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.