the CIC. When utilizing

this penalty, the TECM and CIC select Toeplitz and AR-1, respectively. However, this does not result in a practical problem, as use of these structures resulted in similar values for $\hat{\beta}$ and corresponding SE estimates. The Hardin and Hilbe and Shults and Hilbe penalties with the CIC are larger than Westgate's penalties, as expected. Furthermore, they select independence. Alternatively, if we did not consider independence for selection, then these penalties select AR-1. Finally, the AGP and BGP both select exchangeable.

We point out that our analyses of this dataset extend the analyses of Westgate (2014). First, Westgate (2014) did not apply the Kauermann and Carroll (2001) correction because $N = 50$ is not small and in order to stay consistent with previous articles on correlation selection that did not use this correction. However, we do apply this correction because we found that it had small but notable impact on some standard error estimates. Second, the current manuscript allowed Toeplitz to be considered for selection, whereas Westgate (2014) did not. Although Westgate (2014) concluded that the AR-1 structure may be preferred, this manuscript demonstrates that Toeplitz is also a viable choice when utilizing the TECM with Westgate's penalty. Finally, we point out that Westgate's TECM and CIC penalties for the exchangeable and AR-1 structures were negligible and therefore not actually needed in this setting.

6 Concluding Remarks

When estimating a population-averaged GEE model, a working correlation structure must be selected. To do so, a criterion that penalizes the estimation of correlation parameters can be utilized. The goal of this manuscript was to describe and contrast different penalties, and inherently the criteria that can incorporate these penalties, in order to make data analysts aware of important aspects to consider as to which penalty and criterion combination will ultimately result in the least variable regression parameter estimates via accurate correlation structure selection. In summary, the penalized Gaussian Pseudolikelihood criteria work well, as do the TECM and CIC with Westgate's penalty. Furthermore, if the analyst is willing to not consider independence as a selection option due to its potential inefficiency, the Hardin and Hilbe (2012) and Shults and Hilbe (2014) penalties with the CIC can also work well, and these penalties have a practical advantage in that they can be easily calculated and used regardless of the software that is employed. Finally, the empirical likelihood criteria of Chen and Lazar (2012), when utilizing unstructured as the maximal structure, were found to not work well when the number of subjects was small.

As seen in our simulation study, the smaller the number of independent clusters, or subjects in a longitudinal study, the more notable the difference between distinct penalty and criteria combinations in terms of the resulting variances of estimated regression parameters.

Alternatively, if the sample size is not small, differences may be negligible. However, what constitutes a small or large sample size may not be clear, and the analyst should consider the implications of different criteria and penalties.

In our experience with small-sample settings, we have had difficulty finding true Toeplitz or unstructured correlation matrices for generating outcomes in simulation studies such that the incorporation of corresponding working structures within GEE results in smaller variances

of regression parameter estimates relative to the incorporation of simpler structures such as exchangeable or AR-1. For instance, the AR-1 working structure resulted in the smallest empirical standard deviations of regression parameter estimates in simulation results (not shown) from the setting in which the true structure was unstructured, but for $N = 20$. The reason for this result is because notable covariance inflation arises from the estimation of the nuisance correlation parameters, and therefore it is difficult to find a realistic, non-parsimonious structure that has an efficiency gain that more than offsets the corresponding variance inflation in small-sample settings. Therefore, if N is small, the analyst may want to over-penalize, or even not consider, structures such as Teoplitz or unstructured. To over-penalize, we recommend the use of the Hardin and Hilbe or Shults and Hilbe penalty with the CIC, or use one of these penalties in addition to Westgate's penalty. We briefly note that the relative performances, in terms of resulting empirical standard deviations of regression parameter estimates, of the different criteria and penalty combinations from the simulation setting in which the true structure was unstructured and $N = 20$ were similar to those that were observed when the true structure was AR-1 and $N = 20$.

We note that the inflation of the variances of regression parameter estimates is often negligible for structures such as AR-1 and exchangeable that require only one correlation parameter to be estimated, as can be seen in the simulation study of Westgate (2015). As a result, Westgate's penalty will often be negligible with such structures, as is apparent in Tables 4 and 5. As a practical result, there often is no need to penalize one parameter correlation structures except possibly for when N is very small (Westgate, 2015). However, the definition of very small is vague here, and ideally the data analyst will still apply an appropriate penalty method which can determine an adequate penalty value. Furthermore, the use of one parameter working correlation structures can result in the need to estimate multiple correlation parameters, for instance, when the analyst allows correlation parameters to vary across groups such as trial arms. In such a case, use of a penalty may very well be beneficial.

