STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:859–874 (DOI: 10.1002/sim.1650)

Estimating equations for association structures

Jun Yan[1,3] and Jason Fine[1,2,*,†]

[1] *Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.*
[2] *Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.*
[3] *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa, U.S.A.*

SUMMARY

This paper investigates generalized estimating equations for association parameters, which are frequently of interest in family studies, with emphasis on covariance estimation. Separate link functions are used to connect the mean, the scale, and the correlation to linear predictors involving possibly different sets of covariates, and separate estimating equations are proposed for the three sets of parameters. Simulations show that the robust 'sandwich' variance estimator and the jackknife variance estimator for the correlation parameters are generally close to the empirical variance for the sample size of 50 clusters. The results contradict Ziegler *et al*. and Kastner and Ziegler, where the 'sandwich' estimator obtained from the software MAREG was shown to be unsuitable for practical usage. The problem appears to arise because the MAREG variance estimator does not account for variability in estimation of the scale parameters, but may be valid with fixed scale. We also find that the formula for the approximate jackknife variance estimator in Ziegler *et al.* is deficient, resulting in systematic deviations from the fully iterated jackknife variance estimator. A general jackknife formula is provided and performs well in numerical studies. Data from a study on the genetics of alcoholism is used to illustrate the importance of reliable variance estimation in biomedical applications. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: estimating equations; jackknife; marginal regression; sandwich estimator

1. INTRODUCTION

Generalized estimating equations (GEE) has been a widely used tool in multivariate data analysis in biomedical applications since the seminal work of Liang and Zeger [1]. A major advantage of GEE compared with likelihood based approaches is that the joint distribution is not fully specified. For the first order generalized estimating equations, GEE1, the mean parameters are consistently estimated as long as the mean structure is correctly specified, regardless of whether or not the covariance structure is correctly specified.

*Correspondence to: Jason Fine, Department of Statistics, 1210 West Dayton Street, Madison, Wisconsin 53706, U.S.A.
†E-mail: fine@biostat.wisc.edu

Contract/grant sponsor: NIH; contract/grant number: U10AA08401, U10AA08403

Received December 2001
Accepted March 2003

GEE1 has been extended to GEE2 to incorporate parameters in association structures, which may be of scientific interest, for example, in family studies. We use the terms GEE1 and GEE2 following Reference [2]. See References [3–5], where mean parameters and association parameters may be estimated simultaneously or separately. A disadvantage of simultaneous estimation is that the mean parameter estimates may no longer be consistent if the association structure is incorrectly specified, just as in a likelihood analysis. To maintain the consistency of the mean parameter estimates, the estimating equations must be carefully constructed so that mis-specification of the association structure does not corrupt the mean parameters estimates.

An important applied issue is reliable variance estimation which is essential for valid inferences in small to moderately sized samples. For GEE1, Paik [6] showed that the robust 'sandwich' variance estimator for the mean parameters works well (though slightly conservative) with 50 clusters, and recommended the jackknife variance estimator in small samples with 30 or less clusters. Lipsitz *et al.* [7] showed that the one-step jackknife variance estimator is asymptotically equivalent to the 'sandwich' estimator, and their data example with 81 clusters showed that the robust variance estimator and the jackknife variance estimator tend to agree.

Ziegler *et al.* [2] presented an interesting extension of the jackknife variance estimator for both mean and association parameters using GEE2 with separate estimation of the parameters. Ziegler *et al.* [2] and Kastner and Ziegler [8] reported simulation results with 50 clusters of size 3, which suggested that the 'sandwich' estimator for the association parameters is far too conservative. For example, for cluster constant covariates and exchangeable correlation structure in Table III of Kastner and Ziegler [8], the mean of the 'sandwich' standard error estimate of the association parameter is 0.391, while the empirical standard error is 0.198. This means that the 'sandwich' variance estimate is about 4 times the empirical variance on the average. The jackknife variance estimator was strongly recommended for inference in small samples, and the fully iterated jackknife variance estimator was shown to outperform the approximate jackknife variance estimator.

Let Y_{it} be the response of individual t, $t=1,\ldots,n_i$, from cluster $i=1,\ldots,K$, and μ_{it} be the expectation of Y_{it} conditioning on the covariate matrix X_i for cluster i, with $\mu_{it}=E(Y_{it}\,|\,X_i)=E(Y_{it}\,|\,X_{it})$, where X_{it} is the tth row of X_i. Ziegler *et al.* [2] based their second set of estimating equations on the vector of the pairwise products of the standardized residuals, $z_{ist}=(Y_{it}-\mu_{it})(Y_{is}-\mu_{is})/(\sigma_{it}\sigma_{is})$, where σ_{it}^2 is the conditional variance of Y_{it} given X_i. This formulation is based on the methodology in Reference [3] developed for binary responses. In the generalized linear model framework, σ_{it}^2 is usually expressed as the product of the scale function ϕ_{it} and the variance function V_{it}.

The variance formula in Kastner and Ziegler [8], which originates in Reference [3], is correct if the scale parameter is fixed, for example, at 1, as in the standard binomial or Poisson setup. Their formula is inappropriate either with estimated over- or under-dispersion parameters in discrete data models, or if scale parameters are estimated for continuous responses. The disagreement between the empirical and model based variance estimator for the association parameters reported by Ziegler *et al.* [2] and Kastner and Ziegler [8] appears to be due to the fact that the software MAREG [9] based on Reference [8] does not account for the variation introduced by the estimated scale parameters. In the simulations, the responses are Gaussian variables and a scale parameter is estimated. In fact, Kastner and Ziegler [8] did not explicitly describe the method of estimation of the scale parameter, although it is included in the output of MAREG for continuous responses.

Paik [10] modelled covariate effects on variance heterogeneity with a link function for the scale using a separate estimating equation for scale parameters, but did not permit covariate effects on the correlations. Modelling the correlation parameters may also be of interest, particularly in genetic epidemiology. For example, the collaborative study on the genetics of alcoholism (COGA) is a six-centre program to identify and map genes that increase susceptibility to alcohol dependence and related phenotypes (see Reference [11]). Phenotypes predictive of alcohol dependence may cluster within families and it is important to quantify the strength of the associations in various relative pairs, adjusting for known risk factors, such as age.

In this paper, we model the mean, the scale, and the correlation by separate link functions, as in the generalized linear model framework of Prentice and Zhao [4]. We employ three sets of estimating equations for the three sets of model parameters and utilize a 'sandwich' variance estimator for the correlation parameters which adjusts for the estimated scale parameters. Our scale estimates are robust to mis-specification of the correlation structure. This differs from Prentice and Zhao [4] who used a single estimating equation for scale and correlation.

