Published in final edited form as:

Stat Med. 2012 November 20; 31(26): 3211–3222. doi:10.1002/sim.5394.

A latent-variable marginal method for multi-level incomplete binary data

Baojiang Chen^{a,*} and Xiao-Hua Zhou^b

^aDepartment of Biostatistics, University of Nebraska Medical Center, Omaha, NE 68198, U.S.A

^bDepartment of Biostatistics, University of Washington, Seattle, WA 98195, U.S.A

Abstract

Incomplete multi-level data arise commonly in many clinical trials and observational studies. Because of multi-level variations in this type of data, appropriate data analysis should take these variations into account. A random effects model can allow for the multi-level variations by assuming random effects at each level, but the computation is intensive because high-dimensional integrations are often involved in fitting models. Marginal methods such as the inverse probability weighted generalized estimating equations can involve simple estimation computation, but it is hard to specify the working correlation matrix for multi-level data. In this paper, we introduce a latent variable method to deal with incomplete multi-level data when the missing mechanism is missing at random, which fills the gap between the random effects model and marginal models. Latent variable models are built for both the response and missing data processes to incorporate the variations that arise at each level. Simulation studies demonstrate that this method performs well in various situations. We apply the proposed method to an Alzheimer's disease study.

Keywords

estimating equation; latent variable; missing at random; missing response; multi-level

1. Introduction

Multi-level data arise commonly in many clinical trials and observational studies. For example, in an Alzheimer's disease (AD) study introduced in Section 2, there are four levels: multiple responses, nested within multiple visits, nested within subjects, which, in turn, are nested within AD center (ADC). Multilevel variations arise in this kind of data, and appropriate analysis should take this correlation into account. To deal with multi-level data, common methods include marginal methods and mixed effects models (random effects models).

Marginal methods have seen widespread use in modeling longitudinal data and involve specification of covariate effects on marginal, or 'population averaged', attributes via a specified link function. Liang and Zeger [1] proposed a class of first-order generalized estimating equations (GEEs) for longitudinal data, and Prentice [2] and Zhao and Prentice [3] developed second-order GEEs (GEE2) that increase efficiency and facilitate modeling covariate effects on parameters characterizing associations. Further references about

marginal methods include [4] and [5]. Although inferences based on semiparametric models are attractive because they do not require full model specification, the resulting estimates are not fully efficient (e.g., [6]) when the working correlation matrix is misspecified. In practice, especially in the multi-level clustering data, it is hard and tedious to specify a complicated working correlation matrix or odds ratio matrix [7].

Mixed effects models (e.g., [8–12]) offer an alternative framework for modeling the association between responses over time. With multi-level data, the mixed effects model is appealing in that it allows more variation in the response variable through the assumption that cluster-specific effects are random. In this setting, covariate effects are specified conditionally on a latent subject-specific random effect, and inferences are usually based on the marginal (joint) distribution of responses over time, obtained by marginalizing over the random effect. There are often numerical challenges associated with the computation of the marginal likelihood because higher dimensional integrations are often involved.

Qu *et al.* [13] has developed a latent variable model for correlated binary data. Apparent advantages of the latent variable approach are that it does not restrict the correlation parameter space as does the direct modeling of Liang and Zeger [1] and that it fills the gap between the marginal method and mixed effects models where it cannot only deal with the multi-level data as mixed effects models do, but also have relatively simple estimation procedures. Tan *et al.* [14] showed the robustness of the latent variable model to the misspecification of the latent distribution.

The literature on methods for multi-level incomplete data has primarily focused on longitudinal data via marginal methods (e.g., [15–17]) and multi-level data via mixed effects models (e.g., [18–22]). For marginal methods, the computation is simple, and a population-based interpretation for parameters is desirable; but the resultant estimate can be inefficient. For mixed effects models, the estimate is more efficient when the model is correctly specified; but the computation is intensive, and the interpretation for parameters often conditions on random effects, which is often not desirable. Little work has been devoted to the marginal method for incomplete multi-level data. Yi and Cook [7] developed a marginal method for incomplete clustered binary longitudinal data through specification of odds ratio models within cluster/subject correlation for both the response and missing indicator processes. This method is appealing because the parameter space is no longer constrained, but it is hard to extend this method to multi-level data.

In this paper, we develop a method to deal with multi-level binary data with incomplete responses by using the idea of latent variable models. We build latent variable models for both the response and missing indicator processes; thus, we include the clustering variations in the estimation and inferences. We employ inverse probability weighted GEEs (IPWGEEs; [15]) and second-order estimating equations [2,3] in estimation and inference procedures. A major strength of this approach is that it can handle general types of correlated binary data including clustering, longitudinal data with monotone missingness, and longitudinal data with intermittent missingness. Under the missing-at-random (MAR) mechanism [23], this method not only yields consistent estimators for parameters of the marginal mean but also yields more efficient estimators than the method that ignores the clustering variations in the missing data model. Furthermore, marginal models lead to a population-based interpretation of the parameters, which is more desirable.