Our use of correlation selection criteria and penalties focuses on obtaining the least variable regression parameter estimates. However, other aspects should be considered. Therefore, we refer the reader to Ziegler and Vens (2010) for an informative discussion. Furthermore, we suggested that the analyst may want to consider disregarding independence as a selection option, as it is often no more or less efficient relative to other working structures. We note that independence is a special case of all other structures, and if the estimated correlation values are approximately zero, then independence is still available to the analyst. In situations where independence cannot lose a notable amount of efficiency, an alternative argument may be to simply use independence and ignore any other possible structures (Dahmen and Ziegler, 2004). We refer the reader to articles such as Fitzmaurice (1995), Mancl and Leroux (1996), Wang and Carey (2003), and Dahmen and Ziegler (2004) for details. Additionally, there are situations in which independence should be used, or at least preferred, and we refer the reader to the manuscript by Dahmen and Ziegler (2004) that summarizes such findings.

Acknowledgements

We would like to thank the anonymous associate editor and two reviewers for their constructive comments that helped improve this paper. This publication was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR000117. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would like to thank Dr. Richard J. Kryscio, Dr. Frederick A. Schmitt, and Dr. Erin Abner for allowing us to use the data from the PREADViSE trial, which was supported through a National Institute on Aging grant (R01 AG019241). We also thank Dr. Abner for providing us with the dataset and relevant information.

References

- Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716–723.
- Barnett AG, Koper N, Dobson AJ, Schmiegelow F, and Manseau M (2010). Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution* 1, 15–24.
- Boekamp JR, Strauss ME, and Adams N (1995). 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* 17, 645–653. [PubMed: 8557806]
- Caban-Holt A, Abner E, Kryscio RJ, Crowley JJ, and Schmitt FA (2012). Age-expanded normative data for the Ruff 2&7 Selective Attention Test: evaluating cognition in older males. *The Clinical Neuropsychologist* 26, 751–768. [PubMed: 22651854]
- Carey VJ and Wang Y-G (2011). Working covariance model selection for generalized estimating equations. *Statistics in Medicine* 30, 3117–3124. [PubMed: 21748775]
- Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Weiner MF, and Cullum CM (2005). A total score for the cerad neuropsychological battery. *Neurology* 65, 102–106. [PubMed: 16009893]
- Chen J and Lazar NA (2012). Selection of working correlation structure in generalized estimating equations via empirical likelihood. *Journal of Computational and Graphical Statistics* 21, 18–41.
- Crowder M (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* 82, 407–410.
- Dahmen G and Ziegler A (2004). Generalized estimating equations in controlled clinical trials: hypothesis testing. *Biometrical Journal* 46, 214–232.
- Fan C, Zhang D, and Zhang C-H (2013). A comparison of bias-corrected covariance estimators for generalized estimating equations. *Journal of Biopharmaceutical Statistics* 23, 1172–1187. [PubMed: 23957522]
- Fay MP and Graubard BI (2001). Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics* 57, 1198–1206. [PubMed: 11764261]
- Fitzmaurice GM (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* 51, 309–317. [PubMed: 7766784]
- Fu L, Wang Y-G, and Zhu M (2015). A gaussian pseudolikelihood approach for quantile regression with repeated measurements. *Computational Statistics and Data Analysis* 84, 41–53.
- Genz A and Bretz F (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics Heidelberg: Springer-Verlage. ISBN 978–3-642–01688-2.
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, and Hothorn T (2013). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9–9995 Available from: <http://CRAN.R-project.org/package=mvtnorm>.
- Hardin J and Hilbe J (2012). *Generalized Estimating Equations, Second Edition*. Chapman and Hall/CRC Press, Boca Raton, FL, USA.
- Hin L-Y, Carey VJ, 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. [PubMed: 19065625]