Our estimating equations are presented in Section 2. A simple formula for the joint asymptotic covariance of the mean, scale, and correlation estimators is provided. Jackknife variance estimation for all three sets of parameters is presented in Section 3, extending Ziegler *et al.* [2] who only considered mean and correlation. The jackknife estimator and the 'sandwich' estimator are evaluated under the scenarios in Reference [8] in a simulation study in Section 4. Our variance estimators are generally close to the empirical variance, in contrast to Kastner and Ziegler [8]. The analysis of the COGA data is presented in Section 5. A discussion of the key issues concludes in Section 6. The derivation of the 'sandwich' variance estimator is sketched in Appendix A, and the approximate jackknife variance estimator is given and compared with that in Reference [2] in Appendix B.

2. MODEL AND ESTIMATING EQUATIONS

Consider a sample of K independent observations $Y_i^{\mathrm{T}} = (Y_{i1}, \ldots, Y_{in_i})$, $i = 1, \ldots, K$, of n_i-variate responses. To gain insight into the robust variance estimator for the association parameters, we decompose the association structure into two parts, the variance and the correlation. The covariance of the responses in the ith cluster Y_i is

$$\mathrm{cov}(Y_i) = \mathscr{V}_i^{1/2} \mathscr{R}_i \mathscr{V}_i^{1/2} \tag{1}$$

where $\mathscr{V}_i$ is the diagonal matrix of variances of Y_i, and $\mathscr{R}_i$ is the correlation matrix of Y_i. Each element $\mathscr{V}_{it}$ in the diagonal matrix $\mathscr{V}_i$ can be further factored as the product of the scale function ϕ_{it} and the variance function v_{it}; see Reference [10]. The variance function is often modelled as a function of the mean μ_{it} under the framework of generalized linear models; see Reference [12, pp. 29–30]. The scale function may incorporate covariate effects thereby facilitating modeling of over- or under-dispersion and heteroscedasticity. The correlation matrix $\mathscr{R}_i$ may similarly incorporate covariate effects. Thus, one may explicitly model both scale and correlation parameters. This parameterization is a variation of the abstract covariance model in Reference [4]; see Section 4 of Reference [4].

Let X_{1i}, X_{2i} and X_{3i} be the covariate matrices for the mean, the scale, and the correlation of the response Y_i, with dimensions $n_i \times p$, $n_i \times r$, and $n_i(n_i - 1)/2 \times q$, respectively. Our model is

$$g_1(\mu_i) = X_{1i}\beta \tag{2}$$

$$g_2(\phi_i) = X_{2i}\gamma \tag{3}$$

$$g_3(\rho_i) = X_{3i}\alpha \tag{4}$$

where g_i, $i = 1, 2, 3$, are known link functions, μ_i is a $n_i \times 1$ vector of the mean $E(Y_i \mid X_{1i})$, ϕ_i is a $n_i \times 1$ vector of the scale $\phi_{it} = \mathrm{var}(Y_i \mid X_{2i})/v_{it}$, where $v_{it} = v(\mu_{it})$ is the variance function, and ρ_i is a $n_i(n_i - 1)/2 \times 1$ vector of the pairwise correlation $\mathrm{cor}(Y_{is}, Y_{it} \mid X_{3i})$. The indices β, γ, and α are the mean, the scale, and the correlation parameters of dimension $p \times 1$, $r \times 1$, and $q \times 1$. Hereafter, the link functions g_i, $i = 1, 2, 3$ are referred to as the mean link, the scale link, and the correlation link, respectively.

The mean link function has been well studied. The scale link and the correlation link serve similar purposes. For example, as in Reference [4], the scale link could be log, and the correlation link could be rescaled Fisher's z transformation, in which case the inverse link function is,

$$\rho_{its} = \mathrm{cor}(Y_{is}, Y_{it} \mid X_{3i}) = \frac{\exp(X_{3i(s,t)}\alpha) - 1}{\exp(X_{3i(s,t)}\alpha) + 1} \tag{5}$$

where $X_{3i(s,t)}$ is the row in X_{3i} corresponding to the correlation of Y_{is} and Y_{it}. These links ensure that the scale is positive and the correlation is within $(-1, 1)$. The scale model is useful in situations where parameters are needed for covariate effects on over- or under-dispersion and heteroscedasticity. The choice of the scale link will depend on the issue being addressed. See Reference [12, Chapter 10] and Reference [10] for further discussion. The three link model flexibly incorporates covariates in the mean, scale, and correlation, and may be viewed as a useful specialization of Reference [4].

A convenient set of estimating equations for the three-link model is

$$u(\beta, \gamma, \alpha) = \sum_{i=1}^{K} \begin{pmatrix} D_{1i} & 0 & 0 \\ 0 & D_{2i} & 0 \\ 0 & 0 & D_{3i} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} V_{1i} & 0 & 0 \\ 0 & V_{2i} & 0 \\ 0 & 0 & V_{3i} \end{pmatrix}^{-1} \begin{pmatrix} Y_i - \mu_i \\ s_i - \phi_i \\ z_i - \rho_i \end{pmatrix} \tag{6}$$

where s_i is the $n_i \times 1$ vector of $s_{it} = (Y_{it} - \mu_{it})^2/v_{it}$, z_i is the $n_i(n_i - 1)/2 \times 1$ vector of $z_{its} = (Y_{it} - \mu_{it})(Y_{is} - \mu_{is})/\sqrt{\phi_{it}v_{it}\phi_{is}v_{is}}$, $D_{1i} = \partial\mu_i/\partial\beta^{\mathrm{T}}$, $D_{2i} = \partial\phi_i/\partial\gamma^{\mathrm{T}}$, $D_{3i} = \partial\rho_i/\partial\alpha^{\mathrm{T}}$, and V_{1i}, V_{2i} and V_{3i} are the conditional working covariance matrices of Y_i, s_i, and z_i.

The matrix V_{1i} generally contains scale parameters γ and correlation parameters α. The matrices V_{2i} and V_{3i} may contain other estimated quantities which characterize the third and fourth order moments; see Reference [4]. Paik [10] gave a complicated expression for V_{2i} for general responses, and Prentice [3] gave a complicated expression for V_{3i} for binary responses, both of which may lead to gains in efficiency. In practice, when using (6), to avoid specification of higher order moments, estimation of higher order nuisance parameters, and convergence problems, V_{2i} may be chosen to be a diagonal matrix whose diagonal elements are $2\phi_{it}$, following the independence Gaussian working matrix in Reference [4], and V_{3i} may

be an identity matrix [13, p. 129], at the cost of potential efficiency loss. These simplifications are implemented in the *R* package geepack [14].

