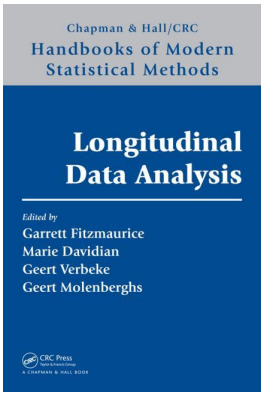

Longitudinal Data Analysis
Handbooks of Modern Statistical Methods

Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, Geert Molenberghs

Generalized estimating equations for longitudinal data analysis

Publication details
https://www.routledgehandbooks.com/doi/10.1201/9781420011579.ch3
Stuart Lipsitz, Garrett Fitzmaurice
Published online on: 11 Aug 2008

CHAPTER 3

Generalized estimating equations for longitudinal data analysis

Stuart Lipsitz and Garrett Fitzmaurice

Contents

3.1 Introduction

In this chapter we discuss the generalized estimating equations (GEE) approach for analyzing longitudinal data. Over the past 20 years, the GEE approach has proven to be an exceedingly useful method for the analysis of longitudinal data, especially when the response variable is discrete (e.g., binary, ordinal, or a count). When the longitudinal response is discrete, linear models (e.g., linear mixed-effects models) are not very appealing for relating changes in the mean response to covariates for at least two main reasons. First, with a discrete response there is intrinsic dependence of the variability on the mean. Second, the range of the mean response (e.g., a proportion or rate for a response that is binary or a count, respectively) is constrained. In the setting of regression modeling of a univariate response, both of these aspects of the response can be conveniently accommodated within generalized linear models via known variance and link functions.

However, a straightforward application of generalized linear models to longitudinal data is not appropriate, due to the lack of independence among repeated measures obtained on the same individual. There has been extensive statistical literature on extending generalized linear models to the longitudinal data setting. One approach for accounting for the within-subject association is via the introduction of random effects in generalized linear models. This leads to a class of models known as generalized linear mixed models (GLMMs); see

Chapter 4 for a very comprehensive and expository discussion of these models. Generalized linear mixed models represent one way to extend generalized linear models to longitudinal data; they extend in a natural way the conceptual approach represented by the linear mixed-effects models for continuous responses (e.g., Laird and Ware, 1982; see also Chapter 1). There is a second approach for extending generalized linear models to longitudinal data that leads to a class of regression models that are known as *marginal models*. The term *marginal* in this context is used to emphasize that the model for the mean response at each occasion depends only on the covariates of interest, and does not incorporate dependence on random effects or previous responses. This is in contrast to GLMMs, where the mean response is modeled not only as a function of covariates but is conditional also on random effects. The most salient feature of marginal models is a regression model, with appropriately specified link function, relating the mean response at each occasion to the covariates. For estimation of the regression model parameters, marginal models do not necessarily require distributional assumptions for the vector of longitudinal responses. When full distributional assumptions for the vector of responses are avoided, the marginal model is said to be *semi-parametric*, and this leads to a method of estimation known as *generalized estimating equations* (Liang and Zeger, 1986), the main focus of this chapter.

Thus, the GEE approach has its basis in one particular type of extension of generalized linear models to longitudinal, or more generally, cluster-correlated data. Before outlining the main features and properties of GEE methods in Section 3.2, it is useful to review some of its predecessors and consider the reasons why this method has been so widely adopted for longitudinal analyses.

Prior to the seminal companion papers on GEE by Liang and Zeger (1986) and Zeger and Liang (1986), statistical methods for the analysis of discrete longitudinal data had lagged somewhat behind the corresponding developments for continuous outcomes. The early foundations for statistical methods for estimation of marginal models for repeated categorical responses can be traced to a general approach developed by Grizzle, Starmer, and Koch (1969); this approach soon became known as the GSK method. Although the GSK method was developed originally as a very general method for the analysis of categorical data, Koch and Reinfurt (1971) and Koch et al. (1977) recognized that the method could be applied to the analysis of repeated measurements. The GSK method is founded on a weighted least-squares (WLS) approach that makes few assumptions about the within-subject association among the repeated categorical outcomes. Specifically, this WLS approach is based on a multinomial sampling model for the joint distribution of the vector of categorical outcomes and relies on the asymptotic normality of estimators of the corresponding multinomial probabilities. Recall that if a categorical outcome, with C levels, is measured repeatedly at n occasions, there are C^n possible response profiles; thus, the joint distribution of the vector of longitudinal responses is multinomial with C^n response probabilities ($C^n - 1$ non-redundant probabilities because the multinomial probabilities are constrained to sum to 1). In general, for a marginal model for the repeated categorical responses, the main interest is focused on the $n \times (C - 1)$ correlated marginal probabilities or transformations of these marginal probabilities (e.g., logit transformations). Moreover, these marginal probabilities can simply be expressed as linear transformations of the underlying multinomial probabilities. In the GSK method, the sample proportions (or means) at each occasion are obtained within each covariate stratum and grouped together to form a sample mean vector of length $n \times (C-1)$; the sample covariance matrix is also estimated. A marginal model is specified by relating the sample means (or known functions of the sample means such as the empirical logits) at each occasion to a linear function of stratum covariates. Appealing to the multivariate central limit theorem, and the use of the so-called delta method, the observed sample means (or functions of the sample means) have approximate normal distributions, given sufficient sample sizes within each stratum. The WLS method, with the sample mean vectors as the

outcome vectors and the inverse of the sample covariance matrices as the weight matrices, is then used to estimate the marginal model regression parameters.

At the time of its introduction, the GSK method provided a method for estimating a very general family of models for repeated categorical data, allowing non-linear link functions to relate the marginal expectations of the longitudinal responses to covariates. For example, the method allows the fitting of logistic regression models to repeated binary outcomes. At the heart of the GSK method are the notions of stratification of subjects on the basis of covariates and non-parametric estimation of the multinomial probabilities for the vector of repeated categorical responses within each stratum. Because the multinomial probabilities are estimated non-parametrically as the sample proportions within strata, asymptotically, the GSK method is equivalent to maximum likelihood (ML) estimation.

However, the GSK method has a number of restrictions that limit its usefulness for the analysis of longitudinal data. Because it relies on stratification, the method requires that all covariates be categorical or the reduction of quantitative covariates to categorical variables. Moreover, it also requires a relatively large number of subjects within each stratum to estimate the C^n multinomial probabilities; note that with $C = 4$ response levels for an outcome measured repeatedly at $n = 6$ occasions, there are $4^6 - 1$ or 4095 non-redundant multinomial probabilities. Implicitly, this means that the method can only be validly applied when (i) the number of repeated measures, n, is relatively small, thereby ensuring that the number of multinomial probabilities to be estimated does not grow exponentially; (ii) the number of covariate strata is relatively small; and (iii) the number of individuals is relatively large, ensuring a sufficient number of subjects within each stratum. Finally, the GSK method, as originally developed, also requires that the longitudinal study design be balanced on time in the sense that all subjects are measured at the same set of occasions; later refinements to the GSK method can handle balanced longitudinal designs with incompleteness due to missing observations.

As an alternative to GSK methods, full likelihood-based methods for estimating marginal models for repeated categorical data can, in principle, accommodate many of these common features of longitudinal studies (e.g., covariates that are both quantitative and categorical, missing data, and so on). However, in practice, ML fitting of marginal models for discrete longitudinal data has proven to be very challenging. Among the many challenges are: (i) it can be conceptually difficult to model higher-order associations in a flexible and interpretable manner that is consistent with the model for the marginal expectations; (ii) given a marginal model for the vector of repeated outcomes, the multinomial probabilities cannot, in general, be expressed in closed form as a function of the model parameters; and (iii) the number of multinomial probabilities grows exponentially with the number of repeated measures. As a result, ML estimation is feasible only for a relatively small number of repeated measures. Thus, although the development of likelihood-based methods for marginal models has been fertile ground for methodological research (e.g., Bahadur, 1961; McCullagh and Nelder, 1989; Zhao and Prentice, 1990; Lipsitz, Laird, and Harrington, 1990; Liang, Zeger, and Qaqish, 1992; Becker and Balagtas, 1993; Fitzmaurice, Laird, and Rotnitzky, 1993; Molenberghs and Lesaffre, 1994; Lang and Agresti, 1994; Glonek and McCullagh, 1995; Bergsma and Rudas, 2002; and many others), to date, ML methods that have been developed have also proven to be of only limited practical use.

Thus, by the end of the 1970s, the GSK method provided the first unified approach for extending generalized linear models to repeated categorical data, based on an application of non-iterative weighted least squares. The method was implemented in widely available statistical software (e.g., `proc catmod` in SAS). However, the method lacked versatility and could not accommodate many of the common features of longitudinal studies, namely mixtures of quantitative and categorical, time-invariant and time-varying, covariates and inherently unbalanced designs with both mistimed measurements and missing data. So, by

the early 1980s, the time was ripe for new methods for the analysis of discrete longitudinal data that could handle all of these aspects of longitudinal data in a relatively flexible way.

Finally, we note that the WLS estimation at the heart of the GSK method is similar in spirit to non-iterative empirical logistic regression for binomial data (see, for example, Cox and Snell, 1989). Empirical logistic regression is also applicable only when individuals can be grouped into strata based on their covariates; at the time of its introduction, this greatly limited the usefulness of empirical logistic regression in practice. However, as computing advanced, empirical logistic (WLS) regression for binomial data was soon replaced by ML methods that were far more versatile. In a similar vein, the GSK method for discrete longitudinal data was replaced not by ML methods, for all the reasons mentioned earlier, but by the GEE approach developed by Liang and Zeger (1986). By and large, the GEE approach has overcome most of the limitations of its predecessors and has fundamentally changed the way empirical researchers can approach the analysis of discrete longitudinal data.

3.2 Generalized estimating equations (GEE) for longitudinal data

In this section we briefly review the main features of marginal models for longitudinal data and discuss the key properties of the GEE approach. Because the GEE approach can be considered a multivariate extension of quasi-likelihood estimation, we first discuss estimation of the regression parameters in generalized linear models for a *univariate* outcome, before considering estimation of the regression parameters in a marginal model for longitudinal outcomes.

3.2.1 Notation

Before we begin our discussion of marginal models and the GEE approach, we introduce some notation. We assume that N subjects are measured repeatedly over time. We let Y_{ij} denote the response variable for the ith subject on the jth measurement occasion ($i = 1, \ldots, N; j = 1, \ldots, n_i$). The response variable can be continuous or discrete (e.g., binary, ordinal, polytomous, or a count). The type of response variable (e.g., binary or count) will have implications for model specification. In general, we do not assume that subjects have the same number of repeated measures or that they are measured at a common set of occasions. To accommodate such *unbalanced* longitudinal data (i.e., repeated measurements that are not obtained at a common set of occasions), we assume that there are n_i repeated measurements of the response on the ith subject and that each Y_{ij} is observed at time t_{ij}. We can group the responses into an $n_i \times 1$ vector of responses denoted by $\boldsymbol{Y}_i$. Finally, associated with each response, Y_{ij}, there is a $p \times 1$ vector of covariates, $\boldsymbol{X}_{ij}$. We note that $\boldsymbol{X}_{ij}$ may include covariates whose values do not change throughout the duration of the study and covariates whose values change over time. The former are referred to as time-stationary or between-subject covariates (e.g., gender and fixed experimental treatments or interventions), whereas the latter are referred to as time-varying or within-subject covariates (e.g., time since baseline, current smoking status, and environmental exposures that can vary over time). We can group the vectors of covariates into an $n_i \times p$ matrix of covariates denoted by X_i.

3.2.2 Defining features of marginal models

As mentioned in Section 3.1, the defining feature of marginal models is a regression model relating the mean response at each occasion, via a suitable link function, to the covariates. With a marginal model, the main focus is on making inferences about population means.

As a result, marginal models for longitudinal data separately model the mean response and the within-subject association among the repeated responses. In a marginal model, the goal is to make inferences about the former, whereas the latter is regarded as a nuisance characteristic of the data that must be taken into account in order to make correct inferences about changes in the population mean response over time.

A marginal model for longitudinal data has the following three-part specification:

1. The conditional expectation of each response, $E(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}$, is assumed to depend on the covariates through a known link function $h^{-1}(\cdot)$, e.g., $\text{logit}(\mu_{ij})$ or $\log(\mu_{ij})$,
$$h^{-1}(\mu_{ij}) = \eta_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta},$$
where $\boldsymbol{\beta}$ is a $p \times 1$ vector of marginal regression parameters.
2. The conditional variance of each Y_{ij}, given $\boldsymbol{X}_{ij}$, is assumed to depend on the mean according to
$$\text{Var}(Y_{ij}) = \phi\, v(\mu_{ij}),$$
where $v(\mu_{ij})$ is a known "variance function" (i.e., a known function of the mean, μ_{ij}) and ϕ is a scale parameter that may be fixed and known or may need to be estimated. Note that dependence of $\text{Var}(Y_{ij})$ on the covariates $\boldsymbol{X}_{ij}$ is suppressed from notation.
3. The conditional within-subject association among the vector of repeated responses, given the covariates, is assumed to be a function of an additional vector of association parameters, say $\boldsymbol{\alpha}$ (and also depends upon the means, μ_{ij}). For example, the components of $\boldsymbol{\alpha}$ might represent the pairwise correlations or log odds ratios among the repeated responses.

This three-part specification of a marginal model makes the extension of generalized linear models to longitudinal data more transparent. The first two parts of the marginal model correspond to the standard generalized linear model, albeit with no explicit distributional assumptions about the responses. It is the third component, the incorporation of the within-subject association among the repeated responses from the same individual, that represents the main extension of generalized linear models to longitudinal data.

It is also worth emphasizing that this three-part specification of a marginal model can, in principle, be extended by making full distributional assumptions about the vector of responses. To do so would require that all two- and higher-way associations be specified in the third component of the model. However, for reasons mentioned in Section 3.1, ML fitting of marginal models for discrete longitudinal data has proven to be very challenging. Consequently, the third component of a marginal model is typically specified in terms of the two-way or pairwise association among the repeated responses from the same individual.

In summary, marginal models are a very natural way to extend generalized linear models to longitudinal responses. Marginal models specify a generalized linear model for the longitudinal responses at each occasion but also include a model for the within-subject association among the responses. A crucial aspect of marginal models is that the mean response and within-subject association are modeled separately. This separation of the modeling of the mean response and the association among responses has important implications for interpretation of the regression parameters in the model for the mean response. In particular, the regression parameters, $\boldsymbol{\beta}$, in the marginal model have so-called *population-averaged* interpretations. That is, they describe how the mean response in the population is related to the covariates. For example, regression parameters in a marginal model might have interpretation in terms of contrasts of the changes in the mean responses in subpopulations (e.g., different treatment, intervention, or exposure groups); see Chapter 7 for a detailed discussion of various aspects of interpretation of marginal model parameters.

3.2.3 Quasi-likelihood and generalized estimating equations

As noted in Section 3.2.2, the three-part marginal model specification does not require full distributional assumptions for the repeated responses, only a regression model for the mean response. The avoidance of distributional assumptions can be advantageous because, in general, there is no convenient specification of the joint multivariate distribution of $\boldsymbol{Y}_i$ for marginal models when the responses are discrete. When full distributional assumptions are avoided, the model is referred to as being *semi-parametric* because there is a parametric component $\boldsymbol{\beta}$ and a non-parametric component determined by the nuisance parameters for the second- and higher-order moments. The avoidance of distributional assumptions for $\boldsymbol{Y}_i$ leads to a method of estimation known as *generalized estimating equations*. Thus, the GEE approach can be thought of as providing a convenient alternative to ML estimation. The GEE approach proposed by Liang and Zeger (1986) is a multivariate generalization of the quasi-likelihood approach for generalized linear models introduced by Wedderburn (1974). To better understand this connection to quasi-likelihood estimation, in the following we briefly outline the quasi-likelihood approach for generalized linear models for a *univariate* response before discussing its extension to multivariate responses.

