A Simple Approach to Fitting Non-Standard GEEs

By M. CONTRERAS

Cornell University, USA

and L. M. RYAN[1]

Harvard School of Public Health and Dana-Farber Cancer Institute, USA

SUMMARY

In this paper we present an easy estimation approach for solving non-standard GEEs based on global optimization methods. Our approach has the added advantage that it can be easily implemented using existing software, such as the subroutine *nlminb* in Splus, which also allows for the inclusion of constraints. To illustrate our approach, we fit a non-linear dose response model with constraints to clustered binary data from a developmental toxicity study.

Keywords:Non-standard GEE's, Global optimization numerical methods, developmental toxicity study

1. Introduction

The use of Generalized Estimating Equations (GEEs) has become increasingly popular in recent years, especially now that software is readily available. For example, the generalized linear models procedure (PROC GENMOD) in the latest version of SAS (SAS 6.12) includes the option to fit GEEs. Among the classic references for the theory of estimating equations are Wedderburn (1974), Liang and Zeger (1986), Zeger and Liang (1986), Godambe (1991), and Diggle, Liang and Zeger (1996). Although standard statistical theory

[1] *Address for correspondence:* Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA.

extends to broader settings, GEEs have been applied primarily in the context of generalized linear models, mostly due to capabilities of existing software.

In this article, we describe a simple approach to fitting non-standard GEEs using a general numerical minimizing routine, such as *nlminb* in Splus. Our approach is based on an estimation method which involves minimizing the inner product of the estimating equations or, more generally, calculating the least squares solution to the estimating equations.

This paper is organized as follows. In sections and , we briefly review the theory of GEEs, discuss some of the algorithms, and motivate the need for alternatives for the non-linear or the non-standard case. Then, we introduce the least squares estimate to the estimating equations, and for completeness present algorithms or iterative procedures to obtain the estimators. In Section , using the Splus subroutine *nlminb*, we illustrate the usefulness and ease of this approach by fitting a non-linear dose response model with constraints to clustered binary data from a developmental toxicity study. We also compare the results obtained via our method with and that of the Maximum Likelihood Estimate.

We conclude with a discussion about the theoretical relationship between the least squares estimate to the estimating equations and standard methodology. In particular we give regularity conditions under which this approach enjoys the same statistical properties (consistency and asymptotic normality) as common methodology for solving the estimating equations.

2. Generalized Estimating Equations (GEEs)

Suppose $\mathbf{Y}_i$ denotes an $n_i \times 1$ vector of outcomes for individual $i, i = 1, ...I$. Depending on the application of interest, the elements of $\mathbf{Y}_i$, $(Y_{i1}, Y_{i2}, ..., Y_{in_i})$, might represent repeated observations over time or measurements on a set of related individuals (e.g. a family) or a set of multiple outcomes measured on the same individual. Let $\mathbf{X}_i$ be a corresponding $n_i \times p$

matrix of covariates associated with the vector $\mathbf{Y}_i$. In the most familiar examples of GEEs, the mean of Y_{ij} is related to a linear function of the covariates $\mathbf{X}_{ij}$ through a link function, h:

$$\mu_{ij} = E\left(Y_{ij}\right) = h(\mathbf{X}_{ij}\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is a vector of unknown regression coefficients and $\mathbf{X}_{ij}$ is the jth row of $\mathbf{X}_{ij}$. Usually, the variance of Y_{ij} is chosen to be a suitable function of μ_{ij}, and the covariance matrix of $\mathbf{Y}_i$ is then written as

$$\mathbf{V}_i \;=\; \mathbf{A}_i^{1/2}\,\mathbf{R}_i\,\mathbf{A}_i^{1/2}, \tag{1}$$

where $\mathbf{A}_i = diag(var(Y_{ij}))$ and $\mathbf{R}_i$ is a correlation matrix. In general, the correlation matrix $\mathbf{R}_i$ will also depend on unknown parameters, which will need to be estimated. Assuming, momentarily, that parameters characterizing $\mathbf{R}_i$ are known, and using the theory of generalized estimating equations (Wedderburn (1974), Liang and Zeger (1986), and Zeger and Liang (1986)) the GEE estimate is obtained by solving

$$\sum_{i=1}^{I} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\,\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{2}$$

where $\boldsymbol{\mu}_i$ is the vector of means, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots \mu_{in_i})^T$.

