Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 1197–1212
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

# A SAS/IML software program for GEE and regression diagnostics

Bradley G. Hammill[a], John S. Preisser[b,*]

[a]*Duke Clinical Research Institute, Duke University, Durham, NC, USA*
[b]*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

Received 19 July 2005; received in revised form 21 November 2005; accepted 21 November 2005
Available online 13 December 2005

**Abstract**

A SAS/IML software program is described that computes regression diagnostics for generalized estimating equations. These diagnostics are computationally efficient and accurate approximations for the effect of deleting one observation or one cluster on individual regression coefficients (DFBETA) or on the overall fit of the model (Cook's Distance). New formulae for the diagnostics are presented which are equivalent to those introduced by Preisser and Qaqish [1996. Deletion diagnostics for generalised estimating equations. Biometrika 83, 551–562]. The new formulae expose the relationships of the diagnostic measures to the GEE score equations and to a bias-corrected GEE variance estimator which is also implemented in the SAS macro. The macro is applied to three clustered data sets.


*Keywords:* Cook's distance; Generalized estimating equations; Leverage; Logistic regression; Poisson regression

## 1. Introduction

The widespread availability of the generalized estimating equation (GEE) method in commercial software has popularized population-averaged model analyses of data that consist of clustered or repeated observations (Horton and Lipsitz, 1999). There are several aspects and advances related to GEE methodology that are often overlooked. These include: (1) Certain model diagnostic measures based upon observation or cluster deletion; (2) a small sample bias-corrected covariance estimator; and for correlated binary responses, (3) restrictions on the elements of the working correlation matrix based upon their permissable ranges.

Regression diagnostics for GEE are increasingly used as evidenced by the considerable attention given to them in a recent monograph on GEE (Hardin and Hilbe, 2003). In particular, it is important for an analyst to be able to quantify the effect that any observation or cluster has on the fitted regression model. Preisser and Qaqish (1996) developed formulae for both observation-level and cluster-level deletion diagnostics for models fitted with GEE. These are computationally efficient "one-step" formulae that estimate the change in regression coefficients due to deletion of an observation or cluster. Observation-level refers to individual records in the data, whether that is one time point for a specific subject in a longitudinal study or one record within a cluster, such as a patient within a hospital. Cluster-level refers, naturally, to the entire group of observations from the same cluster, whether that is a single subject followed over time or a group

* Corresponding author. Tel.: +1 919 966 7265; fax: +1 919 966 3804.
*E-mail addresses:* john_preisser@unc.edu, jpreisse@bios.unc.edu (J.S. Preisser).

0167-9473/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2005.11.016

of patients with the same hospital. These diagnostics are analogous to measures like *DFBETA* that exist and are known for linear models (Belsley et al., 1980). Ziegler and Arminger (1996) concurrently proposed cluster-level diagnostics for GEE including Cook's distance.

As an alternative to the GEE empirical covariance estimator, Mancl and DeRouen (2001) developed a small sample bias-corrected covariance estimator for data with a limited number of clusters. In a simulation study of a binary outcome for a model with two binary covariates, they found that the proposed bias-corrected covariance estimator gave test sizes near the nominal significance level even when the number of clusters was as small as 10 and cluster sizes were unequal. In the same set-up, the GEE empirical covariance estimator gave test sizes that were inflated 2–3 times the nominal level. These results for the bias-corrected covariance estimator may be overly optimistic since continuous covariates with high leverage values, not considered, have been shown to impair the performance of covariance estimators of the sandwich form (Kauermann and Carroll, 2001). Nevertheless, Mancl and DeRouen (2001) showed that bias-correction is important in small samples to help protect against the standard errors of GEE parameter estimates being artificially small.

Finally, Prentice (1988), Qaqish (2003) and others have noted that, for correlated binary data, the elements of the working correlation matrix are bounded by the marginal probabilities. If the working correlation matrix is not subject to these bounds the possibility exists that the joint probability for any two observations from a cluster could be negative. A simple solution is to set the correlation estimate to a value just inside the appropriate bound if it is violated. Shults et al. (2005) consider the implications of such violations when implementing GEE.

This paper presents a macro for estimating population-averaged models using GEE that implements these three methodological considerations using SAS PROC IML (SAS Institute Inc., 1999). Section 2 gives a brief overview of generalized estimating equations and of the concepts and computational formulae for the new features implemented. Section 3 discusses the details and usage of the macro. And Section 4 illustrates the capabilities of the macro using three example data sets, which require a variety of link functions, variance functions, and working correlation structures. The data sets were also selected because they demonstrate different analysis problems that might face statisticians working on clustered data.

## 2. Methods

Generalized estimating equations are useful for data in which observations are correlated due to clustering or repeated measures (Liang and Zeger, 1986). Consider $j = 1, 2, \ldots, n_i$ observations from $i = 1, 2, \ldots, K$ clusters where $y_{ij}$ is the response measure for the $j$th observation in the $i$th cluster. Similarly, $x_{ij}$ is a $p \times 1$ vector of covariates. The mean $\mu_{ij} = E\left(y_{ij} \mid x_{ij}\right)$ is related to the covariates by a link function $g(\cdot)$ as $g\left(\mu_{ij}(\beta)\right) = x'_{ij}\beta$. The variance of the response is $\text{var}\left(y_{ij}\right) = \phi h\left(\mu_{ij}\right)$ where $h$ is a given variance function and $\phi$ is a given or estimated scale parameter. The working covariance matrix for cluster $i$ is $V_i = \phi A_i^{1/2} R_i A_i^{1/2}$ where $A_i = \text{diag}\left[h\left(\mu_{i1}\right), \ldots, h\left(\mu_{in_i}\right)\right]$ and $R_i = R_i(\alpha)$ is the working correlation matrix, which depends on a nuisance parameter vector $\alpha$ and which is assumed not to vary by cluster. GEE estimates are determined by iteratively solving $\sum_{i=1}^{K} D_i' V_i^{-1} r_i = 0$, where $r_i = y_i - \mu_i$ and $D_i = \partial\mu_i/\partial\beta'$. The model-based estimator of the covariance matrix of $\hat{\beta}$ is given by

$$V_{\text{model}}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{K} D_i' V_i^{-1} D_i\right)^{-1} \equiv M^{-1}.$$

The empirical estimator of the covariance matrix of $\hat{\beta}$ is given by

$$V_{\text{emp}}\left(\hat{\beta}\right) = M^{-1}\left(\sum_{i=1}^{K} D_i' V_i^{-1} r_i r_i' V_i^{-1} D_i\right) M^{-1}.$$

The latter estimate is robust to the misspecification of the working correlation matrix in the sense that it is a consistent estimator of the true covariance as long as the model for the marginal mean is correctly specified.

The bias-corrected covariance estimator introduced by Mancl and DeRouen (2001) is also robust to such misspecification. However, it has improved finite sample properties. This estimator is given by

$$V_{\text{bias-corr}}\left(\hat{\beta}\right) = M^{-1}\left(\sum_{i=1}^{K} D_i' V_i^{-1}(I - H_i)^{-1} r_i r_i' (I - H_i)^{-1} V_i^{-1} D_i\right) M^{-1},$$

where $H_i = D_i M^{-1} D_i' V_i^{-1}$ is the cluster leverage matrix.