We organize the remainder of this paper as follows. In Section 2, we introduce a dementia study. In Section 3, we introduce some notation and models. In Section 4, we give the forms of the estimating equations and provide details on estimation and inference. We provide

numerical studies in Section 5 and analyze data arising from an AD in Section 6. We make concluding remarks in Section 7.

2. An Alzheimer's disease study

There are approximately 5 million people in the USA and more than 37 million people worldwide affected by AD. With AD, a person's memory and ability to learn and carry out daily activities such as talking and eating are gradually destroyed. As the disease progresses, individuals may also experience changes in personality and behavior. Unfortunately, there are no cures for AD, and there is no way to predict how fast someone will progress through the stages of the disease. However, early Alzheimer's diagnosis and treatment can slow the progression of Alzheimer's symptoms. So, it is desirable to identify risk factors that affect the occurrence of the disease.

To study AD, the National Institute on Aging funded 32 ADCs, and the National Alzheimer's Coordinating Center (NACC) is responsible for developing and maintaining a database of participant information collected from these ADCs. A standardized clinical data set, called the Uniform Data Set (UDS), was determined and defined. The goal of the UDS is to provide ADC researchers a standard set of assessment procedures, collected longitudinally, to better characterize ADC participants with mild AD and mild cognitive impairment in comparison with nondemented controls [24]. The UDS was implemented from September 2005 as initial visits, and follow-up visits based on Clinical Task Force definitions were developed annually.

In practice, interest often lies in identifying risk factors that affect both the occurrence of dementia and depression status. As generally known, these two responses, the AD and depression status, are correlated, and appropriate analysis should take this correlation into account. Thus, these data have four levels: center, patients, visits, and correlated responses. Data analysis for this example demonstrated that an ordinary PC cannot handle such intensive computation for the mixed effects models (it always gives the buffering errors), and it is desirable to develop other feasible approaches.

Another challenge of AD studies is the missing data problem. There are 9242 subjects from 32 ADCs included at the entry of this study. Follow-up visits for subjects are scheduled at approximately 1-year intervals, with up to four clinical visits at present. During this study, patients may choose when they want to visit clinics for clinical examination according to their previous observed disease status, for example. This creates the incomplete data problem with missing-at-random (not missing-not-at-random) mechanism. There are 6722 subjects with complete data observed. The missing proportion is about 29.1%. An easily implemented method, such as the complete case analysis, generally gives biased results.

3. Notation and model formulation

3.1. Response models

Suppose that there are q levels of subunits, and let $Y_{i1}, \dots, Y_{iq}$ denote the binary outcome for a subunit in the first level unit i , the second level unit $i_2, \dots$, and the q th level unit i_q , $i = 1, \dots, n$, $i_2 = 1, \dots, n_2, \dots$, $i_q = 1, \dots, n_q$. Let $Y_i = (Y_{i1}, \dots, Y_{iq})'$, $i_2 = 1, \dots, n_2, \dots$, $i_q = 1, \dots, n_q$ denote the binary outcome vector in the first level unit i , $i = 1, \dots, n$. Here, for simplicity, we rewrite the vector Y_i as $Y_i = (Y_{i1}, \dots, Y_{ij})'$, where Y_{ij} is the j th sub-cycle observation in unit i . For example, in a cohort study, if there are n clusters, each cluster has three patients, and each patient has two observations. We could denote Y_{i1} to be the first observation for the first patient in the i th cluster and Y_{i6} to be the second observation for the third patient in

the i th cluster, $i = 1, \dots, n$. Similarly, let $X_i = (X'_{i1}, \dots, X'_{ij})'$ be the corresponding covariate vector for cluster i , where X_{ij} is the covariate vector for sub-cycle unit j .

Define $\pi_{ij} = E(Y_{ij} | X_i)$, and let $\pi_i = (\pi_{i1}, \dots, \pi_{ij})'$. Provided that the mean structure of Y_{ij} depends on the covariate vector for subunit j (e.g., [25]), we may consider the model for the mean through the form

$$g(\pi_{ij}) = X'_{ij}\beta \quad (1)$$

for $j = 1, \dots, J_i$, $i = 1, \dots, n$, where β is a vector of regression parameters and g is a monotone link function.