In the following, we now assume N independent observations of a *scalar* response variable, Y_i. Associated with the response, Y_i, there are p covariates, $X_{i1}, \ldots, X_{ip}$. We assume that primary interest is in relating the mean of Y_i, $\mu_i = E(Y_i|X_{i1}, \ldots, X_{ip})$, to the covariates. In generalized linear models, the distribution of the response is assumed to belong to the exponential family of distributions (e.g., normal, Bernoulli, binomial, and Poisson). As described in the first part of the specification of the marginal model in Section 3.2.2, in generalized linear models a transformation of the mean response, μ_i, is linearly related to the covariates via an appropriate link function,

$$h^{-1}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip},$$

where the link function $h^{-1}(\cdot)$ is a known function, such as $\log(\mu_i)$. The assumption that Y_i has an exponential family distribution has implications for the variance of Y_i. In particular, a feature of exponential family distributions is that the variance of Y_i can be expressed in terms of a known function of the mean and a scale parameter,

$$\text{Var}\,(Y_i) = \phi\, v(\mu_i),$$

where the scale parameter $\phi > 0$. The variance function, $v(\mu_i)$, describes how the variance of the response is functionally related to the mean of Y_i; the variance function was previously discussed in the second part of the specification of the marginal model in Section 3.2.2.

Next, we consider estimation of $\boldsymbol{\beta}$. Assuming Y_i follows an exponential family density, with $\text{Var}(Y_i) = \phi\, v(\mu_i)$, the ML estimator of $\boldsymbol{\beta}$ is obtained as the solution to the likelihood score equations,

$$\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)' \frac{1}{\phi\, v(\mu_i)} \{Y_i - \mu_i(\boldsymbol{\beta})\} = \mathbf{0},$$

where $\partial \mu_i / \partial \boldsymbol{\beta}$ is the $1 \times p$ vector of derivatives, $\partial \mu_i / \partial \beta_k, k = 1, \ldots, p$. Interestingly, the likelihood equations for generalized linear models depend only on the mean and variance of the response (and the link function). Consequently, Wedderburn (1974) suggested using them as "estimating equations" for any choice of link or variance function, even when the particular choice of variance function does not correspond to an exponential family distribution. That is, Wedderburn (1974) proposed estimating $\boldsymbol{\beta}$ by solving the quasi-likelihood equations,

$$\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \{Y_i - \mu_i(\boldsymbol{\beta})\} = \mathbf{0}. \tag{3.1}$$

Wedderburn (1974) showed that for any choice of weights, V_i, the quasi-likelihood estimator of $\boldsymbol{\beta}$, say $\widehat{\boldsymbol{\beta}}$, is consistent and asymptotically normal. The choice of weights $V_i = \text{Var}(Y_i)$ yields the estimator with smallest variance among all estimators in this class. Note that in generalized linear models, it is assumed that $V_i = \text{Var}(Y_i) = \phi\, v(\mu_i)$, and this assumption is sufficient to characterize the distribution within the exponential family. Thus, the optimal estimating equations (or quasi-likelihood equations) coincide with the score equations for the case where Y_i is assumed to have an exponential family distribution.

In summary, Wedderburn (1974) proposed estimators of $\boldsymbol{\beta}$ that do not require distributional assumptions on the response. This allows more flexible models for variability, e.g., incorporating overdispersion. Moreover, quasi-likelihood estimation only requires correct specification of the model for the mean to yield consistent and asymptotically normal estimators of $\boldsymbol{\beta}$. That is, a key property of quasi-likelihood estimators is that they are consistent even when the variance of the response has been misspecified, that is, $V_i \neq \text{Var}(Y_i)$. Specifically, it can be shown that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$, the estimator for $\boldsymbol{\beta}$ obtained from (3.1) with a particular choice of V_i, satisfies

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, C_\beta),$$

where

$$C_\beta = \lim_{N\to\infty} I_0^{-1} I_1 I_0^{-1},$$

$$I_0 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right),$$

and

$$I_1 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)' V_i^{-1}\, \text{Var}(Y_i)\, V_i^{-1} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right).$$

Consistent estimators of the asymptotic covariance of the estimated regression parameters can be obtained using the empirical estimator of C_β first suggested by Cox (1961), and later proposed by Huber (1967), White (1982), and Royall (1986). The empirical variance estimator is obtained by evaluating $\partial\mu_i/\partial\boldsymbol{\beta}$ at $\widehat{\boldsymbol{\beta}}$ and substituting $(Y_i - \widehat{\mu}_i)^2$ for $\text{Var}(Y_i)$; this is widely known as the *sandwich* variance estimator. Moreover, it can be shown that the same asymptotic distribution holds when V_i is estimated rather than known, with V_i replaced by estimated weights, say $\widehat{V}_i$.

Next, we consider the multivariate extension of this quasi-likelihood approach to the setting of marginal models for longitudinal responses. In the following, $\boldsymbol{Y}_i$ denotes an $n_i \times 1$ vector of responses (similarly, $\boldsymbol{\mu}_i$ denotes an $n_i \times 1$ vector of means) and X_i is an $n_i \times p$ matrix of covariates. The fundamental idea underlying the GEE approach is to extend the quasi-likelihood equations (or estimating equations) to the multivariate setting by replacing Y_i and μ_i by their corresponding multivariate counterparts ($\boldsymbol{Y}_i$ and $\boldsymbol{\mu}_i$) and using a *matrix* of weights V_i. Thus, the estimating equations for $\boldsymbol{\beta}$ in a marginal model are given by

$$u_\beta(\boldsymbol{\beta}) = \sum_{i=1}^{N} D_i' V_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = \mathbf{0}, \qquad (3.2)$$

where $D_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}$ is now an $n_i \times p$ matrix of derivatives whose kth row is $\partial\boldsymbol{\mu}_i/\partial\beta_k$. As with the quasi-likelihood approach, the optimal choice of V_i is to take $V_i = \text{Cov}(\boldsymbol{Y}_i)$, the $n_i \times n_i$ covariance matrix for $\boldsymbol{Y}_i$. Note that the GEE given by (3.2) is simply the multivariate analog of the quasi-likelihood equations given by (3.1). However, unlike the quasi-likelihood equations given by (3.1), the weight matrix, V_i, depends not only on $\boldsymbol{\beta}$ but also on the pairwise associations among the longitudinal responses.

3.2.4 Estimation: Generalized estimating equations

As mentioned in the previous section, GEE can be regarded as a multivariate extension of the quasi-likelihood estimating equations. Note that the scalar weight in (3.1) is replaced by the weight matrix, V_i, in (3.2). In general, the assumed covariance among the responses can be specified as

$$V_i = \phi A_i^{1/2} R_i(\boldsymbol{\alpha}) A_i^{1/2},$$

where $A_i = \text{diag}\{v(\mu_{ij})\}$ is a diagonal matrix with diagonal elements $v(\mu_{ij})$, which are specified entirely by the marginal means (i.e., by $\boldsymbol{\beta}$), $R_i(\boldsymbol{\alpha})$ is an $n_i \times n_i$ correlation matrix, and ϕ is a dispersion parameter. For the correlation matrix $R_i(\boldsymbol{\alpha})$, $\boldsymbol{\alpha}$ represents a vector of parameters associated with a specified model for $\text{Corr}(\boldsymbol{Y}_i)$, with typical element

$$\rho_{ist} = \rho_{ist}(\boldsymbol{\alpha}) = \text{Corr}(Y_{is}, Y_{it}; \boldsymbol{\alpha}), \quad s \neq t.$$

In the GEE approach, V_i is usually referred to as a "working covariance," where the term "working" is used to emphasize that V_i is only an approximation to the true covariance, say $\Sigma_i = \text{Cov}(\boldsymbol{Y}_i)$. Sometimes, $R_i(\boldsymbol{\alpha})$ is referred to as a "working correlation" matrix; however, this implicitly assumes that the marginal variance assumption, $\text{Var}(Y_{ij}) = \phi\, v(\mu_{ij})$, is correct. Since both $R_i(\boldsymbol{\alpha})$ and $\text{Var}(Y_{ij})$ can be incorrectly specified, we prefer the use of the term "working covariance." Note that if $R_i(\boldsymbol{\alpha}) = I$, the $n_i \times n_i$ identity matrix, then the GEE reduces to the quasi-likelihood estimating equations for a generalized linear model that assume the repeated measures are independent.

Liang and Zeger (1986), and also Prentice (1988), parameterized the within-subject correlations directly as a linear function of $\boldsymbol{\alpha}$ (typically $\rho_{ist}(\boldsymbol{\alpha}) = \alpha_{st}$, although $\rho_{ist}(\boldsymbol{\alpha})$ could, in principle, be allowed to depend on covariates). When $\boldsymbol{\alpha}$ is known, the only unknown quantity in (3.2) is $\boldsymbol{\beta}$ and the solution to (3.2) is a consistent estimator of $\boldsymbol{\beta}$. In the usual case where V_i, and specifically $\boldsymbol{\alpha}$, is unknown, then we must parameterize and estimate $\rho_{ist}(\boldsymbol{\alpha}) = \text{Corr}(Y_{is}, Y_{it})$. Some common examples of models for the correlation include: "exchangeable," in which $\rho_{ist} = \alpha$ for all $s < t$; first-order autoregressive (AR(1)),

$$\rho_{ist} = \alpha^{|t-s|},$$

where $0 < \alpha < 1$, and the correlation decreases as the time between measurements ($|t-s|$) increases; and "unstructured," in which $\rho_{ist} = \alpha_{st}$. For estimation of $\boldsymbol{\alpha}$, let

$$U_{ist}(\boldsymbol{\beta}) = \frac{(Y_{is} - \mu_{is})(Y_{it} - \mu_{it})}{\phi\{v(\mu_{is})v(\mu_{it})\}^{1/2}},$$

which has expected value ρ_{ist}; the U_{ist} can be grouped together to form the $n_i(n_i-1)/2 \times 1$ vector $\boldsymbol{U}_i(\boldsymbol{\beta}) = (U_{i12}, U_{i13}, \ldots, U_{in_{i-1}n_i})'$. Also, let $\boldsymbol{\rho}_i(\boldsymbol{\alpha}) = E(\boldsymbol{U}_i; \boldsymbol{\alpha}) = (\rho_{i12}, \rho_{i13}, \ldots, \rho_{in_{i-1}n_i})'$. Then, a second set of (moment) estimating equations similar to (3.2) can be used to estimate $\boldsymbol{\alpha}$, given by

$$u_\alpha(\boldsymbol{\alpha}) = \sum_{i=1}^{N} E_i' W_i^{-1} \{\boldsymbol{U}_i(\boldsymbol{\beta}) - \boldsymbol{\rho}_i(\boldsymbol{\alpha})\} = \mathbf{0}, \tag{3.3}$$

where $W_i \approx \text{Cov}(\boldsymbol{U}_i)$ and $E_i = \partial \boldsymbol{\rho}_i(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}$. The working covariance matrix for $\boldsymbol{U}_i$ is typically specified as $\text{diag}\{\text{Var}(U_{ist})\}$. In their original paper on GEE, Liang and Zeger (1986) let W_i be the $(n_i \times n_i - 1)/2 \times (n_i \times n_i - 1)/2$ identity matrix, whereas Prentice (1988) suggested letting W_i be a diagonal matrix with the approximate variances of U_{ist} along the diagonal.

For the special case where the outcome is binary, an alternative to the correlation as a measure of association between pairs of binary responses is the odds ratio. The odds ratio has many desirable properties (e.g., see Bishop, Fienberg, and Holland, 1975, Chapter 11) and has a more straightforward interpretation. To estimate the odds ratio as a measure

of association, a second set of estimating equations similar to (3.3) (Lipsitz, Laird, and Harrington, 1991) can be used in combination with (3.2) for the marginal regression parameters. A more efficient second set of estimating equations to estimate the odds ratio parameters, referred to as "alternating logistic regression," was later proposed by Carey, Zeger, and Diggle (1993), and is discussed in greater detail in the next section. We also note that Qu et al. (1992) proposed using tetrachoric correlations as measures of association between pairs of binary responses; the tetrachoric correlation is the correlation between underlying latent normal variables that are assumed to have been dichotomized to form the observed binary outcomes. Additional work extending the GEE approach to longitudinal ordinal and polytomous data has been developed by, for example, Miller, Davis, and Landis (1993), Lipsitz, Kim, and Zhao (1994), Kenward, Lesaffre, and Molenberghs (1994), Gange et al. (1995), Lumley (1996), and Heagerty and Zeger (1996). The main challenge with extending the GEE approach to ordinal and polytomous responses has been that the working covariance for ordinal and polytomous responses (other than "working independence"), in general, requires the specification and estimation of a large number of nuisance parameters.

Given any parameterization of the working covariance, V_i, in (3.2), and using Taylor series expansions similar to Prentice (1988), assuming that the regression model for the mean of $\boldsymbol{Y}_i$ has been correctly specified, $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$, and also $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ has an asymptotic distribution which is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix given by

$$C_\beta = \lim_{N\to\infty} N \left(\sum_{i=1}^{N} D_i' V_i^{-1} D_i\right)^{-1} \left\{\sum_{i=1}^{N} D_i' V_i^{-1} \mathrm{Cov}(\boldsymbol{Y}_i) V_i^{-1} D_i\right\} \left(\sum_{i=1}^{N} D_i' V_i^{-1} D_i\right)^{-1}, \tag{3.4}$$

and if $\mathrm{Cov}(\boldsymbol{Y}_i)$ is correctly specified, so that $V_i = \mathrm{Cov}(\boldsymbol{Y}_i)$, then (3.4) reduces to

$$\lim_{N\to\infty} N \left(\sum_{i=1}^{N} D_i' V_i^{-1} D_i\right)^{-1}. \tag{3.5}$$

Note that a key property of the GEE estimators of $\boldsymbol{\beta}$ is that they are consistent and asymptotically normal, for any choice of working covariance V_i, provided the regression model for the mean response has been correctly specified. Heuristically, using results from the method of moments, $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ regardless of whether V_i is the true covariance matrix of $\boldsymbol{Y}_i$ because

$$E\{u_\beta(\boldsymbol{\beta})\} = E\left[\sum_{i=1}^{N} D_i' V_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\right] = \sum_{i=1}^{N} D_i' V_i^{-1} E\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = \mathbf{0}, \tag{3.6}$$

and we are solving $u_\beta(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ for $\widehat{\boldsymbol{\beta}}$. In particular, for any positive definite and symmetric matrix V_i, the solution to (3.2) is a consistent estimator of $\boldsymbol{\beta}$. This is the key property of the GEE method and implies that V_i does not have to be correctly specified in order to obtain consistent estimators of $\boldsymbol{\beta}$.

Finally, C_β in (3.4) can be consistently estimated by $\widehat{C}_\beta$, which is obtained by replacing $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by their estimates, and also $\mathrm{Cov}(\boldsymbol{Y}_i)$ by $(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i)'$. The estimator $\widehat{C}_\beta$ is a sandwich variance estimator and has the same form as the sandwich variance estimator in our earlier discussion of quasi-likelihood estimation. Some authors refer to this sandwich variance estimator as the "robust" variance estimator because it is consistent for the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ provided $E\{u_\beta(\widehat{\boldsymbol{\beta}})\} = \mathbf{0}$; that is, the sandwich variance estimator can be said to be robust to misspecification of the covariance among the repeated measures.

