Statistical Science
1993, Vol. 8, No. 3, 284–309

Regression Models for Discrete Longitudinal Responses

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

Abstract. In this paper, we review analytic methods for regression models for longitudinal categorical responses. We focus on both likelihood-based approaches and non-likelihood approaches to analysing repeated binary responses. In both approaches, interest is focussed primarily on the regression parameters for the marginal expectations of the binary responses. The association or time dependence between the responses is largely regarded as a nuisance characteristic of the data. We consider these approaches for both the complete and incomplete data cases. We describe the generalized estimating equations (GEE) approach, a non-likelihood approach, and some proposed extensions of it. We also discuss likelihood-based approaches that are based on a log-linear representation of the joint probabilities of the binary responses. We describe how a likelihood-based "mixed parameter" model yields likelihood equations for the regression parameters that are of exactly the same form as the GEE. An outline of the desirable features and drawbacks of each approach is presented. In addition, we provide some comparisons in terms of asymptotic relative efficiency for the complete data case, and in terms of asymptotic bias for the incomplete data case. Finally, we make some recommendations concerning the application of these methods.

Key words and phrases: Correlated binary data, generalized estimating equations, longitudinal binary data, marginal models, repeated measures.

1. INTRODUCTION

In longitudinal studies, repeated observations of a response variable and a set of covariates are made on individuals across occasions. Because repeated measurements are made on the same individual, the response variables will usually be positively correlated. When analysing data from longitudinal studies, this time dependence must be accounted for. The focus of this paper is on longitudinal regression models, in which the expectation of the response is related to a set of covariates by some known link function. For example, when the responses are continuous, a common choice is the identity link function. When the responses are binary, a natural choice is to use a logit link function, although other link functions can be used. In longitudinal studies, the covariates can be both time-stationary, that is, constant across occasions, and time-varying.

In order to fix these ideas, consider the asthma studies conducted by the Environmental Protection Agency's Community Health and Environmental Surveillance System (EPA-CHESS) and described in Korn and Whittemore (1979). These are longitudinal studies of the adverse effects of air pollution on rates of asthma attacks. In these studies, daily records of the presence or absence of an asthma attack were recorded on each participant, in addition to measurements of air pollutants and meteorological variables, such as daily temperature and humidity. Here, the aerometric and meteorological variables are time-varying covariates, since they can change from day to day, whereas individual characteristics, such as (initial) age, sex and ethnicity, are time-stationary. One of the objectives of this study was to determine the effects of outside air quality on rates of asthma attacks. Since we are interested in relating the expectation, here the probability of an asthma attack, to the covariates, a regression model is appropriate. This example is introduced for illustrative purposes only and is not discussed any further; other examples can be found in Liang, Zeger and Qaqish (1992).

We distinguish longitudinal studies, where the association between the vector of responses, Y_i, is of

Garrett M. Fitzmaurice, Nan M. Laird and Andrea G. Rotnitzky are members of the Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115.

scientific interest, from studies where the association parameters are considered to be a nuisance characteristic of the data. In the former, the parameters modelling $\text{cov}(Y_i)$ or the conditional expectations, $E(Y_{it}|y_{is}, s < t)$, are of primary interest, while in the latter interest is focussed primarily on the regression parameters for the marginal expectations, $E(Y_{it})$. Regression or marginal models are the focus of this paper.

When the response is continuous and assumed to be approximately Gaussian, there is a general class of linear models that is suitable for analyses. Ware (1985) provides a comprehensive description of these models. However, when the response variable is categorical, fewer techniques are available. This is due in part to the lack of a discrete multivariate analogue that, as with the multivariate Gaussian, can be parameterized only in terms of mean and covariance parameters that vary independently in separable parameter spaces. In addition, with binary responses, the usual choice of nonlinear models for the means makes the parameters of random effects models difficult to interpret. Random effects models are far less well developed for binary responses than for the case where the responses are continuous. Laird (1991) provides a detailed description of some likelihood-based approaches to modelling categorical data structures. Recently, there has been considerable interest in the generalized estimating equations (GEE) approach to analysing longitudinal binary responses, which does not require the complete specification of the joint distribution of the repeated responses.

The problem of missing or incomplete responses is ubiquitous in longitudinal research. Incomplete or unbalanced data can often arise as a result of attrition, but can also arise by design, for example, in "rotating panel" designs (Laird, 1988). In many situations, it is not possible to identify the missing-data mechanism. A number of alternative approaches has been proposed for analysing incomplete longitudinal binary responses. One very naive approach is to restrict analyses to only those subjects measured at all occasions. However, this approach can lead to substantial biases, and is not usually very efficient (Little and Rubin, 1987). Another approach is to ignore the missing data completely, and base the analyses on only the observed portion of the data. This approach is usually computationally simple, requiring only a slight modification of existing methods for analysing complete data. However, in many instances, this approach may be inefficient, and can also introduce bias. Another alternative for analysing incomplete responses is the likelihood approach. This approach retains the notion of a complete set of data, by imputing the missing responses from their conditional distribution given the observed responses. However, this approach can be sensitive to model misspecification.

In this paper, we review the recent literature on both likelihood-based approaches and non-likelihood approaches to analysing repeated binary responses. The objective is to give a brief survey of the different regression models that have recently been proposed, outlining the desirable features and drawbacks of each approach. Furthermore, we provide some comparisons in terms of asymptotic relative efficiency and bias. In Section 2, we introduce some notation and describe the main features of the different regression models. In subsection 2.1, we briefly describe the generalized estimating equations approach, a non-likelihood approach, and some proposed extensions of it. Next, we consider likelihood-based approaches to analysing longitudinal binary data. In subsection 2.2, we describe two different multinomial representations of the multivariate binary response. In subsection 2.3, we discuss likelihood-based approaches based on a log-linear representation of the joint probabilities of the binary responses. We describe a likelihood-based "mixed parameter" model that yields likelihood equations for the regression parameters that are of the same form as the GEE. In Section 3, we compare the GEE and "mixed parameter" models in terms of their asymptotic relative efficiency. Finally, in Section 4, we consider the behavior of these estimators when there are incomplete or missing responses. We examine their asymptotic bias both as a function of the amount of model misspecification, and as a function of the degree of missingness. In Section 5, we summarize the similarities and differences between the GEE and likelihood-based approaches, and make some recommendations concerning their application.

2. REGRESSION MODELS FOR LONGITUDINAL BINARY RESPONSES

In longitudinal studies, there is an implied ordering of the times of the repeated observations on each individual. Initially, we assume that each of N individuals is observed at the same T occasions. Then, assuming the responses are binary, we can form the $T \times 1$ vector $Y_i = (Y_{i1}, \ldots, Y_{iT})^T$, where the binary random variable $Y_{it} = 1$ if subject i has response 1 (success) at time $T = t$, and 0 otherwise. Each individual has a $P \times 1$ covariate vector x_{it} at occasion t, and we let $X_i = (x_{i1}, \ldots, x_{iT})^T$ represent the $T \times P$ matrix of covariates for individual i. Thus, the data for the ith individual consists of the observation (Y_i, X_i).

The marginal distribution of Y_{it} is Bernoulli, $f(y_{it}|X_i) = \exp[y_{it}\theta_{it} - \log\{1 + \exp(\theta_{it})\}]$, where we assume $\theta_{it} = \log[\mu_{it}/(1 - \mu_{it})] = x_{it}{}^T\beta$, and $\mu_{it} = \mu_{it}(\beta) = E(Y_{it}) = \text{pr}(Y_{it} = 1|x_{it}, \beta)$ is the probability of success at time t; and β is a $P \times 1$ vector of parameters. With binary responses, the logit link function is a natural choice although, in principle, any link function could be chosen. We can group the $\mu_{it}(\beta)$ together to form a vector $\mu_i(\beta)$ containing the marginal probabilities of success, $\mu_i(\beta) = E(Y_i) = (\mu_{i1}, \ldots, \mu_{iT})^T$. In the preceding, the only assumption we have made concerns the marginal distribution of Y_{it}.

If the responses are naively assumed to be independent, then the joint distribution of the binary responses is

$$f(y_i|X_i) = \exp\left[\sum_{t=1}^{T} y_{it}\theta_{it} - \sum_{t=1}^{T}\log\{1 + \exp(\theta_{it})\}\right].$$

Then, an expression for the derivative with respect to β, of the ith individual's contribution to the log-likelihood can be obtained by applying the chain rule,

$$\frac{\partial l_i}{\partial \beta} = \left(\frac{\partial \mu_i}{\partial \beta}\right)^T \frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}.$$

Using properties of exponential family distributions, the derivative of the log-likelihood with respect to the canonical parameters, θ_i, is

$$\frac{\partial l_i}{\partial \theta_i} = y_i - E(Y_i) = y_i - \mu_i,$$

and furthermore,

$$\frac{\partial \mu_i}{\partial \theta_i} = \text{cov}(Y_i).$$

Thus, we can write

$$\frac{\partial l_i}{\partial \beta} = \left(\frac{\partial \mu_i}{\partial \beta}\right)^T \text{var}^{-1}(Y_i)(y_i - \mu_i).$$

Since we have assumed that $\theta_i = X_i\beta$,

$$\left(\frac{\partial \mu_i}{\partial \beta}\right)^T = X_i^T\Delta_i,$$

where $\Delta_i = \text{diag}[\text{var}(Y_{i1}), \ldots, \text{var}(Y_{iT})]$, a $T \times T$ diagonal matrix. Thus, the maximum likelihood estimate (MLE), $\hat{\beta}$, is the solution to

$$\text{(1)} \qquad \begin{aligned} \sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta} &= \sum_{i=1}^{N}\left(\frac{\partial \mu_i}{\partial \beta}\right)^T \text{var}^{-1}(Y_i)(y_i - \mu_i) \\ &= \sum_{i=1}^{N} X_i^T(y_i - \mu_i) = 0, \end{aligned}$$

where μ_i depends on $\hat{\beta}$.

