with (5). In comparison, only Wald- and score-type estimation procedures attend the estimating equations (3). We suspect that this distinction is rather elusive since likelihood ratio procedures and parametric variance formulas, for example, are unlikely to possess the robustness to model misspecification that provides a key motivation for the estimation procedures under discussion.

As a final argument in favor of the use of (3), one can note that even though the paper of FLR and our comments above focus on mean parameter estimation, there are a variety of problems in which response variances, as well as means, are of substantive interest. These include, for example, studies of the dependence among disease rates in pedigree cohort studies, and studies of recombination rates in genetic linkage analysis, and even some problems in longitudinal data analysis. It seems apparent that equations of the form (2) or (3) will be more useful than equations of the form (5) for covariance estimation and covariance model building.

MISSING RESPONSE DATA

As mentioned above, we commend FLR for drawing attention to the missing response data problem, which is common in longitudinal data and in other multivariate response data settings. The missing completely at random (MCAR) special case is typically easily accommodated by available statistical procedures, as it is here by the estimating equations (3). However, the estimate of the mean parameter β from (5) generally ceases to be consistent if elements of y_k are MCAR, owing to the lack of reproducibility of (4), as FLR acknowledge.

The estimation problem becomes conceptually much more difficult if response variables are missing at random (MAR), but not completely at random. Now it is no longer sufficient to specify marginal moments (i.e., means and covariances) as conditional moments for missing components of the response vector, given the value of the corresponding observed components, are required. If each element of the response vector is subject to MAR, there seems little alternative but to fully specify a model for the joint distribution of y_k and use parametric likelihood procedures as FLR have done. One can nevertheless ask which parametric model is likely to be most convenient and useful with MAR data. For example, what advantages or disadvantages would the authors' proposed method based on (4), with $c_k(y_k, \lambda) = w_k\lambda$, have relative to the application of likelihood procedures to (1), with $c_k(y_k) \equiv 0$ or some other specified value. Neither method could ensure consistency of β-estimation under model misspecification. Model specification would presumably be easier based on (1) for reasons described above (i.e., parameter interpretation). There may be differences in computational convenience or in properties such as bias and efficiency. We would like to encourage FLR to pursue such comparisons in order to yield a better understanding of data analysis options in MAR situations.

Comment

Scott L. Zeger, Kung-Yee Liang and Patrick Heagerty

We congratulate Fitzmaurice, Laird and Rotnitzky (hereafter FLR) for their interesting overview of recent work on statistical models for regression analysis with longitudinal binary responses. The paper adopts what we have termed the *marginal* approach to regression where the marginal expectation rather than the conditional expectation given other responses in the vector for an individual is modelled as a function of explanatory variables. Whereas, previous work (e.g., Liang and Zeger, 1986; Prentice, 1988) has focused on the first two moments of the response vector, FLR propose a method in which the entire likelihood is specified. They study a *mixed model* in which the regression parameters describe the marginal means but the association is measured in terms of conditional pairwise odds ratios given the other responses. Alternatively, association can be measured in terms of pairwise correlations or marginal odds ratios. FLR correctly point out the limitations of measuring association between binary observations in terms of correlations.

FLR compare their likelihood approach to a multivariate analogue of quasi-likelihood called *generalized estimating equations* or GEE in which only the first two moments are specified. FLR show that their likelihood formulation leads to using the same GEE with a particular weighting matrix. They compare the asymptotic efficiency of GEE using their weighting matrix and one in which pairwise correlations are assumed to

Scott L. Zeger and Kung-Yee Liang are Professors and Patrick Heagerty is a graduate student, Department of Biostatistics, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205.

be constant across individuals. They argue that GEE with correlations can be highly inefficient, whereas using conditional odds ratios tends to be more efficient. Finally, they explore the effects of missing data on inferences using GEE derived from the full likelihood as opposed to GEE using pairwise correlations.