- Hurvich CM and Tsai C-L (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
- Kauermann G and Carroll RJ (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 96, 1387–1396.
- Liang K-Y and Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.
- Mancl LA and DeRouen TA (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* 57, 126–134. [PubMed: 11252587]
- Mancl LA and Leroux BG (1996). Efficiency of regression estimates for clustered data. *Biometrics* 52, 500–511. [PubMed: 10766502]
- Mathews M, Abner E, Caban-Holt A, Kryscio R, and Schmitt F (2013). CERAD practice effects and attrition bias in a dementia prevention trial. *International Psychogeriatrics* 25, 1115–1123. [PubMed: 23570673]
- McCaffrey DF and Bell RM (2006). Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clusters. *Statistics in Medicine* 25, 4081–4098. [PubMed: 16456895]
- Morel JG, Bokossa MC, and Neerchal NK (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal* 45, 395–409.
- Owen AB (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. [PubMed: 11252586]
- Qin J and Lawless J (1994). Empirical likelihood and general estimating functions. *Annals of Statistics* 22, 300–325.
- R Development Core Team (2011). R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing ISBN 3–900051–07–0. Available from: <http://www.R-project.org>.
- Rotnitzky A and Jewell NP (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 77, 485–497.
- Schwarz G (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.
- Shults J and Hilbe JM (2014). *Quasi-Least Squares Regression*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, Boca Raton, FL.
- Shults J, Sun W, Tu X, Kim H, Amsterdam J, Hilbe JM, 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. [PubMed: 19472307]
- Sutradhar BC and Das K (1999). On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika* 86, 459–465.
- 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 PM (2013). 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. [PubMed: 23255154]
- Westgate PM (2014). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine* 33, 2222–2237. [PubMed: 24504841]
- Westgate PM (2015). A covariance correction that accounts for correlation estimation to improve finite-sample inference with generalized estimating equations: a study on its applicability with structured correlation matrices. *Journal of Statistical Computation and Simulation* doi 10.1080/00949655.2015.1089873
- Xiaolu Z and Zhongyi Z (2013). Comparison of criteria to select working correlation matrix in generalized estimating equations. *Chinese Journal of Applied Probability and Statistics* 29, 515–530.

Ziegler A and Vens M (2010). Generalized estimating equations: notes on the choice of the working correlation matrix. *Methods Inf Med* 49, 421–425. [PubMed: 20871939]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Correlation structure selection frequencies and empirical standard deviations(ESD) of $\hat{\beta}_2$ for the setting in which the true structure is AR-1 ($\rho = 0.5$) and 20 subjects each contribute 4 observations.

Structures of Interest	Criterion	Selection Frequencies					Criterion ESD*
		Independence (ESD=0.40)	Exchangeable (ESD=0.35)	AR-1 (ESD=0.33)	Toeplitz (ESD=0.35)	Unstructured (ESD=2.94**)	
IEATU	TECM _W	49	163	569	177	42	0.34
	CIC _W	55	153	587	181	24	0.35
	CIC _{HH}	420	47	533	0	0	0.36
	CIC _{SH}	597	24	379	0	0	0.37
	EAIC	8	56	304	212	420	0.45
	EBIC	15	84	394	209	298	0.43
	AGP	3	110	801	82	4	0.34
	BGP	15	115	841	28	1	0.34
IEA	TECM _W	58	187	755			0.34
	CIC _W	61	184	755			0.34
	CIC _{HH}	420	47	533			0.36
	CIC _{SH}	597	24	379			0.37
	EAIC	93	193	714			0.34
	EBIC	100	189	711			0.34
	AGP	4	115	881			0.34
	BGP	15	115	870			0.34
EATU	TECM _W		186	591	180	43	0.34
	CIC _W		180	614	181	25	0.34
	CIC _{HH}		167	833	0	0	0.33
	CIC _{SH}		167	833	0	0	0.33
	EAIC		59	308	213	420	0.45
	EBIC		89	399	211	301	0.43
	AGP		110	804	82	4	0.34
	BGP		115	855	29	1	0.34

* empirical standard deviations (ESD) of $\hat{\beta}_2$ resulting from use of the given criterion and penalty

** There was instability with the unstructured working correlation for the analysis of some simulated datasets

TECM_W- trace of the empirical covariance matrix with Westgate penalty

CIC_W- correlation information criterion with Westgate penalty

CIC_{HH}- correlation information criterion with Hardin and Hilbe penalty

CIC_{SH}- correlation information criterion with Shults and Hilbe penalty

EAIC - empirical likelihood AIC-based criterion with penalty

EBIC - empirical likelihood BIC-based criterion with penalty

AGP - AIC-based gaussian pseudolikelihood criterion with penalty

BGP - BIC-based gaussian pseudolikelihood criterion with penalty

IEATU - All 5 structures are considered for selection

IEA - Only independence, exchangeable, and AR-1 are considered for selection

EATU - All structures except for independence are considered for selection

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Correlation structure selection frequencies and empirical standard deviations (ESD) of $\hat{\beta}_2$ for the setting in which the true structure is AR-1 ($\rho = 0.5$) and 100 subjects each contribute 4 observations.