Obtaining GEE estimates requires an iterative algorithm. The structure of the GEE suggests the use of a specific iterative scheme, namely to iterate between estimating $\boldsymbol{\beta}$ (given the current estimate of $\boldsymbol{\alpha}$) as the solution to (3.2), and estimating $\boldsymbol{\alpha}$ (given the current estimate of $\boldsymbol{\beta}$) as the solution to (3.3) until convergence. In particular, the solution $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ to (3.2) and (3.3) can be obtained by a Fisher scoring algorithm. Given a starting value for $\boldsymbol{\beta}$, say under the naive assumption of independence, the solution $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ can be obtained by iterating between

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \left[\sum_{i=1}^{N} D_i^{(m)\prime} \{V_i^{(m)}\}^{-1} D_i^{(m)}\right]^{-1} \sum_{i=1}^{N} D_i^{(m)\prime} \{V_i^{(m)}\}^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}^{(m)})\}, \tag{3.7}$$

and

$$\widehat{\boldsymbol{\alpha}}^{(m+1)} = \widehat{\boldsymbol{\alpha}}^{(m)} + \left[\sum_{i=1}^{N} E_i^{(m)\prime} \{W_i^{(m)}\}^{-1} E_i^{(m)}\right]^{-1} \times \sum_{i=1}^{N} E_i^{(m)\prime} \{W_i^{(m)}\}^{-1} [\boldsymbol{U}_i\{\widehat{\boldsymbol{\beta}}^{(m)}\} - \boldsymbol{\rho}_i\{\widehat{\boldsymbol{\alpha}}^{(m)}\}],$$

until $\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)}$ and $\widehat{\boldsymbol{\alpha}}^{(m+1)} = \widehat{\boldsymbol{\alpha}}^{(m)}$, where $D_i^{(m)} = D_i\{\widehat{\boldsymbol{\beta}}^{(m)}\}$, $V_i^{(m)} = V_i\{\widehat{\boldsymbol{\beta}}^{(m)}, \widehat{\boldsymbol{\alpha}}^{(m)}\}$, $E_i^{(m)} = E_i\{\widehat{\boldsymbol{\beta}}^{(m)}, \widehat{\boldsymbol{\alpha}}^{(m)}\}$, and $W_i^{(m)} = W_i\{\widehat{\boldsymbol{\beta}}^{(m)}, \widehat{\boldsymbol{\alpha}}^{(m)}\}$.

We note that, given the current estimate of $\boldsymbol{\beta}$, the estimator of $\boldsymbol{\alpha}$ proposed by Liang and Zeger (1986) is non-iterative. For example, suppose an "exchangeable" correlation pattern is assumed, in which $\rho_{ist} = \alpha$ for all $s < t$. Then, Liang and Zeger (1986) proposed estimating α, given the current estimate of $\boldsymbol{\beta}$, say $\widehat{\boldsymbol{\beta}}$, by

$$\widehat{\alpha} = \frac{1}{N^*} \sum_{i=1}^{N} \sum_{s<t}^{n_i} \frac{(Y_{is} - \widehat{\mu}_{is})(Y_{it} - \widehat{\mu}_{it})}{\widehat{\phi}\{v(\widehat{\mu}_{is})v(\widehat{\mu}_{it})\}^{1/2}}, \quad \text{where } N^* = \sum_{i=1}^{N} \frac{n_i(n_i-1)}{2};$$

alternatively, various degree-of-freedom corrections to account for the estimation of $\boldsymbol{\beta}$ have been suggested for the denominator N^*, e.g., $N^* - p$.

3.2.5 Properties of GEE estimators

In this section we consider properties of the GEE estimators. Before summarizing the key properties of the GEE approach, we note that there is an implicit assumption in the first component of a marginal model that is often overlooked. Recall that in a marginal model the conditional expectation of each response is assumed to depend on the covariates through a known link function

$$h^{-1}(\mu_{ij}) = \eta_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta}.$$

In marginal models, the primary interest lies in estimation of the regression parameters, $\boldsymbol{\beta}$, in the model for $E(Y_{ij}|\boldsymbol{X}_{ij})$. The key property for (asymptotically) unbiased estimators using GEE is given in (3.6), and can be rewritten as

$$E\big[D_i' V_i^{-1}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\big] = E_{x_i}\big[D_i' V_i^{-1} E_{y_i|x_i}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\big] = \boldsymbol{0}, \tag{3.8}$$

where $E_{x_i}(\cdot)$ denotes expectation with respect to the marginal distribution of X_i and $E_{y_i|x_i}(\cdot)$ denotes expectation with respect to the conditional distribution of $\boldsymbol{Y}_i$ given X_i. Note that the jth element of the vector $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ is $\mu_{ij} = E(Y_{ij}|\boldsymbol{X}_{ij})$, but the elements of $E_{y_i|x_i}(\boldsymbol{Y}_i)$ in (3.8) are $E(Y_{ij}|\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i})$. Thus, for (3.8) to hold, the conditional mean of the jth response, given $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}$, must depend only on $\boldsymbol{X}_{ij}$, that is,

$$E(Y_{ij}|X_i) = E(Y_{ij}|\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}) = E(Y_{ij}|\boldsymbol{X}_{ij}); \tag{3.9}$$

see Fitzmaurice, Laird, and Rotnitzky (1993) and Pepe and Anderson (1994) for a more detailed discussion of this sufficient condition. With time-stationary covariates, this assumption poses no difficulties; it necessarily holds because $\boldsymbol{X}_{ij} = \boldsymbol{X}_{ik}$ for all occasions $k \neq j$. Also, with time-varying covariates that are fixed by design of the study (e.g., time since baseline, treatment group indicator in a crossover trial), the assumption also holds because values of the covariates at any occasion are determined *a priori* by study design and in a manner completely unrelated to the longitudinal response. However, when a time-varying covariate varies randomly over time the assumption made in (3.9) may not hold. For example, the assumption will be violated when the current value of the response, say Y_{ij}, given the current covariates $\boldsymbol{X}_{ij}$, predicts the subsequent value of $\boldsymbol{X}_{i,j+1}$; see Chapter 23 for a more detailed discussion of stochastic time-varying covariates in longitudinal studies. When (3.9) is not satisfied, yet there is subject-matter interest in the dependence of Y_{ij} on $\boldsymbol{X}_{ij}$, Pepe and Anderson (1994) recommend using GEE with a "working independence" assumption. Under a "working independence" assumption, the weight matrix is diagonal and the corresponding estimating equations simplify and are unbiased regardless of whether or not (3.9) is satisfied. In contrast, under alternative choices for the working covariance matrix, the estimating equations are not necessarily unbiased and may yield inconsistent estimators of the regression parameters.

Thus far, we have seen that the GEE approach provides a convenient alternative to ML estimation of the regression parameters in marginal models for longitudinal data, while also retaining a number of appealing properties. First, in many longitudinal designs the GEE estimator $\widehat{\boldsymbol{\beta}}$ is almost as efficient as the ML estimator. For example, consider the case of linear models for continuous responses that are assumed to have a multivariate normal distribution. It can easily be shown that the generalized least-squares estimator of $\boldsymbol{\beta}$ in linear models can be considered a special case of the GEE approach. The GEE also has an expression similar to the likelihood equations for $\boldsymbol{\beta}$ in certain "mixed parameter" models for discrete longitudinal data (e.g., Fitzmaurice and Laird, 1993; Fitzmaurice, Laird, and Rotnitzky, 1993). As a result, for many longitudinal designs, there is little loss of efficiency when the GEE approach is adopted as an alternative to maximum likelihood. Moreover, because all GEE estimators of $\boldsymbol{\beta}$ are consistent and asymptotically normal, it is of interest to consider their efficiency under various working covariance assumptions. Indeed, it has been suggested in the literature that setting $R_i = I$, the identity matrix, leads to an estimator with nearly the same efficiency as the optimally weighted GEE (with $V_i = \Sigma_i$). Interestingly, for a balanced longitudinal design, with no missing data, there are cases where this claim has some justification. Specifically, there is relatively little loss of efficiency either for between-subject effects (where the covariate design $X_{ijk} = X_{ij'k}$ for all occasions $j \neq j'$) or for within-subject effects when the covariate design on time for the latter effects is the same for all subjects (i.e., $X_{ijk} \neq X_{ij'k}$ for some occasions $j \neq j'$, but $X_{ijk} = X_{i'jk}$ for all subjects $i \neq i'$). However, for the case of a within-subject effect when the covariate design on time is not the same for all subjects (i.e., $X_{ijk} \neq X_{ij'k}$ for some occasions $j \neq j'$ and $X_{ijk} \neq X_{i'jk}$ for some subjects $i \neq i'$), there can be a very discernible loss of efficiency under the non-optimal "working independence" assumption for the covariance, especially when the true correlations are moderately large (see, for example, Lipsitz et al., 1994). Heuristically, this result can be explained as follows. When using GEE methods, the estimated $\boldsymbol{\beta}$s in a marginal model can be thought of as weighted averages of between- and within-subject contrasts. In the case of a between-subject effect (e.g., exposure or treatment group) or a within-subject effect when the covariate design on time is the same for all subjects (e.g., time since baseline), the respective between- or within-subject contrasts are weighted approximately equally, regardless of the assumed (or working) covariance. On the other hand, for a within-subject effect when the covariate design on time is not the same for all subjects, the GEE estimate is a weighted average of *both* between-subject

and within-subject contrasts that are weighted differently. Moreover, the weights for these different contrasts depend on the assumed covariance. So, for the latter case, the correlation is more important and determines the optimal weights for combining the contrasts. In addition, in the setting of unbalanced data, with $n_i \neq n$, each individual no longer contributes equally weighted components to even the between-subject effects, and thus we can expect a loss of efficiency for the GEE under "working independence" in that setting as well. So, in general, it can be advantageous to choose a working covariance that closely approximates the true covariance. The closer the working covariance matrix (V_i) approximates the true underlying covariance matrix (Σ_i), the greater the efficiency for estimation of $\boldsymbol{\beta}$.

A second appealing property of the GEE estimator $\widehat{\boldsymbol{\beta}}$ is its robustness, yielding a consistent estimator of $\boldsymbol{\beta}$ even if the within-subject associations among the repeated measures have been misspecified. It only requires that the model for the mean response be correct. This robustness property of GEE is important because the usual focus of a longitudinal study is on changes in the mean response. In general, there is usually far less interest in, and correspondingly less subject-matter knowledge of, the patterns of covariance among the repeated measures. Although the GEE approach yields a consistent estimator of $\boldsymbol{\beta}$ under misspecification of the within-subject associations, the usual standard errors obtained under the misspecified model for the within-subject association are not valid. Fortunately, in many cases, valid standard errors for $\widehat{\boldsymbol{\beta}}$ can be obtained using the so-called sandwich variance estimator. A remarkable property of the sandwich estimator is that it is also robust in the sense that it provides valid standard errors when the assumed model for the covariances among the repeated measures is not correct. That is, with large sample sizes, the sandwich variance estimator yields correct standard errors. However, it is worth emphasizing that this robustness property of the sandwich variance estimator is an asymptotic property. In general, use of the sandwich variance estimator is best suited to balanced longitudinal designs where the number of subjects (N) is relatively large and the number of repeated measures (n) is relatively small. Moreover, the sandwich estimator is less appealing when the design is severely unbalanced and/or when there are few replications to estimate the true underlying covariance matrix. The use of the sandwich estimator implicitly relies on there being many replications of the vector of responses associated with each distinct set of covariate values.

Note that bias-corrected versions of the sandwich variance estimator have been proposed that have somewhat better finite-sample properties, including the jackknife (Paik, 1988; Mancl and DeRouen, 2001). Most of these bias-corrected versions involve a "degrees-of-freedom" correction. In cases where there are few, if any, replications of $\boldsymbol{Y}_i$ associated with each distinct set of covariate values, the use of the sandwich estimator can be problematic. In particular, standard errors based on the sandwich estimator tend to be biased downward. In addition, the sampling variability of the sandwich estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ can be very large, resulting in an unstable estimator of variability. In these settings, it may be preferable to carefully model the covariances among the responses and use the "model-based" estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ given by (3.5) and evaluated at $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$. This estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ is referred to as a "model-based" estimator to remind us that it yields valid standard errors provided that the working covariance matrix, V_i, is a close approximation to the true underlying covariance matrix, Σ_i. In general, unlike the sandwich variance estimator, the "model-based" estimator does not require such a large number of replications; the sampling variability of the "model-based" estimator tends to be smaller than that of the sandwich variance estimator. The key to obtaining approximately unbiased variance estimators using the "model-based" estimator is to choose a model for the working covariance matrix that is close to the true covariance matrix.

Finally, for making inferences about $\boldsymbol{\beta}$, since the GEE approach is not likelihood-based, likelihood ratio tests are not available for hypotheses testing; in addition, this also has ramifications for inferences when there are missing data, a topic that will be discussed in Section 3.4. Instead, inferences typically rely on Wald test statistics based on quadratic forms. For example, if it is of interest to test whether or not the ℓth element of $\boldsymbol{\beta}$ is 0, $\mathrm{H}_0 : \beta_\ell = 0$, then the Wald test statistic,

$$Z_\ell = \frac{\widehat{\beta}_\ell}{\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_\ell)}} \sim N(0,1)$$

can be constructed, where $\widehat{\mathrm{Var}}(\widehat{\beta}_\ell)$ is the ℓth diagonal element of either the model-based or, more typically, the sandwich variance estimate. More generally, for a null hypothesis of the form $H_0 : L\boldsymbol{\beta} = 0$, versus the alternative $H_A : L\boldsymbol{\beta} \neq 0$, for an $r \times p$ matrix L of full rank, $r \leq p$, the Wald test statistic is

$$X^2 = (L\widehat{\boldsymbol{\beta}})'\{L\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}})L'\}^{-1}L\widehat{\boldsymbol{\beta}} \sim \chi^2_r,$$

where χ^2_r denotes a chi-square distribution with r degrees of freedom.

Although this use of the Wald statistic allows for the comparison of nested models, it can be potentially problematic because Wald statistics are known to have less than optimal properties under the alternative in logistic regression models for univariate outcome data (Hauck and Donner, 1977). We conjecture that the same may be true for Wald statistics based on GEE estimates of marginal regression parameters for longitudinal binary data. As an alternative, and to circumvent this potential problem with the Wald statistic, one can base inferences on the score test statistic proposed by Rotnitzky and Jewell (1990) and Boos (1992). Recall the property of the GEE score vector $u_\beta(\boldsymbol{\beta})$ given in (3.6), namely $E\{u_\beta(\boldsymbol{\beta})\} = 0$. Furthermore, it can be shown that

$$V_u(\boldsymbol{\beta}) = \mathrm{Cov}\{u_\beta(\boldsymbol{\beta})\} = \{\mathrm{Cov}(\widehat{\boldsymbol{\beta}})\}^{-1}.$$

Finally, since the score vector is a sum of independent random variables, using the central limit theorem, it can be shown that

$$u_\beta(\boldsymbol{\beta}) \sim N\{0, V_u(\boldsymbol{\beta})\}.$$

Then, the "score" test statistic can be expressed as a quadratic form in the score vector,

$$u_\beta(\tilde{\boldsymbol{\beta}})'\{V_u(\tilde{\boldsymbol{\beta}})\}^{-1}u_\beta(\tilde{\boldsymbol{\beta}}) \sim \chi^2_r, \tag{3.10}$$

where $u_\beta(\tilde{\boldsymbol{\beta}})$ and $V_u(\tilde{\boldsymbol{\beta}})$ are the score vector and its variance under the alternative hypothesis, evaluated at $\tilde{\boldsymbol{\beta}}$, which is the estimate of $\boldsymbol{\beta}$ under the null hypothesis $H_0 : L\boldsymbol{\beta} = 0$.

3.3 Some extensions of GEE methods for longitudinal data

In this section we briefly review some extensions of the standard GEE methods proposed by Liang and Zeger (1986). In particular, we describe alternative estimators of the within-subject association, especially when the longitudinal responses are discrete. We also discuss joint estimation of the marginal mean and within-subject association parameters.