The estimation procedure based on $u(\beta,\gamma,\alpha)$ is robust for β and γ. Let $(\hat{\beta},\hat{\gamma},\hat{\alpha})$ satisfy $u(\hat{\beta},\hat{\gamma},\hat{\alpha})=0$. The estimate $\hat{\beta}$ is consistent with mis-specification of the correlation and scale, which is a well known property of GEE1. The estimate $\hat{\gamma}$ is consistent if the mean and the scale structure are correctly specified, regardless of mis-specification of the correlation, by similar arguments to that for $\hat{\beta}$. Notice that the three-estimating-equations (3EE) approach is more robust than the two-estimating-equations (2EE) approach in Reference [4], where the scale and the correlation parameters are estimated simultaneously. Chaganty and Shults [15] also estimate the mean, scale, and correlation parameters, using the quasi-least square method, without covariates in the scale and correlation, and the estimation procedure is robust, like the 3EE. Hall and Severini [16] simultaneously estimate the mean and association parameters similar to Liang *et al.* [5], but the mean parameter estimates are robust to mis-specification of covariance, and no high order moment specification is needed, unlike the GEE2 in Liang *et al.* [5].

The structure of equation (6) facilitates estimation. The estimating equations for the scale parameters depend only on the estimates of mean parameters, while the estimating equations for the correlation parameters depend on both the mean parameters and the scale parameters. To find the root of the estimating equations, an alternating modified Fisher scoring method may be used, where at step $s+1$ of the iteration,

$$\hat{\beta}_{s+1}=\hat{\beta}_s+\left(\sum D_{1i}^{\mathrm{T}}V_{1i}^{-1}D_{1i}\right)^{-1}\sum D_{1i}^{\mathrm{T}}V_{1i}^{-1}(Y_i-\mu_i) \tag{7}$$

$$\hat{\gamma}_{s+1}=\hat{\gamma}_s+\left(\sum D_{2i}^{\mathrm{T}}V_{2i}^{-1}D_{2i}\right)^{-1}\sum D_{2i}^{\mathrm{T}}V_{2i}^{-1}(s_i-\phi_i) \tag{8}$$

$$\hat{\alpha}_{s+1}=\hat{\alpha}_s+\left(\sum D_{3i}^{\mathrm{T}}V_{3i}^{-1}D_{3i}\right)^{-1}\sum D_{3i}^{\mathrm{T}}V_{3i}^{-1}(z_i-\rho_i) \tag{9}$$

and the second term on the right hand side of each expression is evaluated at the parameter values from step s of the iteration. The iterative procedure begins with starting values $(\hat{\beta}_0,\hat{\gamma}_0,\hat{\alpha}_0)$—for example, those corresponding to constant scale and independent correlation structure—and updates until convergence.

Arguments analogous to those in Reference [4] give that $\{K^{1/2}(\hat{\beta}-\beta)^{\mathrm{T}},K^{1/2}(\hat{\gamma}-\gamma)^{\mathrm{T}},K^{1/2}(\hat{\alpha}-\alpha)^{\mathrm{T}}\}^{\mathrm{T}}$ quite generally has an asymptotic normal distribution, with mean zero and covariance matrix consistently estimated by $\{\hat{\Sigma}_{1K}^{-1}\}\hat{\Sigma}_{2K}\{\hat{\Sigma}_{1K}^{-1}\}^{\mathrm{T}}$ given in Appendix A. Note that $\hat{\Sigma}_{1K}$ is not symmetric, which is in the same spirit as the variance estimator in References [3, 10]. It is worth emphasizing that the variance of $\hat{\gamma}$ depends on $\hat{\beta}$ and the variance of $\hat{\alpha}$ depends on $\hat{\beta}$ and $\hat{\gamma}$. When α or γ or both are not included in the model, such as in the case of independence correlation structure or fixed scale model, the covariance matrix is obtained by deleting the corresponding rows and columns in $\hat{\Sigma}_{1K}^{-1}$ and $\hat{\Sigma}_{2K}$.

3. JACKKNIFE VARIANCE ESTIMATOR

In small samples ($K\leqslant 30$), Paik [6] recommended using the jackknife variance estimator instead of the robust variance estimator for GEE1, which exhibited bias. The jackknife variance

estimator is

$$\frac{K-p}{K}\sum_{i=1}^{K}(\hat{\beta}_{-i}-\hat{\beta})(\hat{\beta}_{-i}-\hat{\beta})^{\mathrm{T}} \tag{10}$$

where p is the number of parameters in the mean structure and $\hat{\beta}_{-i}$ and $\hat{\beta}$ are the estimates of β leaving out the ith cluster and with all observations, respectively [6]. This estimator is asymptotically equivalent to the corresponding 'sandwich' variance estimator; see Reference [7].

Ziegler *et al.* [2] extended the jackknife variance estimator to the correlation parameters based on their estimation procedure in Reference [8]. Applying the same line of reasoning to our generalized estimating equation (6), the jackknife variance estimator is

$$\frac{K-p-r-q}{K}\sum_{i=1}^{K}(\hat{\theta}_{-i}-\hat{\theta})(\hat{\theta}_{-i}-\hat{\theta})^{\mathrm{T}} \tag{11}$$

where p, r, and q are the number of parameters in the mean, scale and correlation models, respectively, $\theta=(\beta^{\mathrm{T}},\gamma^{\mathrm{T}},\alpha)^{\mathrm{T}}$, and $\hat{\theta}_{-i}$ and $\hat{\theta}$ are the parameter estimates leaving out the ith cluster and with all observations, respectively. A simpler one-step jackknife variance estimator may be based on parameter estimates after one Fisher scoring step from the full observation estimates using equation (7)–(9). Both jackknife estimators are equivalent to the 'sandwich' estimator in large samples.

The computation of the delete-1 jackknife takes the full observation estimates as the starting value in the Fisher scoring algorithm. It is helpful to view the system of estimating equations (6) as a whole when applying the one-step Fisher scoring. In this way, a one-step jackknife variance can be obtained. This is called the approximate jackknife variance estimator by Ziegler [17]; see Ziegler *et al.* [2] for a computational formula, which is used in MAREG [9]. Unfortunately, a careful examination of their appendix [2] shows that, even in the case that the scale parameter is known, when applying matrix inversion to get A_{-i}^{-1}, the inverse of A leaving the ith cluster out, the lower-left block is not handled correctly; see details in Appendix B. Therefore, the estimator is correct for the mean parameters, but not for the association parameters. A valid formula for the approximate jackknife variance estimator for $\hat{\theta}$ is given in Appendix B.

We implemented the three jackknife variance estimators discussed above: fully iterated jackknife estimator (FIJ), one-step jackknife estimator (1-SJ), and approximate jackknife estimate (AJS), using the *R* package geepack. All the jackknife variance estimators are based on equation (11). FIJ uses $\hat{\theta}_{-i}$ from full iteration of Fisher scoring, 1-SJ uses $\hat{\theta}_{-i}$ after one step Fisher scoring from the full observation estimates, and AJS uses formula (B2) in Appendix B. For more details about AJS, see Appendix B.

4. SIMULATION STUDY

Ziegler *et al.* [2] and Kastner and Ziegler [8] conducted simulation studies comparing the performance of the robust variance estimator and the jackknife variance estimates under various configurations reflecting different between- and within-cluster variation of the explanatory

Table I. Frequency table for generating within-cluster varying binary covariate.