Structures of Interest	Criterion	Selection Frequencies					Criterion ESD*
		Independence (ESD=0.18)	Exchangeable (ESD=0.15)	AR-1 (ESD=0.14)	Toeplitz (ESD=0.14)	Unstructured (ESD=0.15)	
IEATU	TECM _W	0	19	643	227	111	0.14
	CIC _W	0	16	664	226	94	0.14
	CIC _{HH}	0	21	977	2	0	0.14
	CIC _{SH}	1	22	977	0	0	0.14
	EAIC	0	2	728	162	108	0.14
	EBIC	0	6	959	31	4	0.14
	AGP	0	1	866	115	18	0.14
	BGP	0	4	989	7	0	0.14
IEA	TECM _W	0	32	968			0.14
	CIC _W	0	25	975			0.14
	CIC _{HH}	0	22	978			0.14
	CIC _{SH}	1	22	977			0.14
	EAIC	0	10	990			0.14
	EBIC	0	10	990			0.14
	AGP	0	4	996			0.14
	BGP	0	4	996			0.14
EATU	TECM _W		19	643	227	111	0.14
	CIC _W		16	664	226	94	0.14
	CIC _{HH}		21	977	2	0	0.14
	CIC _{SH}		22	978	0	0	0.14
	EAIC		2	728	162	108	0.14
	EBIC		6	959	31	4	0.14
	AGP		1	866	115	18	0.14
	BGP		4	989	7	0	0.14

* empirical standard deviations (ESD) of $\hat{\beta}_2$ resulting from use of the given criterion and penalty

TECM_W - trace of the empirical covariance matrix with Westgate penalty

CIC_W - correlation information criterion with Westgate penalty

CIC_{HH} - correlation information criterion with Hardin and Hilbe penalty

CIC_{SH} - correlation information criterion with Shults and Hilbe penalty

EAIC - empirical likelihood AIC-based criterion with penalty

EBIC - empirical likelihood BIC-based criterion with penalty

AGP - AIC-based gaussian pseudolikelihood criterion with penalty

BGP - BIC-based gaussian pseudolikelihood criterion with penalty

IEATU - All 5 structures are considered for selection

IEA - Only independence, exchangeable, and AR-1 are considered for selection

EATU - All structures except for independence are considered for selection

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Correlation structure selection frequencies and empirical standard deviations (ESD) of $\hat{\beta}_2$ for the setting in which the true structure is Unstructured and 100 subjects each contribute 3 observations.

Structures of Interest	Criterion	Selection Frequencies					Criterion ESD*
		Independence (ESD=0.21)	Exchangeable (ESD=0.16)	AR-1 (ESD=0.13)	Toeplitz (ESD=0.12)	Unstructured (ESD=0.12)	
IEATU	TECM _W	0	0	60	260	680	0.12
	CIC _W	0	0	155	348	497	0.12
	CIC _{HH}	0	0	232	671	97	0.12
	CIC _{SH}	0	0	326	626	48	0.12
	EAIC	0	0	8	359	633	0.12
	EBIC	0	0	69	566	365	0.12
	AGP	0	0	26	346	628	0.12
	BGP	0	0	48	419	533	0.12
IEA	TECM _W	0	0	1000			0.13
	CIC _W	0	0	1000			0.13
	CIC _{HH}	0	0	1000			0.13
	CIC _{SH}	0	0	1000			0.13
	EAIC	0	0	1000			0.13
	EBIC	0	0	1000			0.13
	AGP	0	0	1000			0.13
	BGP	0	0	1000			0.13
EATU	TECM _W		0	60	260	680	0.12
	CIC _W		0	155	348	497	0.12
	CIC _{HH}		0	232	671	97	0.12
	CIC _{SH}		0	326	626	48	0.12
	EAIC		0	8	359	633	0.12
	EBIC		0	69	566	365	0.12
	AGP		0	26	346	628	0.12
	BGP		0	48	419	533	0.12

* empirical standard deviations (ESD) of $\hat{\beta}_2$ resulting from use of the given criterion and penalty

TECM_W - trace of the empirical covariance matrix with Westgate penalty

CIC_W - correlation information criterion with Westgate penalty

CIC_{HH} - correlation information criterion with Hardin and Hilbe penalty

CIC_{SH} - correlation information criterion with Shults and Hilbe penalty

EAIC - empirical likelihood AIC-based criterion with penalty

EBIC - empirical likelihood BIC-based criterion with penalty

AGP - AIC-based gaussian pseudolikelihood criterion with penalty

BGP - BIC-based gaussian pseudolikelihood criterion with penalty

IEATU - All 5 structures are considered for selection

IEA - Only independence, exchangeable, and AR-1 are considered for selection

EATU - All structures except for independence are considered for selection

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Empirical means of CIC values and corresponding penalties.