Generally, the correlation matrix $\mathbf{R}_i$ will also depend on unknown parameter(s), ϕ, that need to be estimated from the data. Liang and Zeger (1986) and Zeger and Liang (1986) suggest an iterative approach whereby one solves (2) conditional on $\boldsymbol{\phi}^{(0)}$, then updates the correlation estimates by equating the empirical correlation of the Pearson residuals with the assumed correlation, possibly with a degree of freedom adjustment. Iteration proceeds until some established convergence criterion has been met. For instance, suppose $\mathbf{R}_i$ is the exchangeable correlation matrix, so that ϕ is the assumed common correlation parameter. Then, the estimate of ϕ recommended by Liang and Zeger (1986) would be

$$\hat{\phi} = \frac{1}{M-p}\sum_{i=1}^{I}\sum_{j\neq k} r_{ij}r_{ik}, \tag{3}$$

where r_{ij} is the Pearson residual,

$$r_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{var(Y_{ij})}}.$$

If the assumed covariance matrix $\mathbf{V}_i$ is correct, then the variance of the solution to (2) can be consistently estimated by:

$$\mathbf{W} = \frac{1}{I}\sum_{i=1}^{I} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T. \tag{4}$$

However, one of the most widely appreciated aspects of GEEs is that even if the assumed $\mathbf{V}_i$ is incorrect, under the proper regularity assumptions, the solution to (2) remains consistent, and its variance can be estimated by $\mathbf{W}^{-1}\boldsymbol{\Sigma}\mathbf{W}^{-T}$, where $\boldsymbol{\Sigma}$ is an empirical estimate of the variance of the estimating equation,

$$\boldsymbol{\Sigma} = \frac{1}{I}\sum_{i=1}^{I} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-T} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T. \tag{5}$$

Over the past decade, standard GEEs have become so popular that software to fit them is now widely available in packages such as *Splus* and *SAS*. For example, the latest version of SAS's *PROC GENMOD* allows users to "mix and match" with choices of several link functions relating the mean response to a linear function of covariates (logit, probit, complementary log-log, inverse) and variance functions, including the Gaussian ($var(Y_{ij}) = \sigma^2$), binomial($var(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$), and Poisson ($var(Y_{ij}) = \mu_{ij}$). Users can choose between several standard correlation matrices (e.g. exchangeable, autoregressive) or define their own.

3. Non-standard GEEs

In practice, statisticians sometimes are interested in fitting models that fall outside the range of options offered by existing software. One such situation arises when the mean μ_{ij} cannot be expressed as a function of a linear combination of covariates, but instead is some nonlinear function of $\mathbf{X}_{ij}$ and an unknown parameter, $\boldsymbol{\beta}$,

$$\mu_{ij} = h(\mathbf{X}_{ij}, \boldsymbol{\beta}).$$

One approach to fit these type of non-linear GEEs is to use the linearizing technique described by McCullagh and Nelder (1989) for generalized linear models. Ryan (1992) describes the use of this method for fitting non-linear dose response models to teratology data. This linearizing technique is appealing for non-computational experts, since it requires setting up a simple iteration that repeatedly calls standard GEE software. However, from a practical point of view there are several disadvantages. For example the technique does not allow one to incorporate constraints without substantial programming. Also, standard software may not allow easily for the user's choice of link function. For example, this approach will not help in fitting the so called "multi-hit" dose response model which assumes

$$\mu_{ij} = 1 - exp(-\sum_{j=0}^{p} \beta_j x_i^j),$$

where p is a fixed integer (usually 1, 2 or 3) and the β_j's are constrained to be non-negative.

For the teratology application to be discussed in Section , we will be interested in classes of dose-response models of the form

$$\mu_{ij} = h(\beta_0 + \beta_1 x_i^{\beta_2}),$$

where x_i is the dose level assigned to the ith litter and $h(.)$ is some function that takes value between 0 and 1.