Following the work of Qu *et al.* [13], we assume that the observed binary vector Y_i is a manifestation of a latent multivariate normal vector $\eta_i = (\eta_{i1}, \dots, \eta_{ij})'$ with mean $\mu_i = (\mu_{i1}, \dots, \mu_{ij})'$ and covariance matrix $\Omega_i = [\rho_{ijk}]_{J_i \times J_i}$. We assume the marginal distribution for η_{ij} to be $N(\mu_{ij}, 1)$, so Ω_i is a correlation matrix. Following the definition of a standardized threshold model [26], we define the binary response Y_{ij} as 1 if $\eta_{ij} \geq 0$ and 0 otherwise. The latent variable η_{ij} is often expressed as

$$\eta_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} \sim N(0, 1) \quad (2)$$

and the marginal mean is

$$\pi_{ij} = P(\eta_{ij} \geq 0) = \Phi(\mu_{ij}).$$

Hence, the marginal mean $\mu_{ij} = \Phi^{-1}(g^{-1}(X'_{ij}\beta))$. The joint probability $\pi_{ijk} = P(Y_{ij} = 1, Y_{ik} = 1)$ can be given by

$$\pi_{ijk} = P(\eta_{ij} \geq 0, \eta_{ik} \geq 0) = \int_{-\infty}^{\mu_{ij}} \int_{-\infty}^{\mu_{ik}} \varphi_2(u, v, \rho_{ijk}) du dv, \quad (3)$$

where $\varphi_2(u, v, \rho_{ijk})$ is the standard bivariate normal density and ρ_{ijk} is the *tetrachoric correlation* between Y_{ij} and Y_{ik} . The (j, k) th element of the covariance matrix V_i of Y_i is given by $v_{ijk} = \pi_{ijk} - \pi_{ij}\pi_{ik}$.

To help study the variation at each subunit level, we now introduce clustering variables, which are integer valued, following the same notation by Qu *et al.* [13]. We partition the data set into n units or clusters by an equivalence E_1 , which is denoted by indicator I^1 . We may partition each unit into subunits by equivalence E_2 , which is denoted by indicator I^2 . Similarly, we can define the s level subunits by equivalence E_s and denote by indicator I^s , $s = 3, \dots, q$. Two observations Y_{ij} and Y_{ik} belong to the same group under if and only if $I^s_{ij} = I^s_{ik}$ ($s = 1, \dots, q$).

To identify the source of clustering variations for each outcome, the random component ε_{ij} in (2) is assumed to consist of

$$\varepsilon_{ij} = \sum_{s=1}^{q-1} \delta_{is} + e_{ij},$$

[13], where δ_{is} is the random effect under E_s , $\delta_{is} \sim N(0, \sigma_s^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, $\text{cov}(e_{ij}, \delta_{is}) = 0$ for $s = 1, \dots, q-1$, and $\text{cov}(e_{ij}, e_{ik}) = 0$ for $j \neq k$. Thus, the correlation between η_{ij} and η_{ik} is

$$\rho_{ijk} = \sum_{s=1}^{q-1} I\{l_{ij}^s = l_{ik}^s\} \sigma_s^2. \quad (4)$$

Note that $\sum_{s=1}^{q-1} \sigma_s^2 + \sigma_e^2 = 1$. To ensure that all values of σ_s^2 and σ_e^2 are between 0 and 1, we build a multinomial model

$$\log \left(\frac{\sigma_s^2}{\sigma_e^2} \right) = z_s' \phi_s$$

for $s = 1, \dots, q-1$, where z_s is the covariate vector under E_s . By varying z_s , we may obtain different models for σ_s^2 . For example, by letting $z_s = 1$, the variance σ_s^2 does not depend on covariates under each level of clusters. Let $\phi = (\phi_1', \dots, \phi_{q-1}')'$ and $\theta = (\beta', \phi')'$.

3.2. Missing data models

To indicate the availability of the response Y_{ij} , we let $R_{ij} = 1$ if Y_{ij} is observed and 0 if Y_{ij} is missing. Denote $R_i = (R_{i1}, \dots, R_{iJ_i})'$. Let $\tilde{\pi}_{ij} = P(R_{ij} = 1 | Y_i, X_i)$ denote the conditional probability that the response Y_{ij} is observed. Typically, a generalized linear model can be modeled as

$$h(\tilde{\pi}_{ij}) = \omega_{ij}' \alpha, \quad (5)$$

where h is a link function, ω_{ij} is a subset of $\{Y_{ij}, X_{ij}\}$, and α is the corresponding regression coefficient vector. Here, we also make the implicit assumption that $P(R_{ij} = 1 | Y_i, X_i) = P(R_{ij} = 1 | Y_{ij}, X_{ij})$ as is commonly carried out (e.g., [25]).