$X_{i11} - X_{i12} - X_{i13}$	Frequency
1-1-1	10
1-1-0	20
1-0-0	10
0-0-0	10

Table II. Frequency table for generating continuous covariate.

	Within-cluster varying			
Cluster-constant	X_{i21}	X_{i22}	X_{i23}	Frequency
$(0,1)$	$(0,1)$	$(0,2)$	$(1,3)$	5
$(1,3)$	$(1,3)$	$(2,5)$	$(3,4)$	10
$(3,6)$	$(3,6)$	$(5,6)$	$(4,8)$	20
$(6,8)$	$(6,8)$	$(6,8)$	$(8,9)$	10
$(8,9)$	$(8,9)$	$(8,9)$	$(9,9)$	5

variables. In order to compare the two implementations, we simulate under the same setups, with slight modification, and analyse the simulated data using both geepack and MAREG.

The model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

where Y_i, $i = 1, \ldots, 50$, are trivariate response vectors, and $X_{ij} = (X_{ij1}, X_{ij2}, X_{ij3})^{\mathrm{T}}$, $j = 1, 2$, where X_{i1} contains binary covariates and X_{i2} contains continuous covariates. The scalars β_0, β_1, and β_2 are the regression parameters, and ε_i is a standard trivariate normal vector with mean zero and covariance (correlation) matrix $\Sigma(\alpha)$, where α denotes the correlation parameters. The link functions g_1, g_2, and g_3 are identity, identity, and Fisher's z, respectively. We used identity link for the scale since there is no covariate, and the parameter interpretation is more transparent.

The simulations explore the properties of the variance estimators by varying the distributions of X_{i1}, X_{i2} and ε_i. Each covariate is specified to be either constant or non-constant within clusters and the covariance structure is either exchangeable or unstructured, resulting in 8 configurations in total.

For the cluster-constant case, the binary variable equals 1 in 40 clusters and equals 0 in 10 clusters. For within-cluster variation case, X_{i1} was generated from Table I.

The continuous variable is generated from the grouped data frequency distribution in Table II. Within each interval in the table, the support points are equidistant and include the end points. For example, for interval $(0,1)$ with frequency 5, the 5 values are 0, 0.25, 0.50, 0.75, and 1. For the cluster-constant case, a sample is drawn from a permutation of all points in the frequency table. For $X_{i21} \neq X_{i22} \neq X_{i23}$, the covariates are a random permutation of the points generated from the frequency table and the three values are permuted independently.

Table III. Simulation result for cluster-constant binary variable.

Parameter	True value	Mean estimate	Standard error						
			geepack					MAREG	
			EMP	SAN	AJS	1-SJ	FIJ	SAN	AJS
Cluster-constant continuous variable, exchangeable correlation									
α_0	0.7	0.640	0.210	0.197	0.216	0.205	0.204	0.495	0.248
γ_0	1.0	0.967	0.127	0.122	0.129	0.123	0.123		
Cluster-constant continuous variable, unstructured correlation									
α_0	0.5	0.451	0.285	0.282	0.293	0.285	0.286	0.470	0.298
α_1	1.0	0.958	0.325	0.316	0.328	0.320	0.320	0.794	0.392
α_2	0.7	0.654	0.314	0.291	0.301	0.292	0.294	0.597	0.325
γ_0	1.0	0.963	0.130	0.123	0.127	0.121	0.121		
Within-cluster-varying continuous variable, exchangeable correlation									
α_0	0.7	0.659	0.205	0.199	0.212	0.205	0.205	0.462	0.248
γ_0	1.0	0.976	0.122	0.124	0.127	0.123	0.123		
Within-cluster-varying continuous variable, unstructured correlation									
α_0	0.5	0.483	0.295	0.281	0.284	0.279	0.283	0.453	0.299
α_1	1.0	0.985	0.339	0.319	0.324	0.319	0.324	0.762	0.402
α_2	0.7	0.692	0.304	0.295	0.299	0.298	0.294	0.544	0.329
γ_0	1.0	0.975	0.124	0.126	0.126	0.122	0.123		

Note: EMP: empirical; SAN: 'sandwich'; AJS: approximate jackknife; 1-SJ: one-step jackknife; FIJ: fully iterated jackknife.

The regression parameter β is $(1, 3, -0.2)^{\mathrm{T}}$ and the scale parameter γ is 1 in the simulations. Using Fisher's z link function, the true correlation parameters are $\alpha = 0.7$ for exchangeable correlation and $\alpha = (0.5, 1, 0.7)^{\mathrm{T}}$ for unstructured correlation. This gives the correlation matrices

$$\begin{pmatrix} 1.000 & 0.336 & 0.336 \\ 0.336 & 1.000 & 0.336 \\ 0.336 & 0.336 & 1.000 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1.000 & 0.245 & 0.462 \\ 0.245 & 1.000 & 0.336 \\ 0.462 & 0.336 & 1.000 \end{pmatrix}.$$

The parameters are estimated using the generalized estimating equation (6) with working variance matrix V_{2i} being a diagonal matrix whose diagonal elements are $2\phi_{it}$, and working variance matrix V_{3i} being an identity matrix. The results are summarized in Tables III and IV, including results from both geepack and MAREG. The point estimates of the mean and correlation parameter are virtually identical from both implementations. This implies that the empirical variance and the fully iterated jackknife variance estimator for the mean and correlation parameter estimates from both implementations should be the same. Empirically, we also find that the one-step jackknife variance estimators from the two programs differ very little. For this reason, in Tables III and IV, we only include the empirical standard error and standard errors based on the fully iterated jackknife and one-step jackknife from geepack. All variance estimators from geepack are reasonably close to the empirical value. This observation holds for all parameters: mean, scale, and correlation. The results from MAREG indicate that the 'sandwich' variance estimator of association parameters is much too conservative, and the AJS variance estimator is also systematically conservative.

Table IV. Simulation result for within-cluster-varying binary variable.