In spite of the fact that the repeated measurements on the same individual are correlated, ordinary logistic regression maximum likelihood estimation (which assumes the repeated measures are independent) produces estimates which are consistent and asymptotically normal (Liang and Zeger, 1986). However, the joint likelihood under independence ignores the possible correlations among the binary responses. Consequently, the inverse of the estimated information matrix can give inconsistent estimates of the asymptotic variance of estimated parameters. In general, the standard errors of the time-stationary effects tend to be underestimated, while the standard errors of the time-varying effects tend to be overestimated. To circumvent this problem, Liang and Zeger (1986) propose using a "robust" estimate of the variance of the estimated parameters, which is consistent regardless of the true correlation between the responses. This "robust" variance was also proposed by Huber (1967), and more recently by White (1982) and Royall (1986), and is described in the next section. Thus, one simple approach to analysing longitudinal binary responses is to use ordinary logistic regression, followed by a "robust" variance correction. When the correlation between responses is not too high, Zeger (1988) suggests that these estimators should be highly efficient.

2.1 Generalized Estimating Equations Approach

To gain more efficiency in estimating the parameters of the marginal model, Liang and Zeger (1986), Zeger and Liang (1986) and Prentice (1988) have developed moment-based GEE. The GEE approach produces consistent estimators of the regression parameters, under only the correct specification of the form of the mean function, μ_i, of the vector of responses for each individual. The GEE for β are of the form

$$\text{(2)} \qquad U(\beta) = \sum_{i=1}^{N} D_i^T V_i^{-1}(y_i - \mu_i) = 0,$$

where $D_i = \partial\mu_i/\partial\beta^T$, and V_i is a "working" or approximate covariance matrix of Y_i, chosen by the investigator. The "working" covariance matrix in (2) can be expressed in the following form:

$$V_i = \Delta_i^{1/2}\mathcal{R}_i(\alpha)\Delta_i^{1/2},$$

where $\Delta_i = \text{diag}[\text{var}(Y_{i1}), \ldots, \text{var}(Y_{iT})]$, $\mathcal{R}_i(\alpha) = \text{corr}\,(Y_i)$ is a $T \times T$ "working" correlation matrix, and α represents a vector of parameters associated with a specified model for corr(Y_i). Note that the form of the estimating equations in (2) is similar to the quasi-likelihood estimating equations described in McCullagh and Nelder (1989, Ch. 9). With a binary response vector, these equations simply generalize the ordinary logistic regression estimating equations given in (1) by introducing a "working" or approximate correlation matrix, $\mathcal{R}_i(\alpha)$. This leads to estimating equations of the form

$$\text{(3)} \qquad U(\beta) = \sum_{i=1}^{N} X_i^T\Delta_i V_i^{-1}(y_i - \mu_i) = 0.$$

The GEE approach allows the time dependence to be specified in a variety of ways. Some common specifications for corr(Y_i) are as follows:

1. $\mathcal{R}_i(\alpha) = I$, where I is a $T \times T$ identity matrix. This corresponds to the "working independence" assumption, and gives estimating equations identical to (1).

2. Exchangeable correlation: $\text{corr}(Y_{is}, Y_{it}) = \alpha$; $s \neq t$.
3. Autoregressive correlation: $\text{corr}(Y_{is}, Y_{it}) = \alpha^{|s-t|}$; $s \neq t$.
4. Unstructured or pairwise correlation: $\text{corr}(Y_{is}, Y_{it}) = \alpha_{st}$; α is a $T(T-1)/2 \times 1$ vector containing all the pairwise correlations.

Many other correlation structures can be considered, and α can also depend on subject-specific covariates. Thus, the specification of $\mathcal{R}_i(\alpha)$ can be expressed more generally as $h(\mathcal{R}_i) = Z_i\alpha$, where Z_i is a set of subject-specific covariates, and $h(\mathcal{R}_i)$ is some suitable link function (e.g., inverse hyperbolic). Alternatively, Z_i might represent a common design matrix from the time dependence.

Before discussing the estimation of $\mathcal{R}_i(\alpha)$, we note that alternative specifications of the time dependence have been proposed. Lipsitz, Laird and Harrington (1991) and Liang, Zeger and Qaqish (1992) suggest modelling the association by the pairwise marginal odds-ratios, $\gamma_i = (\gamma_{i12}, \gamma_{i13}, \ldots, \gamma_{i(T-1)T})$, where

$$\gamma_{ist} = \frac{E(Y_{is}Y_{it})E[(1-Y_{is})(1-Y_{it})]}{E[Y_{is}(1-Y_{it})]E[Y_{it}(1-Y_{is})]} \qquad s \neq t.$$

With binary responses, the marginal odds-ratios are a natural measure of association, and $\ln(\gamma_i)$ can be modelled as a linear function of covariates. Furthermore, given (μ_i, γ_i), we can always construct $\mathcal{R}_i$ since, given the means, the pairwise correlations are a one-to-one function of the pairwise marginal odds-ratios.

In order to estimate $\mathcal{R}_i$, define a $T(T-1)/2$ vector of empirical correlations, r_i, with elements

$$r_{ist} = \frac{(Y_{is} - \mu_{is})(Y_{it} - \mu_{it})}{[\mu_{is}(1-\mu_{is})\mu_{it}(1-\mu_{it})]^{1/2}}.$$

Note that $E(r_{ist}) = \rho_{ist} = \text{corr}(Y_{is}, Y_{it})$. If α is known, the only unknown quantity in (2) is β and the solution to (2) is a consistent estimate of β. However, α is usually unknown and must be estimated. To estimate α, a second set of moment estimating equations similar to (2) can be used,

$$U(\alpha) = \sum_{i=1}^{N} A_i^T B_i^{-1}[r_i - \rho_i(\alpha)] = 0, \tag{4}$$

in which $\rho_i(\alpha) = (\rho_{i12}, \rho_{i13}, \ldots, \rho_{i(T-1)T})^T$; $A_i = \partial\rho_i(\alpha)/\partial\alpha^T$, and $B_i \approx \text{cov}(r_i)$, a "working" covariance matrix for r_i. Although r_i is a function of β, β is assumed fixed in (4). The estimate $(\hat{\alpha}, \hat{\beta})$ is the solution to (2) and (4), and can be obtained using a modified Fisher scoring algorithm. Liang and Zeger (1986) let B_i be the $T \times T$ identity matrix. In many cases, this leads to simple non-iterative methods for estimating α. For example, if a common pairwise correlation is assumed with $\rho_i(\alpha) = \rho$, then $\hat{\rho} = \sum_{i=1}^{N} r_i/N$. Furthermore, the choice of estimator for α has no effect on the asymptotic efficiency for estimating β (Newey, 1990). Thus, in general, the estimate $(\hat{\alpha}, \hat{\beta})$ can be obtained by iterating between a modified Fisher scoring algorithm for β and moment estimation of α using r_i.

Finally, if the time dependence has been correctly specified, so that $V_i = \text{cov}(Y_i)$, then a consistent estimate of the asymptotic variance of $\hat{\beta}$ is given by $H_1^{-1}(\hat{\beta})$, where

$$H_1(\hat{\beta}) = \sum_{i=1}^{N}(\hat{D}_i^T\hat{V}_i^{-1}\hat{D}_i) = \sum_{i=1}^{N}(X_i^T\hat{\Delta}_i\hat{V}_i^{-1}\hat{\Delta}_iX_i),$$

where $\hat{V}_i$ is V_i evaluated at $(\hat{\beta}, \hat{\alpha})$, and $\hat{D}_i$ and $\hat{\Delta}_i$ are D_i and Δ_i evaluated at $\hat{\beta}$, respectively. However, if the "working" correlation, $\mathcal{R}_i(\alpha)$, is misspecified, $H_i^{-1}(\hat{\beta})$ can give inconsistent estimates. Liang and Zeger (1986) suggest using the following "robust" estimate:

$$H_i^{-1}(\hat{\beta})H_2(\hat{\beta})H_1^{-1}(\hat{\beta}),$$

where

$$H_2(\hat{\beta}) = \sum_{i=1}^{N}(\hat{D}_i^T\hat{V}_i^{-1}(y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)^T\hat{V}_i^{-1}\hat{D}_i)$$
$$= \sum_{i=1}^{N}(X_i^T\hat{\Delta}_i\hat{V}_i^{-1}(y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)^T\hat{V}_i^{-1}\hat{\Delta}_iX_i).$$

This estimate is "robust" since it is consistent even if the "working" covariance, V_i, is not equal to $\text{cov}(Y_i)$.

The GEE approach has a number of attractive features that can be summarized as follows. First, it provides a consistent estimate, $\hat{\beta}$, that only requires that the model for the mean, $\mu_i(\beta)$, is correctly specified. Thus, regardless of whether the "working" correlation is correctly specified, consistent estimates of the regression parameters are obtained. In addition, robust variance estimates, that are consistent even when the "working" correlation is misspecified, can easily be obtained.

Recently, Zhao and Prentice (1990) and Prentice and Zhao (1991) have described extensions of the GEE methodology to allow for joint estimation of the mean and covariance parameters. This leads to estimating equations of the form

$$\sum_{i=1}^{N}\begin{pmatrix} \frac{\partial\mu_i}{\partial\beta} & 0 \\ \frac{\partial\sigma_i}{\partial\beta} & \frac{\partial\sigma_i}{\partial\alpha}\end{pmatrix}^T \begin{pmatrix} V_i & C_i \\ C_i^T & B_i\end{pmatrix}^{-1}\begin{pmatrix} y_i - \mu_i \\ s_i - \sigma_i\end{pmatrix} = 0, \tag{5}$$

where $S_{ist} = (Y_{is} - \mu_{is})(Y_{it} - \mu_{it})$, $\sigma_{ist} = E(S_{ist})$, $C_i \approx \text{cov}(Y_i, S_i)$, and $B_i \approx \text{cov}(S_i)$. The matrices C_i and B_i are "working" covariance matrices, expressed as a function of the first two moments. Liang, Zeger and Qaqish (1992), specifying the time dependence in terms of marginal odds-ratios, describe an equivalent set of estimating equations for jointly estimating the mean and marginal association parameters. The main advantage of these approaches is that they lead to more efficient estimates of both β and α, provided the model for both the mean and the marginal association is correctly specified. However, a serious drawback is that $\hat{\beta}$ may fail to be consistent when the model for the marginal association is misspecified even if the model for the mean is correctly specified. This is in contrast to (2) which only requires the correct model for the mean in order to obtain consistent estimates of β. Furthermore, given (β, α), specification of (C_i, B_i) requires additional assumptions about the third and fourth moments. Thus, in longitudinal studies where interest is focussed primarily on the regression parameters for the marginal expectation, the solution to (5) may yield inconsistent estimates of β.