3.3.1 Alternative estimators of within-subject association parameters

As discussed in the previous section, when V_i is correctly specified in the standard GEE approach, the resulting estimator of $\boldsymbol{\beta}$ is semi-parametric efficient in the sense of having the smallest variance among all estimators in the class given by (3.2). However, the estimators of

the correlation parameters $\boldsymbol{\alpha}$ proposed by Liang and Zeger (1986) are not efficient, in large part due to the use of a suboptimal weight matrix, W_i, in (3.3). This has led researchers to develop alternative estimating equations for $\boldsymbol{\alpha}$. Using the same set of estimating equations for $\boldsymbol{\beta}$ given in (3.2), we now consider two alternative estimating equations for $\boldsymbol{\alpha}$ (and thus for $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$). We note that any set of estimating equations that use (3.2) to estimate $\boldsymbol{\beta}$ are often referred to as first-order GEE or "GEE1". Here, we consider two alternative GEE1s developed with the goal of providing more stable and/or efficient estimates of $\boldsymbol{\alpha}$ than the standard GEE of Liang and Zeger (1986).

One set of estimating equations for $\boldsymbol{\alpha}$ that has generated some interest in the statistical literature is that based on the notion of "Gaussian" estimation (see, for example, Lipsitz, Laird, and Harrington, 1992; Lee, Laird, and Johnston, 1999; Hall and Severini, 1998; Lipsitz et al., 2000; Fitzmaurice, Lipsitz, and Molenberghs, 2001; Wang and Carey, 2003). The key idea here is to base estimation of $\boldsymbol{\alpha}$ on the multivariate normal estimating equations for the correlations. In particular, let

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_i(\boldsymbol{\beta}) = \frac{1}{\sqrt{\phi}} A_i^{-1/2}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i),$$

be the vector of standardized residuals, where A_i is a diagonal matrix with diagonal elements $v(\mu_{ij})$. Note that $\mathrm{Cov}(\boldsymbol{\xi}_i)$ is the correlation matrix of $\boldsymbol{Y}_i$, which we denoted earlier as $R_i = R_i(\boldsymbol{\alpha})$. Then, a second set of (moment) estimating equations can be obtained as the score equations for $\boldsymbol{\alpha}$ under the assumption that $\boldsymbol{\xi}_i \sim N(\mathbf{0}, R_i(\boldsymbol{\alpha}))$. Specifically, the estimating equations for $\boldsymbol{\alpha}$ are given by

$$u_\alpha(\boldsymbol{\alpha}) = \left[\frac{\partial}{\partial \boldsymbol{\alpha}} \left\{ \sum_{i=1}^{N} \log|R_i(\boldsymbol{\alpha})| - \sum_{i=1}^{N} \boldsymbol{\xi}_i'(\boldsymbol{\beta}) R_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{\xi}_i(\boldsymbol{\beta}) \right\}\right] = \mathbf{0}. \tag{3.11}$$

The rth component of $\boldsymbol{u}_\alpha(\boldsymbol{\alpha})$ in (3.11) equals

$$\sum_{i=1}^{N} \mathrm{tr}\big[R_i^{-1}(\boldsymbol{\alpha})\{\boldsymbol{\xi}_i(\boldsymbol{\beta})\boldsymbol{\xi}_i'(\boldsymbol{\beta}) - R_i(\boldsymbol{\alpha})\} R_i^{-1}(\boldsymbol{\alpha}) \dot{\mathrm{R}}_{ir}(\boldsymbol{\alpha})\big],$$

where $\dot{\mathrm{R}}_{ir}(\boldsymbol{\alpha}) = \partial R_i(\boldsymbol{\alpha})/\partial \alpha_r$ and $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix. An appealing feature of the use of the multivariate normal estimating equations is that it ensures that the estimated correlation matrix, $R_i(\widehat{\boldsymbol{\alpha}})$, is non-negative definite when there are unbalanced data (i.e., $n_i \neq n$). In contrast, this is not the case for alternative GEE1 methods. Thus, in general, this leads to more stable estimation of $\boldsymbol{\alpha}$. As a slight modification to these multivariate normal estimating equations, Wang and Carey (2004) proposed estimating the correlation parameters by differentiating the Cholesky decomposition of the working correlation matrix; a similar approach was also used by Ye and Pan (2006). Furthermore, similar Gaussian or "quadratic" estimating equations have been proposed by Crowder (1995) and Qu, Lindsay, and Li (2000).

Motivated by the application of GEE methods to longitudinal binary outcomes, a second alternative set of estimating equations for $\boldsymbol{\alpha}$ were proposed by Carey, Zeger, and Diggle (1993) and Lipsitz and Fitzmaurice (1996). Specifically, they proposed a set of estimating equations based on the *conditional* residuals $\{Y_{it} - E(Y_{it}|Y_{is} = y_{is}, X_i)\}$, that is, deviations about *conditional* expectations. In contrast, note that the second set of estimating equations given by (3.3) are based on the *unconditional* residuals

$$\frac{(Y_{is} - \mu_{is})(Y_{it} - \mu_{it})}{\phi\{v(\mu_{is})v(\mu_{it})\}^{1/2}} - \rho_{ist}.$$

The *conditional* expectations can be grouped together to form the $n_i(n_i - 1)/2 \times 1$ vector of conditional residuals, $(\boldsymbol{\mathcal{U}}_i - \boldsymbol{\eta}_i)$, where

$$\boldsymbol{\mathcal{U}}_i = \{\mathcal{U}_{i12}, \mathcal{U}_{i13}, \ldots, \mathcal{U}_{i,n-1,n}\}', \qquad \boldsymbol{\eta}_i = \{\eta_{i12}, \eta_{i13}, \ldots, \eta_{i,n-1,n}\}',$$

with $\mathcal{U}_{ist} = Y_{it}$ and $\eta_{ist} = E(Y_{it}|Y_{is} = y_{is}, X_i)$, for $s < t$. Note that for the special case where the Y_{it} are binary, the conditional mean equals

$$\begin{aligned}
\eta_{ist}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= E(Y_{it}|Y_{is} = y_{is}, \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= y_{is}E(Y_{it}|Y_{is} = 1, \boldsymbol{\beta}, \boldsymbol{\alpha}) + (1 - y_{is})E(Y_{it}|Y_{is} = 0, \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= y_{is}\Pr(Y_{it} = 1|Y_{is} = 1, \boldsymbol{\beta}, \boldsymbol{\alpha}) + (1 - y_{is})\Pr(Y_{it} = 1|Y_{is} = 0, \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= y_{is}\left\{\Pr(Y_{is} = 1, Y_{it} = 1; \boldsymbol{\beta}, \boldsymbol{\alpha})/\Pr(Y_{is} = 1; \boldsymbol{\beta})\right\} \\
&\quad + (1 - y_{is})\left\{\Pr(Y_{is} = 0, Y_{it} = 1; \boldsymbol{\beta}, \boldsymbol{\alpha})/\Pr(Y_{is} = 0; \boldsymbol{\beta})\right\} \\
&= y_{is}\left(\pi_{ist}/\mu_{is}\right) + (1 - y_{is})\left\{(\mu_{it} - \pi_{ist})/(1 - \mu_{is})\right\},
\end{aligned} \tag{3.12}$$

where $\pi_{ist} = \Pr(Y_{is} = 1, Y_{it} = 1; \boldsymbol{\beta}, \boldsymbol{\alpha})$. Then, a set of estimating equations for $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{u}_\alpha(\boldsymbol{\alpha}) = \sum_{i=1}^{N} E_i' W_i^{-1}\{\boldsymbol{\mathcal{U}}_i - \boldsymbol{\eta}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})\} = \mathbf{0}, \tag{3.13}$$

where $E_i = \partial\boldsymbol{\eta}_i/\partial\boldsymbol{\alpha}$ and $W_i = \text{diag}\{\text{Var}(Y_{it}|Y_{is} = y_{is})\}$. Note that for binary Y_{it}, $\text{Var}(Y_{it}|Y_{is} = y_{is}) = \eta_{ist}(1 - \eta_{ist})$ since $E(Y_{it}^2|Y_{is} = y_{is}) = E(Y_{it}|Y_{is} = y_{is}) = \eta_{ist}$.

Although the *conditional* residuals $(\mathcal{U}_{ist} - \eta_{ist})$ are neither independent nor uncorrelated, Carey, Zeger, and Diggle (1993) argue that the correlations among the *conditional* residuals should be substantially smaller than the correlations among the *unconditional* residuals. Consequently, setting $W_i = \text{diag}\{\eta_{ist}(1 - \eta_{ist})\}$ in (3.13) should provide a closer approximation to the optimal weight matrix. Kuk (2004) proposed a symmetrized version of these estimating equations that may be particularly useful for an exchangeable correlation structure.

We note that after some algebra, (3.12) can be shown to equal

$$E(Y_{it}|Y_{is} = y_{is}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \mu_{it} + (y_{is} - \mu_{is})\rho_{ist}\sqrt{\frac{\text{Var}(Y_{it})}{\text{Var}(Y_{is})}}, \tag{3.14}$$

which is exactly the conditional expectation if (Y_{is}, Y_{it}) are assumed to be bivariate normal. Thus, even though these estimating equations were originally proposed for analysis of longitudinal binary data, the form of the conditional expectation in (3.14) suggests that they can be used also for non-binary longitudinal outcomes. The only additional issue that arises for non-binary outcomes is the specification of $\text{Var}(Y_{it}|Y_{is} = y_{is})$. For binary outcome data, it is straightforward to specify this variance as $\text{Var}(Y_{it}|Y_{is} = y_{is}) = \eta_{ist}(1 - \eta_{ist})$. For non-binary outcomes, one alternative is to use the conditional variance from the bivariate normal, with

$$\text{Var}(Y_{it}|Y_{is} = y_{is}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \text{Var}(Y_{it})\left(1 - \rho_{ist}^2\right).$$

As with the standard GEE1 discussed in Section 3.2.3, the two alternative approaches discussed in this section require an iterative algorithm. That is, both of these GEE1 methods require iterating between estimating $\boldsymbol{\beta}$ (given the current estimate of $\boldsymbol{\alpha}$) as the solution to (3.2), and estimating $\boldsymbol{\alpha}$ (given the current estimate of $\boldsymbol{\beta}$) as the solution to (3.11) or (3.13), until convergence.

3.3.2 Second-order generalized estimating equations (GEE2)

In the previous section, a number of alternative proposals for estimating $\boldsymbol{\alpha}$, and thus $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$, were described. All of these approaches broadly fall within the framework of GEE1.

That is, all of the approaches discussed so far use estimating equations for $\boldsymbol{\beta}$ given by (3.2), and differ only in terms of how $\boldsymbol{\alpha}$, and hence $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$, is estimated. In this section we discuss a second-order extension of generalized estimating equations, hereafter referred to as GEE2 (Zhao and Prentice, 1990; Liang, Zeger, and Qaqish, 1992). To do so, we need to introduce some additional notation. Let

$$\boldsymbol{\mathcal{Y}}_i = (Y_{i1}, \ldots, Y_{in_i}, Y_{i1}Y_{i2}, \ldots, Y_{i,n_i-1}Y_{in_i})' = (\boldsymbol{Y}_i', \boldsymbol{Z}_i')',$$

$\boldsymbol{\Pi}_i = E(\boldsymbol{\mathcal{Y}}_i | X_i, \boldsymbol{\beta}, \boldsymbol{\alpha})$ and, in an ever so slight departure from previous notation, $\mathcal{V}_i \approx \text{Cov}(\boldsymbol{\mathcal{Y}}_i | X_i)$. Then, the GEE2 estimating equations for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ are given by

$$u_2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \mathcal{D}_i' \mathcal{V}_i^{-1} \{\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\Pi}_i(\boldsymbol{\theta})\} = \mathbf{0}, \tag{3.15}$$

where $\mathcal{D}_i = \partial \boldsymbol{\Pi}_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. Note that, unlike GEE1, $\mathcal{D}_i$ (and $\mathcal{V}_i$) in GEE2 is not block-diagonal and thus the estimating equations for $\boldsymbol{\beta}$ depend on $\boldsymbol{\mathcal{Y}}_i = (\boldsymbol{Y}_i', \boldsymbol{Z}_i')'$, and not simply on $\boldsymbol{Y}_i$. As a consequence, the GEE2 estimating equations for $\boldsymbol{\beta}$ given by (3.15) are quite different from the GEE1 estimating equations for $\boldsymbol{\beta}$ given by (3.2). In particular, for the GEE2 estimator of $\boldsymbol{\beta}$ to be consistent, the model for all elements of $\boldsymbol{\mathcal{Y}}_i$ must be correctly specified; that is, the first two moments of $\boldsymbol{Y}_i$ must be correctly specified. GEE2 can yield substantial bias in estimators of both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ (Fitzmaurice, Lipsitz, and Molenberghs, 2001) if assumptions about second moments are misspecified. This is in contrast to GEE1 where a consistent estimator of $\boldsymbol{\beta}$ is obtained provided only that the mean of $\boldsymbol{Y}_i$ has been correctly specified.

The main appeal of GEE2 is that it is almost fully efficient for both the marginal regression parameters, $\boldsymbol{\beta}$, and the within-subject associations, $\boldsymbol{\alpha}$. Note that the working covariance matrix $\mathcal{V}_i$ is usually obtained by making additional assumptions about the third and fourth moments of $\boldsymbol{Y}_i$. Some alternative specifications for the third and fourth moments are discussed in Zhao and Prentice (1990) and Liang, Zeger, and Qaqish (1992). Thus, a notable feature of GEE2 is that it makes additional assumptions about higher-order moments. Furthermore, specification of these higher-order moments can greatly increase the computational complexity and make implementation of the method somewhat more difficult; it is notable that none of the major statistical software packages currently have options for GEE2.

Finally, we note that some authors refer to the solution to (3.2) and (3.13) as GEE2; in contrast, we prefer to reserve the term GEE2 to refer to estimators that require the first two moments of $\boldsymbol{Y}_i$ to be correctly specified in order to yield consistent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. That is, in contrast to GEE1 estimators, GEE2 estimators of $\boldsymbol{\beta}$ are not robust to misspecification of the second moments.

3.4 GEE with missing data

Although most longitudinal studies are designed to collect complete data on all participants, missing data very commonly arise and must be properly accounted for in the analysis; otherwise, biased estimators of longitudinal change can result. When longitudinal data are missing, the data set is necessarily unbalanced over time because not all individuals have the same number of repeated measurements at a common set of occasions. This feature of missingness creates no difficulties for the standard GEE method as it can handle the unbalanced data without having to discard data on individuals with any missing data. That is, when some individuals' response vectors are only partially observed, the standard GEE approach circumvents the problem of missing data by simply basing inferences on the *observed* responses. However, the validity of this method of analysis will require that certain assumptions about the reasons for any missingness, often referred to as the *missing-data mechanism*, are tenable.

The missing-data mechanism can be thought of as a model that describes the probability that a response is observed or missing at any occasion. Here, we briefly review and distinguish two general types of missing-data mechanisms; see Chapter 17 for a more detailed description. The two missing-data mechanisms are referred to as *missing completely at random* (MCAR) and *missing at random* (MAR) (Rubin, 1976; Laird, 1988). These two mechanisms differ in terms of assumptions concerning whether or not missingness is related to responses that have been observed. The distinction between these two mechanisms determines the appropriateness of standard GEE methods. Specifically, the standard GEE method yields consistent marginal regression parameter estimators provided the responses are MCAR (see, for example, Laird, 1988).