Parameter	True value	Mean estimate	Standard error						
			geepack					MAREG	
			EMP	SAN	AJS	1-SJ	FIJ	SAN	AJS
Cluster-constant continuous variable, exchangeable correlation									
α_0	0.7	0.656	0.208	0.197	0.211	0.204	0.205	0.386	0.245
γ_0	1.0	0.972	0.126	0.122	0.126	0.122	0.122		
Cluster-constant continuous variable, unstructured correlation									
α_0	0.5	0.461	0.301	0.281	0.286	0.280	0.283	0.443	0.294
α_1	1.0	0.995	0.345	0.321	0.327	0.322	0.325	0.611	0.400
α_2	0.7	0.684	0.311	0.294	0.299	0.294	0.298	0.427	0.326
γ_0	1.0	0.975	0.129	0.125	0.126	0.122	0.122		
Within-cluster-varying continuous variable, exchangeable correlation									
α_0	0.7	0.677	0.212	0.198	0.206	0.203	0.206	0.369	0.250
γ_0	1.0	0.974	0.123	0.123	0.124	0.122	0.121		
Within-cluster-varying continuous variable, unstructured correlation									
α_0	0.5	0.482	0.304	0.283	0.283	0.280	0.286	0.431	0.297
α_1	1.0	1.016	0.329	0.321	0.320	0.320	0.327	0.578	0.395
α_2	0.7	0.691	0.318	0.293	0.292	0.290	0.297	0.398	0.325
γ_0	1.0	0.979	0.126	0.126	0.124	0.121	0.122		

Note: EMP: empirical; SAN: 'sandwich'; AJS: approximate jackknife; 1-SJ: one-step jackknife; FIJ: fully iterated jackknife.

To test whether the problems with the 'sandwich' and the approximate variance estimator were caused by small samples, as reported by Ziegler *et al*. [2] and Kastner and Ziegler [8], we also ran the simulation with 200 clusters. The results are the same as those for 50 clusters (hence not reported here), except that all the standard errors are halved.

In conclusion, our simulation results suggest strongly that: (1) When the scale is not fixed and needs to be estimated, the 'sandwich' variance estimator for association parameters is not as conservative as reported by Ziegler *et al*. [2] and Kastner and Ziegler [8]; and (2) The formula for the approximate jackknife variance estimator is deficient in Reference [2].

5. COGA DATA

We now apply the model in Section 2 to study the association of phenotypes of alcohol dependence among family members in the COGA data. We compare the results from our implementation to those from MAREG.

Probands with alcohol dependence were systematically selected from inpatient and outpatient treatment units. Families were required to have three or more available sibs (including the proband) if both parents were available. Four or more sibs were required if one or both parents were missing or unavailable. The proband could belong either to the offspring or parental generation. At least two first-degree relatives were required to live in the catchment area of a COGA center. Probands were not included in the study if they were uncooperative, were positive for the HIV virus, were intravenous drug users, had a fatal illness, or were unable to

participate in the protocol. First-degree relatives were personally interviewed and administered a personality trait inventory interview and families with three or more affected members were selected. Bilineal families were not included in the study. The data set contains 987 individuals from 105 multigenerational families, selected through 23 female and 82 male probands.

Our focus is the three variables which measure psychological predisposition to alcohol dependence. The outcomes are HA (Harm Avoidance scale), NS (Novelty Seeking scale), and RD (Reward Dependence scale). The ranges of these outcomes in the observed data are: HA 0–33 with mean 13.70 and standard deviation 6.98, NS 2–31 with mean 15.19 and standard deviation 5.39, and RD 5–29 with mean 18.49 and standard deviation 4.31. A person with higher values of these outcomes is more likely to develop alcohol dependence. Omitting the individuals with any missing outcomes, and leaving out 3 pedigrees of size 1, gives 95 pedigrees and 727 individuals (346 males and 381 females). The pedigree sizes range from 2 to 19, with median 7 and mean 7.653.

We fit models (2)–(4) for each measurement, with identity link for the mean and the scale and Fisher's z link for correlations. Since there was either no covariate or only binary covariates in the scale model, the identity link was used for scale for its ease of interpretation. With continuous covariates, the log link might be more appropriate. The results are summarized in Table V. The relevant prognostic factors in the data set are gender ($1 =$ male and $0 =$ female), age (in years), and smoke ($1 =$ smoking and $0 =$ non-smoking), at the interview time. The mean models in Table V give results from fitting reduced models which include factors significant at level 0.05 with all three factors in the full model. The scale model has only an intercept, since MAREG does not allow covariates in the scale. The covariates in the correlation structure are indicators of same gender siblings (sgsib), different gender siblings (dgsib), same gender parent–child (sgpc), different gender parent–child (dgpc), spouse pairs (spouse), and a numerical variable age difference between siblings (sibagedif) for RD only. There is no intercept in the correlation model and all other relative pairs are lumped together (other).

From Table V, we observe that men have lower HA and RD, and higher NS, than do women. Increasing age decreases NS and RD. Smoking leads to larger values of HA and NS. After adjusting for the factors in the mean, the correlation between same gender sibs and different gender sibs are significant for HA and RD. Same gender parent–child correlation is significant for HA and NS. Spouse correlation is marginally significant for HA. The age difference in sibs has significant effect on correlations for RD, with sibs closer in age being more alike. We have chosen to omit sibagedif from the correlation models for HA and NS. For HA, including sibagedif caused instability in the model fitting and sgsib, dgsib, and sibagedif were not significant. For NS, none of the sib correlations are significant in the model without sibagedif. Hence, the sibagedif is excluded from the NS correlation model. The interpretation of the parameters in the correlation models depends on the correlation link, which is Fisher's z in this analysis. Consider a sib pair of same gender with age difference of 3 years. By equation (5), the estimated correlation of RD is 0.204, conditional on their genders and ages, which are accounted for by the mean model.

In this example, the difference between our implementation and MAREG in the 'sandwich' variance estimator is not large for most correlation parameters for NS. However, there are noticeable differences in the variance estimates for correlation parameters for HA and RD, which may be substantial in some instances. For example, in the analysis of RD, for sgsib, dgsib, and sibagedif, MAREG yielded a large 'sandwich' variance estimator and somewhat large AJS estimator (for sgsib and dgsib) relative to our implementation. The 'sandwich'

Table V. Generalized estimating equation result (parameter estimates and standard errors) for the model fitted for 95 pedigrees and 727 individuals in the COGA data.