The observation leverage for the $j$th observation of the $i$th cluster, $HOBS_{ij}$, is defined as the value of the $j$th diagonal element of the cluster-level $H_i$ matrix. Cluster leverage for cluster $i$, $HCLS_i$, is defined as the sum of the observation leverages (Preisser and Qaqish, 1996). This has implications when comparing leverage values between clusters of unequal size, as we expect, on average, higher $HCLS_i$ values for larger clusters even in data without high-leverage observations. In general, all of the cluster-level diagnostics presented are expected to have higher values as cluster size increases. Because of this, when examining diagnostics in data with unbalanced clusters, it is important to stratify by or otherwise account for cluster size.

The diagnostic measures proposed for GEE are similar to those that exist for generalized linear models: *DFBETA*, Cook's *D*, and leverage. The diagnostic purpose of each measure is similar as well: *DFBETA* is a measure of the influence that any observation or cluster has on each $\hat{\beta}$ (Belsley et al., 1980); Cook's *D* is a measure of the influence of any observation or cluster on the overall fit of the model (Cook, 1977); and leverage is a measure of how extreme an observation or cluster is with respect to the predictors (Belsley et al., 1980). *DFBETA* and Cook's *D* are referred to as deletion diagnostics because the magnitude of each is related to changes in the fit of the model after a particular observation or cluster is removed compared to the fit of the model on the full data.

Similar to the diagnostics proposed by Pregibon (1981) for logistic regression (see Preisser and Garcia (2005) for generalized linear model computations using SAS), the GEE diagnostics are one-step approximations to the fully iterated "exact" quantities. Since both types of models are estimated using iterative methods, dropping each case one at a time from the data and fully re-estimating the model is computationally costly. Pregibon proposed removing each case, one at a time, from the data and iterating the model one step beyond where it initially converged using the full data. That is the theoretical basis for the diagnostics; in practice, it is interesting to note that with the computational formulae, another iteration of the model is not required to calculate the full set of diagnostics. These one-step diagnostics for GEE are therefore simple to compute and have been shown to be accurate approximations of the fully iterated versions (Preisser and Qaqish, 1996).

The regression diagnostics provided by the macro are just a few of the diagnostic tools available for the assessment of population-averaged models fitted with GEE. Hardin and Hilbe's text provides a more complete review. We note that these authors have reproduced some formulae (i.e., equations 4.48, 4.49 and 4.51) of Preisser and Qaqish (1996; i.e., Eqs. (5), (6) and Corollary 2.2, respectively) incorrectly. In this paper, we have introduced a notation for the *DFBETA* measures that is different from that of Preisser and Qaqish (1996) and more compatible with much of the GEE literature that we hope clarifies their computation.

If $\hat{\beta}_h$ represents the $h$th parameter estimate from the model using the full data set and $\hat{\beta}_{h[i]}$ represents the $h$th parameter estimate from the model after omitting cluster $i$, then let $DFBETACh_i = \hat{\beta}_h - \hat{\beta}_{h[i]}$ be the effect that cluster $i$ has on $\hat{\beta}_h$. Similarly, if $\hat{\beta}_{h[ij]}$ represents the parameter estimate from the model after dropping observation $j$ from cluster $i$, then let $DFBETAOh_{ij} = \hat{\beta}_h - \hat{\beta}_{h[ij]}$ be the effect that observation $j$ from cluster $i$ has on $\hat{\beta}_h$. The vector of all $DFBETACh_i$s for cluster $i$ can be calculated as

$$DFBETAC_i = M^{-1} D_i' V_i^{-1}(I - H_i)^{-1} r_i.$$

We show in the appendix that this formula is equivalent to an alternative representation given in Corollary 1.1 of Preisser and Qaqish (1996). These diagnostics are closely linked to the bias-corrected variance estimator:

$$V_{\text{bias-corr}}\left(\hat{\beta}\right) = \sum_{i=1}^{K} (DFBETAC_i)\,(DFBETAC_i)'$$

The formula for the vector of observation-level $DFBETAOh_{ij}$s is similar, but more complicated due to the intra-cluster correlation.

$$DFBETAO_{ij} = M^{-1}\tilde{D}'_{ij}\,\frac{\tilde{r}_{ij}}{\tilde{V}_{ij}\left(1-\tilde{H}_{ij}\right)},$$

where

$$\begin{aligned}
\tilde{D}_{ij} &= D_{ij} - V_{ij[j]}V_{i[j]}^{-1}D_{i[j]},\\
\tilde{r}_{ij} &= r_{ij} - V_{ij[j]}V_{i[j]}^{-1}r_{i[j]},\\
\tilde{V}_{ij} &= V_{ij} - V_{ij[j]}V_{i[j]}^{-1}V_{i[j]j},\\
\tilde{H}_{ij} &= \tilde{D}_{ij}M^{-1}\tilde{D}'_{ij}\tilde{V}_{ij}^{-1},
\end{aligned}$$

and $V_{ij[j]}$, $V_{i[j]j}$, etc. above are submatrices of the partitioned cluster-level $V_i$ matrix that do not contain the row and/or column corresponding to the $j$th observation of that cluster, as

$$V_i = \begin{pmatrix} V_{ij} & V_{ij[j]} \\ V_{i[j]j} & V_{i[j]} \end{pmatrix}.$$

Note that $\tilde{r}_{ij}$, $\tilde{V}_{ij}$ and $\tilde{H}_{ij}$ are scalar quantities. These *DFBETA* measures are unstandardized. Standardization is simple and is achieved by dividing each element of the $DFBETAC_i$ or $DFBETAO_{ij}$ vector by the standard error of the corresponding parameter estimate from the model based on the full data. These standard errors may be derived from any of the estimated covariance structures: model-based, empirical, or bias-corrected.

Cook's $D$ for a cluster or an observation is calculated by using the vector of $DFBETAC_i$ or $DFBETAO_{ij}$ values, respectively. Let $DCLS_i$ be the cluster-level Cook's $D$ for cluster $i$, which can be calculated as

$$DCLS_i = (DFBETAC_i)'M\,(DFBETAC_i)\,/\,p,$$

where $p$ is the number of predictors in the model, including the intercept. Similarly, let $DOBS_{ij}$ be the observation-level Cook's $D$ for the $j$th observation in cluster $i$, which can be calculated as

$$DOBS_{ij} = \left(DFBETAO_{ij}\right)'M\left(DFBETAO_{ij}\right)/\,p.$$

In this macro, the following specifications of $M$ are available: (1) $M$ = the inverse of the model-based covariance estimate, (2) $M$ = the inverse of the empirical covariance estimate, and (3) $M$ = the inverse of the bias-corrected covariance estimate.