Data are said to be MCAR when the probability that responses are missing is unrelated to either the specific values that, in principle, should have been obtained (the *missing* responses) or the set of observed responses. Thus, longitudinal data are MCAR when missingness in $\boldsymbol{Y}_i$ is simply the result of a chance mechanism that does not depend on either observed or unobserved components of $\boldsymbol{Y}_i$. To better understand this missing-data process, consider the simple case where there are only two measurement occasions with the outcome fully observed on all subjects at the first occasion, and missing on some subjects at the second occasion. Because the outcome can be missing at the second occasion only, we can define the single indicator random variable R_{i2}, which equals 1 if Y_{i2} is observed and 0 if Y_{i2} is unobserved. The missing data are said to be MCAR if

$$\Pr(R_{i2} = 1|Y_{i1}, Y_{i2}, X_i) = \Pr(R_{i2} = 1|X_i);$$

that is, the probability that Y_{i2} is missing does not depend on the observed value of Y_{i1} or the possibly missing value of Y_{i2}. We note that the use of the term MCAR is sometimes restricted to the case where

$$\Pr(R_{i2} = 1|Y_{i1}, Y_{i2}, X_i) = \Pr(R_{i2} = 1);$$

this distinction only becomes important when an analysis is based on a subset of the covariates in X_i that excludes a covariate that is predictive of R_{i2}. The essential feature of MCAR is that the observed data can by thought of as a random sample of the complete data. As a result, all of the moments of the observed data do not differ from the corresponding moments of the complete data. This property provides the validity for the standard GEE method that bases inferences on the observed responses.

In contrast to MCAR, data are said to be MAR when the probability that responses are missing depends on the set of observed responses, but is unrelated to the specific missing values that, in principle, should have been obtained. For the simple bivariate response example considered previously, the missing data are said to be MAR if

$$\Pr(R_{i2} = 1|Y_{i1}, Y_{i2}, X_i) = \Pr(R_{i2} = 1|Y_{i1}, X_i); \tag{3.16}$$

that is, the probability that Y_{i2} is missing depends on Y_{i1} (and X_i), but is conditionally independent of the possibly missing value of Y_{i2}. Because the missing-data mechanism now depends upon observed responses, the sample moments based on the available data are biased estimates of the corresponding moments in the target population. Consequently, when data are MAR, the standard GEE method that bases inferences on the observed responses can yield biased regression parameter estimates. Finally, if $\Pr(R_{i2} = 1|Y_{i1}, Y_{i2}, X_i)$ depends in any way on the possibly missing outcome Y_{i2}, the missing data are said to be *not missing at random* (NMAR) or oftentimes referred to as being "non-ignorably" missing. The case of NMAR is beyond the scope of this chapter; see Chapters 20 and 22 for a more in-depth discussion of this topic.

When missing data are assumed to be MAR, there are two general approaches for handling this problem within the GEE framework: multiple imputation (e.g., Paik, 1997) and

weighted estimating equations (Robins, Rotnitzky, and Zhao, 1995). The idea behind multiple imputation is very simple: substitute or fill in the values that were not recorded with imputed values. However, to reflect the uncertainty inherent in the imputation of the unobserved responses, the imputation process is repeated multiple times. The imputations are typically based on some assumed model for the missing data given the observed data. The attractive feature of imputation methods is that, once a filled-in data set has been constructed, the standard GEE method for complete data can be applied. The multiple filled-in data sets produce different sets of parameter estimates and their standard errors that are then appropriately combined to provide a single estimate of the parameters of interest, together with standard errors that reflect the uncertainty inherent in the imputation of the unobserved responses. There is an extensive literature on multiple imputation, and the reader can find an excellent review of this topic in Chapter 21. We do not discuss multiple imputation for GEE any further because, to date, there is no overwhelming concensus on the best way to impute discrete longitudinal data; see Chapter 21 for more details.

The second general approach for handling data that are MAR is via weighted estimating equations. In one of the simplest versions of the weighted estimating equations approach, an individual's contribution to the standard GEE is weighted inversely by the probability of being observed at the given times. The key idea behind weighting methods is that the underrepresentation of certain response profiles in the observed data is taken into account and corrected. The weighted estimating equations approach is best suited to the case of monotone missing-data patterns as might arise when missingness is due to attrition or dropout. A variety of different weighting methods that adjust for dropout have been proposed. These approaches are often called propensity weighted or inverse probability weighted methods; see Chapter 20 for an excellent discussion of these methods. In all of these methods the underlying idea is to base estimation on the observed responses but weight them to account for the probability of remaining in the study.

Inverse probability weighted methods were first proposed in the sample survey literature, where the weights are known and based on the survey design (e.g., Horvitz and Thompson, 1952). In the weighted estimating equations approach, however, the weights are not known but must be estimated based on an assumed model for dropout. The propensities for dropout can be estimated as a function of the observed responses prior to dropout, and also as a function of the covariates and any extraneous variables that are thought likely to predict dropout.

Consider the simple example introduced earlier where there are only two measurement occasions and any missingness is restricted to the second occasion. In this simple case, the missing-data pattern is monotone. Furthermore, let us assume the data are MAR as in (3.16) and let

$$\pi_i = \pi_i(\gamma) = \Pr(R_{i2} = 1 | Y_{i1}, X_i, \gamma),$$

a function of possibly unknown parameters γ. The basic idea underlying the weighted GEE method is to weight each individual's contribution to the standard GEE by the inverse probability of being observed, thereby accounting for those subjects with the same history of responses and covariates (y_{i1}, X_i), but who were missing Y_{i2}. Specifically, the GEE in (3.2) is weighted by the inverse probabilities to yield a simple "weighted GEE,"

$$\sum_{i=1}^{N} \left(\frac{R_{i2}}{\pi_i} + \frac{1 - R_{i2}}{1 - \pi_i} \right) \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} \{ (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} = \mathbf{0}. \tag{3.17}$$

Note that when $R_{i2} = 1$, then $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ and, similarly, $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})$; but when $R_{i2} = 0$, then $\boldsymbol{Y}_i = Y_{i1}$ and, similarly, $\boldsymbol{\mu}_i = \mu_{i1}$.

The intuition behind this approach is that it reweights the observed data to mimic what would likely be seen in a data set without missing data, thereby producing unbiased

estimating equations and consistent estimators for $\boldsymbol{\beta}$. In particular, under MAR as in (3.16),

$$\pi_i = \pi_i(\gamma) = \Pr(R_{i2} = 1 | Y_{i1}, Y_{i2}, X_i, \gamma) = \Pr(R_{i2} = 1 | Y_{i1}, X_i, \gamma),$$

or, equivalently,

$$E\left(\frac{R_{i2}}{\pi_i} \,\middle|\, Y_{i1}, Y_{i2}, X_i \right) = E\left(\frac{R_{i2}}{\pi_i} \,\middle|\, Y_{i1}, X_i \right) = 1.$$

Therefore, the estimating equations given by (3.17) are unbiased for $\mathbf{0}$ at the true $\boldsymbol{\beta}$,

$$\begin{aligned}
& E\left[\left(\frac{R_{i2}}{\pi_i} + \frac{1 - R_{i2}}{1 - \pi_i} \right) \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} \right] \\
&= E\left[E\left\{ \left(\frac{R_{i2}}{\pi_i} + \frac{1 - R_{i2}}{1 - \pi_i} \right) \,\middle|\, Y_{i1}, Y_{i2}, X_i \right\} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) \right] \\
&= E\left[\left(\frac{\pi_i}{\pi_i} + \frac{1 - \pi_i}{1 - \pi_i} \right) \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} \right] \\
&= 2E\left[\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} \right] \\
&= \mathbf{0}.
\end{aligned}$$

As the estimating equations are unbiased for $\mathbf{0}$, using results from method of moments, the solution $\widehat{\boldsymbol{\beta}}$ to (3.17) defines a consistent estimator for $\boldsymbol{\beta}$.

Unlike in the survey sampling setting where π_i is known, here π_i is unknown and will need to be replaced in (3.17) with an estimate; this estimate can be obtained using, for example, a logistic regression model with "outcome" R_{i2} and "predictors" (y_{i1}, X_i). In general, the validity of most weighting methods requires that the model for the missingness probabilities, π_i, has been correctly specified. We note that the weighted GEE given above is a very simple special case of a general class of weighted estimators. In particular, Robins, Rotnitzky, and Zhao (1995) discuss more general weighted estimating equations for longitudinal data and the construction of "semi-parametric efficient" weighted estimating equations; see Chapter 20 for an excellent summary of these developments. Finally, Robins (2000) and Robins and Rotnitzky (2001) have also recently developed so-called "doubly robust" weighted estimators that relax the assumption that the model for the missingness probabilities, π_i, has been correctly specified, albeit requiring additional assumptions on the model for $\boldsymbol{Y}_i$ given X_i. Doubly robust methods require the specification of two models, one for the missingness probabilities and another for the distribution of the complete data. For doubly robust estimation, the weighted GEE is augmented by a function of the response. When this augmentation term is selected and modeled correctly according to the distribution of the complete data, the estimator of $\boldsymbol{\beta}$ is consistent even if the model for missingness is misspecified. On the other hand, if the model for missingness is correctly specified, the augmentation term does not need to be correctly specified to yield consistent estimators of $\boldsymbol{\beta}$ (Scharfstein, Rotnitzky, and Robins, 1999, see Section 3.2 of Rejoinder; Robins and Rotnitzky, 2001; van der Laan and Robins, 2003; Bang and Robins, 2005; see also Lipsitz, Ibrahim, and Zhao, 1999; Lunceford and Davidian, 2004). Thus, the appealing property of doubly robust methods is that they yield estimators that are consistent when either, but not necessarily both, the model for the missingness mechanism or the model for the distribution of the complete data has been correctly specified. These estimators are doubly robust in the sense of providing double protection against model misspecification. See Chapter 23 for a more detailed discussion of doubly robust estimators.

As noted previously, the weighted estimating equations approach is best suited to the case of monotone missing-data patterns. However, in many longitudinal studies an individual's response can be missing at one follow-up time, and be measured at the next follow-up time, resulting in a large class of distinct missingness patterns. This is often referred to as "intermittent" missingness or non-monotone missingness. In that setting, approximate methods have been proposed for handling missing data that are assumed to be MAR but not MCAR. For example, via simulations and asymptotic studies, Lipsitz et al. (2000) and Fitzmaurice, Lipsitz, and Molenberghs (2001) show that the GEE1 using the second set of "Gaussian" estimating equations described in Section 3.3.1 yields estimators of $\boldsymbol{\beta}$ with relatively small bias in many settings. However, this method does require that the within-subject association, usually considered a nuisance characteristic of the data, is correctly specified. The method does not, however, require estimation of the missingness probabilities, π_i.

3.5 Goodness-of-fit and model diagnostics

As is the case for any generalized linear model, model checking is an important aspect of the fitting of marginal models to longitudinal data. However, because GEE methods are not likelihood-based, many of the standard goodness-of-fit statistics and model diagnostics are not immediately available. In this section we very briefly review some recent literature on this topic; further research on model diagnostics is needed.

Recall that in linear regression with independent univariate outcome data, Cook's distance and so-called "DFBETAs" are widely used measures of influence (see, for example, Belsley, Kuh, and Welsch, 1980; Cook and Weisberg, 1982). Specifically, DFBETAs assess how each coefficient is changed by including the observation from the ith subject, thereby measuring the influence or effect that the ith subject has on the estimate of the regression parameters $\boldsymbol{\beta}$. It is calculated by deleting the ith observation, and recomputing the ordinary least-squares estimate of $\boldsymbol{\beta}$. Because ordinary least squares is non-iterative, fast computer algorithms have been developed to recompute the estimate of $\boldsymbol{\beta}$ for each subject. Pregibon (1981) extended DFBETAs to logistic regression using one-step estimators of $\boldsymbol{\beta}$. Preisser and Qaqish (1996) extended DFBETAs in a similar way within the GEE approach. Their one-step DFBETAs for GEE require deletion of the n_i repeated measures for the ith subject. That is, conditional on the estimate $\boldsymbol{\alpha}$, a "one-step" estimate of $\boldsymbol{\beta}$, say $\widehat{\boldsymbol{\beta}}_{-i}$, is obtained by excluding the n_i repeated measures for the ith subject and performing one step of the Fisher scoring algorithm described in Equation (3.7) with the full-data estimate $\widehat{\boldsymbol{\beta}}$ as the starting value,

$$\widehat{\boldsymbol{\beta}}_{-i} = \widehat{\boldsymbol{\beta}} + \left[\sum_{k=1,k\neq i}^{N} D_k'(\widehat{\boldsymbol{\beta}}) \{V_k(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})\}^{-1} D_k(\widehat{\boldsymbol{\beta}}) \right]^{-1} \sum_{k=1,k\neq i}^{N} D_k'(\widehat{\boldsymbol{\beta}}) V_k^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}) \{\boldsymbol{Y}_k - \boldsymbol{\mu}_k(\widehat{\boldsymbol{\beta}})\}.$$

To obtain a unitless, composite measure of the influence of each subject on the entire set of regression coefficients, a statistic similar to Wilks's (1963) statistic for multivariate outliers,

$$w_i = (\widehat{\boldsymbol{\beta}}_{-i} - \widehat{\boldsymbol{\beta}})' \{\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}})\}^{-1} (\widehat{\boldsymbol{\beta}}_{-i} - \widehat{\boldsymbol{\beta}}),$$

can be used. A large value of w_i (relative to the other w_is) indicates that the ith subject may have unusually high influence. Using Wilks's (1963) criterion as a very rough guide, $\{(N-p-1)(N-1)/(Np)\}w_i$ has an approximate F-distribution with p (the dimension of $\boldsymbol{\beta}$) and $N-p-1$ degrees of freedom (for example, see Lipsitz, Laird, and Harrington, 1992).

Goodness-of-fit criteria for generalized estimating equations have also been developed. For example, Pan (2001) proposed an extension of Akaike's information criterion (AIC) to GEE. For the special case of longitudinal binary data, Barnhart and Williamson (1998) and Horton et al. (1999) proposed goodness-of-fit tests for GEE that are extensions of

the Hosmer–Lemeshow goodness-of-fit test for logistic regression (Hosmer and Lemeshow, 1980). Here, we briefly review goodness-of-fit statistics for GEE. Suppose it is of interest to determine whether or not the marginal model

$$\mu_{ij} = \frac{\exp(\boldsymbol{X}'_{ij}\boldsymbol{\beta})}{1+\exp(\boldsymbol{X}'_{ij}\boldsymbol{\beta})}$$

provides an adequate fit to the data. A standard approach is to fit a broader model (e.g., a model with interactions and/or polynomial and higher-order terms) and test whether the additional terms are significantly different from zero. Alternatively, a "global goodness-of-fit" statistic can be obtained by extending the Hosmer–Lemeshow statistic (Hosmer and Lemeshow, 1980; see also Tsiatis, 1980). Following the suggestion of Hosmer and Lemeshow for ordinary logistic regression, G (usually 10) groups can be formed based on combinations of the covariates $\boldsymbol{X}_{ij}$ in the logistic regression model. A test for goodness of fit is constructed by testing whether the additional regression coefficients for the $G-1$ indicator variables differ from zero. Following Hosmer and Lemeshow, Horton et al. (1999) suggest forming groups based on deciles of the predicted probabilities from the given model,

$$\tilde{\mu}_{ij} = \frac{\exp(\boldsymbol{X}'_{ij}\widehat{\boldsymbol{\beta}})}{1+\exp(\boldsymbol{X}'_{ij}\widehat{\boldsymbol{\beta}})}.$$

Note that each subject has n_i separate estimates of risk ($\tilde{\mu}_{ij}$s) and that there are $\sum_{i=1}^{N} n_i$ observations in total. Horton et al. (1999) suggest forming 10 groups of approximately equal size from the deciles of these predicted probabilities. For example, the first group contains the $\sum_{i=1}^{N} n_i/10$ $(Y_{ij}, \boldsymbol{X}_{ij})$s with the smallest values of $\tilde{\mu}_{ij}$, and the last group contains the $\sum_{i=1}^{N} n_i/10$ $(Y_{ij}, \boldsymbol{X}_{ij})$s with the largest values of $\tilde{\mu}_{ij}$. Finally, defining the $G-1$ group indicators

$$I_{ijg} = \begin{cases} 1 \text{ if } \tilde{\mu}_{ij} \text{ is in group } g, \\ 0 \text{ otherwise,} \end{cases} \qquad g = 1, \ldots, G-1,$$

where the groups are based on "percentiles of risk," a test for goodness of fit is obtained by considering the alternative model

$$\text{logit}(\mu_{ij}) = \boldsymbol{X}'_{ij}\boldsymbol{\beta} + \gamma_1 I_{ij1} + \ldots + \gamma_{G-1} I_{ij,G-1}.$$

Specifically, the score statistic given in (3.10) can be used to test the null hypothesis

$$\text{H}_0 : \gamma_1 = \ldots = \gamma_{G-1} = 0.$$

Finally, graphical displays are very useful techniques for conveying information about the most salient features of longitudinal data. They can provide insights about patterns of change in the mean response over time (e.g., linearity or the lack thereof) and the choice of suitable functional forms for covariates. Graphical techniques, especially those based on residuals, are especially useful for assessing the adequacy of any postulated model for longitudinal data. They are also useful for identifying observations and individuals that are potential outliers. With appropriate transformations, residual diagnostics developed for standard linear regression can be extended to the longitudinal setting. However, an acknowledged difficulty with conventional residual diagnostics is that they are somewhat subjective in nature. What appears to be a random scatter to one individual might be considered evidence of systematic trend to another. That is, it can be very difficult to discern whether an apparent trend in a scatter plot of the residual reflects some aspect of model misspecification or is simply a reflection of natural variation. McCullagh and Nelder (1989, pp. 392–393) aptly summarize this problem when they state that "the practical problem is that any finite set of residuals can be made to yield some kind of pattern if we look hard enough, so that we have to guard against over-interpretation."