In the next two sections, we consider likelihood-based approaches to analysing longitudinal binary data. Before describing likelihood-based models, we first discuss different multinomial representations of the multivariate binary response.

2.2 Multinomial Models for Multivariate Binary Responses

The GEE approaches described above are not likelihood-based, and do not require the complete specification of the joint distribution of the repeated responses. They only require specification of the form of the mean function. Next, we consider likelihood-based approaches that are based on a complete representation of the joint probabilities of the binary responses.

The joint distribution of an individual's binary responses at the T times is multinomial with a 2^T probability vector, $m_i = \{m_{ij_1j_2\cdots j_T}\}$. The means, μ_{it}, are related to m_i by $E(Y_{it}) = \mu_{it} = \sum_{j_k \neq t} m_{ij_1j_2\cdots j_T}$. The fully parameterized distribution has $2^T - 1$ parameters. In this section we outline two particular parametric descriptions of the joint distribution of a set of binary responses. The first, suggested by Bahadur (1961), describes the joint distribution in terms of the marginal means, μ_i, and the marginal correlations, ρ_i. An alternative way to describe the joint distribution is in terms of the canonical parameters, that is, conditional logits and log odds-ratios. We refer to the latter as the log-linear representation of the joint distribution.

2.2.1 *Bahadur representation*

Bahadur's representation of the joint distribution for Y_i can be written as

$$\prod_{j=1}^{T} \mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}} \Big(1 + \sum_{j<k} \rho_{ijk} e_{ij} e_{ik} + \sum_{j<k<l} \rho_{ijkl} e_{ij} e_{ik} e_{il} + \cdots + \rho_{i12\ldots T} e_{i1} e_{i2} \cdots e_{iT}\Big),$$

where $e_{ij} = (Y_{ij} - \mu_{ij})/[\mu_{ij}(1-\mu_{ij})]^{1/2}$; and $\rho_{ijk} = E(e_{ij}e_{ik}), \ldots, \rho_{i12\ldots T} = E(e_{i1}e_{i2}\cdots e_{iT})$. Thus, in terms of the $2^T - T - 1$ *marginal* correlations, $\rho_i = (\rho_{i12}, \rho_{i13}, \ldots, \rho_{i12\ldots T})$, the joint distribution of the responses can be evaluated in closed form. For example, with a *trivariate* response, $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})^T$, the joint probability $\text{pr}\{Y_{i1} = 1, Y_{i2} = 1, Y_{i3} = 1\}$ is represented as

$$\mu_{i1}\mu_{i23} + \mu_{i2}\mu_{i13} + \mu_{i3}\mu_{i12} - 2\mu_{i1}\mu_{i2}\mu_{i3} + \rho_{123}\{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2})\mu_{i3}(1-\mu_{i3})\}^{1/2},$$

where $\mu_{ist} = \text{pr}\{Y_{is} = 1, Y_{it} = 1\} = \mu_{is}\mu_{it} + \rho_{st}\{\text{B}\mu_{is}(1-\mu_{is})\mu_{it}(1-\mu_{it})\}^{1/2}$. The parameter ρ_{123} can be thought of as a "three-way" association parameter. The above specification allows for varying degrees of dependence among the Y_{it}. For example, if all of the pairwise and "higher-way" marginal correlations are set to zero, we have the "independence" structure. A feature of Bahadur's representation, however, is that the marginal correlations must satisfy certain linear inequalities determined by the marginal probabilities. That is, the marginal correlations are constrained by the marginal probabilities.

2.2.2 *Log-linear representation*

An alternative to Bahadur's representation is the log-linear specification (e.g., Cox, 1972 and Bishop, Fienberg and Holland, 1975, Ch. 2.5), which assumes that the joint distribution of Y_i is of the form

$$(6) \quad f(y_i, \Psi_i, \Omega_i) = \exp\{\Psi_i^T y_i + \Omega_i^T w_i - A(\Psi_i, \Omega_i)\},$$

where $W_i = (Y_{i1}Y_{i2}, \ldots, Y_{iT-1}Y_{iT}, \ldots, Y_{i1}Y_{i2}\cdots Y_{iT})^T$ is a $(2^T - T - 1) \times 1$ vector of two- and higher-way cross-products of Y_i, $\Psi_i = (\psi_{i1}, \ldots, \psi_{iT})^T$ and $\Omega_i = (\omega_{i12}, \ldots, \omega_{iT-1,T}, \ldots, \omega_{i12\ldots T})^T$ are vectors of canonical parameters, and $A(\Psi_i, \Omega_i)$ is a normalizing constant, $\exp\{A(\Psi_i, \Omega_i)\} = \sum \exp(\Psi_i^T y_i + \Omega_i^T w_i)$, where summation is over all 2^T possible values of Y_i. The parameters of Ψ_i have interpretations in terms of conditional probabilities, $\psi_{ir} = \text{logit}\{\text{pr}(Y_{ir} = 1 \mid Y_{is} = 0, s \neq r)\}$, while the parameters of Ω_i can be interpreted in terms of log conditional odds-ratios and contrasts of log conditional odds-ratios. That is,

$$\exp(\omega_{irs}) = \frac{\text{pr}(Y_{ir} = 1, Y_{is} = 1 \mid Y_{it} = 0, t \neq r, s)}{\text{pr}(Y_{ir} = 0, Y_{is} = 0 \mid Y_{it} = 0, t \neq r, s)} \times \frac{\text{pr}(Y_{ir} = 1, Y_{is} = 0 \mid Y_{it} = 0, t \neq r, s)}{\text{pr}(Y_{ir} = 0, Y_{is} = 1 \mid Y_{it} = 0, t \neq r, s)},$$

and so on. Note that μ_i is a function of both Ψ_i and Ω_i.

The above form of the joint distribution allows for varying degrees of dependence among the Y_{it}. For example, if all of the two- and higher-way association parameters are set to zero, we have the independence model. On the other hand, if $\Omega_i = (\omega_{i12}, \ldots, \omega_{iT-1,T}, \ldots, \omega_{i12\ldots T})^T$, we have a model that is saturated in the association parameters. Between these extremes, we can consider parsimonious models for the time dependence. An important special case of (6) is the "quadratic exponential family" or pairwise model, obtained by fixing the three- and higher-way association parameters of Ω_i at some value, typically zero (Zhao and Prentice, 1990). Markov structures are also easily modelled in this framework. An advantage of this representation is that the time dependence parameters are not constrained by the marginal probabilities. However, a drawback of this representation is that, unlike Bahadur's representation, it is not "reproducible." That is, if Y_i^* is a $T^* \times 1$ subset of Y_i, where $T^* < T$, then

$$(7) \qquad f(y_i^*) \neq \exp\{\Psi_i^{*T} y_i^* + \Omega_i^{*T} w_i^* - A(\Psi_i^*, \Omega_i^*)\},$$

where Ψ_i^* is the corresponding $T^* \times 1$ subset of Ψ_i, and Ω_i^* and w_i^* are the corresponding subsets of Ω_i and w_i respectively. In Bahadur's representation, the parameters are marginal moments and thus it is reproducible, while the parameters of the log-linear representation are not expressible in terms of the marginal moments. That is, $f(y_i^*)$ depends on the entire (Ψ_i, Ω_i).

2.3 Likelihood-Based Approaches

Zhao and Prentice (1990) propose an approach that is based on the "quadratic exponential family," with the three- and higher-way association parameters set to zero. When these parameters are set to zero, (6) holds with $W_i = (Y_{i1}Y_{i2}, \ldots, Y_{iT-1}Y_{iT})^T$ a $T(T-1)/2 \times 1$ vector of two-way cross-products of Y_i, and $\Psi_i = (\psi_{i1}, \ldots, \psi_{iT})^T$ and $\Omega_i = (\omega_{i12}, \ldots, \omega_{iT-1,T})^T$ the vectors of canonical parameters. Zhao and Prentice (1990) propose to model the mean, μ_i, and the covariance of the response as a function of covariates by some specified link function. They make a one-to-one transformation from (Ψ_i, Ω_i) to the moment parameters (μ_i, σ_i) in order to obtain the likelihood equations. Using this approach, Zhao and Prentice (1990) derive the following set of likelihood equations for β and α, the parameters indexing μ_i and σ_i, respectively,

$$(8) \qquad \sum_{i=1}^{N} \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ \frac{\partial \sigma_i}{\partial \beta} & \frac{\partial \sigma_i}{\partial \alpha} \end{pmatrix}^T \begin{pmatrix} V_i & K_i \\ K_i^T & U_i \end{pmatrix}^{-1} \begin{pmatrix} y_i - \mu_i \\ s_i - \sigma_i \end{pmatrix} = 0,$$

where $S_{ist} = (Y_{is} - \mu_{is})(Y_{it} - \mu_{it})$, $\sigma_{ist} = E(S_{ist})$, $K_i = \text{cov}(Y_i, S_i)$, and $U_i = \text{cov}(S_i)$. Note, a 1-1 transformation from (Ψ_i, Ω_i) to (μ_i, σ_i), where σ_i is the vector of pairwise covariances, results in a set of likelihood equations identical to (5). The equations given by (8) yield pseudo-maximum likelihood (PML) estimates $(\hat{\beta}, \hat{\alpha})$, which are maximum likelihood estimates when the true three- and higher-way association parameters are zero. A serious drawback of this approach, however, is that consistency of $\hat{\beta}$ (and $\hat{\alpha}$) requires the correct specification of the model for both the mean and the pairwise marginal correlations. Thus, this pseudo-maximum likelihood approach suffers from the same drawback as the joint generalized estimating equations for (β, α). Namely, estimates of β may be asymptotically biased when the time dependence is misspecified.