Similarly, we also assume that the binary vector R_i is a manifestation of a latent multivariate normal vector $\tilde{\eta}_i = (\tilde{\eta}_{i1}, \dots, \tilde{\eta}_{iJ_i})'$ with mean $\tilde{\mu}_i = (\tilde{\mu}_{i1}, \dots, \tilde{\mu}_{iJ_i})'$ and covariance matrix $\tilde{\Omega}_i = [\tilde{\rho}_{ijk}]_{J_i \times J_i}$ and that the binary response R_{ij} is defined as 1 if $\tilde{\eta}_{ij} \geq 0$ and 0 otherwise. We also express the latent variable $\tilde{\eta}_{ij}$ as

$$\tilde{\eta}_{ij} = \tilde{\mu}_{ij} + \tilde{\varepsilon}_{ij}, \quad \text{where } \tilde{\varepsilon}_{ij} \sim N(0, 1), \quad (6)$$

and hence, the marginal mean gives the following form:

$$\tilde{\pi}_{ij} = P(\tilde{\eta}_{ij} \geq 0) = \Phi(\tilde{\mu}_{ij})$$

and $\tilde{\mu}_{ij} = \Phi^{-1}(g^{-1}(\omega'_{ij}\alpha))$. The joint probability $\tilde{\pi}_{ijk} = P(R_{ij} = 1, R_{ik} = 1)$ can be given by

$$\tilde{\pi}_{ijk} = P(\tilde{\eta}_{ij} \geq 0, \tilde{\eta}_{ik} \geq 0) = \int_{-\infty}^{\tilde{\mu}_{ij}} \int_{-\infty}^{\tilde{\mu}_{ik}} \varphi_2(u, v, \tilde{\rho}_{ijk}) du dv, \quad (7)$$

where $\tilde{\rho}_{ijk}$ is the *tetrachoric correlation* between R_{ij} and R_{ik} . The (j, k) th element of the covariance matrix $\tilde{V}_i$ of R_i is given by $\tilde{v}_{ijk} = \tilde{\pi}_{ijk} - \tilde{\pi}_{ij}\tilde{\pi}_{ik}$.

We assume the random component $\tilde{e}_{ij}$ in (6) to consist of q random effects; that is,

$$\tilde{e}_{ij} = \sum_{s=1}^{q-1} \tilde{\delta}_{is} + \tilde{e}_{ij},$$

where $\tilde{\delta}_{is}$ is the random effect under E_s , $\tilde{\delta}_{is} \sim N(0, \tilde{\sigma}_s^2)$, $\tilde{e}_{ij} \sim N(0, \tilde{\sigma}_e^2)$, $\text{cov}(\tilde{e}_{ij}, \tilde{\delta}_{is}) = 0$ for $s = 1, \dots, q-1$, and $\text{cov}(\tilde{e}_{ij}, \tilde{e}_{ik}) = 0$ for $j \neq k$. Thus, the correlation between $\tilde{\eta}_{ij}$ and $\tilde{\eta}_{ik}$ is

$$\tilde{\rho}_{ijk} = \sum_{s=1}^{q-1} I\{l_{ij}^s = l_{ik}^s\} \tilde{\sigma}_s^2. \quad (8)$$

Because $\sum_{s=1}^{q-1} \tilde{\sigma}_s^2 + \tilde{\sigma}_e^2 = 1$, we may similarly build a multinomial model

$$\log\left(\frac{\tilde{\sigma}_s^2}{\tilde{\sigma}_e^2}\right) = \tilde{z}_s' \tilde{\phi}_s$$

for $s = 1, \dots, q-1$, where $\tilde{z}_s$ is the covariate vector under E_s . By varying $\tilde{z}_s$, we may obtain different models for $\tilde{\sigma}_s^2$. For example, by letting $\tilde{z}_s = 1$, the variance $\tilde{\sigma}_s^2$ does not depend on covariates under each level of clusters. Let $\tilde{\phi} = (\tilde{\phi}_1', \dots, \tilde{\phi}_{q-1}')'$ and $\gamma = (\alpha', \tilde{\phi}')'$.

In this paper, we assume a missing-at-random mechanism [23], where we require

$$P(R_{ij} = 1 | Y_i, X_i) = P(R_{ij} = 1 | Y_i^{(o)}, X_i),$$

and $\tilde{\sigma}_s^2$ does not depend on the unobserved data $Y_i^{(m)}$, where $Y_i = (Y_i^{(o)}, Y_i^{(m)})'$ and $Y_i^{(o)}$ and $Y_i^{(m)}$ denote the observed and missing parts of the vector Y_i , respectively. Here, we consider an intermittent missing data pattern although it can apply to monotone missing patterns in a similar spirit. For intermittently MAR data, it is often convenient to adopt a further assumption that the missing data indicators at time j depend only on the previously observed outcomes and covariates [27].