Recently, Lin, Wei, and Ying (2002) developed model-checking techniques based on "cumulative sums" and "moving sums" of residuals that help discern the "signal" from the "noise." The basic idea is to aggregate the residuals over certain coordinates. The coordinates typically used for these sums of residuals are the individual covariates (e.g., X_{ijk}, the kth covariate) and the fitted values, $h(\boldsymbol{X}'_{ij}\widehat{\boldsymbol{\beta}})$. These authors demonstrated that a key advantage of working with sums of residuals, rather than crude residuals, is that a reference distribution is available to ascertain their natural variation. That is, the *observed* sums of the residuals can be compared, both graphically and numerically, to a reference distribution under the assumption of a correctly specified marginal model for the mean. This allows for the determination of whether any apparent pattern is evidence of a systematic trend or simply due to natural variation. This model-checking technique developed by Lin, Wei, and Ying (2002) removes a large degree of subjectivity from the assessment of graphical displays of residuals and places residual diagnostics on a more objective footing.

Specifically, if the assumed model for the mean response is correct, then the cumulative sums of residuals are centered at zero. Moreover, the distribution of the cumulative sum can be approximated by that of a Gaussian process with zero mean whose realizations can be generated via computer simulation. It is relatively straightforward to generate realizations from the distribution of the cumulative sum, under the assumption that the model for the mean is correct (the details are omitted here and the interested reader is referred to Lin, Wei, and Ying, 2002). Thus, to assess whether any apparent trend in the *observed* cumulative sum of residuals reflects systematic trends rather than chance fluctuations, a number of realizations from the appropriate Gaussian process can be superimposed. To the extent that the curves generated from the null distribution tend to be closer to and intersect zero more often than the observed curve, this provides evidence of lack of fit. This assessment can be put on a more formal footing by comparing the maximum absolute value of the observed cumulative sum to a large number of realizations (say 10,000) from the null distribution.

An appealing feature of the graphical and numerical methods based on cumulative and moving sums of residuals is that they are valid regardless of the true joint distribution of the longitudinal response vector; in particular, they do not require correct specification of the covariance among the responses. As such, these graphical and numerical techniques for assessing the model for the mean response are relatively robust to assumptions about the distribution of the responses and assumptions about the covariance among the repeated measures.

3.6 Case study

In this section we present results of analyses of cardiovascular abnormalities from a longitudinal study of children infected with HIV-1 to illustrate some of the main ideas highlighted in earlier sections. Results from various cross-sectional and short-term longitudinal studies (Lipshultz et al., 1998) have suggested that children infected with HIV-1 might have higher risks of cardiovascular abnormalities. We consider this hypothesis using data from the Pediatric Pulmonary and Cardiac Complications (P2C2) of Vertically Transmitted HIV Infection Study (Lipshultz et al., 1998). This was a large, prospective longitudinal study designed to monitor heart disease and the progression of cardiac abnormalities in children born to HIV-infected women. In the P2C2 study, a birth cohort of 401 infants born to women infected with HIV-1 were scheduled to have their cardiovascular function measured approximately every year from birth to age 6, producing up to seven repeated measurements of cardiovascular function on each child. Of the 401 infants who participated in this study, 74 (18.8%) were HIV positive, and 319 (81.2%) were HIV negative.

Table 3.1 Data from 10 Randomly Selected Children from the P2C2 Study

Subject	HIV[b]	Mom Smoked[c]	Gest. Age (wks)	Low Birth Weight[d]	Heart Pumping Ability at Age[a]						
					Birth	1	2	3	4	5	6
1	1	0	41	0	0	0	0	0	0	0	.
2	1	1	34	0	1	.	0	0	1	.	.
3	0	1	40	0	1	0	0	.	.	.	.
4	1	0	40	0	0	.	0	0	0	1	.
5	0	1	39	0	.	1	0	.	.	.	.
6	0	1	35	0	1	.	.	.	.	.	.
7	0	0	36	0	.	0	0	.	.	.	.
7	1	0	33	1	.	1	1	1	.	.	.
8	0	0	36	1	0	0	.	.	.	.	.
9	0	0	41	1	.	.	.	.	0	.	.
10	0	1	34	1	.	0	0	.	0	1	0

(. = missing)
[a] 1 = abnormal, 0 = normal.
[b] 1 = HIV positive, 0 = not HIV positive.
[c] 1 = mother smoked during pregnancy, 0 mother did not smoke.
[d] 1 = low birth weight for age, 0 = normal birth weight.

The question of main scientific interest is to determine if children infected with HIV-1 have worse heart function over time. In particular, we are interested in modeling the pumping ability of the heart (left ventricular fractional shortening) over time. The outcome variable for this analysis is a binary response denoting abnormally low left ventricular fractional shortening (1 = low fractional shortening, 0 = normal) at each occasion. For these data, we are interested in modeling the marginal means or, equivalently, the probabilities of abnormal heart function over time, and relating changes in these marginal probabilities to covariates. The main covariate is the indicator of HIV infection; it is of interest to estimate the effect of HIV infection on heart function over time. Previous results (Lipshultz et al., 1998, 2000, 2002) from the P2C2 study have shown that subclinical cardiac abnormalities develop early in children born with HIV, and that they are frequent, persistent, and often progressive.

Thus, for these analyses we are interested in assessing whether children with HIV have progressively worse heart function over time compared to non-HIV children. In the regression model for abnormal heart function, this translates into a test of the time-by-HIV status interaction. For these analyses, potential confounding variables that must be adjusted for include mother's smoking status during pregnancy (coded 1 = yes, 0 = no), gestational age (in weeks), and birth weight standardized for age (coded 1 = abnormal, 0 = normal). Data from 10 randomly selected children are displayed in Table 3.1.

We note that there is a substantial amount of missing data. Although each child was scheduled to have an echocardiogram every year for the first 6 years of life, including at birth, only 1 (0.25%) of the 401 study participants had outcomes measured at all seven occasions; see Table 3.2 for the frequency distribution of the number of repeated echocardiograms. Table 3.3 shows the number of subjects with echocardiogram measurements at each of the seven measurement occasions. From Table 3.2, we see that only 91 subjects (22.7%) were seen on more than three occasions. From Table 3.3, we see that 276 of the 401 children (68.8%) have baseline measurements; after birth, the percentage of children with echocardiogram measurements declines until only 7 (1.7%) of the 401 subjects have measurements at 6 years of age. For illustrative purposes, the analyses presented here assume that the missing responses are MCAR (Rubin, 1976; Laird, 1988). As discussed in Section 3.4, when the outcome data are MCAR, all GEE approaches yield consistent estimators of

Table 3.2 Frequency Distribution of the Number of Echocardiograms for Children in the P2C2 Study

Number of Echocardiograms	Number of Subjects	Percentage
1	148	36.91
2	104	25.94
3	58	14.46
4	50	12.47
5	30	7.48
6	10	2.49
7	1	0.25
Total	401	100.00

the regression parameters provided the model for the mean is correctly specified. However, we caution the reader that a substantive analysis of these data would require a far more careful treatment of the missing data (see the chapters in Part 5 for detailed discussions of this topic).

To examine the effect of HIV-1, we considered the following marginal logistic regression model for the probability of abnormal heart function at time t_j, denoted μ_{ij}:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 t_j + \beta_2 \text{HIV}_i + \beta_3 t_j \text{HIV}_i + \beta_4 \text{smoke}_i + \beta_5 \text{age}_i + \beta_6 \text{wt}_i,$$

for $j = 1, \ldots, 7$, where $t_j = j - 1$, HIV_i equals 1 if the ith child is born with HIV-1 and equals 0 otherwise; smoke_i equals 1 if the mother smoked during pregnancy and 0 otherwise; age_i is the gestational age (in weeks); and wt_i equals 1 if the child's birth weight for gestational age was abnormal and 0 otherwise. Because there are so few children with echocardiogram measurements after age 4, there are insufficient data to fit an unstructured correlation matrix with $(7 \times 6)/2 = 21$ parameters,

$$\rho_{ist} = \text{Corr}(Y_{is}, Y_{it}; \boldsymbol{\alpha}) = \rho_{st}.$$

Instead, to account for the within-subject association among the binary responses, we consider two possible "working correlation" patterns: "exchangeable," in which $\rho_{ist} = \alpha$ for all $s < t$, and AR(1), with

$$\rho_{ist} = \alpha^{|t-s|},$$

where $0 < \alpha < 1$.

Finally, to illustrate similarities and differences between various GEE estimation techniques, we compare the estimates of $(\boldsymbol{\beta}, \alpha)$ obtained using six approaches: (1) GEE under "working independence" (equivalent to ordinary logistic regression); (2) standard GEE1

Table 3.3 Frequency Distribution of the Number of Children with Echocardiograms at Each Occasion

Age at Visit (Years)	Number of Subjects	Percentage
Birth	276	68.83
1	267	66.58
2	154	38.40
3	123	30.67
4	83	20.70
5	37	9.23
6	7	1.75

with estimation of α based on *unconditional* residuals; (3) GEE1 with estimation of α based on *conditional* residuals; (4) GEE1 with Gaussian estimation of α; (5) GEE2 with $\mathcal{V}_i$ in (3.15) specified using the Bahadur distribution (Bahadur, 1961); and (6) maximum likelihood using the parametric Bahadur distribution, which is based on two- and higher-order correlations.

Because the Bahadur distribution is used in both the GEE2 and ML approaches, we provide a brief description here. We define the standardized binary variable S_{ij} as

$$S_{ij} = \frac{Y_{ij} - \mu_{ij}}{\{\mu_{ij}(1-\mu_{ij})\}^{1/2}}.$$

The pairwise correlation between Y_{ij} and Y_{ik} is $\rho_{jk} = E(S_{ij}S_{ik})$, and the Rth-order correlation between the first R responses is defined as $\rho_{12\ldots R} = E(S_{i1}S_{i2}\ldots S_{iR})$. The Rth-order correlation between any R of the n repeated binary responses is defined similarly. The Bahadur representation of the $2^n - 1$ multinomial probabilities corresponding to the joint distribution of $(Y_{i1}, Y_{i2}, \ldots, Y_{in})$ is

$$\Pr\{Y_{i1} = y_1, Y_{i2} = y_2, \ldots, Y_{in} = y_n | X_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} = \left\{\prod_{j=1}^{n} \mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}}\right\}$$
$$\times \left\{1 + \sum_{j>k} \rho_{jk} s_{ij} s_{ik} + \sum_{j>k>l} \rho_{jkl} s_{ij} s_{ik} s_{il} + \cdots + \rho_{1\ldots n} s_{i1} \cdots s_{in}\right\}.$$

For both GEE2 and ML approaches, we assumed all fifth- and higher-order correlations are 0 ($\rho_{jklmn} = \cdots = \rho_{1\ldots n} = 0$). In addition, we assumed that all fourth-order correlations are the same, regardless of the sets of times ($\rho_{jklm} = \rho_{j'k'l'm'}$ for all $jklm \neq j'k'l'm'$), and that all third-order correlations are the same, regardless of the sets of times ($\rho_{jkl} = \rho_{j'k'l'}$ for all $jkl \neq j'k'l'$). The two models for the pairwise correlations ρ_{st} (exchangeable and autoregressive) are the same for all six estimation methods.

The logistic regression parameter estimates and standard errors are presented in Table 3.4 for the "working independence" GEE and GEE1 with exchangeable correlation estimated using *unconditional* residuals. Table 3.4 presents both model-based and empirical (or so-called sandwich) variance estimates. As might be expected, the most discernible differences between the model-based and empirical standard errors occur for the "working independence" GEE. This is because the naive assumption of independence among repeated binary responses obtained from the same child is "furthest" from the true underlying within-subject correlation. In general, when the true correlation among repeated measures is high, the differences between the model-based and empirical standard errors for the "working independence" GEE can be substantial; the former provide unrealistic estimates of the sampling variability. For the data from the P2C2 study, the estimated correlation is relatively modest (assuming an exchangeable correlation, the estimated correlation is 0.15); consequently, the differences between the model-based and empirical standard errors are not too extreme. For example, for the "working independence" GEE, the largest differences are in the estimated variance of the time-by-HIV infection interaction, with the empirical variance being almost 1.5 (or $[0.16/0.13]^2$) times larger than the model-based variance, and in the gestational age effect where the empirical variance is almost 1.4 (or $[0.36/0.31]^2$) times larger than the model-based variance. We note that for the "working independence" GEE the model-based variance is not necessarily always smaller than the empirical variance; their relative size depends on whether the parameter of interest is the effect of a time-varying (or within-subject) or time-stationary (or between-subject) covariate, the magnitude of the correlation, and the proportion of missing data over time (see Mountford et al., 2007, for a more detailed discussion). We also note that for the GEE1 with exchangeable correlation estimated using

Table 3.4 Regression Parameter Estimates with Model-Based and Empirical Standard Errors (SE) for Independence GEE and GEE1 with Exchangeable Correlation Estimated Using Unconditional Residuals

Effect	Method	Estimate	Model SE	Empirical SE	Empirical Wald Z	p-Value
Intercept	IND	2.284	1.236	1.508	1.51	0.130
	GEE1	1.763	1.416	1.468	1.20	0.230
Time	IND	−0.600	0.076	0.097	−6.16	<0.001
	GEE1	−0.630	0.078	0.096	−6.54	<0.001
HIV	IND	−0.069	0.252	0.266	−0.26	0.796
	GEE1	−0.081	0.255	0.266	−0.30	0.761
Time*HIV	IND	0.204	0.130	0.159	1.28	0.200
	GEE1	0.283	0.125	0.149	1.90	0.058
Mom Smoke	IND	−0.196	0.153	0.175	−1.12	0.263
	GEE1	−0.219	0.175	0.168	−1.31	0.191
Gest. Age	IND	−0.058	0.031	0.038	−1.52	0.130
	GEE1	−0.044	0.036	0.037	−1.19	0.233
Low Birth Wt.	IND	0.048	0.163	0.198	0.24	0.810
	GEE1	0.096	0.186	0.186	0.52	0.604

unconditional residuals, there are fewer differences between the model-based and empirical standard errors. However, the small number of discernible differences (e.g., the estimated variances of the time-by-HIV interaction) suggest that the model for the correlation could potentially be improved. That is, if the working model for the correlation has been correctly specified, in general, we would expect the model-based and empirical standard errors to be relatively similar.