To avoid this problem, Fitzmaurice and Laird (1993) proposed a "mixed parameter" model based on the general log-linear representation

$$f(y_i, \Psi_i, \Omega_i) = \exp\{\Psi_i^T y_i + \Omega_i^T w_i - A(\Psi_i, \Omega_i)\},$$

where $W_i = (Y_{i1}Y_{i2}, \ldots, Y_{iT-1}Y_{iT}, \ldots, Y_{i1}Y_{i2}\ldots Y_{iT})^T$ is a $(2^T - T - 1) \times 1$ vector of two- and higher-way cross-products of Y_i, $\Psi_i = (\psi_{i1}, \ldots, \psi_{iT})^T$ and $\Omega_i = (\omega_{i12}, \ldots, \omega_{iT-1,T}, \ldots, \omega_{i12\ldots T})^T$ are vectors of canonical parameters. With their approach, a model for the mean, μ_i, and the canonical association parameters, Ω_i, is assumed and likelihood equations are obtained via the 1-1 transformation from (Ψ_i, Ω_i) to (μ_i, Ω_i). That is, the "mixed parameter" model is parameterized in terms of the mixed mean and canonical association parameters. This model can be viewed as a special case of the class of partly exponential models introduced by Zhao, Prentice and Self (1992).

The mean parameters are modelled, in the usual way, as a logit function of a set of covariates, while the canonical association parameters can be expressed as a function of a parameter vector α and a set of subject-specific, time-stationary, covariates Z_i. Alternatively, Z_i might represent the design matrix for an association structure that is assumed to be the same for all individuals. In principle, any link function could be used; a natural one in this setting is a linear link function, $\Omega_i = Z_i\alpha$. Finally, note that although μ_i and Ω_i completely specify the joint distribution of the binary responses, there is, in general, no closed form expression representing the joint probabilities as a function of μ_i and Ω_i (Bishop, Fienberg and Holland, 1975, Ch. 3.4.2). However, Fitzmaurice and Laird (1993) describe how this problem can be circumvented using the iterative proportional fitting algorithm (Deming and Stephan, 1940).

Fitzmaurice and Laird (1993) derive the following set of likelihood equations for (β, α):

$$(9) \qquad \sum_{i=1}^{N} \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ 0 & \frac{\partial \Omega_i}{\partial \alpha} \end{pmatrix}^T \begin{pmatrix} V_i^{-1} & 0 \\ -F_i V_i^{-1} & I \end{pmatrix} \begin{pmatrix} y_i - \mu_i \\ w_i - v_i \end{pmatrix} = 0,$$

where $v_i = E(W_i)$, $V_i = \text{cov}(Y_i)$ and $F_i = \text{cov}(W_i, Y_i)$. Note that V_i is not a "working" covariance; rather it is the covariance between responses under the specified model. This yields the following separate likelihood equations for β and α:

$$(9.1) \qquad \sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V_i^{-1}(y_i - \mu_i) = 0,$$

$$(9.2)\quad \sum_{i=1}^{N}\left(\frac{\partial\Omega_i}{\partial\alpha}\right)^T [w_i - v_i - F_iV_i^{-1}(y_i - \mu_i)] = 0,$$

respectively. Note that the likelihood equations for β, given in (9.1), are identical in form to the generalized estimating equations given in (2). In the GEE for β, V_i is the "working" or approximate covariance between the responses. Thus, the "mixed parameter" model and GEE only differ in how they compute the "weight" matrix, V_i.

The asymptotic covariance of $(\hat{\beta}, \hat{\alpha})$ can be approximated by the inverse of the Fisher information matrix,

$$\text{cov}(\hat{\beta}, \hat{\alpha}) \approx \begin{bmatrix} \left\{\sum_{i=1}^{N}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\frac{\partial\mu_i}{\partial\beta}\right\}^{-1} & 0 \\ 0 & \left\{\sum_{i=1}^{N}\left(\frac{\partial\Omega_i}{\partial\alpha}\right)^T (G_i - F_iV_i^{-1}F_i^T)\frac{\partial\Omega_i}{\partial\alpha}\right\}^{-1} \end{bmatrix},$$

where $G_i = \text{cov}(W_i)$. Note that the (β, α) component of the Fisher information matrix is zero, implying that β and α are orthogonal (Cox and Reid, 1987, 1989). In general, with mixed parameterizations of exponential models, the mean components are orthogonal to the canonical components (Barndorff-Nielsen, 1978, Ch. 9.8). Finally, as an alternative to the parametric variance given above, a "robust" variance for β can also be constructed in the usual way.

This "mixed parameter" model has a number of desirable features:

1. The choice of parameterization yields maximum likelihood estimates of the marginal mean parameters that are robust to misspecification of the time dependence. That is, $\hat{\beta}$ is asymptotically unbiased even if the model for the time dependence is misspecified.
2. The orthogonality of the mixed parameter space implies that knowledge of the time dependence parameters does not asymptotically add information about β, that is, the asymptotic variance of $\hat{\beta}$ remains the same whether α is known or estimated. Furthermore, as noted by Newey (1990), the limiting distribution of $N^{1/2}(\hat{\beta} - \beta)$ remains the same if in (9.1) α is replaced by any $N^{1/4}$—consistent estimate of it. Thus, $\hat{\beta}$ remains asymptotically normal even when the number of time dependence parameters increases with the sample size at an appropriate rate.
3. Given (β, α), G_i, F_i, and V_i can be calculated in a straightforward manner, using an iterative proportional fitting algorithm.
4. Unlike marginal correlations and odds-ratios, the conditional association parameters in the "mixed parameter" model are not constrained.
5. Using a full likelihood approach yields likelihood ratio tests and parametric variance formulae. When the assumed model is correct, the parametric variance estimators will be more efficient than the "robust" variance estimators.

There are two potential drawbacks with the "mixed parameter" model. First, the association parameters, the log conditional odds-ratios, do not have the attractive interpretation that marginal association parameters have. However, this may not be such an issue in longitudinal studies when interest is focussed primarily on the regression parameters, and the association parameters are considered to be nuisance parameters. Second, a more important drawback of the "mixed parameter" model, as indicated by (7), is that the distribution is not "reproducible." Thus, the "mixed parameter" model is not appropriate for analyzing data from clusters of unequal size.

2.4 Summary

In this section, we have reviewed some recently developed methods for analyzing longitudinal binary responses. We have noted that the likelihood equations for β from the "mixed parameter" model are identical in form to the GEE for β. Assuming that the mean structure has been correctly specified, these "estimating equations" yield estimates, $\hat{\beta}$, that are asymptotically unbiased and robust to misspecification of the time dependence. In contrast, the pseudo-maximum likelihood approach proposed by Zhao and Prentice (1990) and the extensions of the GEE to allow joint estimation of (β, α) require the correct specification of the model for both the mean and the pairwise marginal correlations in order to yield consistent estimates of β. That is, estimates of β may be asymptotically biased when the time dependence is misspecified.

In general, however, the time dependence and hence $\text{cov}(Y_i)$ is unknown, and must be estimated. Therefore, models for the time dependence that closely approximate $\text{cov}(Y_i)$ will lead to near efficient estimation of β. In the next section, we compare the asymptotic efficiency of estimators from the "mixed parameter" model to the GEE estimators.

3. ASYMPTOTIC RELATIVE EFFICIENCY WITH COMPLETE DATA

Here we address the issue of the asymptotic efficiency of estimators from the likelihood-based "mixed parameter" model and GEE estimators of the marginal parameters in models for longitudinal binary responses. In particular, we compare the asymptotic efficiency of the "optimal" estimators, under a fixed fully parametric model, to both GEE estimators and likelihood-based estimators from a "mixed parameter" model that assume an incorrect likelihood.

In the following, we assume that the mean structure has been correctly specified, but that the time dependence may be incorrectly specified. When the model for the means is correctly specified, but that for the time dependence parameters is not, both the "mixed parameter" model and GEE estimators are consistent. The time dependence might be incorrectly specified by assuming a parsimonious association structure between the responses, for example, *pairwise* or *exchangeable* association. Next, let $\hat{\beta}_W$ be the solution to the GEE or likelihood equations under some working assumption about the time dependence. Then, $\hat{\beta}_W$ normalized has asymptotic covariance matrix V_W given by

$$V_W = \lim_{N\to\infty} N[H_1^{-1}(\beta_W)H_2(\beta_W)H_1^{-1}(\beta_W)],$$

where

$$H_1(\beta_W) = \sum_{i=1}^{N}\left(X_i^T\Delta_i V_i^{-1}\Delta_i X_i\right)$$

and

$$H_2(\beta_W) = \sum_{i=1}^{N}\left(X_i^T\Delta_i V_i^{-1}\mathrm{cov}(Y_i)V_i^{-1}\Delta_i X_i\right).$$

Let $\hat{\beta}_{opt}$ be the solution to the likelihood equations when $V_i = \mathrm{cov}(Y_i)$, the true covariance of Y_i. This is the "optimal" or semiparametric efficient estimator of β, which has minimum asymptotic variance among all estimators guaranteed to be asymptotically normal and unbiased under only the restrictions on the marginal means (Chamberlain, 1987). The estimator $\hat{\beta}_{opt}$ normalized has asymptotic covariance matrix V_{opt} given by

$$V_{opt} = \lim_{N\to\infty} N\left(H_1^{-1}(\beta_{opt})\right),$$

where

$$H_1(\beta_{opt}) = \sum_{i=1}^{N}(X_i^T\Delta_i \mathrm{cov}^{-1}(Y_i)\Delta_i X_i).$$

Note, if the true joint distribution of the binary responses is a "mixed parameter" model, then $\hat{\beta}_{opt} = \hat{\beta}_{ML}$, the maximum likelihood estimate.

The asymptotic relative efficiency (ARE) of $\hat{\beta}_W$ versus $\hat{\beta}_{opt}$ is given by the diagonal elements of

$$\mathrm{ARE} = \mathrm{diag}(V_W)(\mathrm{diag}(V_{opt}))^{-1}.$$

To distinguish between the GEE and likelihood approaches for calculating V_i, we let $\hat{\beta}_{GEE}$ denote the estimate obtained using the GEE and $\hat{\beta}_{PML}$ denote the estimate obtained using the "mixed parameter" likelihood approach.