4. Estimation and inference

In practice, the interest often lies in estimation of θ , and γ is considered as a nuisance parameter vector.

4.1. Estimating equations for response parameters

To obtain a consistent estimator for the mean parameter β , we may follow the IPWGEEs approach of Robins *et al.* [15] to solve estimating equations

$$\sum_{i=1}^n U_{1i}(\theta, \gamma) = 0, \quad (9)$$

where $U_{1i}(\theta, \gamma) = D_i \Delta_i(\gamma) V_i^{-1} (Y_i - \pi_i)$ with $D_i = \pi_i' / \beta'$ being a $p \times J_i$ derivative matrix and

$$\Delta_i(\gamma) = \text{diag} \left(\frac{I(R_{ij}=1)}{\tilde{\pi}_{ij}}, \quad j=1, \dots, J_i \right).$$

Let $C_i = (c_{ijk}, j < k)'$ be a vector containing up diagonal elements of the empirical covariance matrix whose (j, k) th element is given by $c_{ijk} = (Y_{ij} - \pi_{ij})(Y_{ik} - \pi_{ik})$, and let $v_i = (v_{ijk}, j < k)$ be a vector containing up diagonal elements of V_i . To obtain a consistent estimator of the random effects parameter ϕ , we may employ the second-order estimating equations (e.g., [2])

$$\sum_{i=1}^n U_{2i}(\theta, \gamma) = 0, \quad (10)$$

where $U_{2i}(\theta, \gamma) = E_i \Delta_i^*(\gamma) W_i^{-1} (C_i - v_i)$, $E_i = v_i' / \phi'$,

$$\Delta_i^*(\gamma) = \text{diag} \left(\frac{I(R_{ij}=1, R_{ik}=1)}{\tilde{\pi}_{ijk}}, \quad j < k \right),$$

and W_i is a working covariance matrix for C_i . Following Qu *et al.* [13], $W_i = \text{diag}(w_{ijk}, j < k)$ with

$$w_{ijk} = \text{var}(c_{ijk}) = (1 - 2\pi_{ij})(1 - 2\pi_{ik})v_{ijk} + v_{ijj}v_{ikk} - v_{ijk}^2.$$

Let $U_i(\theta, \gamma) = (U_{1i}'(\theta, \gamma), U_{2i}'(\theta, \gamma))'$. When γ is specified, we can solve $\sum_{i=1}^n U_i(\theta, \gamma) = 0$ for the estimate $\hat{\theta}$. In practice, γ is not available, and we must get an estimate of it. The following section introduces the estimation procedure for the missing data model parameters.

4.2. Estimation for missing data parameters

Let $\tilde{D}_i = \tilde{\pi}_i' / \alpha'$. Then the estimating functions for α are $\sum_{i=1}^n \tilde{U}_{1i}(\gamma)$, where $\tilde{U}_{1i} = \tilde{D}_i \tilde{V}_i^{-1} (R_i - \tilde{\pi}_i)$.

Let $\tilde{C}_i = (\tilde{c}_{ijk}, j < k)'$ be a vector containing up diagonal elements of the empirical covariance matrix whose (j, k) th element is given by $\tilde{c}_{ijk} = (R_{ij} - \tilde{\pi}_{ij})(R_{ik} - \tilde{\pi}_{ik})$, and let $\tilde{v}_i = (\tilde{v}_{ijk}, j < k)$ be a vector containing up diagonal elements of $\tilde{V}_i$. To construct estimating

equations, we let $\tilde{E}_i = \tilde{\gamma}_i' \tilde{\phi}'$ and let $\tilde{W}_i = \text{diag}(\tilde{w}_{ijk}, j < k)$ be a working covariance matrix for $\tilde{C}_i$, where

$$\tilde{w}_{ijk} = \text{var}(\tilde{c}_{ijk}) = (1 - 2\tilde{\pi}_{ij})(1 - 2\tilde{\pi}_{ik})\tilde{v}_{ijk} + \tilde{v}_{ijj}\tilde{v}_{ikk} - \tilde{v}_{ijk}^2.$$

Then the estimating functions for $\tilde{\phi}$ are given by $\sum_{i=1}^n \tilde{U}_{2i}(\gamma)$, where $\tilde{U}_{2i}(\gamma) = \tilde{E}_i \tilde{W}_i^{-1}(\tilde{C}_i - \tilde{\gamma}_i)$, and the estimating equations for γ are given by $\sum_{i=1}^n \tilde{U}_i(\gamma) = 0$, where $\tilde{U}_i(\gamma) = (\tilde{U}_{1i}(\gamma), \tilde{U}_{2i}(\gamma))'$.