Finally, Prentice (1988) pointed out that, for correlated binary data, the elements $\{\delta_{kl}\}$ of the working correlation matrix $R_i$ are bounded, with limits based on the observed data and the predicted probabilities produced by the model. Here $\delta_{kl}$ denotes the correlation between observations $k$ and $l$ in the same cluster. The joint probability for any two observations in the same cluster is given by

$$\Pr(Y_k, Y_l) = p_k^{Y_k}q_k^{1-Y_k}p_l^{Y_l}q_l^{1-Y_l}\left[1+\delta_{kl}(p_kq_kp_lq_l)^{-\frac{1}{2}}(Y_k-p_k)(Y_l-p_l)\right],$$

where $p_k = \Pr(Y_k = 1)$, $q_k = 1 - p_k$, $p_l = \Pr(Y_l = 1)$, and $q_l = 1 - p_l$. In order to ensure that this joint probability is non-negative, $\delta_{kl}$ must be in the range from $\max\left[-(q_kq_l/p_kp_l)^{1/2}, -(p_kp_l/q_kq_l)^{1/2}\right]$ to $\min\left[(p_kq_l/q_kp_l)^{1/2}, (q_kp_l/p_kq_l)^{1/2}\right]$. Shults et al. (2005) suggest some approaches to overcoming the problem of violation of bounds on $\delta_{kl}$ (and, thus, on $\alpha$). First, they suggest that such a violation can be addressed by the use of an alternative correlation structure. For example, violation of bounds often occur for an exchangeable correlation structure (i.e., $\{\delta_{kl} \equiv \alpha\}$) in the context of a marginal mean model with continuous covariates because, then, the distinctly unique computed values for $\Pr(Y_k, Y_l)$ over all observation pairs and clusters may be quite numerous affording much opportunity to estimate some of them outside of the valid (0, 1) range for probabilities. For longitudinal data, Shults et al. (2005) suggest that one way to incorporate dependence of the correlation structure on covariates is to use an AR(1) working correlation that models the correlation between observations as a function of measurementoccasion. For clustered data, allowing correlations to vary between groups or depend upon cluster-level covariates in general ways can be implemented with

GEE methods that allow greater flexibility in modelling intra-cluster correlations (eg., the method of Prentice (1988) for which the authors provide software at http://www.bios.unc.edu/~jpreisse/software.html). Shults et al. (2005) also suggest that the estimates of $\alpha$ could be adjusted so that the $\delta_{kl}$ are within bounds. The impact of violation of bounds for a data set can be assessed by computing GEE estimates with and without the adjustment.

There are similar, but more computationally difficult, bounds for marginal joint probabilities of three or more observations in a cluster. Checks for the existence of all trivariate (or higher order) marginals are not incorporated in the macro. One of the inherent difficulties with GEE for larger clusters of correlated binary data is the question of the existence of a full multivariate binary distribution given a set of lower order marginals (a distribution exists if there is a compatible set of $2^n$ joint probabilities satisfying the usual constraints- values between 0 and 1 summing to 1—for a multinomial(n) distribution). While the satisfying of bounds imposed by the marginal means and the positive definiteness of $R_i$ are necessary, they are not sufficient conditions for existence. Nonetheless, checks on these conditions are helpful for checking the validity of GEE. Positive definiteness can be checked by determining that the minimum eigenvalue of $R_i$ is greater than 0. Generally, computing eigenvalues is computer-intensive $\left(\text{of order } O\left(n^3\right)\right)$. However, easy-to-check ranges on $\alpha$ for positive definiteness for specific correlation structures can be easily implemented as described in the next section.

## 3. Macro details

The macro uses GEE for the estimation of the population-averaged models and implements the above formulae in SAS/IML. As such, before the macro is called, there is some minor data preparation that must occur. The data set may not contain any records with missing data for the dependent variable or for any of the independent variables. Any categorical variable with $C > 2$ levels needs to be recoded into $C - 1$ binary indicator variables. If an intercept term is desired in the model, a variable, equal to one for all records, must be explicitly created. If a time or order variable is used to indicate within-cluster ordering, the possible values of that variable must be consecutive integers starting at 1. And finally, if there is no time or order variable being used, the data must be pre-sorted by cluster before being passed to the macro. This allows for any implicit ordering within cluster, if set up by the user, to be maintained by the macro.

There are a number of options related to model estimation that the user must or may specify. The macro requires the user to provide a data set name, a dependent or outcome variable, a list of covariates (which includes an intercept term, if desired), a cluster identification variable, a link function, a variance function, and a working correlation structure. These last three options are identical to those in SAS PROC GENMOD. The most commonly specified functions and structures have been implemented in this macro. The link functions available are: Identity, Logarithm, Logit, and Reciprocal; the variance functions available are: Normal, Poisson, Binomial, and Gamma; and the correlation structures available are: Independent, Stationary $M$-dependent, Non-stationary $M$-dependent, Exchangeable, Autoregressive(1), Unstructured, and User-specified. Naturally, if either $M$-dependent structure is requested, $M$ should be specified or 1-dependence is assumed.

Additionally, there are a number of optional model-related specifications. A denominator variable can be given for binomial data, but if not specified is assumed to equal 1. A fixed scale parameter can be specified; otherwise it is assumed to equal 1 for binary data and estimated for other types of data. The estimated scale parameter is assumed to be constant across all observations. Starting values for the regression coefficient estimates may be given; otherwise generalized linear model estimates are computed in the initial step. And a variable containing an offset may be specified, for example, for use in Poisson regression with unequal exposure periods.

With respect to the mechanics of estimation, the user can control the maximum number of iterations allowed, can specify the magnitude of the convergence criteria to be used, and can print out details ($\hat{\beta}$s, working correlation matrix, etc.) of each iteration. Convergence is determined by the maximum relative (or absolute, if desired) change in the $\hat{\beta}$s between iterations. If the working correlation matrix is not positive definite (p.d.), the macro will fail. This can be prevented for a few specific correlation structures that have known bounds on $\alpha$. The p.d. range for exchangeable correlation is $\alpha \in (-1/(n-1), 1)$ and the p.d. range for stationary 1-dependent is $\text{abs}(\alpha) < [-2\cos\{n\pi/(n+1)\}]^{-1}$, where $n$ is max $(n_i)$, the maximum cluster size. These two cases are checked at every iteration; a message is printed for a violation and an adjustment is made so that the correlation is in the valid p.d. range. And finally, for correlated binary data, the bounds on each $\delta_{kl}$—that ensure non-negative joint probabilities for all within-cluster pairs of observations—are enforced by default. At each iteration, the lower and upper bound for each pairwise correlation are checked and if the

estimated correlation is smaller (larger) than the lower (upper) bound it is set to the bound. A user may opt to turn this check off.

The only macro option related to computation of the diagnostic measures is the designation of which covariance estimator is to be used for the standardization of Cook's distance and the *DFBETA*s. All of the covariance estimators computed by the macro—model-based, empirical, or bias-corrected—are available for this purpose. If unspecified, the model-based covariance estimator is used for standardization.

A number of output data sets can be requested by the user. Data sets containing beta estimates, standard errors, and any of the computed covariance matrices are available. These are useful for constructing contrasts and hypothesis tests using SAS/IML. And data sets for both observations and clusters containing the regression diagnostics previously discussed are available. In fact, an output data set for the regression diagnostics must be requested if the user wants to see this information, as the macro does not print these measures to the screen or output file. Since the macro does not reorder the raw data, a simple one-to-one merge between the observation-level diagnostic results and the raw data may be performed to combine the diagnostic information with the raw data for further investigation. Cluster-level diagnostic results can be merged with the raw data using the cluster identification variable.