4. Numerical estimation

There is a large literature on numerical techniques for solving general non-linear equations of the form $\mathbf{U}(\boldsymbol{\beta}) = 0$. Detailed discussion can be found in a variety of places, including the classic textbooks on iterative solutions to non-linear equations by Ortega and Rheinboldt (1970) and Dennis and Schnabel (1986). Among the most popular iterative estimation approaches is Newton's method which involves an iterative algorithms of the form

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k - \mathbf{J}(\boldsymbol{\beta}_k)^{-1}\mathbf{U}(\boldsymbol{\beta}_k) \tag{6}$$

where $\boldsymbol{\beta}_k$ is the estimated value of $\boldsymbol{\beta}$ at iteration k and $\mathbf{J}$ is the Jacobian matrix of $\mathbf{U}$ (that is, the derivative matrix of $\mathbf{U}$ with respect to $\boldsymbol{\beta}$).

Newton's method is known to converge to the solution to the estimating equation under standard regularity assumptions, see Dennis and Schnabel (1986). Nonetheless, a disadvantage is that depending on the form of the mean function μ_{ij}, setting up the iteration correctly may require extensive programming, and it is generally not a recommended avenue for the non-computational expert. Furthermore, even a relatively slight modification of the mean function may require substantial programming changes, particularly if constraints are involved since then (6) must be extended to accommodate constraints via the inclusion of Lagrange multipliers both in the computation of the Jacobian and estimating equations.

5. Proposed approach

We now describe an approach to solving non-linear GEEs that involves the use of global optimization methods.

As described by Dennis and Schnabel (1986, Section 6.5, page 147), solving $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ can be conveniently translated to the problem of minimizing over $\boldsymbol{\beta}$ the quadratic form

$$||\mathbf{U}(\boldsymbol{\beta})||^2 := \mathbf{U}(\boldsymbol{\beta})^T\mathbf{U}(\boldsymbol{\beta}), \tag{7}$$

which is the inner product of the estimating equations.

More is said about (7) in Section . Note, however, that from a practical perspective, once the inner product has been formed, then it can be specified as the objective function for a suitable non-linear global optimization program. An obvious advantage of this approach is that it can be implemented using any computer package that has a global optimization routine. In *Splus*, for example, the subroutine *nlminb* can be used. Moreover, this subroutine has the added advantage that it allows the easy incorporation of constraints.

In addition, (7) provides a simple framework for estimating the correlation parameters

and for modeling them as a function of covariates. The idea is to recognize that the estimator $\hat{\phi}$, from (3), can also be expressed as the solution to a set of estimating equations. Hence, one simply has to add this equation into the inner product to be fed into the optimization program being used. Further discussion about the idea of setting up additional equations to estimate the unknown correlation parameters can be found in Liang, Zeger and Qaqish (1992) who refer to this method of estimating the correlation parameters as GEE2.

To illustrate how this estimation approach works, we next consider a concrete example from the teratology literature. Using *nlminb*, we present the numerical results of this method to those obtained via maximum likelihood theory or the MLE.

6. Application

The need to fit non-linear models via generalized estimating equations arises often in the analysis of teratology data. In a typical teratology study (depicted in Figure 1), pregnant dams (usually mice, rats or sometimes rabbits) are randomized to a control group or one of 3 or 4 exposed groups. Dams are exposed to the test substance during the period of major oganogenesis when the developing offspring are likely to be most sensitive to insult. Just prior to normal delivery, the dams are sacrificed and the uterine contents examined for defects. A typical study might have 25 to 30 dams per group, with anywhere from 1 to 20 offspring per litter.

Insert figure 1 about here

For the teratology example, Y_{ij} might represent the weight of the jth pup from the ith litter, or alternatively, might be a binary indicator of whether or not the pup was defective (e.g. dead or malformed). Generally, x_i will denote the dose level for the ith litter, though it is also possible that pup-specific covariates might be included. Figure 2 provides a graphical representation of data from a study in DEHP, an industrial plasticizer. Each dot corresponds to the response rate for a particular litter, while the crosses show the overall response rate

within each dose group. The study involved a total of 131 dams, including 30 controls and 101 exposed to one of 4 different dose groups ranging from .044mg/kg to .292 mg/kg. The lines shown in the Figure correspond to various fitted values, to be discussed presently. The full data set was given by Chen and Kodell (1989).

Because litter mates tend to respond more similarly than non-litter mates, it is important to use statistical methods that properly allow for within litter correlations. Earlier suggestions (e.g. Williams (1975)) involved use of a beta-binomial distribution, mainly because of its conceptual simplicity and a certain biological appeal. More recently, however, GEEs have become popular because of their robustness properties and ease of use.