4.3. Estimation and inference

In practice, we may employ a two-stage estimation procedure to obtain the estimates. In the first stage, we estimate the missing data parameter γ . Let

$$\tilde{M}_i(\gamma) = \begin{pmatrix} \frac{\partial \tilde{U}_{1i}(\gamma)}{\partial \alpha'} & 0 \\ \frac{\partial \tilde{U}_{2i}(\gamma)}{\partial \alpha'} & \frac{\partial \tilde{U}_{2i}(\gamma)}{\partial \phi} \end{pmatrix}.$$

Given an initial value $\gamma^{(0)}$, we employ the iterative equation

$$\gamma^{(t)} = \gamma^{(t-1)} - \left(\sum_{i=1}^n \tilde{M}_i(\gamma^{(t-1)}) \right)^{-1} \cdot \sum_{i=1}^n \tilde{U}_i(\gamma^{(t-1)}), \quad t=1, \dots$$

to obtain the estimate $\hat{\gamma}$.

In the second stage, we put the estimate $\hat{\gamma}$ into (9) and (10) and estimate the response parameter θ . Let

$$M_i(\theta, \gamma) = \begin{pmatrix} \frac{\partial U_{1i}(\theta, \gamma)}{\partial \beta'} & 0 \\ \frac{\partial U_{2i}(\theta, \gamma)}{\partial \beta'} & \frac{\partial U_{2i}(\theta, \gamma)}{\partial \phi} \end{pmatrix}.$$

Given an initial value $\theta^{(0)}$, the updated value of $\theta^{(t)}$ is

$$\theta^{(t)} = \theta^{(t-1)} - \left(\sum_{i=1}^n M_i(\theta^{(t-1)}, \gamma) \right)^{-1} \cdot \sum_{i=1}^n U_i(\theta^{(t-1)}, \hat{\gamma}), \quad t=1, \dots$$

Denote the limit value to be $\hat{\theta}$.

Here, we comment that, in general, the two-stage iteration estimate differs from the simultaneous estimate [15]. However, the updated values from these two procedures converge to the same limit under mild regularity conditions. Whereas the two-stage

procedure is much easier to use for estimation of the parameter $\hat{\theta}$, the simultaneous estimation method is more useful for developing the asymptotic distribution for $\hat{\theta}$.

To obtain the asymptotic covariance matrix of $\hat{\theta}$, we let $H_i(\theta, \gamma) = (U_i'(\theta, \gamma), \tilde{U}_i'(\gamma))'$. Note that $E[H_i(\theta, \gamma)] = 0$ when the response and missing data models are correctly specified. Hence, by Theorem 3.4 of Newey and McFadden [28], under some standard regularity conditions, there is a unique solution $(\hat{\theta}', \hat{\gamma}')'$ to the equation $\sum_{i=1}^n H_i(\theta, \gamma) = 0$ with probability approaching 1, which satisfies

$$n^{1/2} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\gamma} - \gamma \end{pmatrix} = -\{E[\partial H_i(\theta, \gamma) / \partial(\theta', \gamma')]\}^{-1} \cdot n^{-1/2} \sum_{i=1}^n H_i(\theta, \gamma) + o_p(1).$$

For the estimator $\hat{\theta}$ of central interest, we have

$$n^{1/2}(\hat{\theta} - \theta) = -\Gamma^{-1} n^{-1/2} \cdot \sum_{i=1}^n Q_i(\theta, \gamma) + o_p(1),$$

where $\Gamma = E[U_i(\theta, \gamma) / \partial \theta']$ and

$$Q_i(\theta, \gamma) = U_i(\theta, \gamma) - E[\partial U_i(\theta, \gamma) / \partial \gamma'] \cdot [E(\partial \tilde{U}_i(\gamma) / \partial \gamma')]^{-1} \cdot \tilde{U}_i(\gamma).$$

The central limit theorem then leads to the asymptotic distribution for $n^{1/2}(\hat{\theta} - \theta)$, which is normal with mean 0 and asymptotic variance $\Gamma^{-1} \Sigma [\Gamma^{-1}]'$, where $\Sigma = E[Q_i(\theta, \gamma) Q_i'(\theta, \gamma)]$.

5. Simulation studies

Set $n = 100$. In each cluster, we assume that there are three subjects and that each subject has three assessments. Then $y_i = (y_{i1}, \dots, y_{i9})'$, $i = 1, \dots, n$. We also did simulation results for $n = 50$ (the results are similar) but do not report them here because of the limited space. The response model is given by

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 x_{ij}, \quad (11)$$

where x_{ij} is a time-varying binary covariate generated from $\text{Bin}(1, 0.5)$, associated with response y_{ij} , $j = 1, \dots, 9$. Set $\beta_0 = -0.5$ and $\beta_1 = 1.0$. We assume that the cluster level variance $\sigma_1^2 = 0.3$ and 0 and that the subject level variance $\sigma_2^2 = 0.6$ and 0.