3.1 Comparison of Asymptotic Efficiencies

In this section, we consider two different design configurations with a *trivariate* binary response. For the two designs, we compare the efficiency of $\hat{\beta}_{opt}$ to (i) estimators from a "mixed parameter" model, under a misspecified association structure, and (ii) estimators based on the GEE, under various "working" correlation structures. Specifically, we consider a simple two-group configuration, and allow for two different covariate designs, one with group as a time-stationary covariate (design A), the other with group as a time-varying covariate (design B). When group is a time-stationary covariate, subjects are assumed to belong to either group with equal probability. When group is a time-varying covariate, each of the 8 possible covariate configurations is assumed to have equal probability of occurrence. We chose design B because it is one where ordinary least squares (OLS) is grossly inefficient for estimating the group effect. However, this type of design is related to those used in higher-order cross-over designs for carryover effects (for an excellent description of cross-over designs, see Jones and Kenward, 1989). For example, in the case of a 2 × 2 cross-over design, there are two treatments **T1** and **T2**, and subjects in one sequence group receive **T1** → **T2**, while subjects in the other sequence group receive **T2** → **T1**. *Baalam's* design is a variant of the 2 × 2 cross-over design that allows more efficient estimation of carryover effects (Baalam, 1968). In *Baalam's* design, subjects receive treatments in one of the following sequences:

Sequence	*Period I*	*Period II*
1	**T1**	**T1**
2	**T2**	**T2**
3	**T1**	**T2**
4	**T2**	**T1**

Thus, "treatment" in this context can be thought of as an *external* time-varying covariate, in the sense described by Kalbfleisch and Prentice (1980).

In order to calculate the relative efficiency, we need to specify the true joint distribution of the binary responses. We assume the following model for the marginal expectation:

$$\mathrm{logit}(\mu_{it}) = \beta_0 + \beta_1 x_{it} + \beta_2(t-1); \qquad t = 1,2,3.$$

Here, x_{it} is a dichotomous covariate indicating group membership for the ith individual at the tth occasion. We characterize the "true" time dependence either in terms of (i) *marginal* correlations, ρ_i, or (ii) *conditional* log odds-ratios, Ω_i. The parameters of the "true" model are as follows. The marginal parameters are

$$\beta_0 = 0; \beta_1 = 0.5; \beta_2 = 0.5;$$

and two different sets of association parameters are selected:

(i) Marginal Correlations: $\rho_i = (\rho_{i12}, \rho_{i13}, \rho_{i23}, \rho_{i123}) = (\rho, \rho^2, \rho, 0)$, where $\rho \in (0, 0.45)$;
(ii) Conditional Log Odds-Ratios: $\Omega_i = (\omega_{i12}, \omega_{i13}, \omega_{i23}, \omega_{i123}) = (\omega, \omega/2, \omega, \kappa)$, where $\omega \in (0, 10)$ and $\kappa = 3$.

Thus, in both parameterizations, the "three-way" association parameter is held fixed, while the "two-way" association parameters vary across a range of values. Note that given the covariate values, ρ and β were chosen so that they satisfied the necessary constraints in the Bahadur representation for the eight cell probabilities, $\mathrm{pr}\{Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}\}$, to be positive and to sum to 1.

We compare the asymptotic relative efficiency of $\hat{\beta}_{opt}$ versus $\hat{\beta}_{PML}$ and $\hat{\beta}_{GEE}$ estimated under the following assumptions about the "true" time dependence:

1. Pairwise: $\hat{\beta}_{GEE}$ estimated assuming "working" pairwise correlation, $\hat{\beta}_{PML}$ estimated assuming $\Omega_i = (\omega_{12}, \omega_{13}, \omega_{23}, 0)$.
2. Exchangeable: $\hat{\beta}_{GEE}$ estimated assuming "working" exchangeable correlation, $\hat{\beta}_{PML}$ estimated assuming $\Omega_i = (\omega, \omega, \omega, 0)$.
3. Independence: $\hat{\beta}_{GEE}$ estimated assuming "working" independence, (identical to $\hat{\beta}_{PML}$ estimated assuming $\Omega_i = 0$).

Note that the assumption of a common pairwise conditional odds-ratio association does not imply a common pairwise correlation. Similarly, a common exchangeable conditional odds-ratio association does not imply a common exchangeable correlation.

3.2 Results

In this section, we compare the asymptotic relative efficiency for both the group and time effects. Results are not reported for the intercept term, since the intercept is usually regarded as a nuisance parameter in this setting. First, we consider design A, the covariate design with group as a time-stationary covariate. Regardless of whether the "true" time dependence was parameterized in terms of conditional odds-ratios or marginal correlations, the asymptotic relative efficiencies of both estimators were never less than 0.95, and in most cases were very close to 1.0 for both the group and time effects. These results are consistent with the relative efficiencies for binary data reported by Liang and Zeger (1986). Thus, for this type of design, with group as a time-stationary covariate, there does not appear to be any discernible loss in efficiency when the time dependence has been misspecified. This is the case regardless of how the "true" dependence between the responses has been parameterized. Thus, the estimator that assumes independence between responses has asymptotic variance close to that of estimators that allow for up to $T(T-1)/2$ association parameters.

For design B, the covariate design with group as a time-varying covariate, the asymptotic relative efficiency for the time effect is never less than 0.95 when

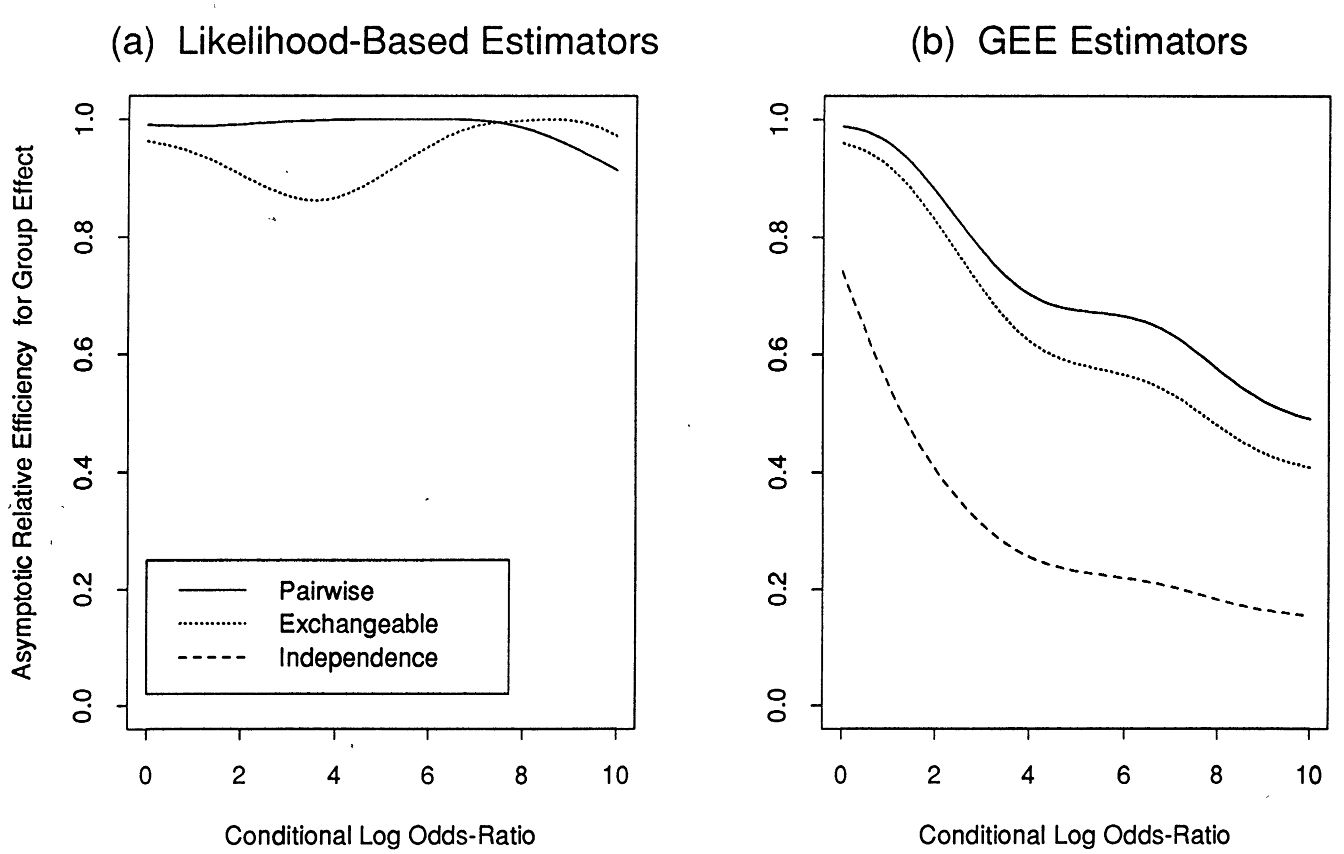

FIG. 1. *Asymptotic efficiency of the likelihood-based and GEE estimators relative to the optimal estimator in design B, when the true underlying joint distribution has a log-linear representation.*

the "true" time dependence is parameterized in terms of conditional odds-ratios. When parameterized in terms of marginal correlations, the asymptotic relative efficiency for the time effect is very close to 1.0. However, for the group effect, there is substantial loss of efficiency depending on how the "true" time dependence is parameterized. When the time dependence is parameterized in terms of conditional odds-ratios, the GEE estimators perform poorly. In Figure 1, the asymptotic relative efficiency of the likelihood-based and GEE estimators are plotted against the time dependence parameter. In Figure 1(b), there is a notable loss of efficiency for the "working independence" estimator, while the efficiency of the "pairwise" GEE estimator can drop as low as 50%. In contrast, in Figure 1(a) the likelihood-based estimators that assume "pairwise," and to a lesser extent those that assume "exchangeable" association, have quite high efficiency.

In Figure 2, we note that the performance of the likelihood-based estimators is quite comparable to that of the GEE estimators when the "true" time dependence is parameterized in terms of marginal correlations. Thus, the likelihood-based methods seem to perform well even when the "true" time dependence is parameterized in terms of constant marginal correlations. Finally, note that the efficiency of the "working independence" estimator decreases rapidly with increasing dependence between the responses. This result is important since pairwise correlations of 0.45 and higher are not unusual in longitudinal studies.