		geepack					MAREG		
	Variables	PM	SAN	AJS	1-SJ	FIJ	PM	SAN	AJS
HA: harm avoidance scale									
Mean:	Intercept	14.370	0.429	0.412	0.412	0.420	14.370	0.428	0.393
	Gender	−2.563	0.539	0.516	0.516	0.553	−2.564	0.537	0.531
	Smoke	1.249	0.563	0.540	0.540	0.594	1.249	0.561	0.542
Scale:	Intercept	46.512	2.536	2.444	2.432	2.428	46.705		
corr:	sgsib	0.222	0.105	0.103	0.103	0.105	0.221	0.119	0.104
	dgsib	0.255	0.097	0.093	0.095	0.095	0.254	0.112	0.095
	sgpc	0.333	0.127	0.123	0.123	0.123	0.332	0.145	0.128
	dgpc	0.147	0.112	0.108	0.116	0.119	0.147	0.114	0.104
	spouse	0.632	0.329	0.319	0.326	0.328	0.630	0.376	0.311
	other	0.069	0.053	0.053	0.054	0.055	0.069	0.056	0.056
NS: novelty seeking scale									
Mean:	Intercept	18.241	0.645	0.618	0.618	0.640	18.240	0.645	0.639
	Gender	0.842	0.384	0.366	0.366	0.393	0.842	0.383	0.367
	Age	−0.107	0.012	0.012	0.012	0.012	−0.107	0.012	0.012
	Smoke	1.934	0.359	0.344	0.344	0.358	1.933	0.360	0.372
Scale:	Intercept	25.078	1.259	1.206	1.199	1.199	25.217		
corr:	sgsib	0.173	0.121	0.119	0.118	0.118	0.172	0.118	0.124
	dgsib	0.127	0.113	0.110	0.111	0.111	0.126	0.115	0.112
	sgpc	0.302	0.141	0.135	0.132	0.137	0.300	0.161	0.135
	dgpc	−0.031	0.096	0.092	0.091	0.092	−0.031	0.093	0.093
	spouse	−0.027	0.217	0.210	0.207	0.213	−0.027	0.215	0.206
	other	−0.008	0.050	0.054	0.054	0.055	−0.008	0.058	0.052
RD: reward dependence scale									
Mean:	Intercept	21.153	0.568	0.540	0.540	0.541	21.153	0.566	0.702
	Gender	−1.703	0.317	0.301	0.301	0.307	−1.703	0.316	0.347
	Age	−0.044	0.012	0.012	0.012	0.012	−0.044	0.012	0.012
Scale:	Intercept	17.503	1.043	1.006	1.000	1.001	17.576		
corr:	sgsib	0.536	0.150	0.145	0.146	0.146	0.534	0.294	0.164
	dgsib	0.402	0.157	0.154	0.154	0.155	0.400	0.211	0.179
	sgpc	0.219	0.128	0.125	0.123	0.125	0.218	0.140	0.151
	dgpc	0.136	0.145	0.142	0.142	0.142	0.136	0.149	0.144
	spouse	0.190	0.239	0.232	0.231	0.232	0.190	0.297	0.241
	other	0.084	0.072	0.070	0.069	0.070	0.084	0.096	0.081
	sibagedif	−0.041	0.018	0.017	0.017	0.018	−0.041	0.026	0.015

Note: PM: parameter estimate; SAN: 'sandwich'; AJS: approximate jackknife; 1-SJ: one-step jackknife; FIJ: fully iterated jackknife.

variance estimator for these three parameters from MAREG is large enough that these effects become insignificant at level 0.05. The differences between the approximate and fully iterated jackknife variance estimators are not large in this data analysis.

Table VI. Scale model results (parameter estimates and standard errors) from geepack for HA and RD fitted with 95 pedigrees and 727 individuals in the COGA data.

	Variables	PM	SAN	AJS	1-SJ	FIJ
HA: harm avoidance scale						
Scale:	Intercept	42.722	3.385	3.267	3.250	3.252
	Smoke	8.249	4.199	4.049	4.033	4.042
RD: reward dependence scale						
Scale:	Intercept	16.179	1.176	1.134	1.127	1.128
	Gender	2.783	1.474	1.411	1.404	1.401

Note: PM: parameter estimate; SAN: 'sandwich'; AJS: approximate jackknife; 1-SJ: one-step jackknife; FIJ: fully iterated jackknife.

We also fit models with individual level covariates in the scale using geepack to illustrate the facility. Smoking is found to lead to a significant increase on the scale of HA. None of the factors is significant in the scale model for NS. Hence, the intercept model appears to be sufficient. Gender has a marginally significant effect on the scale of RD, with males having greater variability. The results for HA and RD are presented in Table VI. The modeling of scale with covariates changes the mean and the correlation only slightly, in both point estimates and standard errors. This may be due to the lack of severe heteroscedasticity.

6. DISCUSSION

In this paper, we extended the standard 2EE GEE2 to a 3EE GEE2 with separate estimating equations for the scale and the correlation. Covariate effects are allowed for the mean, scale, and correlation with different link functions. The 3EE procedure is more robust than the 2EE in Reference [4], in that the scale parameter estimates are consistent regardless of mis-specification of the correlation structure. When the scale parameter is fixed, our 'sandwich' variance estimator reduces to that in Reference [2], which may be inappropriate when the scale parameters are estimated. In the binomial and Poisson cases where the scale parameter is ordinarily fixed at 1, our model can be used to evaluate systematic overdispersion which often arises in practice. In addition, including the scale parameters in the working covariance matrices may lead to more efficient estimation of the mean and the correlation parameters by accounting for the heterogeneity of the scale [10]. Thus, this strategy may be useful when the correlation and scale parameters are of scientific interest.

We also generalized the jackknife variance estimator for the mean and the association estimates proposed by Ziegler *et al*. [2] to our setup for the mean, scale, and correlation, and similarly derived the approximate jackknife variance estimator, which greatly simplifies the computations associated with the leave one out process. We point out in Appendix B that, even when the scale is fixed, the formula for the approximate jackknife variance estimator supplied in Reference [2] fails to incorporate the variation caused by the estimated mean parameters adequately.

Using our 3EE procedure, we investigated the performance of robust and jackknife variance estimators for GEE2 with 50 clusters of size 3. For the mean parameters β_0, β_1 and β_2, the

empirical standard errors are well approximated by each of the four approaches to variance estimations. This is in line with the findings of Lipsitz *et al.* [7] and Ziegler *et al.* [2]. For the correlation parameters, our simulation result is quite different from that in References [2, 8]. The variance estimates from geepack all tend to agree with the empirical variance. The 'sandwich' variance estimator does not perform badly, as reported in Reference [8]. We believe that this is due to the incorporation of the dependence of correlation parameter estimates on scale parameter estimates. When the scale is fixed, such as at 1 as in the binomial and Poisson cases [3], this issue is handled automatically since there is no variation associated with the scale parameters in generating z_i's for the estimating equation for correlation parameters. However, when the scale parameter is unknown and needs to be estimated as in the normal case, this variation must be taken into account. Prentice and Zhao [4] directly modelled covariance rather than scale and correlation, and hence do not have this problem with 2EE.

The fully iterated jackknife and the one step jackknife variance estimators did not outperform the approximate jackknife variance estimator for the correlation parameters in our simulations. Kastner and Ziegler [8] reported that the approximate jackknife standard error may be up to 20 per cent higher than the fully iterated jackknife standard error (the same is observed in our simulations using MAREG [9]). In the appendix of Ziegler *et al.* [2], the approximation formula did not handle the lower-left block of the matrix A appropriately when applying matrix inversion, and this affected the validity of the approximation. Our implementation gives results for AJS which are comparable to the other variance estimators.

The results indicate that at this sample size, all variance estimators are suitable for inference. This suggests that the 'sandwich' estimator and the approximate jackknife estimator are sufficient for inferential purposes, and it is unnecessary to use the fully iterated and one step jackknife estimators.