## 4. Examples

### 4.1. GUIDE data

The data for this example come from the Guidelines for Urinary Incontinence Discussion and Evaluation (GUIDE) study, which has been previously presented and discussed by Preisser and Qaqish (1999). The data are available at `http://www.bios.unc.edu/~jpreisse/data.html`. GUIDE was a randomized controlled trial that assessed if provider adherence to a set of guidelines for treatment of patients with urinary incontinence (UI) affected patient outcomes. Data were collected on 139 elderly patients from 38 medical practices. The number of patients per practice (i.e. cluster) ranged from 1 to 8 and the median was 4 patients. The interest of the present analysis is to determine what predicts whether or not a patient considers their UI a problem that interferes with their daily life. Covariates include gender, age, the number of daily UI accidents, the severity of the loss of urine, and the number of daily visits to the toilet. Details about the coding for the outcome variable and for the model covariates are shown in Table 1. The binary response was modeled with GEE using a logit link function, the binomial variance function, and an exchangeable correlation structure. Two patients with incomplete data for the model variables were dropped from the analysis.

The results from the regression are shown in Table 2. There are two interesting things to note in these results. First, it seems clear that the relatively small number of clusters had some impact on the magnitude of the standard errors, as the bias-corrected standard error estimates range from 8% to 23% larger than the empirical standard error estimates. Second, the exchangeable correlation is estimated as 0.0145, which is less than the previously reported correlation of 0.10 (Preisser and Qaqish, 1999). The estimated correlation is smaller because 0.10 would have resulted in some intra-cluster joint predicted probabilities being negative. The correlation was reduced so that it would fall within the feasible range. Violation of bounds did not have a large practical impact in this example, as estimated coefficients based upon adjusting the correlation estimate are similar to those in Table 2 of Preisser and Qaqish (1999) that did not adjust.

Table 1
Variable coding in GUIDE data

| | |
|---|---|
| *Outcome* | |
| BOTHERED | ("Do you consider this accidental loss of urine a problem that interferes with your day to day activities or bothers you in other ways?"): 1 if Yes; 0 if No |
| *Covariates* | |
| FEMALE | 0 if Male; 1 if Female |
| AGE | Standardized age: (Age in years − 76)/10 |
| DAYACC | Patient report of the number of leaking accidents they experience in an average day (derived from number of accidents reported per week) |
| SEVERE | Severity of the loss of urine: 1 if there is only some moisture; 2 if the patient wet the underwear; 3 if the urine trickled down the thigh ; 4 if the patient wet the floor |
| TOILET | Patient report of the number of times during the day they usually go to the toilet to urinate |

Table 2
GEE results for GUIDE data

| Variable | Parameter estimate | Empirical | | Bias-corrected | |
|---|---|---|---|---|---|
| | | Std error | $Z$-value | Std error | $Z$-value |
| Intercept | −3.2512 | 0.9645 | −3.371 | 1.0560 | −3.079 |
| FEMALE | −0.6853 | 0.6185 | −1.108 | 0.6945 | −0.987 |
| AGE | −0.6467 | 0.5712 | −1.132 | 0.6302 | −1.026 |
| DAYACC | 0.4111 | 0.0969 | 4.241 | 0.1059 | 3.881 |
| SEVERE | 0.8260 | 0.3647 | 2.265 | 0.3944 | 2.094 |
| TOILET | 0.1104 | 0.1004 | 1.099 | 0.1236 | 0.893 |

Working exchangeable correlation = 0.015.

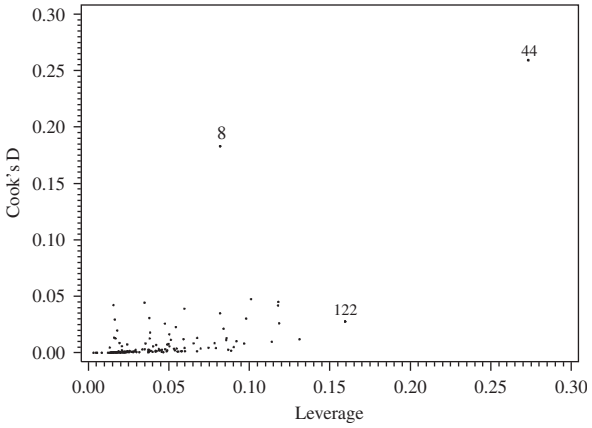


Fig. 1. Observation-level Cook's $D$ by leverage for GUIDE data.

In Fig. 1, which displays observation-level leverage by Cook's $D$, three observations stand out from the rest: Observation 8 has a high Cook's $D$ value, but normal leverage; observation 44 has both a high Cook's $D$ value and high leverage; and observation 122 has a normal Cook's $D$ value, but high leverage. Observation 8 is a 77 year old male who does not report being bothered by his UI, even though he has a high frequency of daily visits to the toilet (10) and a high number of daily accidents (just over 9). This discordance is the factor behind both a large standardized residual (−3.3) and his high $DOBS_{ij}$ value. His low $HOBS_{ij}$ indicates that his moderately high frequency of toileting or daily accidents did not contribute to high leverage. Observation 44 is a 77 year old female that has the highest observed frequency of toilet visits in a day (20). Combined with the fact that she indicates that this does not bother her leads to a high $DOBS_{ij}$ value. Observation 122 is a 79 year old female that is not bothered by her UI, which happens to take on a very infrequent, but severe character. She is an outlier for the fact that she has the lowest number of accidents (1 per week) but is one of only seven patients to report that this accident is most severe.

Fig. 2 shows cluster-level Cook's $D$ values and Fig. 3 shows cluster-level leverage values. Again, since we expect larger clusters to have diagnostic measure values that are higher than those for smaller clusters, we show each measure stratified by cluster size and look for clusters that have high values for their size. It is clear from the plots that clusters 27, 41, 107, and 156 have relatively large Cook's $D$ values, while only clusters 107 and 125 have large leverage values—although the leverage for cluster 125 might actually be as expected, which is hard to know since there are no other clusters containing eight patients. These plots illustrate an interesting point about high cluster diagnostic values. While it is possible for one very extreme observation to affect the diagnostic measure for an entire cluster—note that cluster 27 contains observation 8 and cluster 107 contains observation 44—it is also possible for a cluster to be an outlier or influential because of a lot of moderate, but not extreme, cases. Cluster 156 is a good example of this. All three of the patients in this cluster report being bothered by their UI, even though all indicate low severity of their accidents and have frequencies of visits to the toilet that are below the median. Two of these three cases have standardized residuals

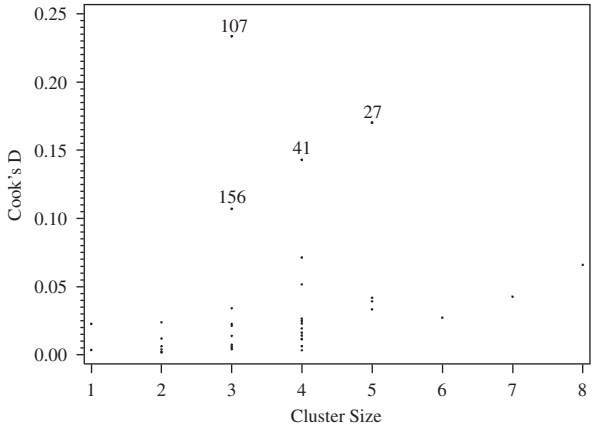


Fig. 2. Cluster-level Cook's *D* for GUIDE data.

