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

causal interpretation even if the marginal models are correctly specified. In addition, both the GEE and mixed parameter model estimators of the parameters of correctly specified marginal models may be inconsistent, regardless of whether or not the parameters have a causal interpretation.

To illustrate our points, consider the study described by Drum and McCullagh, but suppose now that the outcomes of interest are duration of pregnancy, Z, and infant morbidity at the first month of life, Y. For simplicity of exposition, we consider Z and Y to be dichotomous indicators of premature birth and child illness, respectively, and that the study was restricted to women of the same age, so that X_2 is constant. Suppose now that in addition to the smoking status during pregnancy, X_{11}, the study also collects information on maternal smoking status during the first month after childbirth, X_{12}. For simplicity, we assume X_{11} and X_{12} are binary variables. As indicated by Drum and McCullagh, a marginal regression model for the dependence of Z on X_{11} and Y on X_{11} and X_{12} is often used to answer public-policy-related questions. Thus, suppose that we have correctly assumed the logistic regression models

$$(1) \qquad E(Z|X_{11}) = [1 + \exp(-\alpha_1 - \alpha_2 X_{11})]^{-1}$$

$$(2) \qquad E(Y|X_{11}, X_{12}) = [1 + \exp(-\beta_1 - \beta_2 X_{11} - \beta_3 X_{12})]^{-1}.$$

The benefits of a maternal smoking reduction intervention program are quantified by the causal effects of maternal smoking on infant morbidity.

One approach to specifying causal effects in this setting is through the use of counterfactual variables (Rubin, 1978; Robins, 1989). Consider the, possibly counterfactual, variable $Y_i^{(0)}$ to be the health status of the ith child at one month of age had his or her mother never smoked since the start of pregnancy. Similarly, define $Y_i^{(1)}$ to be the ith child's health status at one month of age had his or her mother continuously smoked since the start of pregnancy. The morbidity rate at age one month had all mothers smoked continuously since the beginning of pregnancy is given by $p^{(1)} = E(Y_i^{(1)})$. Similarly, $p^{(0)} = E(Y_i^{(0)})$ is the morbidity rate in the absence of smoking. Robins (1989) suggests specifying the causal effects of maternal smoking as some function of $p^{(1)}$ and $p^{(0)}$, such as $(p^{(1)} - p^{(0)})$ or the log odds ratio (logit $p^{(1)}$ − logit $p^{(0)}$). Since $Y_i^{(0)}$ and $Y_i^{(1)}$ are not simultaneously observed, $p^{(0)}$ and $p^{(1)}$ are not identified without additional assumptions. Robins (1987, 1989) shows that $p^{(0)}$ and $p^{(1)}$ are identified provided X_{11} is independent of $Y_i^{(0)}$ and $Y_i^{(1)}$, and X_{12} is independent of $Y_i^{(0)}$ and $Y_i^{(1)}$ conditional on X_{11} and Z. Robins (1993) and Robins and Hu (1993) refer to this as the assumption of no unmeasured confounders.

If duration of pregnancy is not a predictor of child morbidity conditional on smoking history, that is,

$$(3) \qquad E(Y|Z, X_{11}, X_{12}) = E(Y|X_{11}, X_{12}),$$

or duration of pregnancy is not a predictor of smoking after pregnancy conditional on pregnancy smoking status, that is,

$$(4) \qquad E(X_{12}|Z, X_{11}) = E(X_{12}|X_{11}),$$

then, as pointed out by Robins and Hu (1993), the parameters $(\beta_1, \beta_2, \beta_3)$ of the logistic regression model (2) have a causal interpretation. The causal effect of continuous smoking exposure compared to the absence of maternal smoking exposure is to increase the log odds ratio, (logit $p^{(1)}$ − logit $p^{(0)}$), by the amount $\beta_2 + \beta_3$.

When (4) holds, the time-dependent exposure X_{1t} $(t = 1, 2)$ is said to be an external covariate process (Kalbfleisch and Prentice, 1980). In this case, the GEE estimators for the parameters of models (1) and (2) are consistent. Furthermore, the mixed parameter model estimators are consistent, when the likelihood which explicitly assumes (4) is correctly specified. However, when (4) is not true but (3) is true, Robins and Hu (1993) showed that the GEE estimators of β are consistent only when the working correlation structure is the identity matrix, and in this case they are maximum likelihood under an independence model. When both (3) and (4) are false, the parameters β do not have a causal interpretation even when model (1) and (2) are correctly specified, since duration of pregnancy is both a predictor of exposure to maternal smoking and an independent risk factor for child illness. When risk factors for morbidity reduce subsequent exposure, exposure-specific morbidity rates tend to underestimate the true effect of exposure on the probability of illness. For example, the observed proportion of sick children of mothers who continuously smoked, that is, with $X_{11} = X_{12} = 1$, will be an underestimate of $p^{(1)}$ if mothers of premature children tend to stop smoking after birth and premature children have higher morbidity rates than full-term children. Approaches to estimating the causal effects of exposure, $(p^{(1)} - p^{(0)})$, under the assumption of no unmeasured confounders are described in Robins (1987, 1989, 1993) and Robins and Hu (1993).