Setting	Working Structure	Mean of Unpenalized CIC	Mean of CIC_W	Mean of CIC_{HH}	Mean of CIC_{SH}	Mean of Westgate's Penalty	Hardin & Hilbe Penalty	Shults & Hilbe Penalty
1	Independence	4.05	4.05	4.80	5.38	0.00	0.75	1.33
	Exchangeable	3.62	3.68	4.96	5.77	0.06	1.33	2.14
	AR-1	3.32	3.41	4.65	5.46	0.09	1.33	2.14
	Toeplitz	3.24	3.94	6.47	7.90	0.70	3.23	4.67
	Unstructured*	-	-	-	-	-	9.00	12.22
2	Independence	4.05	4.05	4.17	4.26	0.00	0.13	0.21
	Exchangeable	3.62	3.63	3.83	3.94	0.01	0.21	0.32
	AR-1	3.33	3.35	3.54	3.65	0.02	0.21	0.32
	Toeplitz	3.31	3.37	3.76	3.92	0.06	0.45	0.61
	Unstructured	3.29	3.45	4.29	4.53	0.16	1.00	1.24
3	Independence	4.13	4.13	4.25	4.34	0.00	0.13	0.21
	Exchangeable	3.32	3.33	3.53	3.64	0.01	0.21	0.32
	AR-1	2.80	2.81	3.01	3.12	0.01	0.21	0.32
	Toeplitz	2.62	2.75	2.94	3.07	0.13	0.32	0.45
	Unstructured*	-	-	-	-	-	0.45	0.61

* Due to instability in some simulated CIC values, the resulting distorted empirical means are not given

Setting 1: True correlation structure is AR-1, and $N=20$ subjects contribute 4 observations

Setting 2: True correlation structure is AR-1, and $N=100$ subjects contribute 4 observations

Setting 3: True correlation structure is Unstructured, and $N=100$ subjects contribute 3 observations

CIC_W - correlation information criterion with Westgate penalty

CIC_{HH} - correlation information criterion with Hardin and Hilbe penalty

CIC_{SH} - correlation information criterion with Shults and Hilbe penalty

Table 5:

Parameter estimates, standard error estimates, and correlation selection criteria values and penalties from analyses of the PREADViSE trial data.

Parameter	Ind $\hat{\beta}$ (SE)	Exch $\hat{\beta}$ (SE)	AR-1 $\hat{\beta}$ (SE)	Toeplitz $\hat{\beta}$ (SE)	Un $\hat{\beta}$ (SE)
β_0	48.5 (1.42)	48.8 (1.42)	48.7 (1.38)	48.6 (1.35)	48.0 (1.51)
β_1	2.10 (0.38)	1.79 (0.37)	1.76 (0.40)	1.88 (0.39)	2.11 (0.50)
β_2	0.24 (0.16)	0.24 (0.16)	0.20 (0.16)	0.21 (0.16)	0.17 (0.16)
β_3	-0.57 (0.23)	-0.55 (0.23)	-0.54 (0.23)	-0.53 (0.23)	-0.44 (0.27)
β_4	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)
Criterion					
TECM No Penalty	2.25	2.24	2.12	2.06	2.06
TECM _W	2.25	2.24	2.13	2.07	2.63
CIC No Penalty	11.61	11.64	11.33	10.96	9.76
CIC _W	11.61	11.66	11.47	11.49	12.94
CIC _{HH}	12.29	12.62	12.31	12.72	13.23
CIC _{SH}	12.59	12.97	12.66	13.21	13.98
AGP	1,001	908	919	911	947
BGP	1,011	920	930	926	968

Ind - independence; Exch - exchangeable; Un - unstructured

TECM_W - trace of the empirical covariance matrix with Westgate penalty

CIC_W - correlation information criterion with Westgate penalty

CIC_{HH} - correlation information criterion with Hardin and Hilbe penalty

CIC_{SH} - correlation information criterion with Shults and Hilbe penalty

AGP - AIC-based gaussian pseudolikelihood criterion with penalty

BGP - BIC-based gaussian pseudolikelihood criterion with penalty