For the missing data model, we consider the following dropout model:

$$\text{logit } \tilde{\pi}_{ij} = \alpha_0 + \alpha_1 x_{ij}^* + \alpha_2 y_{ij}^*, \quad (12)$$

where $\tilde{\pi}_{ij} = P(R_{ij} = 1 | R_{ij}^* = 1, Y_i, X_i)$, y_{ij}^* denotes the previously observed response of y_{ij} , x_{ij}^* denotes the previously observed covariate of x_{ij} , and R_{ij}^* denotes the previously observed

missing indicator. We assume that the cluster level variance $\tilde{\sigma}_1^2=0.3$ and 0.6 . Set $\alpha_0 = -0.5$ and $\alpha_1 = 1.0$, and we change the value of α_2 .

We compare the proposed method with other methods that are often used in practice. The first, called GEE-I, is the proposed method where we ignore the cluster level correlation in the missing data model; that is, we set $\tilde{\sigma}_1^2=0$. The second method, called 'CC', is an analysis of only the subjects with complete data. We report the results in Tables I and II, where bias is the percent relative bias, SD is the standard deviation for 1000 simulations, and CP is the empirical coverage probability for the 95% confidence interval. It is observed that the CC method yields larger biases and poor coverage probabilities, whereas the other two methods yield negligible biases and good coverage probabilities. The proposed method gives more efficient estimates than the GEE-I method when there is clustering level variation ($\tilde{\sigma}_1^2 \neq 0$) in the missing data model, and the missing data mechanism is MAR ($\alpha_2 = 0$). As the variation $\tilde{\sigma}_1^2$ increases, the efficiency gain for the proposed method compared with GEE-I method increases. As the missing proportion increases (α_2 decreases), the efficiency gain increases. Particularly, in some scenarios, the proposed method gives much efficiency improvement compared with the GEE-I method. For example, when $\tilde{\sigma}_1 = 0.3$, $\sigma_1^2=0.3$, $\sigma_2^2=0.6$, and $\alpha_2 = -1$, the variance ratio of the estimates between the proposed methods and the GEE-I method is 0.75 for β_1 ; that is, the variance of the proposed method estimate reduced to about 25% compared with that of the GEE-I method estimate.

6. Application to an Alzheimer's disease study

We apply the proposed method to the NACC UDS. The two responses of interest are the diagnosis of dementia (yes/no) and depression (or dysphoria; *yes* if the total global deterioration scale score is ≥ 5 and *no* otherwise). The potential covariates of interest that may influence the status of dementia and depression include sex, congestive heart failure (CVCHF; yes/no), family history of dementia (FHDEM; yes/no), diabetes (yes/no), hypertension (yes/no), education (years), Mini-mental State Exam (MMSE) score, and age. There were 9242 subjects from 32 ADCs included at the entry of this study. Follow-up visits for subjects were scheduled at approximately 1-year intervals, with up to four clinical visits at present. So, there are four levels in this process: the center, subject, visit, and response levels. However, patients may miss a clinic visit or refuse to undergo a clinic examination during the clinic visit, leading to incomplete responses. We think that the MAR mechanism may be reasonable, because, for example, patients' visit to the clinic may depend on their previous observed diagnosis of disease status: if they were diagnosed with AD at the last visit, they may be likely to attend the clinic visit this time in order to control the disease efficiently.

Consider the following regression model for the response process:

$$\text{logit } \pi_{ij} = X'_{ij}\beta,$$

where X_{ij} is the covariate vector for sub-cycle unit j , which includes the function of sex, CVCHF, FHDEM, diabetes, hypertension, education, MMSE, and age. Here, we consider three levels of random variations: the center, subject, and visit levels. We use σ_1^2 , σ_2^2 , and σ_3^2 to denote the variances of their random effects, respectively. We assume that the two responses for a patient are independent given random effects in the center, subject, and visit levels.

For missing indicators, we build the following regression model:

$$\text{logit } \tilde{\pi}_{ij} = \omega'_{ij} \alpha,$$

where ω_{ij} includes sex, CVCHF, FHDEM, diabetes, hypertension, education, MMSE, age, and the previously observed response. Similarly, we also consider three levels of random variations (the center, subject, and visit levels) in this model and use $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, and $\tilde{\sigma}_3^2$ to denote the variances of the corresponding random effects, respectively.