Based on the regression parameter estimates and empirical standard errors from the GEE1 with exchangeable correlation, there is the suggestion that the pattern of change in risk of abnormal heart function differs by HIV-1 infection group ($Z = 1.90, p < 0.06$). Specifically, after adjustment for maternal smoking during pregnancy, gestational age, and birth weight, the risk decreases at a slower rate in the HIV-1 infected children. Note that the interpretation of the results is very similar to the interpretation of results from an ordinary logistic regression model for cross-sectional data. For example, children with HIV-1 infection have $\exp(\widehat{\beta}_2 + \widehat{\beta}_3 t_j) = \exp(-0.081 + 0.285 t_j)$ times the odds of having an abnormal pumping ability compared to children without HIV-1 infection at time t_j. Specifically, at birth ($t_1 = 0$), the ratio of their odds of having an abnormal pumping ability is approximately 1 ($\exp(-0.081) = 0.92$), while at year 6 the ratio of their odds is approximately 5.0 ($\exp(-0.081 + 6 \times 0.285) = 5.1$).

For illustrative purposes, Table 3.5 presents the parameter estimates (and empirical standard error estimates) obtained from all five methods for a "working exchangeable" correlation. Note that the regression parameter estimates and estimated standard errors from the three GEE1 methods are very similar to each other. In contrast, the regression parameter estimates from GEE2 are somewhat different. Recall that GEE2 requires the stronger assumption that *both* the first and second moments must be correctly specified. For these data, the empirical standard errors for GEE2 estimates are similar to those for GEE1; however, this result is not to be expected in general. Finally, as expected, ML yields the smallest estimated variances for the estimated regression parameters. The largest gain in efficiency is for the time-by-HIV infection interaction, the effect of greatest subject-matter interest.

Table 3.5 Comparison of Parameter Estimates (and Empirical Standard Errors) under an Exchangeable Correlation

Effect	Method	Estimate	SE	Wald Z	p-Value
Intercept	GEE1 (uncond)	1.763	1.468	1.20	0.230
	GEE1 (cond)	1.851	1.470	1.26	0.208
	GEE1 (Gaussian)	1.822	1.469	1.24	0.216
	GEE2	2.126	1.443	1.47	0.141
	ML	1.733	1.336	1.30	0.195
Time	GEE1 (uncond)	−0.630	0.096	−6.54	<0.001
	GEE1 (cond)	−0.626	0.096	−6.49	<0.001
	GEE1 (Gaussian)	−0.627	0.096	−6.51	<0.001
	GEE2	−0.583	0.095	−6.16	<0.001
	ML	−0.678	0.084	−8.10	<0.001
HIV	GEE1 (uncond)	−0.081	0.266	−0.30	0.761
	GEE1 (cond)	−0.079	0.265	−0.30	0.765
	GEE1 (Gaussian)	−0.080	0.265	−0.30	0.764
	GEE2	−0.047	0.274	−0.17	0.865
	ML	−0.037	0.255	−0.15	0.884
Time*HIV	GEE1 (uncond)	0.283	0.149	1.90	0.058
	GEE1 (cond)	0.271	0.151	1.80	0.072
	GEE1 (Gaussian)	0.275	0.150	1.83	0.068
	GEE2	0.264	0.156	1.70	0.089
	ML	0.287	0.122	2.35	0.019
Mom Smoke	GEE1 (uncond)	−0.219	0.168	−1.31	0.191
	GEE1 (cond)	−0.215	0.168	−1.28	0.201
	GEE1 (Gaussian)	−0.216	0.168	−1.29	0.198
	GEE2	−0.249	0.166	−1.50	0.134
	ML	−0.241	0.159	−1.51	0.131
Gest. Age	GEE1 (uncond)	−0.044	0.037	−1.19	0.233
	GEE1 (cond)	−0.047	0.037	−1.25	0.211
	GEE1 (Gaussian)	−0.046	0.037	−1.23	0.219
	GEE2	−0.055	0.036	−1.52	0.128
	ML	−0.043	0.034	−1.27	0.204
Low Birth Wt.	GEE1 (uncond)	0.096	0.186	0.52	0.604
	GEE1 (cond)	0.089	0.187	0.48	0.632
	GEE1 (Gaussian)	0.092	0.186	0.49	0.622
	GEE2	0.042	0.180	0.23	0.816
	ML	0.155	0.168	0.93	0.356
ρ	GEE1 (uncond)	0.153	0.120	1.28	0.201
	GEE1 (cond)	0.194	0.078	2.47	0.013
	GEE1 (Gaussian)	0.165	0.049	3.37	0.001
	GEE2	0.174	0.059	2.95	0.003
	ML	0.151	0.034	4.42	<0.001

The estimated relative efficiency for GEE1 with exchangeable correlation versus ML is $(0.12/0.15)^2 = 66\%$. However, a potential drawback of ML is that the full joint distribution of the binary outcomes must be correctly specified; thus, the assumptions made about all fifth- and higher-order correlations being zero must be correct for ML to yield asymptotically unbiased estimators of the regression parameters.

Although all GEE1 methods produce similar standard errors for the estimated regression parameters, this is not the case for the estimated correlation parameter. Specifically, the GEE1 based on conditional residuals greatly reduces the estimated variance of the correlation compared to GEE1 with unconditional residuals, while GEE1 based on Gaussian

Table 3.6 Comparison of Parameter Estimates (and Empirical Standard Errors) under an AR(1) Correlation

Effect	Method	Estimate	SE	Wald Z	p-Value
Intercept	GEE1 (uncond)	2.018	1.506	1.34	0.180
	GEE1 (cond)	1.817	1.508	1.20	0.228
	GEE1 (Gaussian)	1.978	1.506	1.31	0.190
	GEE2	2.118	1.429	1.48	0.138
	ML	1.763	1.352	1.30	0.193
Time	GEE1 (uncond)	−0.634	0.097	−6.50	<0.001
	GEE1 (cond)	−0.661	0.098	−6.72	<0.001
	GEE1 (Gaussian)	−0.639	0.098	−6.55	<0.001
	GEE2	−0.519	0.095	−5.48	<0.001
	ML	−0.637	0.088	−7.28	<0.001
HIV	GEE1 (uncond)	−0.072	0.264	−0.27	0.785
	GEE1 (cond)	−0.075	0.264	−0.28	0.777
	GEE1 (Gaussian)	−0.073	0.264	−0.28	0.783
	GEE2	−0.073	0.299	−0.24	0.808
	ML	−0.037	0.269	−0.14	0.891
Time*HIV	GEE1 (uncond)	0.242	0.157	1.54	0.123
	GEE1 (cond)	0.273	0.155	1.76	0.078
	GEE1 (Gaussian)	0.248	0.156	1.58	0.114
	GEE2	0.173	0.168	1.03	0.302
	ML	0.213	0.140	1.53	0.128
Mom Smoke	GEE1 (uncond)	−0.199	0.173	−1.15	0.250
	GEE1 (cond)	−0.200	0.172	−1.16	0.246
	GEE1 (Gaussian)	−0.199	0.173	−1.15	0.250
	GEE2	−0.266	0.174	−1.53	0.127
	ML	−0.206	0.172	−1.20	0.231
Gest. Age	GEE1 (uncond)	−0.050	0.038	−1.31	0.190
	GEE1 (cond)	−0.044	0.038	−1.15	0.249
	GEE1 (Gaussian)	−0.049	0.038	−1.28	0.202
	GEE2	−0.056	0.036	−1.55	0.122
	ML	−0.043	0.034	−1.26	0.207
Low Birth Wt.	GEE1 (uncond)	0.076	0.194	0.39	0.694
	GEE1 (cond)	0.100	0.192	0.52	0.603
	GEE1 (Gaussian)	0.081	0.193	0.42	0.677
	GEE2	0.040	0.188	0.21	0.830
	ML	0.096	0.173	0.55	0.581
ρ	GEE1 (uncond)	0.167	0.083	2.02	0.043
	GEE1 (cond)	0.289	0.063	4.63	<0.001
	GEE1 (Gaussian)	0.192	0.059	3.23	0.001
	GEE2	0.167	0.076	2.21	0.027
	ML	0.206	0.050	4.08	<0.001

estimation reduces the estimated variance even further. Note that GEE1 based on Gaussian estimation yields a smaller variance estimate than GEE2, although this result cannot be expected in general.

In Table 3.6, we consider estimates based on the five methods presented in Table 3.5 when the "working correlation" is assumed to be first-order autoregressive (AR(1)). In general, the pattern of results in Table 3.6 is similar to that in Table 3.5. With a modest correlation, it is perhaps not surprising that the estimates of the regression parameters under AR(1) or exchangeable should be so similar. The imbalance in the data due to missingness probably accounts for some of the small differences when the results in Table 3.5 and Table 3.6 are

compared. Also, we remind the reader that all analyses have made the strong assumption that the missing data are MCAR.

In earlier sections, we remarked that for the GEE2 estimate of $\boldsymbol{\beta}$ to be consistent the model for the first two moments of $\boldsymbol{Y}_i$ must be correctly specified. This is in contrast to GEE1 where a consistent estimator of $\boldsymbol{\beta}$ is obtained provided only that the mean of $\boldsymbol{Y}_i$ has been correctly specified. To illustrate the sensitivity of GEE2 estimates of $\boldsymbol{\beta}$ to misspecification of the second moments, we fixed the exchangeable correlation at larger and smaller values than the estimate of $\widehat{\rho} = 0.174$ obtained in Table 3.5 and re-estimated $\boldsymbol{\beta}$. For example, when ρ is fixed at 0.7, the GEE2 estimate of the time-by-HIV status interaction, the parameter of main interest, is 0.390 (SE = 0.156), which is discernibly different from the estimate of 0.264 obtained in Table 3.5. Similarly, when ρ is fixed at 0.065, the GEE2 estimate of the time-by-HIV status interaction is 0.075 (SE = 0.192). This highlights how GEE2 estimates of $\boldsymbol{\beta}$ are sensitive to misspecification of the within-subject association.

Finally, in Section 3.2.4 we mentioned that for the special case where the outcome is binary, an alternative to the correlation as a measure of within-subject association is the odds ratio. The odds ratio has many desirable properties and, unlike the correlation, is not constrained by the marginal probabilities, μ_{ij}. For illustrative purposes, we replicated the GEE1 analyses in Table 3.5 and Table 3.6 where the within-subject association was parameterized in terms of log odds ratios rather than correlations. Recall that the joint distribution of Y_{is} and Y_{it} depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, with

$$\pi_{ist} = E(Y_{is}Y_{it}|\boldsymbol{X}_{is}, \boldsymbol{X}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \Pr(Y_{is} = 1, Y_{it} = 1|\boldsymbol{X}_{is}, \boldsymbol{X}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}).$$

This joint probability can be modeled in terms of either the marginal correlation or the marginal odds ratio. The marginal correlation between the responses at times s and t is

$$\rho_{ist} = \rho_{ist}(\boldsymbol{\alpha}) = \mathrm{Corr}(Y_{is}, Y_{it}; \boldsymbol{\alpha}) = \frac{\pi_{ist} - \mu_{is}\mu_{it}}{\{\mu_{is}(1-\mu_{is})\mu_{it}(1-\mu_{it})\}^{1/2}}.$$

Note that the joint probability π_{ist} can be written in terms of the correlation coefficient as

$$\pi_{ist} = \mu_{is}\mu_{it} + \rho_{ist}\{\mu_{is}(1-\mu_{is})\mu_{it}(1-\mu_{it})\}^{1/2}.$$

Alternatively, instead of modeling the association between pairs of binary responses in terms of the marginal correlation, Lipsitz, Laird, and Harrington (1991), Liang, Zeger, and Qaqish (1992), and Carey, Zeger, and Diggle (1993) propose using the marginal odds ratio. The odds ratio between the responses at times s and t is

$$\psi_{ist} = \psi_{ist}(\boldsymbol{\alpha}) = \frac{\pi_{ist}(1 - \mu_{is} - \mu_{it} + \pi_{ist})}{(\mu_{is} - \pi_{ist})(\mu_{it} - \pi_{ist})}.$$

In terms of the odds ratio, the probability π_{ist} can then be written as

$$\pi_{ist} = \begin{cases} \dfrac{a_{ist} - \{a_{ist}^2 - 4\psi_{ist}(\psi_{ist}-1)\mu_{is}\mu_{it}\}^{1/2}}{2(\psi_{ist}-1)} & \text{if } \psi_{ist} \neq 1, \\ \mu_{is}\mu_{it} & \text{if } \psi_{ist} = 1, \end{cases}$$

where $a_{ist} = 1 - (1 - \psi_{ist})(\mu_{is} + \mu_{it})$. An "exchangeable" odds ratio model is given by $\psi_{ist} = \alpha$. As a binary data analog of the AR(1) model for correlation, Fitzmaurice and Lipsitz (1995) proposed the following serial pattern model for the odds ratio,

$$\psi_{ist} = \alpha^{1/|t-s|},$$

where $1 < \alpha < \infty$. Note that as $|t - s| \to 0$, $\psi_{ist} \to \alpha^{\infty}$ and there is perfect association. When observations are far apart in time, then as $|t - s| \to \infty$, $\psi_{ist} \to \alpha^0 = 1$ and the pairs of observations are independent. This serial odds ratio model mimics an AR(1) pattern with the strength of association declining with increasing time separation. Thus, we can express

both the "exchangeable" and "serial" odds ratio models as linear models for the log odds ratio, with

$$\log(\psi_{ist}) = \left(\frac{1}{|t-s|}\right)\log(\alpha)$$

for the "serial" odds ratio model, and

$$\log(\psi_{ist}) = \log(\alpha)$$

for the "exchangeable" odds ratio model.

