Comment

Ross L. Prentice and Lloyd A. Mancl

We would like to make two general points as well as a few details concerning the stimulating paper by Garrett Fitzmaurice, Nan Laird and Andrea Rotnitzky (hereafter FLR). Our first general point will be to argue in favor of joint estimating equations for mean and covariance parameters, as compared to estimating equations for mean and "mixed model" canonical parameters, for "marginal" parameter estimation in multivariate response regression situations. Our reasons include reproducibility of the estimation procedure, ease of calculations and, importantly, ease of regression modelling and parameter interpretation. Second, we will argue that the problem of response variables missing at random, but not completely at random, introduces an order of magnitude greater complexity into the estimation problem. We appreciate the authors drawing attention to this important topic, but believe that additional work is required to compare possible data analysis methods.

ESTIMATING EQUATIONS FOR MEAN PARAMETERS AND DEPENDENCE PARAMETERS

The issues discussed and reviewed by FLR do not seem intimately tied to binary, as opposed to more general, multivariate response vectors, nor do they seem intimately tied to longitudinal data. Hence, our comments will address regression methods for general multivariate response vectors.

First consider quadratic exponential models (e.g., Prentice and Zhao, 1991) for a response vector $y_k^T = (y_{k1}, \ldots, y_{kn_k})$ having mean vector $\mu_k^T(\beta) = (\mu_{k1}, \ldots, \mu_{kn_k})$ and covariance vector $\sigma_k(\beta, \alpha) = (\sigma_{k11}, \sigma_{k12}, \ldots, \sigma_{k22}, \ldots, \sigma_{kn_kn_k})$, where β and α are parameter vectors to be estimated. The class of quadratic exponential models can be written

$$(1)\qquad \begin{aligned} &P_k\{y_k; \mu_k, \sigma_k, c_k(\cdot)\} \\ &\quad = \Delta_k^{-1}\exp\{y_k^T\theta_k + w_k^T\lambda_k + c_k(y_k)\}, \end{aligned}$$

where $w_k^T = (y_{k1}^2, y_{k1}y_{k2}, \ldots, y_{k2}^2, y_{k2}y_{k3}, \ldots)$, $c_k(y_k)$ is a shape parameter function involving products of three or more elements of y_k; $\theta_k = \theta_k(\mu_k, \sigma_k)$ and $\lambda_k = \lambda_k(\mu_k, \sigma_k)$ are canonical parameters; and $\Delta_k = \Delta_k[\theta_k, \lambda_k, c_k(\cdot)]$ is a normalizing constant. The marginal distribution of any subset of y_k can easily be shown to have the same form as (1), hence the semiparametric class of models (1) has a desirable reproductive property. Also for any specified shape function $c_k(\cdot)$, (1) is parametrized only in terms of the mean and covariance parameters for y_k, yielding the type of analogue of the multivariate Gaussian distribution for discrete or mixed data assumed to be lacking by FLR.

The nice feature of the class of models (1) is that joint score equations for the parameter vectors β and α that characterize the mean and covariance of y_k can be written, in a notation slightly more general than that used by FLR, as

$$(2)\qquad \sum_{k=1}^{n} D_k^T V_k^{-1} f_k = 0$$

where

$$D_k = \begin{pmatrix} \partial\mu_k/\partial\beta^T & 0 \\ \partial\sigma_k/\partial\beta^T & \partial\sigma_k/\partial\alpha^T \end{pmatrix},$$

$$V_k = \begin{pmatrix} \operatorname{var} y_k & \operatorname{cov}(y_k, \sigma_k) \\ \operatorname{cov}(s_k, y_k) & \operatorname{var} s_k \end{pmatrix}, \quad f_k = \begin{pmatrix} y_k - \mu_k \\ s_k - \sigma_k \end{pmatrix},$$

and where $s_k^t = (s_{k11}, s_{k12}, \ldots, s_{kn_kn_k})$ and $s_{kij} = s_{kij}(\beta) = (y_{ki} - \mu_{ki})(y_{kj} - \mu_{kj})$. Direct application of these score equations is hampered by the fact that V_k is typically a complicated function of all model parameters, and one would need to specify the shape function $c_k(\cdot)$ and apply a double iterative computational procedure. Instead one can specify "working" models for $V_{k12} = \operatorname{cov}(y_k, s_k)$ and $V_{k22} = \operatorname{var} s_k$, giving rise to a conceptually and computationally simple estimation procedure that generalizes the seminal results of Liang and Zeger (1986) from mean parameters to mean and covariance parameters. As we have noted previously (e.g., Zhao and Prentice, 1990; Prentice and Zhao, 1991), the solution $\hat\beta$ to (2) need not be consistent under covariance model misspecification. We agree with FLR that this is a serious drawback, and one that applies also to FLR's mixed model approach if elements of the response vector are missing or if response vectors are of variable dimension. However, as we have noted in the references just cited, the inconsistency in $\hat\beta$ solving (2) is easily eliminated by requiring D_{k21} and V_{k12} to be identically zero, in which case (2) reduces to the pair of equations

Ross L. Prentice is Director of the Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, and Professor of Biostatistics, University of Washington, Seattle, Washington 98105. Lloyd A. Mancl is Postdoctoral Fellow, Department of Dental Public Health Sciences, University of Washington, Seattle, Washington 98105.

$$(3)\qquad \begin{aligned}&\sum_{k=1}^{n} D_{k11}^{T}\, V_{k11}^{-1}(y_k-\mu_k)=0,\\ &\sum_{k=1}^{n} D_{k22}^{T}\, V_{k22}^{-1}(s_k-\sigma_k)=0,\end{aligned}$$

that formalize the approach advocated by Liang and Zeger (1986) in regard to covariance parameter estimation. We believe the solutions $(\hat{\beta}, \hat{\alpha})$ to (3) to have broad applicability for "population-averaged" regression parameter estimation with multivariate data.