EFFICIENCY RESULTS

We completely agree with Zeger, Liang and Heagerty that the results in Section 3 illustrate how the degree of efficiency of the GEE estimator is a function of how close the assumed or "working" covariance is to the true covariance between the responses. Zeger, Liang and Heagerty criticize our choice of Ω in Figure 1 as being unrealistic. While our choice of values was very extreme, Figure 1 nevertheless provides a striking illustration of the loss of efficiency for the GEE estimator when the assumed covariance is substantially different than the true covariance. Indeed, a more "rea-

sonable" choice of values would yield correlations that do not vary so substantially across individuals, and estimators that assume constant correlation across individuals will be nearly efficient in this case. Note, however, that the "working independence" estimators will still tend to be quite inefficient.

In Figure 2, results are presented for quite a narrow range of constant correlations (0–0.45). This choice of values was constrained by the means of the binary responses. That is, the correlations were constrained to this narrow range by the choice of marginal probabilities. However, even with this restricted range of constant correlations, the conditional odds ratios did vary across individuals. For example, when $\rho = 0.45$ the conditional odds ratio, $\mathrm{OR}(y_{i1}y_{i2}|y_{i3} = 0)$, varied from 2.9 to 7.8. We agree with Zeger, Liang and Heagerty that the challenge to the estimators that assume constant conditional odds ratios is not quite as extreme as that posed in Figure 1. We also agree that the potential inefficiency of the estimators that assume constant conditional odds ratios could be demonstrated by allowing the conditional odds ratios to vary with covariates. However, we want to emphasize that the availability of likelihood-ratio tests for the association parameters would minimize misspecification of this kind in practice.

Finally, we wish to emphasize that the degree of efficiency of the GEE estimator is both a function of how close the assumed covariance is to the true covariance *and* a function of the design matrix. For many designs, there may be no discernible loss of efficiency even when the assumed covariance structure is independence between the responses. However, for designs which include time-varying covariates, it appears to be much more important to obtain a close approximation to the true covariance in order to obtain high efficiency.

PARAMETER INTERPRETATION

Prentice and Mancl raise concerns about the mixed parameter model because of the lack of reproducibility of its association parameters. Their point is an important one when modelling data from clusters of unequal size. In that setting, we agree that one would not want to use the mixed parameter model. Because the association parameters have interpretation in terms of *conditional* probabilities, their interpretation rests on there being the same number of responses available to each experimental unit. However, the approach outlined in our paper is presented in the context of longitudinal studies where the number of responses for each individual is the same. In this setting, both the mixed parameter model and the quadratic exponential model are potentially useful models and the choice between them ought to be determined in part by the scientific question one wants to answer. If the focus is on answering questions concerning the association among the repeated outcomes, then we agree that the mixed parameter model should not be used, although it is not evident to us that the quadratic exponential model is best for this case. However, if the goal is to estimate the marginal means of the repeated outcomes, as was the focus of our paper, then one is concerned with two issues: robustness of estimation to misspecification of second and higher order moments and associations, and the efficiency of the estimators.

As we argued in our paper, the MLE of the mean parameters under the mixed parameter model is consistent even when the model for the second and higher order associations is misspecified. In contrast, the solution of (2) (here and throughout this section, equation numbers refer to those in the Prentice and Mancl discussion) will fail to converge in probability to the true mean parameters when the model for the second moments is incorrect. If the focus is on efficiency in the estimation of the mean parameters, then the MLE under the mixed parameter model will be fully efficient of course if the model is correctly specified. If the investigator firmly believes that the quadratic exponential model is true and that the model for the variance–covariance parameters is correct, then equation (2) should be used for estimating the marginal mean parameters, since the resulting estimator will exploit all the information about the mean parameters in the second moment parameters.