We briefly summarize our comments on the following three topics:

1. *Likelihood justification for* GEE *with binary data:* One important contribution of this paper is to give a fully parametric justification of the GEE approach for the analysis of multivariate binary data. FLR have shown that the score equations for the marginal regression coefficients β in the log-linear model with a particular parameterization are identical to the equations used in the GEE approach. Hence, the full power of the likelihood paradigm can be invoked in problems where the FLR parameterization is appropriate as discussed further below.
2. *Parameterizing* $\mathrm{Var}(Y_i)$ *in* GEE: The variance matrix V_i in equation 9.1 can be parameterized in terms of pairwise correlations, conditional odds ratios or marginal odds ratios. We concur with FLR that the use of correlations to measure association among binary responses can be problematic because the marginal means constrain the space of possible correlation values. The most serious limitation of using conditional odds ratios is that they make sense only for block designs in which the observation times are the same for all subjects. For many problems, marginal odds ratios may be a sensible compromise.
3. *Efficiency results:* We disagree with the implications of the efficiency discussion in subsection 3.1. The range of conditional odds ratios where GEE with constant correlations does poorly is somewhat unrealistic as we shall illustrate below. On the other hand, the challenge put to the FLR approach is much weaker.

The remainder of this discussion expands upon points 2 and 3.

PARAMETERIZING VAR(Y_i) IN GEE

FLR study the estimation of regression coefficients β using GEE where the covariance matrix V_i used to weight the residual vector $Y_i - \mu_i$ is parameterized in terms of conditional odds ratios $\omega_{ijk} = \mathrm{OR}(y_{ij}, y_{ik} \mid y_{il} = 0, l \neq j, k)$ and analogous higher order conditional moments. These parameters Ω are the natural parameters in the log-linear model for contingency tables.

The first significant advantage of this approach is the availability of a likelihood function. This permits: appropriate modelling of non-ignorable missing values; likelihood ratio tests and intervals as alternatives to Wald inferences; and goodness of fit tests. A second important advantage is that the parameter space for Ω is unconstrained by the marginal means μ and asymptotically the estimates of β and Ω are independent.

A major disadvantage is that the interpretation of the elements of Ω is specific to the sampling design and in particular the number of observations per person. This is clear since for example, $\mathrm{OR}(y_{i1}, y_{i2} \mid y_{i3} = 0) \neq \mathrm{OR}(y_{i1}, y_{i2} \mid y_{i3} = y_{i4} = 0)$. Hence, the same parameters cannot be used to measure association between y_{i1} and y_{i2} if some people have three observations while others have four. The application of the FLR approach is therefore limited to studies with observations collected (when not missing) at the same times for every individual. A second related issue is the interpretation of the conditional odds ratios. When the association is itself the focus of the investigation, conditional odds ratios may be less useful because of their dependence on the number of observations per subject.

Using correlations is one alternative but FLR correctly point out the problems with measuring association between binary observations with correlations. The range of possible values can be severely constrained by the means of the binary responses as detailed by Prentice (1988) and Carey (1992). The constraints become more severe as the absolute difference between the means increases and as the means deviate from 0.5. For example, when the means for two binary responses are 0.5 and 0.1, their correlation must lie in the range from −0.33 to 0.33 (Carey, 1992).

In a regression context where the marginal means are assumed to change across individuals with their changing x's, the assumption that the pairwise correlations are constant is easily violated since the range of the correlations may itself be changing. This is especially a problem when the correlation is very high. The solution is either to model the dependence of the correlation on x as suggested by Prentice (1988) or to use another measure of association.

One possibility which avoids the limitations of the conditional odds ratios and some of the constraint problems associated with correlations is to parameterize V_i in terms of marginal odds ratios. This has been suggested by Lipsitz, Laird and Harrington (1991) and Liang, Zeger and Qaqish (1992). Carey, Zeger and Diggle (1993) show how to easily implement GEE models so that both the marginal mean and odds ratios can be expressed as a function of covariates.