One might be concerned that the solutions to (3) may be markedly less efficient for β estimation than are the solutions to (2). While one can undoubtedly construct examples where such is the case, in most applications there is likely to be little usable information about the mean parameter β in the vector s_k of "empirical covariances." Specifically, in a rather extensive set of simulation exercises (Mancl, 1992) there appeared to be very little potential for efficiency gain in using (2) rather than (3). In fact, some circumstances were encountered in which (2) involved noticeable efficiency loss for β-estimation as compared to (3) even when the variance model σ_k (or V_{k11}) was correctly specified. This apparently happened because of inappropriate working models for V_{k12} and V_{k22} which assigned undue weight to the contribution of s_k, $k = 1, 2, \ldots, n$ in β-estimation. Hence, even though the solutions to (3) do not have a known maximum likelihood interpretation, we believe they enjoy considerable efficiency robustness, particularly if σ_k is fairly accurately specified.

In order to obtain a maximum likelihood interpretation for the first set of equations in (3), FLR consider a partly exponential or mixed parameter model, which in the notation of Zhao, Prentice and Self (1992) can be written

$$(4)\qquad P_k(y_k;\mu_k,\lambda)=\Delta_k^{-1}\exp\{y_k^T\theta_k+c_k(y_k,\lambda)\},$$

where c_k is now a parameter function involving products of two or more elements of y_k and a fixed (independent of k) parameter λ. This semiparametric family is generally not reproductive. However, score equations for the mean parameter β are precisely the first set of equations in (3). In order to reduce (4) to a model with a finite parameter vector FLR, following Fitzmaurice and Laird (1993), set $c_k(y_k, \lambda) = w_k\lambda$ and propose that β and λ be estimated by solving the likelihood score equations

$$(5)\qquad \begin{aligned}&\sum_{i=1}^{n} D_{k11}^{T} V_{k11}^{-1}(y_k-\mu_k)=0,\\ &\sum_{k=1}^{n}\frac{\partial E(w_k)}{\partial\lambda}^{T}\{w_k-E(w_k)-F_kV_k^{-1}(y_k-\mu_k)\}=0,\end{aligned}$$

where $F_k = \operatorname{cov}(w_k, y_k)$. We would like to argue the merits of (3) rather than (5).

First, and most important, (3) involves the regression modelling of marginal means and covariances for which considerable intuition may be available, whereas (5) involves the regression modelling of a fixed parameter λ of rather difficult interpretation, as FLR note. With binary response data λ can be given a log conditional odds ratio interpretation, but this interpretation may not help very much, particularly if the dimension of the response vector, and hence if the number of binary variates conditioned upon, is variable for $k = 1, \ldots, n$. If the elements of y_k are continuous on the entire real line, then λ can be given an interpretation in terms of the inverse of the variance of y_k, but such inverse is not so readily modelled as var y_k itself and can only implausibly be modelled using a fixed parameter λ if the dimension of y_k is variable.

A second, possibly minor, reason for preferring (3) to (5) is computational. Upon specifying the mean and covariance matrix and a working variance model V_{k22}, the solution to (3) involves a simple Newton–Raphson iteration, whereas a maximum likelihood solution to (5) evidently requires some more complex procedure such as double iteration or the use of an iterative proportional fitting algorithm.

The estimates of mean parameters β are consistent under (3) even if the model is otherwise misspecified, but under (5) only if the response vectors are of common dimension. Furthermore, it is evident from the first equations in (3) and (5) that these estimators are identical aside from different estimators of the weight matrix var $y_k = V_{k11}$. Since asymptotic relative efficiencies for β will quite generally be independent of the precision with which parameters characterizing V_{k11} are estimated, it seems evident that (3) and (5) are distinguished primarily in terms of the appropriateness of the form assumed for V_{k11}. We see no reason to expect that a more accurate approximation to the form of V_{k11} can be achieved by introducing a rather difficult to interpret parameter λ, as does (4) and (5), as compared to direct modelling of V_{k11}, as does (3), the authors' efficiency calculations notwithstanding.

On the other side of the argument, FLR draw a distinction between (3) and (5) in terms of constraints on parameters (with binary response data). Given the mean vector, the covariance vector for multivariate binary data does satisfy a complicated set of constraints that need to be kept in mind when modelling σ_k in (3). These same constraints must, of course, be met by the values of cov y_k corresponding to (4). It seems unlikely that a fixed parameter λ, with $c_k(y_k, \lambda) = w_k\lambda$ can span the space of all possible covariance matrices, given the corresponding mean vectors, and, furthermore, we question whether all possible real values for λ are permissible if the dimension of y_k varies.

Finally, FLR point to the ability to use likelihood ratio tests and other parametric likelihood procedures 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.