Instead, Prentice and Mancl recommend using the estimating equation given by (3) arguing that based on a simulation study little loss of efficiency was found. We believe that these findings heavily depend on the choice of design matrix and parameter values. A closer look at (3) reveals that the estimator of the mean parameters is solely obtained from the first estimating equation which in turn is identical to the score equation for the mean parameters under the mixed parameter model. The difference now is how the unknown parameters in the *true* variance–covariance matrix are estimated. The investigator who uses the model that better approximates the true variance–covariance matrix will obtain the most efficient estimator of the mean parameters. Prentice and Mancl use the estimator of the variance–covariance parameters resulting from the second set of equations in (3). This estimator will be consistent, and the resulting estimator of the mean parameters can be more efficient than the mixed parameter model estimator, only when the assumed model for the second moments is correctly specified. As was demonstrated in our paper (Figure 1), this estimator can be highly inefficient when the mixed parameter model holds instead.

Finally, Prentice and Mancl argue that our method does not provide an attractive approach to covariance

model building. It is true that models for the association parameters as functions of covariates will not in general translate into easily interpretable models for the covariance parameters. However, we emphasize that the focus of our approach is on obtaining efficient estimates of the mean parameters, while the association parameters are regarded as nuisance characteristics of the data.

MISSING DATA

We would like to respond to Prentice and Mancl on the problem of inference with missing data by addressing two issues raised by these authors. The first concerns the advantage of using the mixed parameter model as opposed to the quadratic exponential model with regards to protection against misspecification bias. The answer here is simple. With repeated binary data the latter is a special case of the former. Thus, for example, model (1) with $c_k(y_k) = 0$ is the same as model (4) with $c_k(y_k, \lambda) = w_k\lambda$ and $\lambda = 0$. Thus, model (4) is preferable since misspecification of model (4) implies that model (1) is misspecified but the opposite is not true.

Our second point concerns the comments Prentice and Mancl made regarding the need to use parametric likelihood procedures when the data is missing at random. In recent work, Robins and Rotnitzky (1992, 1993) and Robins, Rotnitzky and Zhao (1993) describe a new class of semiparametric estimators that are consistent for estimating the mean parameters when the data is missing at random. Their estimators are based on inverse probability weighted estimating equations and can be viewed as an extension of the GEE estimators of Liang and Zeger (1986). In contrast to a likelihood approach, with their method one does not need to specify the complete data likelihood. However, their approach requires correct specification of the probability of response given the observed data. Thus, their method will be particularly appealing for analysing repeated binary data in settings where there is little knowledge about the model linking the second and higher order associations or moment parameters, but a model for the nonresponse process can be satisfactorily posed.

MISCELLANEOUS COMMENTS

In response to Prentice and Mancl, we do believe that many of the issues discussed in our paper are tied specifically to binary data and are not directly relevant to the general continuous case. Although, in principle, we can specify likelihoods following the mixed parameter representation, when the data are binary, the problem has a number of unique features:

1. The covariance parameters are functions of the marginal means, and hence the regression parameters; this introduces dependencies in the parameter space which are not necessarily present with continuous outcomes.
2. The λ parameters have interpretation in terms of conditional log odds-ratios.
3. The multinomial distribution can be fully specified with only a finite number of λ parameters.
4. The iterative proportional fitting algorithm provides a convenient computational tool.

Finally, in passing, we note that λ is not assumed constant but can depend on individual-level covariates.

In conclusion, we thank, once again, all of the discussants for their thoughtful and constructive comments.

ADDITIONAL REFERENCES

CAREY, V. C. (1992). Regression analysis for large binary clusters. Ph.D. dissertation, Dept. Biostatistics, Johns Hopkins Univ.

CAREY, V. C., ZEGER, S. L. and DIGGLE, P. F. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika.* To appear.

MANCL, L. (1992) Regression analysis of correlated discrete and continuous data: Evaluation of an estimating equation approach. Ph.D. dissertation, Dept. Biostatistics, Univ. Washington.

ROBINS, J. M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases* 40 139S-161S.

ROBINS, J. M. (1989). The control of confounding intermediate variables. *Statistics in Medicine* 8 679-701.

ROBINS, J. M. (1993). Analytic methods for HIV treatment and cofactor effects. In *Methodological Issues of AIDS Behavioral Research* (D. G. Ostrow and R. Kessler, eds.) Plenum, New York.

ROBINS, J. M. and HU, F.-C. (1993). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *J. Amer. Statist. Assoc.* To appear.

ROBINS, J. and ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodological Issues* (N. P. Jewell, K. Dietz, and V. Farewell, eds.) 297-331. Birkhäuser, Boston.

ROBINS, J. and ROTNITZKY, A. (1993). Semiparametric efficiency in multivariate regression with missing data. *J. Amer. Statist. Assoc.* To appear.

ROBINS, J., ROTNITZKY, A. and ZHAO, L. (1993). Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. *J. Amer. Statist. Assoc.* To appear.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* 6 34-58.