To estimate $\boldsymbol{\beta}$ and α, under an "exchangeable" or "serial" odds ratio pattern, we can use the estimating equations proposed by Lipsitz, Laird, and Harrington (1990) based on unconditional residuals or the estimating equations proposed by Carey, Zeger, and Diggle (1993) based on conditional residuals; in general, the latter yield more efficient estimators of $\boldsymbol{\alpha}$. The results of fitting the two models for the odds ratio pattern using both methods of estimation are presented in Table 3.7. The pattern of results in Table 3.7 is similar to

Table 3.7 Comparison of Parameter Estimates (and Empirical Standard Errors) Using Log Odds Ratio as a Measure of Within-Subject Association, Estimated Using Unconditional and Conditional Residuals

Effect	Method	Odds Ratio Model	Estimate	SE	Wald Z	p-Value
Intercept	GEE1 (uncond)	exc	1.870	1.491	1.25	0.210
	GEE1 (cond)	exc	1.797	1.492	1.20	0.228
	GEE1 (uncond)	serial	2.003	1.503	1.33	0.183
	GEE1 (cond)	serial	1.810	1.506	1.20	0.230
Time	GEE1 (uncond)	exc	−0.618	0.096	−6.46	<0.001
	GEE1 (cond)	exc	−0.623	0.096	−6.51	<0.001
	GEE1 (uncond)	serial	−0.618	0.097	−6.41	<0.001
	GEE1 (cond)	serial	−0.637	0.097	−6.58	<0.001
HIV	GEE1 (uncond)	exc	−0.075	0.265	−0.28	0.777
	GEE1 (cond)	exc	−0.078	0.265	−0.30	0.768
	GEE1 (uncond)	serial	−0.061	0.263	−0.23	0.816
	GEE1 (cond)	serial	−0.060	0.263	−0.23	0.819
Time*HIV	GEE1 (uncond)	exc	0.257	0.150	1.72	0.086
	GEE1 (cond)	exc	0.269	0.149	1.80	0.071
	GEE1 (uncond)	serial	0.232	0.154	1.51	0.132
	GEE1 (cond)	serial	0.256	0.152	1.68	0.093
Mom Smoke	GEE1 (uncond)	exc	−0.202	0.169	−1.20	0.232
	GEE1 (cond)	exc	−0.203	0.168	−1.21	0.228
	GEE1 (uncond)	serial	−0.194	0.172	−1.13	0.258
	GEE1 (cond)	serial	−0.194	0.171	−1.13	0.257
Gest. Age	GEE1 (uncond)	exc	−0.047	0.038	−1.24	0.214
	GEE1 (cond)	exc	−0.045	0.038	−1.19	0.234
	GEE1 (uncond)	serial	−0.050	0.038	−1.31	0.189
	GEE1 (cond)	serial	−0.044	0.038	−1.16	0.244
Low Birth Wt.	GEE1 (uncond)	exc	0.101	0.188	0.54	0.589
	GEE1 (cond)	exc	0.110	0.187	0.59	0.555
	GEE1 (uncond)	serial	0.087	0.192	0.45	0.652
	GEE1 (cond)	serial	0.111	0.190	0.58	0.559
Log(OR)	GEE1 (uncond)	exc	0.859	0.931	0.92	0.356
	GEE1 (cond)	exc	1.051	0.281	3.74	<0.001
	GEE1 (uncond)	serial	0.788	0.999	0.79	0.430
	GEE1 (cond)	serial	1.369	0.301	4.54	<0.001

those in Table 3.5 and Table 3.6, confirming that GEE1 methods are relatively insensitive to choice of models and methods of estimation for the within-subject association. Interestingly, the potential benefits of GEE1 based on conditional residuals for estimation of the within-subject association are highlighted in Table 3.7 where the standard error for the estimated log odds ratio is discernibly smaller than for GEE1 based on unconditional residuals. For the parameter of main interest, the time-by-HIV status interaction, the estimates of effects are similar to those found in Table 3.5 and Table 3.6.

3.7 Discussion and future directions

As discussed in the previous sections, Liang and Zeger (1986) developed generalized estimating equations to estimate the parameters of marginal models; the latter class of models represent a particular extension of generalized linear models to longitudinal data. Liang and Zeger (1986) introduced a class of moment-based estimating equations that yield consistent estimators of the regression parameters, and their variances, under relatively mild conditions. Over the past 20 years, the GEE approach has been the dominant method for estimation of marginal models for longitudinal data; so much so, indeed, that some authors confusingly refer to marginal models as "GEE models." However, it is worth emphasizing that GEE methods are but one approach for estimating marginal model parameters. Moreover, GEE methods should not be regarded as conjoined at the hip to marginal models; the estimating equation approach can be applied equally to alternative classes of models for longitudinal data, such as so-called transitional or response-conditional models (e.g., Markov models for longitudinal data) and GLMMs.

The GEE method is semi-parametric, in that the estimating equations are derived without fully specifying the joint distribution of the vector of repeated measures. This is a very appealing feature of the GEE approach, especially for the analysis of discrete longitudinal data, because for the latter case the total number of parameters in the saturated model for the joint distribution of the vector of responses grows exponentially with the number of repeated measures. In particular, GEE methods only require specification of the form of the first two moments of the outcome vector, and provide consistent estimators of the marginal regression parameters under the weak condition that only the first moment is correctly specified. That is, provided the marginal model for the mean response (the first moment) is correctly specified, the estimating equations yield estimators of the marginal regression parameters that are consistent and asymptotically normal, regardless of whether the second moments have been correctly specified.

Similar to the quasi-likelihood estimating equations originally proposed by Wedderburn (1974), the optimal choice of weights is to take $V_i = \text{Cov}(\boldsymbol{Y}_i)$. Naturally, there are certain trade-offs associated with the use of less than optimal weights. However, in general, if the correlation among repeated measures is not too high and/or the effects of primary interest are between-subject effects, then the efficiency of GEE estimators even under the naive assumption of independence is remarkably high (Liang and Zeger, 1986). However, when there is interest in the effects of non-stochastic time-varying covariates and/or the repeated measures are highly correlated, then suboptimal choices of weights (i.e., misspecification of the covariance) can result in a discernible loss of efficiency. Similarly, with inherently unbalanced longitudinal data (e.g., widely varying n_i), suboptimal choices of weights can also result in loss of efficiency. In general, more efficient estimators of the marginal regression parameters can be obtained by specifying and estimating the covariance among the repeated measures; however, the potential gains in efficiency will depend in a subtle way on both the magnitude of the correlation and the covariate design.

As mentioned earlier, the major impetus for the GEE approach was that it provided a convenient alternative to maximum likelihood estimation of marginal models. For the

case of continuous outcomes, likelihood-based methods are widely used for the analysis of longitudinal data, for example, general linear models and linear mixed-effects models. In general, likelihood-based estimation of marginal models for discrete longitudinal data has proven to be very challenging, in large part because there is no simple unified joint likelihood for discrete longitudinal data. Unlike the multivariate normal distribution, which is completely specified by the first two moments of the outcome vector, for discrete longitudinal data complete specification of the joint distributions requires specifying third- and higher-order moments. Although various likelihood approaches have been proposed — for example, models based on second- and higher-order correlations (Bahadur, 1961; Zhao and Prentice, 1990) and models based on second- and higher-order odds ratios (McCullagh and Nelder, 1989; Lipsitz, Laird, and Harrington, 1990; Liang, Zeger, and Qaqish, 1992; Becker and Balagtas, 1993; Fitzmaurice, Laird, and Rotnitzky, 1993; Molenberghs and Lesaffre, 1994; Lang and Agresti, 1994; Glonek and McCullagh, 1995; Bergsma and Rudas, 2002) — none of these likelihood-based models have proven to be of real practical use except in relatively limited settings. As the number of repeated measures increases, the number of parameters that need to be specified and estimated proliferates rapidly for any of these joint distributions, and a solution to the likelihood equations quickly becomes intractable. Thus, in general, full likelihood approaches require the specification of too many nuisance parameters, are complicated algebraically, and ML estimation can be computationally prohibitive. Furthermore, to obtain unbiased estimators of the marginal regression parameters, the full joint distribution of the data must, in general, be correctly specified. In contrast, with GEE methods, only the first moment needs to be correctly specified in order to obtain unbiased estimators; moreover, the GEE approach is not computationally demanding.

We note that Heagerty (1999) and Heagerty and Zeger (2000) have recently developed a likelihood-based approach that combines the versatility of GLMMs for modeling the within-subject association with a marginal regression model for the mean response. They refer to these models as *marginalized* random-effects models; they are closely related to other approaches that formulate conditional models subject to marginal specification (e.g., Fitzmaurice and Laird, 1993; Azzalini, 1994; Heagerty, 2002; see also Wang and Louis, 2003). Recall that in the standard GLMM, the marginal means obtained by integrating over the random effects, in general, no longer follow a generalized linear model, due to the non-linearity of the link function typically adopted in regression models for discrete responses. In contrast, the *marginalized* random-effects model is specifically formulated such that the marginal mean follows a generalized linear model. Estimation for these *marginalized* random-effects models is as computationally demanding as for GLMMs, and ML estimation is, so far, limited to relatively low-dimensional random-effects distributions. However, these models appear to have the potential to overcome many of the limitations of previously proposed likelihood-based methods and to provide a likelihood-based alternative to GEE; further research is needed.

Finally, although an appealing feature of the GEE approach is its robustness to misspecification of the within-subject association, there are settings where it can be appealing to model the covariance. To date, the implementations of GEE in standard statistical software packages provide only very limited options for modeling the covariance. In particular, there are few choices of models for the "working covariance" when the data are highly unbalanced and irregularly spaced in time. This is in contrast to models for continuous responses (e.g., general linear models and linear mixed-effects models), where there are a broad class of models for the covariance. Future work is needed in both the formulation and implementation of flexible models for the working covariance in GEE methods.

Acknowledgments

This work was supported by grants AI 60373, GM 29745, and MH 54693 from the U.S. National Institutes of Health.

References

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767–775.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (ed.), *Studies in Item Analysis and Prediction*, pp. 158–168. Palo Alto, CA: Stanford University Press.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973.

Barnhart H. X. and Williamson J. M. (1998). Goodness-of-fit tests for GEE modeling with binary responses. *Biometrics* **54**, 720–729.

Becker, M. P. and Balagtas, C. C. (1993). Marginal modeling of binary cross-over data. *Biometrics* **49**, 997–1009.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics*. New York: Wiley.

Bergsma, W. P. and Rudas, T. (2002). Marginal models for categorical data. *Annals of Statistics*, **30**, 140–159.

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

Boos, D. D.(1992). On generalized score tests. *American Statistician* **46**, 327–333.

Carey, V., Zeger, S. L., and Diggle, P. J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80**, 517–526.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.

Cox, D. R. (1961). Tests of separate families of hypotheses. In J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematics, Probability and Statistics*, Vol. 1, pp. 105–123. Berkeley: University of California Press.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*, 2nd ed. London: Chapman & Hall.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* **82**, 407–410.

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

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statistical Science* **8**, 248–309.

Fitzmaurice, G. M. and Lipsitz, S. R. (1995). A model for binary time series data with serial odds ratio patterns. *Applied Statistics* **44**, 51–61.

Fitzmaurice G. M., Lipsitz S. R., and Molenberghs G. (2001). Bias in estimating association parameters for longitudinal binary responses with drop-outs. *Biometrics* **57**, 15–21.

Gange S. J., Linton, K. L., Scott, A. J., DeMets, D. L., and Klein, R. (1995). A comparison of methods for correlated ordinal measures with ophthalmic applications. *Statistics in Medicine* **14**, 1961–1974.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* **57**, 533–546.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics* **15**, 489–504.

Hall, D. B. and Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association* **93**, 1365–1375.

Hauck, W. W. and Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis, *Journal of the American Statistical Association* **72**, 851–853.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55**, 688–698.

Heagerty, P. J. (2002). Margininalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* **91**, 1024–1036.

Heagerty, P. J. and Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science* **15**, 1–26.

Horton, N. J., Bebchuk, J. D., Jones, C. L., Lipsitz, S. R., Catalano, P., and Fitzmaurice, G. M. (1999). Goodness of fit for GEE: An example with quality of life and prostate cancer. *Statistics in Medicine* **18**, 213–222.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd ed. New York: Wiley.

Huber, P. J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In L. LeCam and J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 221–233. Berkeley: University of California Press.

Kenward, M., Lesaffre, E., and Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945–953.

Koch, G. G. and Reinfurt, D. W. (1971). The analysis of categorical data from mixed models. *Biometrics* **27**, 57–173.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* **33**, 133–158.

Kuk, A. Y. C. (2004). Permutation invariance of alternating logistic regression for multivariate binary data. *Biometrika*, **91**, 758–761.

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

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lang, J. B. and Agresti, A. (1994). Simultaneous modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* **89**, 625–632.

Lee, H., Laird, N. M., and Johnston, G. (1999). Combining GEE and REML for estimation of generalized linear models with incomplete multivariate data. Unpublished manuscript.

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). *Journal of the Royal Statistical Society, Series B* **54**, 3–40.

Lin, D. Y., Wei, L. J., and Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

Lipshultz, S. E., Easley, K. A., Orav, E. J., Kaplan, S., Starc, T. J., Bricker, J. T., Lai, W.W., Moodie, D. S., McIntosh, K., Schluchter, M. D., and Colan, S. D. (1998). Left ventricular structure and function in children infected with human immunodeficiency virus: The Prospective P2C2 HIV Multicenter Study. Pediatric Pulmonary and Cardiac Complications of Vertically Transmitted HIV Infection (P2C2 HIV) Study Group. *Circulation* **97**, 1246–1256.

Lipshultz, S. E., Easley, K. A., Orav, E. J., Kaplan, S., Starc, T. J., Bricker, J. T., Lai, W.W., Moodie, D. S., Sopko, G., and Colan, S. D. (2000). Cardiac dysfunction and mortality in HIV-infected children: The Prospective P2C2 HIV Multicenter Study. Pediatric Pulmonary and Cardiac Complications of Vertically Transmitted HIV Infection (P2C2 HIV) Study Group. *Circulation* **102**, 1542–1548.

Lipshultz, S. E., Easley, K. A., Orav, E. J., Kaplan, S., Starc, T. J., Bricker, J. T., Lai, W.W., Moodie, D. S., Sopko, G., Schluchter, M. D., and Colan, S. D. (2002). Cardiovascular status of infants and children of women infected with HIV-1 (P^2C^2 HIV): A cohort study. *Lancet* **360**, 368–373.

Lipsitz, S. R. and Fitzmaurice, G. M. (1996). Estimating equations for measures of association between repeated binary responses. *Biometrics* **52**, 903–912.

Lipsitz, S. R., Ibrahim, J. G., and Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association* **94**, 1147–1160.

Lipsitz, S. R., Kim, K., and Zhao, L. P. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* **13**, 1149–1163.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1990). Maximum likelihood regression methods for paired binary data. *Statistics in Medicine* **9**, 1517–1525.

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 assocition. *Biometrika* **78**, 153–160.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1992). A three-stage estimator for studies with repeated and possibly missing binary outcomes. *Applied Statistics* **41**, 203–213.

Lipsitz, S. R., Fitzmaurice, G. M., Orav, E. J., and Laird, N. M. (1994). Performance of generalized estimating equations in practical situations. *Biometrics* **50**, 270–278.

Lipsitz S. R., Molenberghs G., Fitzmaurice G. M., and Ibrahim J. G. (2000). GEE with Gaussian estimation of the correlations when data are incomplete. *Biometrics* **56**, 528–536.

Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics* **52**, 354–361.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* **23**, 2937–2960.

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57**, 126–134.

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

Miller, M. E., Davis, C. S., and Landis, J. R. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics* **49**, 1033–1044.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.

Mountford, W. K., Lipsitz, S. R., Lackland, D., Fitzmaurice, G. M., and Carter, R. E. (2007). Estimating the variance of estimated trends in proportions when there is no unique subject identifier. *Journal of the Royal Statistical Society, Series A* **170**, 185–193.

Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics — Simulation and Computation* **17**, 1155–1171.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of American Statistical Association* **92**, 1320–1329.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics — Simulation and Computation* **23**, 939–951.

Pregibon, D. (1981). Logistic regression diagnostics, *Annals of Statistics* **9**, 705–724.

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

Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* **83**, 551–562.

Qu, A., Lindsay, B. G., and Li, B. (2000) Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Qu, Y., Williams, G. W., Beck, G. J., and Medendorp, S. V. (1992). Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics* **48**, 1095–1102.

Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the 1999 Joint Statistical Meetings*.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

Robins, J. M. and Rotnitzky, A. (2001). Comment on "Inference for semiparametric models: Some questions and an answer," by P. J. Bickel and J. Kwon. *Statistica Sinica* **11**, 920–936.

Rotnitzky A. and Jewell N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* **54**, 221–226.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94**, 1096–1120 (with rejoinder, 1135–1146).

Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika* **67**, 250–251.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.

Wang, Y. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* **90**, 29–41.

Wang, Y. and Carey, V. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *Journal of the American Statistical Association* **99**, 845–853.

Wang, Z. and Louis, T. A. (2003). Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika* **90**, 765–775.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

White, H. (1982). Maximum likelihood estimation under mis-specified models. *Econometrica* **50**, 1–26.

Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhyā, Series B* **25**, 407–426.

Ye, H. and Pan, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* **93**, 927–941.

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.