4. BIAS OF THE ESTIMATORS WITH INCOMPLETE RESPONSES

The problem of missing or incomplete responses is ubiquitous in longitudinal research. Incomplete or unbalanced data can arise as a result of attrition, when, for example, individuals drop out of a clinical trial because they have not responded to the treatment. Alternatively, unbalanced data can arise by design, in studies which allow individuals to leave the study for a specified period and then return. These "rotating panel" designs are often used to reduce respondent burden and discourage uncontrolled dropouts (Laird 1988). In many situations, it is not possible to identify or distinguish between the different missing-data mechanisms.

The GEE and likelihood-based methods differ in their approaches for analysing incomplete data. The GEE approach ignores the missing data completely, and bases the analyses only on the observed portion of the data. This approach is usually computationally simple,

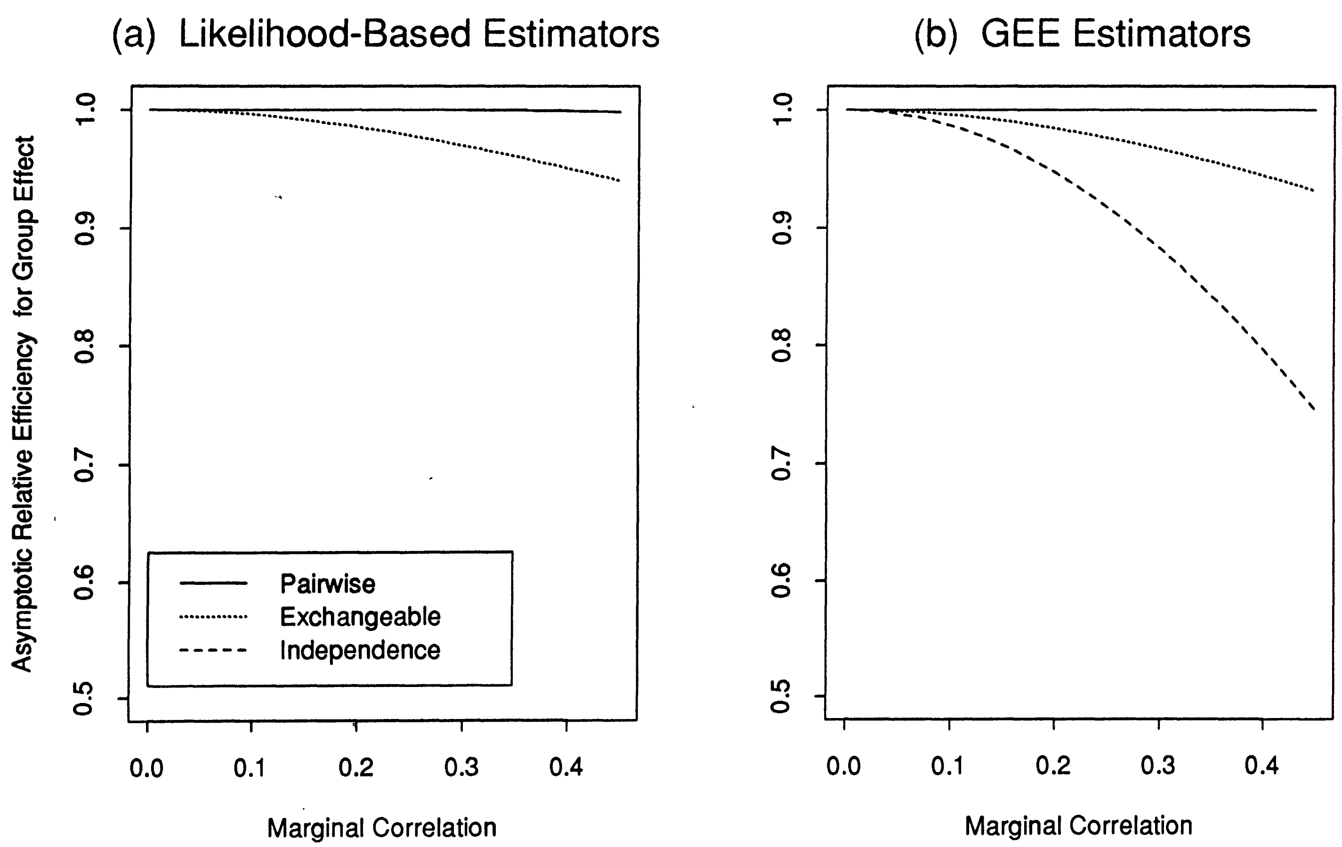

FIG. 2. *Asymptotic efficiency of the likelihood-based and GEE estimators relative to the optimal estimator in design B, when the true underlying joint distribution has Bahadur's representation.*

requiring only a slight modification of the methods for analysing complete data. However, in many instances, this approach may be inefficient and can also introduce bias. The likelihood approach retains the notion of a complete set of data, by imputing the missing responses from their conditional distribution given the observed responses. This approach, however, can be sensitive to model misspecification.

In this section, we examine the asymptotic bias of estimators based on the "mixed parameter" model and GEE estimators when there are incomplete responses. In the following, we assume the model for the marginal expectation has been correctly specified, but that for the time dependence has been misspecified. With complete data, both the GEE and "mixed parameter" models, described in the previous sections, lead to estimators of the marginal regression effects that are consistent regardless of how the time dependence has been specified. Before considering the bias of these estimators with incomplete or missing responses, we need to introduce some additional notation and describe the form of the estimators.

4.1 Notation

Adopting the standard notation of Little and Rubin (1987) and Laird (1988), let R_i denote a $T \times 1$ vector of indicator variables for the ith subject, where $R_{it} = 1$ if Y_{it} is observed, and $R_{it} = 0$ if Y_{it} is missing. Given R_i, the complete data vector Y_i can be partitioned into $Y_i = (Y_{oi}, Y_{mi})$, where Y_{oi} are the components of Y_i which are observed ($R_{it} = 1$), and Y_{mi} denotes the values of Y_i that are missing ($R_{it} = 0$). We can also partition μ_i, X_i^T, V_i and Δ_i in a similar fashion to obtain (μ_{oi}, μ_{mi}), (X_{oi}^T, X_{mi}^T), and so on. Note that the covariate matrix, X_i, is assumed to contain no missing values.

Next, let φ denote the vector of parameters of the nonresponse model, and $f(r_i|y_i, \varphi)$ denote the joint distribution of R_i given y_i and φ. If the nonresponse mechanism is independent of Y_i, $f(r_i|y_i, \varphi) = f(r_i|\varphi)$ and the responses are said to be *missing completely at random* (MCAR). When the probability of nonresponse depends on the observed response, Y_{oi}, but not on the missing values, Y_{mi}, the responses are said to be *missing at random* (MAR) and $f(r_i|y_i, \varphi) = f(r_i|y_{oi}, \varphi)$ (Little and Rubin, 1987). Intuitively, MCAR means that missingness is unrelated to the response, while MAR implies that missingness is related to the observed responses but unrelated to any responses that are missing. Clearly, the former is a special case of the latter.

4.2 GEE and Likelihood Equations with Missing Responses

With missing responses, the GEE for β is based only on the observed portion of the data. The same basic set of estimating equations are used, except Y_i, μ_i, X_i, V_i and Δ_i are replaced by their "observed" counterparts to yield

$$U_o(\beta) = \sum_{i=1}^{N} X_{oi}{}^T \Delta_{oi} V_{oi}^{-1}(y_{oi} - \mu_{oi}) = 0. \tag{10}$$

Note that the dimensionalities of y_{oi}, μ_{oi}, X_{oi}, V_{oi} and Δ_{oi} are all conformable, but potentially different for each i depending on the pattern of missing data. With incomplete or missing responses, these equations yield consistent estimates of β if the responses are MCAR, since $E(Y_{oi}|R_i) = \mu_{oi}$. However, the GEE estimators will be biased when the responses are MAR, since $E(Y_{oi} - \mu_{oi}) \neq 0$, conditional on the response.

Next, we consider the form of the likelihood equations for the "mixed parameter" model when there are missing responses. Let $f(y_i)$ denote the joint distribution of Y_i, then the observed data, (Y_{oi}, R_i), has joint distribution

$$f(y_{oi}, r_i|\varphi) = \sum f(y_{mi}|y_{oi}) f(y_{oi}) f(r_i|y_i, \varphi),$$

where summation is over all possible values of Y_{mi}. If the responses are assumed to be MAR,

$$f(y_{oi}, r_i|\varphi) = f(y_{oi}) f(r_i|y_{oi}, \varphi).$$

Since we are not interested in making likelihood-based inferences about φ, the contribution of $f(r_i|y_{oi}, \varphi)$ to the likelihood can be ignored. Thus, when responses are MAR, and φ is independent of the parameters of $f(y_i)$, the nonresponse mechanism is said to be *ignorable*. Note, however, that Y_{oi} and R_i are not independent, and that the expectation of Y_{oi}, given r_i depends on the nonresponse model (Laird, 1988).

When the nonresponse mechanism is assumed to be *ignorable*, the objective is to maximize the incomplete-data likelihood. For the "mixed parameter" model, it can be shown that the derivative of the log of the incomplete-data likelihood with respect to β and α is,

$$\sum_{i=1}^{N} \mathcal{E}\left[\frac{\partial l_i(\Psi_i, \Omega_i|y_i)}{\partial(\beta,\alpha)}\right] = \sum_{i=1}^{N} \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ 0 & \frac{\partial \Omega_i}{\partial \alpha} \end{pmatrix}^T \begin{pmatrix} V_i^{-1} & 0 \\ -F_i V_i^{-1} & I \end{pmatrix} \begin{pmatrix} \mathcal{E}(y_i) - \mu_i \\ \mathcal{E}(w_i) - v_i \end{pmatrix} = 0, \tag{11}$$

where expectation is taken with respect to Y_{mi} given Y_{oi}, that is,

$$\mathcal{E}(y_i) = E_{(Y_{mi}|y_{oi})}(y_i) = (y_{oi}, \mathcal{E}(y_{mi}))$$
$$\mathcal{E}(w_i) = E_{(Y_{mi}|y_{oi})}(w_i) = (w_{oi}, \mathcal{E}(w_{mi})).$$

Note that (11) is identical to the complete data likelihood equations except that y_i and w_i are replaced by $\mathcal{E}(y_i)$ and $\mathcal{E}(w_i)$. This yields the following likelihood equations for β:

$$\sum_{i=1}^{N}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}(\mathcal{E}(y_i)-\mu_i)$$
$$=\sum_{i=1}^{N}X_i^T\Delta_i V_i^{-1}(\mathcal{E}(y_i)-\mu_i)=0.$$

Note that the likelihood equations above use X_{mi}, the rows of X_i corresponding to the missing responses.