The *R* package geepack [14], computationally intensive is freely available from any CRAN site (for example, http://cran.r-project.org). For multivariate data, geepack allows covariates in the mean, scale, and correlation structures by separate link functions, and provides 'sandwich' and three versions of jackknife variance estimators for all the parameter estimates. geepack can also handle clustered ordinal responses, using the method in Reference [18].

APPENDIX A: 'SANDWICH' VARIANCE FORMULAS

Under regularity conditions, $\{K^{1/2}(\hat{\beta}-\beta)^{\mathrm{T}}, K^{1/2}(\hat{\gamma}-\gamma)^{\mathrm{T}}, K^{1/2}(\hat{\alpha}-\alpha)^{\mathrm{T}}\}^{\mathrm{T}}$ can, using a Taylor expansion, be approximated by

$$\begin{bmatrix} -K^{-1}\dfrac{\partial U_1(\beta,\gamma,\alpha)}{\partial \beta^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_1(\beta,\gamma,\alpha)}{\partial \gamma^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_1(\beta,\gamma,\alpha)}{\partial \alpha^{\mathrm{T}}} \\ -K^{-1}\dfrac{\partial U_2(\beta,\gamma,\alpha)}{\partial \beta^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_2(\beta,\gamma,\alpha)}{\partial \gamma^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_2(\beta,\gamma,\alpha)}{\partial \alpha^{\mathrm{T}}} \\ -K^{-1}\dfrac{\partial U_3(\beta,\gamma,\alpha)}{\partial \beta^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_3(\beta,\gamma,\alpha)}{\partial \gamma^{\mathrm{T}}} & -K^{-1}\dfrac{\partial U_3(\beta,\gamma,\alpha)}{\partial \alpha^{\mathrm{T}}} \end{bmatrix}^{-1} \begin{bmatrix} -K^{-1/2}U_1(\beta,\gamma,\alpha) \\ -K^{-1/2}U_2(\beta,\gamma,\alpha) \\ -K^{-1/2}U_3(\beta,\gamma,\alpha) \end{bmatrix} \quad \text{(A1)}$$

where

$$\begin{aligned} U_1(\beta,\gamma,\alpha) &= \sum_{i=1}^{K} U_{1i} = \sum_{k=1}^{K} D_{1i}^{\mathrm{T}} V_{1i}^{-1}(Y_i - \mu_i) \\ U_2(\beta,\gamma,\alpha) &= \sum_{i=1}^{K} U_{2i} = \sum_{k=1}^{K} D_{2i}^{\mathrm{T}} V_{2i}^{-1}(s_i - \phi_i) \\ U_3(\beta,\gamma,\alpha) &= \sum_{i=1}^{K} U_{3i} = \sum_{k=1}^{K} D_{3i}^{\mathrm{T}} V_{3i}^{-1}(z_i - \rho_i) \end{aligned} \tag{A2}$$

The random quantities $[K^{-1/2}U_1(\beta,\gamma,\alpha)^{\mathrm{T}}, K^{-1/2}U_1(\beta,\gamma,\alpha)^{\mathrm{T}}, K^{-1/2}U_1(\beta,\gamma,\alpha)^{\mathrm{T}}]$ have a multivariate normal distribution as $K \to \infty$ with mean zero and with variance matrix $\Sigma_{2K} = \lim_{K\to\infty} \hat{\Sigma}_{2K}$, where

$$\hat{\Sigma}_{2K} = K^{-1}\sum_{i=1}^{k} \begin{bmatrix} D_{1i}^{\mathrm{T}} V_{1i}^{-1} \hat{\mathrm{cov}}(Y_i) V_{1i}^{-1} D_{1i} & D_{1i}^{\mathrm{T}} V_{1i}^{-1} \hat{\mathrm{cov}}(Y_i, s_i) V_{2i}^{-1} D_{2i} & D_{1i}^{\mathrm{T}} V_{1i}^{-1} \hat{\mathrm{cov}}(Y_i, z_i) V_{3i}^{-1} D_{3i} \\ D_{2i}^{\mathrm{T}} V_{2i}^{-1} \hat{\mathrm{cov}}(s_i, Y_i) V_{1i}^{-1} D_{1i} & D_{2i}^{\mathrm{T}} V_{2i}^{-1} \hat{\mathrm{cov}}(s_i) V_{2i}^{-1} D_{2i} & D_{2i}^{\mathrm{T}} V_{2i}^{-1} \hat{\mathrm{cov}}(s_i, z_i) V_{3i}^{-1} D_{3i} \\ D_{3i}^{\mathrm{T}} V_{3i}^{-1} \hat{\mathrm{cov}}(z_i, Y_i) V_{1i}^{-1} D_{1i} & D_{3i}^{\mathrm{T}} V_{3i}^{-1} \hat{\mathrm{cov}}(z_i, s_i) V_{2i}^{-1} D_{2i} & D_{3i}^{\mathrm{T}} V_{3i}^{-1} \hat{\mathrm{cov}}(z_i) V_{3i}^{-1} D_{3i} \end{bmatrix} \tag{A3}$$

and $\hat{\mathrm{cov}}(t) = [t - E(t)][t - E(t)]^{\mathrm{T}}$, $\hat{\mathrm{cov}}(s,t) = [s - E(s)][t - E(t)]^{\mathrm{T}}$, all quantities evaluated empirically at $(\hat{\beta}, \hat{\gamma}, \hat{\alpha})$.

The first matrix in brackets in (A1) evaluated at $(\hat{\beta}, \hat{\gamma}, \hat{\alpha})$, $\hat{\Sigma}_{1K}$, converges in probability to the limit as $K \to \infty$ of the block triangular matrix

$$K^{-1} \begin{bmatrix} A & 0 & 0 \\ -B & C & 0 \\ -D & -E & F \end{bmatrix} \tag{A4}$$

where

$$A = \sum_{i=1}^{K} D_{1i}^{\mathrm{T}} V_{1i}^{-1} D_{1i} \quad B = \sum_{i=1}^{K} D_{2i}^{\mathrm{T}} V_{2i}^{-1} \partial s_i / \partial \beta^{\mathrm{T}} \quad C = \sum_{i=1}^{K} D_{2i}^{\mathrm{T}} V_{2i}^{-1} D_{2i}$$

$$D = \sum_{i=1}^{K} D_{3i}^{\mathrm{T}} V_{3i}^{-1} \partial z_i / \partial \beta^{\mathrm{T}} \quad E = \sum_{i=1}^{K} D_{3i}^{\mathrm{T}} V_{3i}^{-1} \partial z_i / \partial \gamma^{\mathrm{T}} \quad F = \sum_{i=1}^{K} D_{3i}^{\mathrm{T}} V_{3i}^{-1} D_{3i}$$

and where

$$\partial s_{it}/\partial \beta = \frac{1}{v_{it}^2}\left[2(Y_{it} - \mu_{it})(-D_{1it})v_{it} - (Y_{it} - \mu_{it})^2 \frac{\mathrm{d}v_{it}}{\mathrm{d}\mu_{it}} D_{1it}\right]$$

$$\partial z_{its}/\partial \beta = \frac{1}{\sqrt{v_{it}v_{is}}}\left[-D_{1it}(Y_{is} - \mu_{is}) - D_{1it}(Y_{it} - \mu_{it}) - \frac{1}{2}s_{it}s_{is}\left(\frac{\mathrm{d}v_{it}}{\mathrm{d}\mu_{it}} D_{1it} v_{is} - \frac{\mathrm{d}v_{is}}{\mathrm{d}\mu_{is}} D_{1is} v_{it}\right)\right]$$

$$\partial z_{its}/\partial \gamma = -\frac{s_{it}s_{is}}{2}[\phi_{it}^{-1} D_{2it} + \phi_{is}^{-1} D_{2is}]$$