Many authors have discussed the importance of having flexible classes of dose response models. A popular one is

$$\mu_i = h(\beta_0 + \beta_1 x_i^{\beta_2}), \tag{8}$$

where $h(.)$ is a cdf. The most familiar example of a logistic model corresponds to $h(x) = exp(x)/(1 + exp(x))$ and $\beta_2 = 1$. A particularly popular choice (corresponding to the one-hit model used in carcinogenesis when $\beta_2 = 1$) is $h(x) = 1 - exp(-x)$ with the additional constraint that β_0 and β_1 are positive. As discussed in previous sections standard GEE programs do not allow one to easily fit this mean function, that is, specialized programming is required.

Chen and Kodell (1989) fit a beta-binomial model to the DEHP data assuming the one-hit model with a power parameter, that is,

$$\mu_i = 1 - \exp\left(-\beta_0 - \beta_1 x_i^{\beta_2}\right). \tag{9}$$

The beta-binomial has an additional parameter ϕ which characterizes intra-litter correlations. In Splus, the beta-binomial model can be easily fitted using *nlminb* by specifying the negative log-likelihood as the objective function to be minimized. Table 1 summarizes the results of fitting a beta-binomial model with mean given by (8), assuming first a common correlation

parameter and then allowing the correlation to vary with dose group. The second set of results correspond exactly to those obtained by Chen and Kodell (1989). Using our proposed procedure, we also fitted the same dose response model using GEEs with an exchangeable correlation assumed for $\mathbf{R}_i$. Two versions of the model were fitted: one with a common correlation parameter for all dose groups and another that allowed the correlation to change with dose. Because the mean of Y_{ij} is the same for all litter mates, fitting model (8) assuming an exchangeable correlation matrix involves solving

$$\mathbf{U}_1 := \sum_{i=1}^{I} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \, [\mu_i(1-\mu_i)]^{-1} \mathbf{R}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{10}$$

For the specific model (9), this becomes

$$\sum_{i=1}^{I} \begin{pmatrix} 1 \\ x_i^{\beta_2} \\ \beta_1 x_i^{\beta_2} log(x_i) \end{pmatrix} \mu_i^{-1} \mathbf{1}_i^T \mathbf{R}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{11}$$

where $\mathbf{R}_i = (1-\phi_i)\mathbf{I}_{n_i} + \phi_i \mathbf{J}_{n_i}$, and ϕ_i denotes the correlation parameter for the i^{th} litter which will be determined by its dose group. Using well-known results about the form of the inverse of an exchangeable correlation matrix (see Searle and Henderson (1979)), leads to the following simplification,

$$\sum_{i=1}^{I} \begin{pmatrix} 1 \\ x_i^{\beta_2} \\ \beta_1 x_i^{\beta_2} log(x_i) \end{pmatrix} \frac{(r_i - n_i\mu_i)}{\mu_i[1+\phi_i(n_i-1)]} = 0, \tag{12}$$

where $r_i = \sum_{j=1}^{n_i} Y_{ij}$ is the number of abnormal pups among the n_i in the ith litter. The correlation coefficient can be estimated by solving

$$\mathbf{U}_2 := \sum_{i=1}^{I} [\frac{(r_i - n_i\mu_i)^2}{n_i\mu_i(1-\mu_i)[1+\phi_i(n_i-1)]} - 1] = 0. \tag{13}$$

In (13), $\mathbf{U}_2$ comes from the fact that the expected value of $(r_i - n_i\mu_i)^2$ is the variance of r_i, that is, it equals $n_i\mu_i(1-\mu_i)[1+\phi_i(n_i-1)]$. Thus our final estimates are obtained by solving the estimating equation

$$\mathbf{U} := (\mathbf{U}_1, \mathbf{U}_2)^T = (0,0)^T,$$

for the expanded parameter vector $(\boldsymbol{\beta}, \phi)$.

7. Discussion

We presented an alternative estimation method for general GEE's based on minimizing the inner product of the estimating equations. Such method can conveniently be implemented using existing software, such as the Splus subroutine, *nlminb*. In addition, through the analysis of a real data set from the teratology literature, we showed that this approach yields estimates which are comparable to those obtained via the method of Maximum Likelihood, see Table 1. Statistical properties such as consistency and asymptotic normality of the estimates have been established by Hall (1993). To see this, we recognize (7) as a special case of their more general quadratic form which can be found on page 400 with $W_n = I$ and $g_n = \mathbf{U}$.