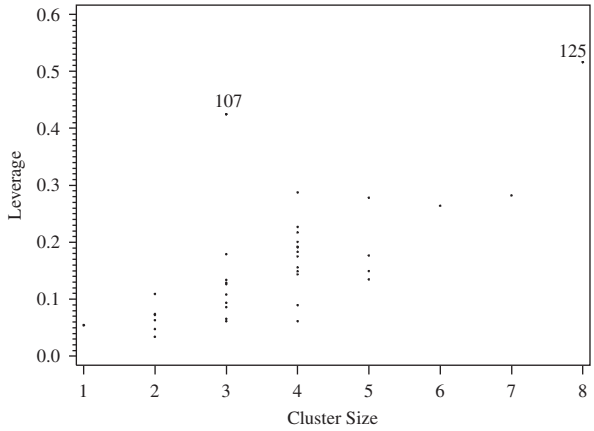


Fig. 3. Cluster-level leverage for GUIDE data.

that are greater than 2.0. The third has a high number of daily accidents that is more consistent with her response. In data situations like this where a cluster is an aggregation of single observations rather than repeated measures on one individual, cluster diagnostics can be a useful tool. High diagnostic values may indicate a measurement issue specific to the cluster; or it might reflect some unmeasured quantity that all the observations from that cluster share.

### *4.2. Drinking data*

The data for the next example come from an evaluation of the Enforcing Underage Drinking Laws Program, which was a non-randomized community trial (Preisser et al., 2003). The question under study was whether or not discretionary grants given to communities to help curb underage drinking had an impact on self-reported alcohol use among 16–20 year olds. Grants were awarded to 61 communities, 52 of which were matched, using propensity scores, with control communities to be used in the evaluation. In each of the 104 total matched communities, telephone surveys were conducted with 15–30 youths at different points in time: pre-intervention, one year later, and two years later. The sample youths were not the same at each interview wave. The analysis below presents results for the 1868 interviews conducted at one year that were not missing response data. The response of interest is how many times in the last 30 days that a youth has ridden in a car with someone who has been drinking. In addition to the intervention—whether or not the community received a grant—other covariates include age, gender, and a measure of religiosity. Details about the coding for the outcome variable and for the model covariates are shown in Table 3. The response was modeled with

Table 3
Variable coding in drinking data

| | |
|---|---|
| *Outcome* | |
| RIDECAR | Number of times in past 30 days respondent has ridden in a car with someone who has been drinking |
| *Covariates* | |
| GRANT | Intervention: Did community receive grant to support local efforts to prevent underage drinking? 1 if Yes; 0 if No |
| MALE | 1 if Male; 0 if Female |
| AGE18PLUS | Age in years: 1 if 18–20; 0 if 16–17 |
| RELIGIOUS | Does respondent regularly attend religious services? 1 if Yes; 0 if No |

Table 4
GEE results for drinking data

| Variable | Parameter estimate | Empirical | |
|---|---|---|---|
| | | Std error | $Z$-value |
| Intercept | −1.0083 | 0.2516 | −4.008 |
| GRANT | −0.2949 | 0.2090 | −1.411 |
| AGE18PLUS | 0.8876 | 0.1644 | 5.399 |
| MALE | 0.1898 | 0.1871 | 1.015 |
| RELIGIOUS | −0.4333 | 0.1869 | −2.318 |
| Scale | 2.3927 | | |

Working exchangeable correlation = 0.0127.

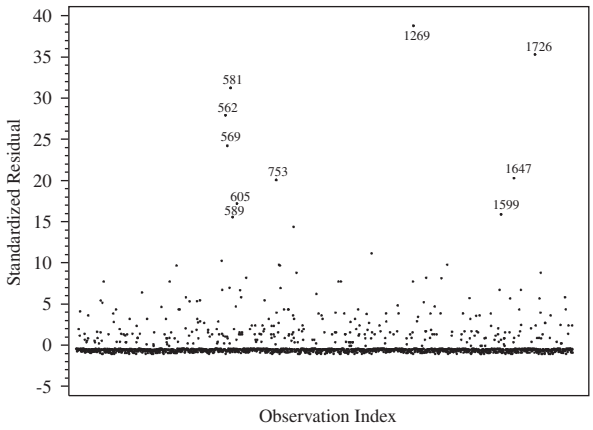


Fig. 4. Observation-level standardized residuals for drinking data.

GEE using a log link function, the Poisson variance function, and an exchangeable correlation structure. In the final data set, the number of youths in each community ranged from 14 to 29, with a mean of 18.

The GEE results are shown in Table 4. The main interest of the model is the treatment effect of the grant. The results show that this effect is in the correct direction, but is not significant at the $\alpha = 0.05$ level. The standardized residuals, shown in Fig. 4, may indicate that this model does not work well for this data. Of the 74 observations with a response value greater than or equal to 3 rides, 73 have residuals greater than 2.0. The excessive number of zero rides reported is a challenge for a Poisson model run on data that exhibits such range of response. Dobbie and Welsh (2001) presented a two-component approach to modelling zero-inflated Poisson data using GEE that may give better results; but estimation of such a model is beyond the scope of this macro at present. Regardless, it is still interesting to look at the diagnostics of the model that was run.

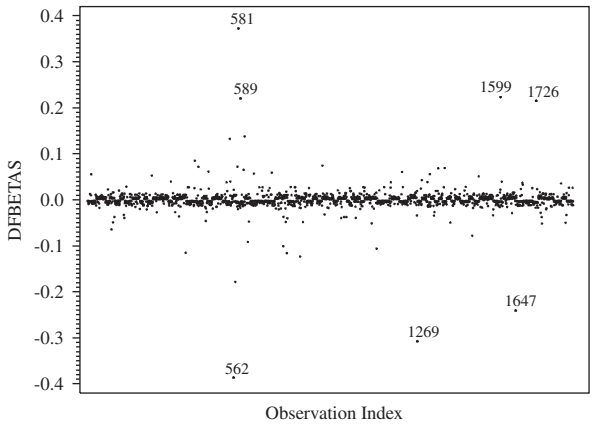


Fig. 5. Observation-level standardized *DFBETA*s for intervention for drinking data.

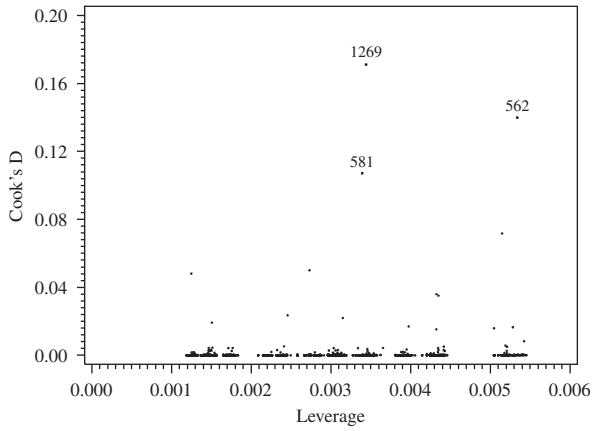


Fig. 6. Observation-level Cook's *D* by leverage for drinking data.