The 'sandwich' variance estimator is $\{\hat{\Sigma}_{1K}^{-1}\}\hat{\Sigma}_{2K}\{\hat{\Sigma}_{1K}^{-1}\}^{\mathrm{T}}$.

APPENDIX B: APPROXIMATE JACKKNIFE VARIANCE ESTIMATE

Let H be the matrix in (A4), $U=(U_1^{\mathrm{T}}, U_2^{\mathrm{T}}, U_3^{\mathrm{T}})^{\mathrm{T}}$ and $U_i=(U_{1i}^{\mathrm{T}}, U_{2i}^{\mathrm{T}}, U_{3i}^{\mathrm{T}})^{\mathrm{T}}$ from (A2). Regarding the estimating equation (6) as a system and leaving the ith cluster out, a one step Fisher scoring starting from the full observation estimate is

$$\hat{\theta}_{-i}=\hat{\theta}+H_{-i}^{-1}U_{-i}=\hat{\theta}-H_{-i}^{-1}U_i \tag{B1}$$

where H_{-i} and U_{-i} are the quantities H and U leaving the ith cluster out. The second equation holds because $U=0$ evaluated at $\hat{\theta}$. This gives $\hat{\theta}_{-i}-\hat{\theta}=-H_{-i}^{-1}U_i$, which can be substituted into equation (11) to obtain the approximate jackknife variance estimator,

$$\frac{K-p-r-q}{K}\sum_{i=1}^{K}\{H_{-i}^{-1}\}U_iU_i^{\mathrm{T}}\{H_{-i}^{-1}\}^{\mathrm{T}} \tag{B2}$$

Note that, we only need to store the matrix H from the robust variance calculation, and update it by subtracting the term contributed by the ith cluster to obtain H_{-i}.

Next, we demonstrate the difference in the approximate jackknife variance estimate between our implementation and MAREG. To make it comparable, we exclude the scale part, as in MAREG, by dropping the corresponding rows and columns in H. Our approximate jackknife variance estimator is based on

$$\hat{\theta}_{-i}-\hat{\theta}=-H_{-i}^{-1}U_i=-\left[H-\begin{pmatrix} D_{1i}^{\mathrm{T}}V_{1i}^{-1}D_{1i} & 0 \\ D_{3i}^{\mathrm{T}}V_{3i}^{-1}\partial z_i/\partial\beta^{\mathrm{T}} & D_{3i}^{\mathrm{T}}V_{3i}^{-1}D_{3i}\end{pmatrix}\right]^{-1}\begin{pmatrix} U_{1i} \\ U_{3i}\end{pmatrix} \tag{B3}$$

Some algebra shows that the formula in Reference [2] leads to

$$\hat{\theta}_{-i}-\hat{\theta}=-H_{-i}^{-1}U_i=-\left[H-\begin{pmatrix} D_{1i}^{\mathrm{T}}V_{1i}^{-1}D_{1i} & 0 \\ 0 & D_{3i}^{\mathrm{T}}V_{3i}^{-1}D_{3i}\end{pmatrix}\right]^{-1}\begin{pmatrix} U_{1i} \\ U_{3i}\end{pmatrix} \tag{B4}$$

where the lower-left block is missing in the subtraction from H. This does not affect the variance for mean parameters, but does affect the variance for association parameters.

ACKNOWLEDGEMENTS

The authors are grateful to the two referees whose comments have significantly improved the presentation of this paper. They also acknowledge COGA supported by NIH grants U10AA08401 and U10AA08403 (NIAAA).

REFERENCES

1. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1):13–22.
2. Ziegler A, Kastner C, Brunner D, Blettner M. Familial association of lipid profile: a generalized estimating equations approach. *Statistics in Medicine* 2000; **19**(24):3345–3357.
3. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics* 1988; **44**:1033–1048.
4. Prentice RL, Zhao LP. Estimating equations for parameters in mean and covariance of multivariate discrete and continuous responses. *Biometrics* 1991; **47**:825–839.
5. Liang KY, Zeger SL, Qaqish B. Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society B* 1992, **54**:3–24.
6. Paik MC. Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics—Simulation and Computation* 1988; **17**:1155–1171.
7. Lipsitz SR, Dear BG, Zhao LP. Jackknife estimators of variance for parameter estimated from estimating equations with applications to clustered survival data. *Biometrics* 1994; **50**:842–846.
8. Kastner C, Ziegler A. A comparison of jackknife estimators of variance for GEE2, SFB 386 *Discussion Paper* # 167, Ludwig–Maximilians University of Munich. Available at http://www.stat.uni-muenchen.de/sfb386/publikation.html, 1999 (28 November 2001).
9. Kastner C, Fieger A, Heumann C. MAREG and WinMAREG—a tool for marginal regression models. *Statistical Software Newsletter in Computational Statistics and Data Analysis* 1997; **24**:237–241.
10. Paik MC. Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics* 1992; **48**:19–30.
11. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally M, Nurnberger JI, Tischfield JA, Crowe RR, Cloninger R, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H. Genome-wide search for genes affecting the risk for alcohol dependence. *American Journal of Medical Genetics* 1998; **81**:207–215.
12. McCullagh P, Nelder J. *Generalized Linear Models* (2nd edn). Chapman & Hall: London, 1989.
13. Ziegler A, Kastner C, Blettner, M. The generalised estimating equations: An annotated bibliography. *Biometrical Journal* 1998; **40**:115–139.
14. Yan J. geepack: Yet another package for generalized estimating equations. *R News* 2002; **2**(3):12–14.
15. Chaganty NR, Shults J. On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *Journal of Statistical Planning and Inference* 1999; **76**:145–161.
16. Hall DB, Severini TA. Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association* 1998; **93**:1365–1375.
17. Ziegler A. Practical considerations of the jackknife estimator of variance for generalized estimating equations. *Statistical Papers* (*Formerly: Statistischen Hefte*) 1997; **38**:363–369.
18. Heagerty PJ, Zeger SL. Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* 1996; **91**:1024–1036.