This result establishes the most general form of the *EM* algorithm (Dempster, Laird and Rubin, 1977), with the expectation-step imputing values for the missing data, given the current values of the parameters (β,α); while the maximization-step solves the usual scoring equations using the adjusted values

$$(y_{oi}, \mathcal{E}(y_{mi}), w_{oi}, \mathcal{E}(w_{mi})),$$

as if it were a sample of complete data. The likelihood-based estimates are easily calculated using the *EM* algorithm, since the expectations needed at the expectation-step are readily computed from the information available at each iteration.

For the likelihood-based estimators, orthogonality between the mean and time dependence parameters no longer holds when there are missing responses. These estimators will be biased if the time dependence has been misspecified, regardless of whether the missing data mechanism is MCAR or MAR. One exceptional case is the likelihood-based estimator that incorrectly assumes the responses to be independent. This estimator, which is identical to the "working independence" GEE estimator, is unbiased when the responses are MCAR.

In the next subsection, we describe how the asymptotic bias of the estimators based on (10) and (11) can be computed, when the likelihood or time dependence has been misspecified.

4.3 Computing Asymptotic Bias

Let $\hat{\beta}$ denote the solution to $\sum_{i=1}^{N}S_i(\beta)=0$, where $S_i(\beta)$ denotes the ith individual's contribution to the GEE or likelihood equations for β, under some working assumption about the time dependence. Then, with complete data and assuming that the model for the marginal expectations have been correctly specified, we have that under regularity conditions

$$E\left[\sum_{i=1}^{N}S_i(\beta)\right]=0 \Rightarrow \hat{\beta}\xrightarrow{P}\beta,$$

even when the time dependence has been incorrectly specified. Note that expectation is taken with respect to the true process generating the data. However, with incomplete data,

$$E\left[\sum_{i=1}^{N}S_i(\beta)\right]\neq 0,$$

when either (i) $S_i(\beta)$ denotes the β-component of the joint score, $S_i(\beta,\alpha)=[\partial\log L(y_{oi})]/[\partial(\beta,\alpha)]$, and the time dependence is misspecified, or (ii) $S_i(\beta)$ denotes the ith individual's contribution to the GEE, $U_o(\beta)$, and the responses are MAR. Instead,

$$E\left[\sum_{i=1}^{N}S_i(\beta^*)\right]=0 \Rightarrow \hat{\beta}\xrightarrow{P}\beta^*.$$

We are interested in assessing ($\beta^*-\beta$), the asymptotic bias of $\hat{\beta}$. But

$$E\left[\sum_{i=1}^{N}S_i(\beta)\right]=0,$$

does not in general have a closed-form analytic solution as a function of β. However, as noted in Rotnitzky and Wypij (1992), since for any fixed β, $\Sigma_{i=1}^{N}S_i(\beta)$ is a function of (Y_i, R_i, X_i), it has expectation given by the sum of all its possible realizations weighted by their respective probabilities. Thus, in order to solve

$$E\left[\sum_{i=1}^{N}S_i(\beta^*)\right]=0,$$

we can simply consider an artificial sample comprised of one observation for each possible realization of (Y_i, R_i, X_i). Then, we can solve for β^* in the usual way, except that we weight each $S_i(\beta)$ by its respective probability.

The probability of (Y_i, R_i, X_i) can be written as

$$\text{pr}(Y_i,R_i,X_i)=\text{pr}(R_i|Y_i,X_i)\text{pr}(Y_i|X_i)\text{pr}(X_i)$$

and is then fully determined by specifying:

1. The probability distribution of the covariates, X_i.
2. The model for the marginal probability of Y_i given X_i and β; and the model for the time dependence.
3. The missing data mechanism. That is, the model for the probability of missingness and a value of φ, $\text{pr}(R_i|Y_i,X_i,\varphi)$.

For specifying 1 and 2 above, we use the same models and designs as in subsection 3.1. For 3, we assume a *monotone* missing data pattern. That is, if Y_{ij} is missing, then Y_{ik} is also missing for every $k>j$. Furthermore, we consider two missing data mechanisms. The first assumes that the responses are MCAR, the second assumes that the responses are MAR. Specifically, we assume that the binary response on the first occasion is always observed, that is, $R_{i1}=1$, and that responses at times 2 and 3 are (i) MCAR, and (ii) MAR. When the responses are MCAR,

$$\text{pr}(R_{i2}=1|Y_{i1})=\text{pr}(R_{i2}=1)=(1-\varphi),$$
$$\text{pr}(R_{i3}=1|R_{i2},R_{i2})=\text{pr}(R_{i3}=1|R_{i2})=R_{i2}(1-\varphi);$$

and when the responses are MAR,

$$\text{pr}(R_{i2} = 1 | Y_{i1}) = (1 - \varphi)^{(1-Y_{i1})},$$

$$\text{pr}(R_{i3} = 1 | R_{i2}, R_{i2}) = R_{i2}(1 - \varphi)^{(1-Y_{i2})};$$

where $\varphi = (0.1, 0.2, 0.5)$.

4.4 Results

In this section, we compare the asymptotic bias for both the group and time effects when there are incomplete responses. First, we consider the case where the missing-data mechanism is MCAR. As noted previously, the GEE estimators are asymptotically unbiased when the responses are MCAR. For both design A and B, the bias for estimators based on "mixed parameter" models that assume an incorrect likelihood is minimal. The relative bias of the group effect is less than 1% even when the probability of dropout, φ, is as large as 0.5. The bias tends to be slightly larger for the group effect in design B. However, the form of the true joint probability representation has little impact on the bias for the group effect.

For the time effect, the relative bias is less than 1% when $\varphi = 0.1$, and is not more than 2% even when $\varphi = 0.5$. In general, when the responses are MCAR, the relative bias is small, and appears to be more a function of the degree of missingness than the amount of misspecification. Recall that one extreme form of misspecification, "working independence," yields asymptotically unbiased estimators.

Next, we consider the case where the missing-data mechanism is MAR. In Tables 1–4, we present the asymptotic relative bias for estimators of the group and time effects in designs A and B. For designs A and B, the relative bias for estimators of the group effect under the "mixed parameter" model tends to increase monotonically with increases in the degree of missingness, regardless of the form of the true joint probability representation. For design A, the relative bias is not more than 1% when $\varphi = 0.1$, and is less than 4% even when $\varphi = 0.5$. For design B, the relative bias is less than 2% when $\varphi = 0.1$, and is not more than 8% even when $\varphi = 0.5$. For design A, the degree and pattern of bias for the GEE estimators is very similar to that for the estimators based on the "mixed parameter" model. However, for design B, the relative bias of the group effect is discernibly larger for the GEE estimators when the true joint probability has a log-linear representation. For this design, the relative bias of the GEE estimator, assuming "exchangeable" or "pairwise" correlation, is about 4% when $\varphi = 0.1$, and approximately 30% when $\varphi = 0.5$, while the bias of the "working independence" GEE estimator is considerably smaller. Thus, in situations where the asymptotic efficiency of the GEE estimator is poor, the asymptotic bias is also noticeably larger.

For the time effect, biases are very dependent on the assumed model. When the true joint probability has a log-linear representation with $\omega = 5$, the relative bias of "working independence" estimators is approximately 16% when $\varphi = 0.1$, 35% when $\varphi = 0.2$, and over 110% when $\varphi = 0.5$. When the true joint probability has Bahadur's representation with $\rho = .45$, the relative bias is approximately 7% when $\varphi = 0.1$, 16% when $\varphi = 0.2$, and almost 50% when $\varphi = 0.5$. With the

TABLE 1
Percent asymptotic relative bias for GEE and PML estimators in design A, when the true underlying joint distribution has a log-linear representation and the missing data mechanism is MAR

Effect	ω	φ	Independence GEE (PML)	Exchangeable		Pairwise	
				GEE	PML	GEE	PML
Group	0	0.1	0.9	0.3	0.4	0.3	0.4
		0.2	1.7	0.6	0.8	0.6	0.9
		0.5	2.7	1.5	2.1	1.5	2.1
	2	0.1	1.0	0.0	0.0	0.0	0.0
		0.2	1.8	0.0	0.0	0.0	0.0
		0.5	3.1	−0.1	0.1	0.1	0.1
	5	0.1	1.1	0.0	0.0	0.0	0.0
		0.2	2.0	−0.1	0.0	0.0	0.0
		0.5	3.7	−0.2	0.0	0.0	0.0
Time	0	0.1	8.2	−1.3	−1.4	0.1	−0.1
		0.2	17.8	−2.3	−2.9	0.3	−0.4
		0.5	60.2	−2.9	−9.2	2.8	−3.5
	2	0.1	14.2	−1.3	−1.0	0.0	0.0
		0.2	30.7	−2.8	−2.6	0.0	0.0
		0.5	102.4	−7.9	−6.5	0.0	−0.4
	5	0.1	15.7	−1.0	−0.1	0.0	0.0
		0.2	34.1	−2.2	−0.2	0.0	0.0
		0.5	113.1	−8.3	−0.6	0.0	0.0

TABLE 2
Percent asymptotic relative bias for GEE *and* PML *estimators in design B, when the true underlying joint distribution has a log-linear representation and the missing data mechanism is* MAR