Since the effect of the grants is the focus of the analysis, it makes sense to look at an index plot for the observation-level standardized $DFBETAO_{ij}$ values. Shown in Fig. 5, there are about 7 observations that have substantial impact on the estimate. Perhaps not surprising, these 7 observations all had response values greater than or equal to 15 rides, and represent 7 of the 8 observations with response values this high. The observation-level leverage and Cook's *D* values, shown in Fig. 6, show that of all the observations with extreme response values, only three had values of Cook's *D* that were much larger than average, while none had higher leverage than other observations. The lack of high leverage points is a function of the covariates, all of which were binary. There were only 16 possible combinations of the covariates and all of these 16 groups were well-populated, none having fewer than 70 observations. The observations with high $DOBS_{ij}$ values were the three observations with the most number of rides reported. Observations 581 and 1269, in addition, both responded "Yes" to the religion question, which was somewhat incongruous with their responses, given the strong negative parameter estimate for the religion variable. The cluster-level diagnostics do not yield any additional information, since only clusters that contained an observation with a high number of rides had high diagnostic measure values.

### *4.3. CARDIA data*

The third example uses data from the Coronary Artery Risk Development in Young Adults (CARDIA) study, analyzed previously by Preisser et al. (2000). The data are available at `http://www.bios.unc.edu/~jpreisse/data.html`. CARDIA is an epidemiological cohort study that was funded to document the preva-

Table 5
Dropout patterns and smoking trends in CARDIA data

| Pattern | *N* | % of subgroup | Smoking rates by year (%) | | | |
|---|---|---|---|---|---|---|
| | | | 0 | 2 | 5 | 7 |
| 0 | 27 | 6.7 | 25.9 | | | |
| 2 | 45 | 11.2 | 40.0 | 51.1 | | |
| 5 | 56 | 13.9 | 41.1 | 43.9 | 42.9 | |
| 7 | 275 | 68.2 | 31.6 | 34.4 | 36.8 | 36.4 |
| Overall | 403 | 100 | 33.5 | 37.8 | 38.0 | 36.4 |

Table 6
Variable coding in CARDIA data

| | |
|---|---|
| *Outcome* | |
| SMOKE | Smoked at least five cigarettes per week, almost every week, for at least three months: 1 if Yes, 0 if No |
| *Covariates in basic model* | |
| AGE | Age in years |
| ED_LEHS | Highest educational attainment was a high school diploma or less: 1 if Yes, 0 if No |
| ED_SMCL | Highest educational attainment was some college (but no degree): 1 if Yes, 0 if No |
| YR_FU | Year of follow-up: 0, 2, 5 or 7 |
| *Additional covariates in pattern mixture model* | |
| PATTERN0 | Last year of follow-up was Year 0: 1 if Yes, 0 if No |
| PATTERN2 | Last year of follow-up was Year 2: 1 if Yes, 0 if No |
| PATTERN5 | Last year of follow-up was Year 5: 1 if Yes, 0 if No |
| YR_FU x PATTERN2 | Interaction between year of follow-up and Pattern 2 |
| YR_FU x PATTERN5 | Interaction between year of follow-up and Pattern 5 |

lence of both coronary artery disease (CAD) and the suspected risk factors for CAD in a population of young adults over time. Baseline data were collected on 5115 men and women aged 18–30 years in 1985 and 1986. This cohort was followed and interviewed every subsequent few years. The analysis below analyzes the relationship between pattern of longitudinal survey participation and smoking trends across the first four waves of data collection: Years 0 (baseline), 2, 5 and 7. Only data for the 403 black males in the cohort born between 1955 and 1958 will be included, because this is a group with especially high dropout rates. The first few columns of Table 5 show the response patterns in this subgroup. Pattern 0 refers to the group of respondents who dropped out after Year 0 data collection; Pattern 2 refers to the group of respondents who dropped out after Year 2 data collection; Pattern 5 refers to the group of respondents who dropped out after Year 5 data collection; and Pattern 7 is the group of respondents who had completed Year 7 data collection. By year 7, only 68% of these men were still being followed. The last four columns of Table 5 show the smoking rates by response pattern. It is clear from this table that smoking has slightly increased over time for all patterns of dropout.

Regression models will be used to assess if the increase is significant over all groups and whether or not it is significant by pattern. Therefore, two models are presented below. The response for both is a binary variable representing whether or not the person smoked regularly. The covariates for the first model are age, education attainment, and year of follow-up. The covariates for the second model are the same, but also include variables representing when a person dropped out of the study and an interaction between this dropout pattern and the year of follow-up. To ensure that model parameters are identifiable, we assume a linear effect for year of dropout and we do not fit a separate slope for subjects with Pattern 0 since these had only one observation. This second model is known as a pattern-mixture model and enables us to decompose the effect of year of follow-up on smoking trends by dropout pattern (Little, 1995; Little and Rubin, 2002). This will enable us to understand whether dropping out of the study was related at all to smoking or the onset of smoking. If it is, the observed rise in smoking across the 7 years of the study may well have been higher if participation had been complete. Coding for all model variables is shown in Table 6. Both models use GEE with a

Table 7
GEE results for CARDIA data: basic model

| Variable | Parameter estimate | Empirical | |
|---|---|---|---|
| | | Std error | $Z$-value |
| Intercept | −5.0816 | 1.4064 | −3.613 |
| AGE | 0.1299 | 0.0691 | 1.880 |
| ED_LEHS | 2.3812 | 0.4642 | 5.130 |
| ED_SMCL | 1.4812 | 0.4737 | 3.127 |
| YR_FU | 0.0342 | 0.0147 | 2.328 |

Table 8
GEE results for CARDIA data: pattern mixture model

| Variable | Parameter estimate | Empirical | |
|---|---|---|---|
| | | SE | $Z$ |
| Intercept | −5.4657 | 1.4606 | −3.742 |
| PATTERN0 | −0.5977 | 0.4629 | −1.291 |
| PATTERN2 | 0.2673 | 0.3334 | 0.802 |
| PATTERN5 | 0.3400 | 0.3186 | 1.067 |
| AGE | 0.1459 | 0.0714 | 2.045 |
| ED_LEHS | 2.3889 | 0.4661 | 5.126 |
| ED_SMCL | 1.4590 | 0.4752 | 3.070 |
| YR_FU | 0.0324 | 0.0156 | 2.076 |
| YR_FU × PATTERN2 | 0.2238 | 0.1044 | 2.143 |
| YR_FU × PATTERN5 | −0.0211 | 0.0507 | −0.417 |

| Working correlation matrix | | | | |
|---|---|---|---|---|
| | Year 0 | Year 2 | Year 5 | Year 7 |
| Year 0 | 1 | 0.724 | 0.724 | 0.630 |
| Year 2 | 0.724 | 1 | 0.872 | 0.748 |
| Year 5 | 0.724 | 0.872 | 1 | 0.912 |
| Year 7 | 0.630 | 0.748 | 0.912 | 1 |

logit link function, binomial variance function, and an unstructured working correlation matrix. Because the missing data within each pattern group is not always monotone (i.e. there are some intermittent missing cases), a time variable is used to properly inform the structure of the correlation matrix.