EFFICIENCY RESULTS

We are not convinced by evidence presented in Section 3 that determining the covariance matrix V_i from the likelihood expressed in terms of conditional odds ratios and analogous higher order moments leads to greater efficiency than some other approach to estimating V_i. As the authors point out, the proposed likelihood estimator is a GEE estimator with a particular choice of variance matrix to weight the residuals. If the assumed variance matrix is close to the true variance matrix, the estimator will be nearly efficient; the degree of inefficiency is a simple function of the degree of misweighting as detailed in the paper.

In Figure 1, FLR present the degree of inefficiency that results from assuming the correlation is constant across individuals when it is not. First note that the conditional log odds ratio ω ranges from 0 to 10 in this illustration so that the odds ratios range between 1 and 22,026. Also, the third order term is fixed at $\kappa = 3$ so that the pairwise correlations are substantial at every value of ω. For example, the correlation between the first two observations ρ_{12} ranges as a function of the x values between 0.43 and 0.55 when $\omega = 0$ and between 0.60 and 0.98 when $\omega = 6$. Note in Figure 1 that assuming constant correlation with $\omega = 0$ (true correlations between 0.43 to 0.55) gives a nearly efficient estimate. This is because the correlations do not vary substantially and so the working variance matrix is nearly correct. Assuming the correlations are constant when they vary as a function of x between 0.6 and 0.98 ($\omega = 6$) leads to inefficient estimates. This should be no surprise. When correlations are this high and vary so dramatically with x, they must be modelled as a function of x to get reasonably efficient inferences as has been done in Liang, Zeger and Qaqish (1992) and Carey, Zeger and Diggle (1993).

To produce the high degree of correlation and dependence on x, we believe an unrealistic dependence structure has been assumed. FLR have set the third order term $\kappa = 3$. This means that when $\omega = 0$, $\mathrm{OR}(y_{i1}, y_{i2} \mid y_{i3} = 0) = 1$ and $\mathrm{OR}(y_{i1}, y_{i2} \mid y_{i3} = 1) = \exp(3) = 20$. Hence there is no association between the first two observations if the third value is 0 but an enormous positive association if the third value is 1. Is this realistic?

The challenge given to the FLR estimator which assumes constant conditional odds ratios is to assume constant correlations that range from 0 to 0.45. Note that the entire x-axis in Figure 2 corresponds to correlations that are smaller than the left most point of Figure 1 ($\omega = 0$). To illustrate the potential inefficiency of the FLR estimator, we must let the conditional odds ratios vary with the xs and use an estimator that assumes they are constant. An arbitrary degree of inefficiency can be produced in this way.

To recap our comments on efficiency, the FLR likelihood estimator is a special case of the GEE approach where the variance matrix has been specified in terms of conditional moments in such a way that the resulting equation is the score equation for a log-linear model. As a GEE estimator, it will be efficient when the assumed covariance matrix is close to the truth and inefficient when not. The same is true for any GEE estimator regardless of the approach to specifying the weighting matrix.

We once again congratulate FLR for their interesting and important paper. We look forward to the opportunity to use their methodology to analyze balanced data sets in problems where the regression parameters are the focus. Clinical trials is an area of application where this approach can be particularly important. We also concur with them that ignoring correlation when it is substantial is problematic even if robust variances are estimated. Their subsection 3.1 shows that grossly misspecifying the weighting matrix when using GEE can lead to inefficient estimates. We look forward to additional efficiency studies based upon more realistic data sets. Finally, while we have not addressed the missing data issue, we are aware of interesting recent work by one of the authors (Rotnitzky) and coworkers on handling missing at random data in the general GEE framework.

Rejoinder

Garrett M. Fitzmaurice, Nan M. Laird and Andrea G. Rotnitzky

We thank all of the discussants for their contributions. We will restrict most of our comments to four issues.

MARGINAL REGRESSION MODELS WITH STOCHASTIC TIME-VARYING COVARIATES

We are in complete agreement with the comments on the role of marginal models made by Drum and McCullagh. A related issue is the role of covariates in a longitudinal study. Our paper focused on nonstochastic covariates and the discussants' comments relate to settings where the covariates are time-stationary. However, when the covariates are both time-varying and stochastic, new issues arise regarding the interpretation and the estimation of the parameters of marginal models. These parameters may not have the implied