The necessary regularity conditions relating the standard approach to solving GEE's and our proposed method can be seen by considering the *first order necessary conditions* for (7). That is, solve for $\boldsymbol{\beta}$

$$\mathbf{J}^T(\boldsymbol{\beta})\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}, \tag{14}$$

where $\mathbf{J}$ is as defined in (6).

If at the computed solution, $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ then this is certainly a solution to (14). However, the reverse need not necessarily hold unless we have that the Jacobian, $\mathbf{J}$, at the solution is invertible or of full rank. Once again, we see that (14) is just a special case of the estimating equation considered by Hall (1993)(see page 400, equation (3.2) with $W_n = I$, $G_n = \mathbf{J}$, and $g_n = \mathbf{U}$).

Acknowledgments

The authors thank Charles McCulloch for valuable discussions. This work was supported by NIH grant CA48061, NSF grant DMS-9710081, and Hatch Project NYC-151413.

References

Chen, J. J., and Kodell, R. L. (1989) Quantitative risk assessment for teratologic effects. *J. Amer. Statist. Assoc.,* **84**, 966-971.

Dennis Jr., J.E. and Schnabel, R.B. (1983) *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs: Prentice Hall.

Diggle, P.J., Liang, K.Y, and Zeger, S.L. (1996) *Analysis of longitudinal data.* Oxford: Oxford University Press.

Godambe, V.P. (Editor) (1991) *Estimating Functions* Oxford: Oxford University Press.

Hall, A. (1993) *Econometrics.* Amsterdam: Elsevier Science Pub.

Henderson, H.V., and Searle, S.R. (1979) Dispersion matrices for variance components models. *J. Amer. Statist. Assoc.,* **74**, 366, 465-470.

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

Liang, K.Y, Zeger, S.L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data. *J.R. Statist. Soc. B* **54**, 1, 3-40.

McCullagh, P. and Nelder, J.A. (1989) *Generalized linear models.* London: Chapman and Hall.

Ortega, J.M. and Rheinboldt, W.C. (1970) *Iterative solution of nonlinear equations in several variables.* New York: Academic Press.

Ryan, L.M. (1992) The use of generalized estimating equations for risk assessment in developmental toxicity. *Risk Analysis,* **12**, 439-447.

Wedderburn, R.W.M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 3, 439.

Williams, D. A. (1975) The analysis of binary responses from toxicology experiments involving reproduction and teratogenicity. *Biometrics*, **31**, 949-952.

Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121-130.

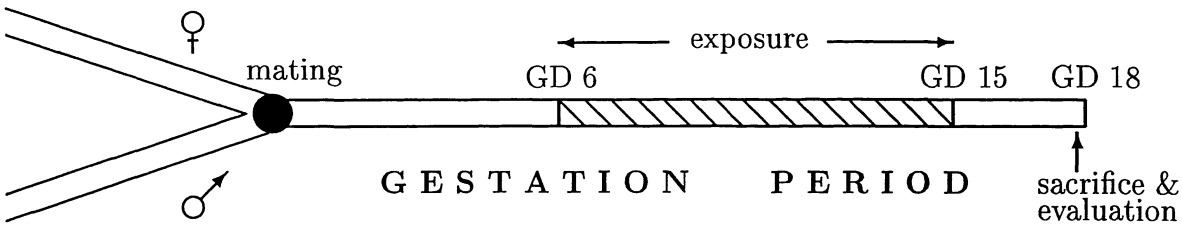

Figure 1: **Chronological events in a developmental toxicity study**

Table 1: Model fitting for DEHP data

	Chen & Kodell				Estimating Equations			
	Common ϕ		Dose-specific ϕ		Common ϕ		Dose-specific ϕ	
	Est	(se)	Est	(se)	Est	(se)	Est	(se)
β_0	.182	(.02)	.141	(.023)	.169	(.030)	.131	(.029)
β_1	168.547	(96.57)	154.05	(93.4)	162.620	(114.794)	162.654	(113.561)
β_2	3.012	(.359)	2.965	(.382)	3.051	(.464)	2.996	(.459)
ϕ	.19		(.19, .01, .09, .38, .28)		.20		(.37, .00, .11, .40, .34)	