Table 7 shows the results from the basic model; and Table 8 shows the results from the pattern mixture model. A Wald-type test that the five additional terms in the pattern mixture model are simultaneously equal to zero gives a test of whether smoking trends vary by pattern of dropout. The observed $\chi^2$ for this test is 10.7 with 5 degrees of freedom ($p=0.057$). While not statistically significant at the $\alpha=0.05$ level, this result is nonetheless suggestive of a relationship between smoking and dropout indicating that the data missing completely at random (MCAR) assumption of GEE is violated. The effect of year of follow-up in the basic model is statistically significant and the odds ratio of being a regular smoker for each additional year of follow-up is 1.036. In the pattern mixture model, the effect of year of follow-up for the complete-cases is still significant with a slightly lower odds ratio than before (1.033). The effect of year of follow-up for the men who dropped out after Year 2, however, is quite substantial and is statistically significant. The odds ratio for smoking given an additional year of follow-up among this group is 1.29, which leads to suppose that the trend of increased smoking in this subgroup of young black men across the seven years of this study might have been more dramatic had there been no dropouts.

The purpose here is to consider influential data on the pattern mixture model results. It follows that, had there been no dropouts, the overall smoking trend, based on all dropout patterns combined, may have exhibited a larger increase over time than the observed trend reflected in the last row of Table 5 and estimated by the year effect in Table 7. Such a result

would be consistent with the weighted GEE analysis of Preisser et al. (2000) based upon the selection modeling of dropouts. Fitzmaurice and Laird (2000) discuss how to combine slopes across dropout patterns in a modeling approach for handling nonignorable dropouts based on GEE.

Cluster-level diagnostics are most salient for this analysis because it is of interest to know if any individuals in the study are particularly influential. Figs. 7–9 show the standardized *DFBETA*s by number of survey waves completed (i.e. cluster size) for each of the variables in the pattern mixture model that involve year of follow-up: YR_FU, the YR_FU x PATTERN2 interaction term, and the YR_FU x PATTERN5 interaction term. None of the clusters look to have had an unusually large impact on the *DFBETA* for the overall year of follow-up; but there do appear to be a handful of clusters that had relatively large impacts on the *DFBETA*s for year of follow-up for the Pattern 2 and Pattern 5 respondents. Further investigation shows that each of these influential clusters were the respondents with that pattern of dropout whose smoking status changed over the course of the study. As such, to consider dropping them from a re-analysis of the data would be problematic, since it would lead to biased estimates of change in the prevalence of smoking over time. Similar to the data from the second example involving underage drinking, there is limited variation in the covariates among these data.

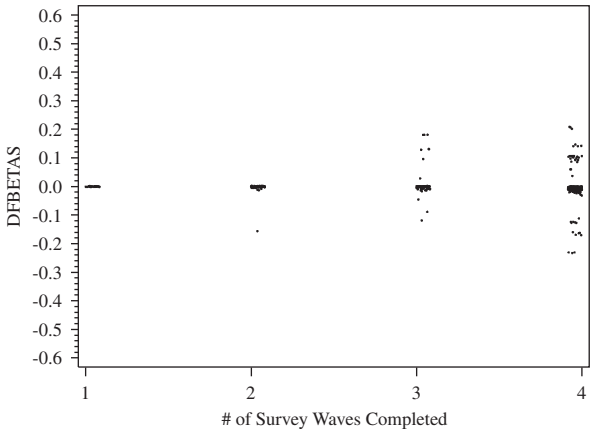


Fig. 7. Cluster-level standardized *DFBETA*s for year of follow-up in CARDIA data.

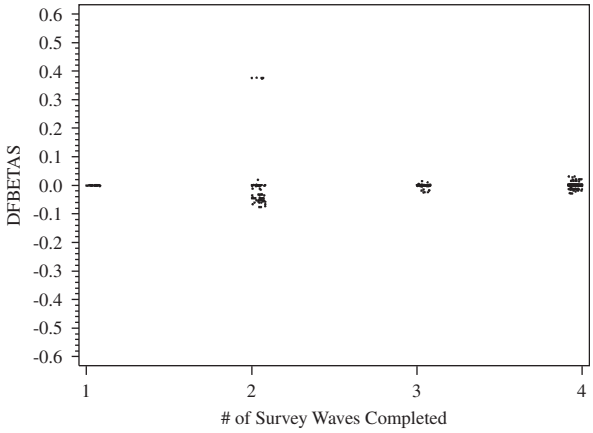


Fig. 8. Cluster-level standardized *DFBETA*s for year of follow-up in CARDIA data (Pattern 2).

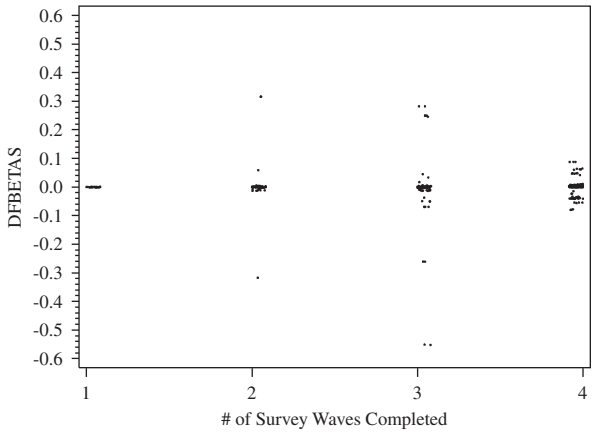


Fig. 9. Cluster-level standardized *DFBETA*s for year of follow-up in CARDIA data (Pattern 5).

## 5. Discussion

The SAS/IML macro presented in this paper goes beyond the existing capabilities of SAS (v9) PROC GENMOD in estimating population-averaged models by enabling analysts to request model diagnostics; by presenting the bias-corrected covariance estimator, which is useful in situations with limited numbers of clusters; and by enforcing restrictions on the elements of the working correlation matrix for binary data. As a result, this macro is an important tool for analysts that are using GEE methods to analyze clustered or repeated observations.

While the macro provides deletion diagnostics to assess the effect on individual clusters and observations on the fit of a model, other techniques and measures to assess a model fitted with GEE were not incorporated. For example, Hardin and Hilbe (2003, Section 4.4) review procedures that have been proposed for goodness-of-fit testing. While the existing procedures are helpful in some circumstances, they are imperfect. Indeed, the goodness-of-fit problem is a difficult one that is expected to undergo further consideration in coming years. At present there does not seem to be one broadly accepted GEE goodness-of-fit method that exemplifies state-of-the-art practice. Therefore, we opted not to include a goodness-of-fit test in the software at this time.

To elaborate, we point to recent work by Lee and Qaqish (2004) who discuss some of the difficulties in developing GEE goodness-of-fit procedures, and who propose a "goodness of marginal fit" test for correlated binary responses when all the covariates are categorical and the contingency table has sufficient sample size to justify large sample results. This offers a limited scenario, as in the asymptotically related weighted least squares approach to fitting linear models to functions of cell probabilities (see appendix to Chapter 13 in Stokes et al., 2000), where the goodness-of-fit test is well-defined. Generally, however, when there are continuous covariates or when the data are otherwise sparse, assessing goodness-of-fit is difficult even for logistic regression (i.e., independent responses) as pointed out by McCullagh and Nelder (1989, Section 4.4.5). Lee and Qaqish (2004) point out that such limitations apply to the correlated data case of GEE as well. Following McCullagh and Nelder, the best we may do is to assess the significance of adding additional terms that have scientific meaning to the model. The macro presented here enables the user to construct hypothesis tests to compare nested models by providing an option to output $\hat{\beta}$ and its estimated covariance matrix (NCOVOUT, RCOVOUT, BCOVOUT depending upon the variance estimator desired).