Effect	ω	φ	Independence GEE (PML)	Exchangeable GEE	Exchangeable PML	Pairwise GEE	Pairwise PML
Group	0	0.1	−0.6	−1.7	−0.8	−1.5	−0.5
		0.2	−1.0	−3.6	−1.6	−3.2	−0.9
		0.5	−1.6	−9.6	−4.5	−8.6	−2.8
	2	0.1	−1.6	−3.8	−0.7	−3.5	−0.2
		0.2	−2.9	−7.9	−1.6	−7.4	−0.4
		0.5	−5.1	−23.0	−4.5	−21.2	−0.4
	5	0.1	−1.8	−4.4	0.8	−4.3	1.8
		0.2	−3.8	−9.9	0.6	−9.8	1.4
		0.5	−7.7	−29.6	−0.2	−29.4	2.0
Time	0	0.1	8.0	−1.2	−1.3	0.1	−0.1
		0.2	17.6	−2.2	−2.8	0.3	−0.4
		0.5	59.4	−2.7	−8.7	2.8	−3.6
	2	0.1	14.0	−1.3	−1.0	0.0	0.0
		0.2	30.3	−2.8	−2.1	−0.1	−0.1
		0.5	100.7	−8.1	−6.5	−0.7	−0.1
	5	0.1	15.8	−1.1	−0.2	−0.1	−0.2
		0.2	34.1	−2.4	−0.4	−0.2	0.2
		0.5	112.9	−9.1	−0.9	−1.2	3.3

TABLE 3
Percent asymptotic relative bias for GEE *and* PML *estimators in design A, when the true underlying joint distribution has Bahadur's representation and the missing data mechanism is* MAR

Effect	ω	φ	Independence GEE (PML)	Exchangeable GEE	Exchangeable PML	Pairwise GEE	Pairwise PML
Group	0	0.1	0.0	0.0	0.0	0.0	0.0
		0.2	0.0	0.0	0.0	0.0	0.0
		0.5	0.0	0.0	0.0	0.0	0.0
	0.1	0.1	0.1	0.0	0.1	0.0	0.1
		0.2	0.1	0.0	0.1	0.0	0.1
		0.5	−0.1	0.0	0.3	0.0	0.4
	0.3	0.1	0.3	0.0	0.1	0.0	0.2
		0.2	0.5	0.0	0.3	0.1	0.4
		0.5	0.3	0.2	1.1	0.5	1.2
	0.45	0.1	0.5	0.0	0.1	0.0	0.2
		0.2	0.9	0.0	0.3	0.2	0.5
		0.5	1.4	0.7	1.6	1.1	1.9
Time	0	0.1	0.0	0.0	0.0	0.0	0.0
		0.2	0.0	0.0	0.0	0.0	0.0
		0.5	0.0	0.0	0.0	0.0	0.0
	0.1	0.1	1.3	−0.3	−0.3	0.0	0.0
		0.2	2.7	−0.7	−0.7	0.0	0.0
		0.5	8.3	−1.7	−1.9	0.0	0.0
	0.3	0.1	4.4	−0.9	−0.8	0.0	0.0
		0.2	9.5	−1.7	−1.8	0.0	0.0
		0.5	29.6	−4.6	−5.4	0.4	−0.4
	0.45	0.1	7.4	−1.1	−1.0	0.0	0.0
		0.2	16.0	−2.2	−2.2	0.1	−0.1
		0.5	50.5	−6.0	−7.0	1.0	−0.9

TABLE 4
Percent asymptotic relative bias for GEE *and* PML *estimators in design B, when the true underlying joint distribution has Bahadur's representation and the missing data mechanism is* MAR

Effect	ω	φ	Independence GEE (PML)	Exchangeable		Pairwise	
				GEE	PML	GEE	PML
Group	0	0.1	0.0	0.0	0.0	0.0	0.0
		0.2	0.0	0.0	0.0	0.0	0.0
		0.5	0.0	0.0	0.0	0.0	0.0
	0.1	0.1	0.1	0.0	0.1	0.0	0.1
		0.2	0.2	0.0	0.2	0.0	0.2
		0.5	0.5	−0.1	0.5	0.0	0.5
	0.3	0.1	0.3	0.0	0.4	0.0	0.4
		0.2	0.7	−0.1	0.8	0.1	0.8
		0.5	2.2	−0.3	1.9	−0.1	2.0
	0.45	0.1	0.5	0.0	0.8	0.0	0.8
		0.2	1.2	−0.1	1.6	0.0	1.7
		0.5	4.3	−0.7	4.2	−0.4	4.5
Time	0	0.1	0.0	0.0	0.0	0.0	0.0
		0.2	0.0	0.0	0.0	0.0	0.0
		0.5	0.0	0.0	0.0	0.0	0.0
	0.1	0.1	1.3	−0.3	−0.3	0.0	0.0
		0.2	2.7	−0.7	−0.7	0.0	0.0
		0.5	8.3	−1.7	−1.9	0.0	0.0
	0.3	0.1	4.4	−0.9	−0.8	0.0	0.0
		0.2	9.5	−1.7	−1.8	0.0	0.0
		0.5	29.6	−4.6	−5.4	0.4	−0.4
	0.45	0.1	7.4	−1.1	−1.0	0.0	0.0
		0.2	16.0	−2.2	−2.2	0.1	−0.1
		0.5	50.3	−6.0	−7.1	1.0	−1.1

pairwise association assumption, the biases are generally negligible for both the GEE and PML estimators. In general, when the responses are MAR, the relative bias appears to be a function of both the degree of missingness and the amount of model misspecification.

5. CONCLUSION

In the previous sections, we have reviewed both likelihood-based and non-likelihood approaches to analysing longitudinal binary responses. The GEE methodology yields consistent estimates of the regression parameters provided that the model for the mean has been correctly specified. "Robust" estimates of the variance of the estimated parameters are easily obtained. These are consistent regardless of how the correlation between responses has been specified. In an effort to obtain more efficient estimates of both the mean and association parameters, extensions of the GEE methodology to allow joint estimation of the regression and association parameters have been proposed. A limitation of the proposed extensions of the GEE is that consistency of the estimates of the regression parameters requires the correct specification of the model for both the mean and the association. Thus, estimates of the regression parameters may be asymptotically biased when the model for the mean is correctly specified, but that for the association is misspecified.

The GEE approaches are not likelihood-based methodologies; that is, they do not require the complete specification of the joint distribution of the repeated binary responses. We also described a "mixed parameter" model that is based on a log-linear specification of the joint distribution. This model yields likelihood equations for the regression parameters that are identical in form to the GEE. Assuming that the mean structure is correctly specified, the "mixed parameter" model provides consistent estimates of the regression parameters that are robust to misspecification of the association between responses. Other likelihood-based approaches, that parameterize the association in terms of the *marginal* associations, do not share this desirable property.

Comparing these two approaches, we find that the simple "working independence" estimator is highly efficient for traditional longitudinal designs, when the data is complete and balanced. However, when the design includes time-varying covariates, it becomes much more important to obtain a close approximation to $\mathrm{cov}(Y_i)$ in order to achieve high efficiency. In this regard, the "mixed parameter" model seems to offer a more flexible approach to modelling $\mathrm{cov}(Y_i)$ for different representations of the association.

Although all the estimators perform well in terms of bias when the data are MCAR, in practice, when there are missing responses, the distinction between responses MCAR and MAR can usually not be made with much certainty. With incomplete responses, the GEE approach performs remarkably well when the responses are MAR and the design does not include time-varying covariates. For both the GEE and likelihood-based approaches, there is a substantial reduction in the bias of the time effects when a close approximation to cov(Y_i) is obtained. Thus, when interest is focussed primarily on the time effects and there are missing responses, the "working independence" estimators cannot be recommended. For group effects, this distinction is not quite so clear, and it seems to depend on whether group is a time-stationary or time-varying covariate. Finally, although in many instances the asymptotic biases of the GEE and likelihood-based approaches are comparable, there may be substantial differences in terms of efficiency.

In conclusion, the importance of accurately modelling the correlation among the repeated responses in a longitudinal study will depend on a number of factors: the design of the study, the parameters of interest and whether or not there are missing data. Fortunately, for many practical situations, it appears that nearly efficient and unbiased estimates of the regression parameters for the marginal expectation can be obtained even when the true association between the responses is only crudely approximated.

ACKNOWLEDGMENT

This research was supported by Grants GM29745 and MH17119 from the National Institutes of Health.

REFERENCES

Baalam, L. N. (1968). A two-period design with t^2 experimental units. *Biometrics* **24** 61–73.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction. Stanford Mathematical Studies in the Social Sciences VI* (H. Solomon, ed.) 158–168. Stanford Univ. Press.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restriction. *J. Econometrics* **34** 305–324.

Cox, D. R. (1972). The analysis of multivariate binary data. *J. Roy. Statist. Soc. Ser. C* **21** 113–120.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. C* **49** 1–39.

Cox, D. R. and Reid, N. (1989). On the stability of maximum-likelihood estimators of orthogonal parameters. *Canad. J. Statist.* **17** 229–233.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11** 427–444.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80** 141–151.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

Jones, B. and Kenward, M. (1989). *Design and Analysis of Cross-Over Trials.* Chapman and Hall, London.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Korn, E. L. and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* **35** 795–802.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7** 305–315.

Laird, N. M. (1991). Topics in likelihood-based methods for longitudinal data analysis. *Statist. Sinica* **1** 33–50.

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 (with discussion). *J. Roy. Statist. Soc. Ser. B* **54** 3–40.

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* **78** 153–160.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd ed. Chapman and Hall, New York.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5** 99–135.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44** 1033–1048.

Prentice, R. L., and Zhao, L. P. (1991). Estimating equations for parameters in mean and covariances of multivariate discrete and continuous responses. *Biometrics* **47** 825–839.

Rotnitzky, A. and Wypij, D. (1992). A note on the bias of estimators with missing data. *Amer. Statist.* To appear.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *Internat. Statist. Rev.* **54** 221–226.

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *Amer. Statist.* **39** 95–101.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–26.

Zeger, S. L. (1988). The analysis of discrete longitudinal data: commentary. *Statistics in Medicine* **7** 161–168.

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

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77** 642–648.

Zhao, L. P., Prentice, R. L. and Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *J. Roy. Statist. Soc. Ser. B* **54** 805–811.