The macro addresses some special issues involving the feasibility of working correlation parameter estimates in applying GEE to correlated binary data. Specifically, at each iteration in the fitting algorithm, the macro computes the lower and upper bounds of the elements in the correlation matrix based upon the current estimates of the marginal means, and sets the estimates to the appropriate bounds if they are violated. It is not clear whether violation of bounds causes problems in practice. Rochon (1998) states that it appears to cause no difficulty, whereas Chaganty and Joe (2004) argue that ignoring such violation may lead to incorrect analysis. Shults et al. (2005) report on results of a simulation study and surmise "the impact of a violation of bounds on estimation of the regression parameter may not be severe, although this may not be true in other situations." As an alternative to GEE, these authors have suggested quasi-least squares (Shults and Morrow, 2002), a method that uses GEE-estimation of $\beta$ but offers an alternative approach to

estimation of the correlation. Finally, a referee has pointed out that a different implementation of GEE for correlated binary data uses pairwise odds ratios to model the within-cluster association structure (Lipsitz et al., 1991). Unlike correlations, there are no bounds imposed by the marginal means. However given that GEE does not specify a full likelihood, similar issues regarding its existence persist.

## Appendix

### A.1. Equivalence of $DFBETAC_i$ to formula of Preisser and Qaqish (1996)

In their Corollary 1.1, Preisser and Qaqish (1996) give the one-step approximation for $\hat{\beta} - \hat{\beta}_{[i]}$ as

$$M^{-1}X_i'\left(W_i^{-1} - Q_i\right)^{-1}E_i,$$

where $W_i = L_i^{-1}V_i^{-1}L_i^{-1}$, $Q_i = X_i M^{-1}X_i'$, $E_i = L_i r_i$, $L_i = \text{Diag}\left\{\partial\eta_i/\partial\mu_i\right\}$ and $\eta_i = X_i\beta$. Noting that $Q_i = L_i H_i V_i L_i$ and $D_i = X_i' L_i^{-1}$ it is easy to show that $X_i'\left(W_i^{-1} - Q_i\right)^{-1}E_i = D_i' V_i^{-1}(I - H_i)^{-1}r_i$ which proves the equivalency of Corollary 1.1 to $DFBETAC_i$ in Section 2.

### A.2. Sample SAS code

The software is available at `http://www.bios.unc.edu/~jpreisse/software.html`. Instructions for using the software and specific details about options are documented in the macro itself. Sample code for including and calling the macro follow:

```
%include "diag102.sas";
%gee(data =mydata,
yvar =response,
xvar =inter age ed_lehs ed_smcl yr_fu,
time =timevar,
id =patid,
link =3,
vari =3,
corr =6,
clsout =geecls,
obsout =geeobs,
rcovout =geercov);
```

## References


Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Chaganty, N.R., Joe, H.J., 2004. Efficiency of generalized estimating equations for binary responses. J. Roy. Stat. Soc. Ser. B 66, 851–860.

Cook, R.D., 1977. Detection of influential observations in linear regression. Technometrics 19, 15–18.

Dobbie, M.J., Welsh, A.H., 2001. Modelling correlated zero-inflated count data. Aust. N. Z. J. Stat. 43, 431–444.

Fitzmaurice, G.M., Laird, N.M., 2000. Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. Biostatistics 1, 141–156.

Hardin, J.W., Hilbe, J.M., 2003. Generalized Estimating Equations. Chapman & Hall/CRC, Boca Raton.

Horton, N.J., Lipsitz, S.R., 1999. Review of software to fit generalized estimating equation (GEE) regression models. Am. Stat. 53, 160–169.

Kauermann, G., Carroll, R.J., 2001. A note on the efficiency of sandwich covariance matrix estimation. J. Am. Stat. Assoc. 96, 1387–1398.

Lee, J.-H., Qaqish, B.F., 2004. Modified GEE and goodness of the marginal fit (GOMF) test with correlated binary responses for contingency tables. Biometrical J. 46, 675–686.

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

Lipsitz, S.R., Laird, N.M., Harrington, D.P., 1991. Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. Biometrika 78, 153–160.

Little, R.J.A., 1995. Modelling the drop-out mechanism in repeated measures studies. J. Am. Stat. Assoc. 88, 125–134.
Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data. second ed. John Wiley, New York.
Mancl, L.A., DeRouen, T.A., 2001. A covariance estimator for GEE with improved small-sample properties. Biometrics 57, 126–134.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. second ed. Chapman & Hall, London.
Pregibon, D., 1981. Logistic regression diagnostics. Ann. Stat. 9, 705–724.
Preisser, J.S., Garcia, D.I., 2005. Alternative computational formulae for generalized linear model diagnostics: identifying influential observations with SAS software. Comput. Stat. Data Anal. 48, 755–764.
Preisser, J.S., Qaqish, B.F., 1996. Deletion diagnostics for generalised estimating equations. Biometrika 83, 551–562.
Preisser, J.S., Qaqish, B.F., 1999. Robust regression for clustered data with application to binary responses. Biometrics 55, 574–579.
Preisser, J.S., Galecki, A.T., Lohman, K.K., Wagenknecht, L.E., 2000. Analysis of smoking trends with incomplete longitudinal binary responses. J. Am. Stat. Assoc. 95, 1021–1031.
Preisser, J.S., Young, M.L., Zaccaro, D.J., Wolfson, M., 2003. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. Stat. Med. 22, 1235–1254.
Prentice, R.L., 1988. Correlated binary regression with covariates specific to each binary observation. Biometrics 44, 1033–1048.
Qaqish, B.F., 2003. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. Biometrika 90, 455–463.
Rochon, J., 1998. Application of GEE procedures for sample size calculations in repeated measures experiments. Stat. Med. 17, 1643–1658.
SAS Institute Inc., 1999. SAS/IML User's Guide, Version 8. SAS Institute Inc., Cary, NC, 846pp.
Shults, J., Morrow, A.L.M., 2002. The use of quasi-least squares to adjust for two sources of association. Biometrics 58, 521–530.
Shults, J., Sun, W., Tu, X., Amsterdam, J., 2005. On the violation of bounds for the correlation in generalized estimating equation analyses of binary data from longitudinal trials. Technical Report 200501, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine.
Stokes, M.E., Davis, C.S., Koch, G.G., 2000. Categorical Data Analysis Using the SAS System. second ed. SAS Institute Inc., Cary, NC.
Ziegler, A., Arminger, G., 1996. Parameter estimation and regression diagnostics using generalized estimating equations. In: Faulbaum, F., Bandilla, W. (Eds.), SoftStat '95. Advances in Statistical Software 5. Lucius & Lucius, Stuttgart, pp. 229–237.