In line with simulation studies, we also use three methods to analyze the data. The first is the proposed method; the second, called GEE-I, is the proposed method where we ignore the cluster correlations in the missing data model; and the third is the complete case analysis (CC). We report the results in Table III. We can observe that the proposed and GEE-I methods give similar results, but they have different results from the complete case analysis results. This makes sense because the missing completely at random is not feasible (Table IV). The proposed method estimate is more efficient than the GEE-I method estimate. All three methods reveal that CVCHF and education have no significant effect on dementia, that MMSE has a positive effect to protect the onset of dementia, that diabetes and hypertension have positive effects to protect the onset of dementia, and that age has a negative effect on the onset of dementia. For sex and FHDEM, the proposed and GEE-I methods give different results from the CC method. The proposed and GEE-I methods demonstrate that women are less likely to acquire dementia than men, but the complete case analysis indicates that it is not significant. The proposed and GEE-I methods indicate that people have a higher probability to acquire dementia if they have FHDEM, but the complete case analysis indicates that it is not significant.

For risk factors on the impact of depression, all methods reveal that MMSE, hypertension, and education have significant effects on the depression status and that FHDEM has no significant effect on the depression status. For sex, the proposed and GEE-I methods demonstrate that it has no significant effect on the depression status, but the complete case analysis demonstrates that it is significant. The proposed and GEE-I methods indicate that people are easily depressed when they have CVCHF disease or diabetes, but the complete case analysis demonstrates that they are not significant. The random effect in the cluster level (σ_1^2) is not significant, whereas random effects in the subject (σ_2^2) and visit levels (σ_3^2) are significant in the proposed method.

For the missing data model, we carry out the generalized score test for the fit of regression models by comparing a model with an expanded model. Here, we only list the results for the final model. Significance of sex, CVCHF, MMSE, FHDEM, diabetes, hypotension, education, age, and the previous observed demential status indicates that the missing-at-random assumption may be plausible: the random effect in the cluster level ($\tilde{\sigma}_1^2$) is not significant, the random effect in the subject level ($\tilde{\sigma}_2^2$) is significant, and the random effect in the visit level ($\tilde{\sigma}_3^2$) is moderately significant.

7. Discussion

In this paper, we introduce a marginal method to deal with incomplete multi-level binary data, which fills the gap between the marginal method and the random effects model. The appealing features of the proposed method are that it can not only handle general types of correlated binary data including clustering, longitudinal data with monotone missingness, and longitudinal data with intermittent missingness, but also have simple computation

procedures. For the missing data process, we build a similar latent variable model to incorporate multi-level variations for the missing data indicators. Simulation studies demonstrate that the proposed method works well in a variety of situations when data are missing at random.

There are alternative popular approaches for dealing with missing response data including multiple imputation [29–31]. Multiple imputation is a sophisticated simulation procedure that may require, in some cases, more complete specification of models, and in the contexts of multi-level data, complete specification of the models are very difficult. In general, misspecification of these models can give invalid inference. For the proposed method, we are able to conduct semiparametric inference when the response is missing at random. Although inverse weighting requires modeling the missing data process, it can be carried out using standard methods of model assessment such as the generalized score test, as well as sensitivity analyses, to provide reassurance to analysts that their inferences are reasonable.

This paper only considers incomplete responses problem. In practice, multi-level clustered data often feature missing responses or/and missing covariates. So, it is desirable to extend the proposed method to deal with both missing responses and missing covariates problem. One may employ the same idea of latent variable models and the idea of Chen *et al.* [32] and Chen and Zhou [33] to build weighted GEEs; the weights are derived from the joint models for the missing response and missing covariate indicators, where additional latent variable model can be built to measure the associations between the missing indicators for the response and covariates. This is a challenging work and still under investigation.

Although we only consider binary data, the proposed method can be extended to ordinal data. Specifically, we can modify the binary logistic regression model to incorporate the ordinal nature of a dependent variable by defining a binary random variable that equals 1 for an event and all events that are ordered before it and 0 otherwise. Then the following steps follow those in Section 3.

Acknowledgments

Dr. Xiao-Hua Zhou, PhD, is presently a Core Investigator and Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA, USA. Dr. Zhou's work was supported in part by the National Science Foundation of China (NSFC 30728019) and the US Department of Veterans Affairs, Veterans Affairs Health Administration, HSR&D grants. Both Drs. Zhou and Chen were supported in part by the National Institute on Aging (grant U01AG016976). This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

References

1. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
2. Prentice RL. Correlated binary regression with covariate specific to each binary observation. *Biometrics*. 1988; 44:1022–1048.
3. Zhao LP, Prentice RL. Correlated binary regression using a quadratic exponential model. *Biometrika*. 1990; 77:642–648.
4. Rajan KB, Leurgans SE. Joint modeling of missing data due to non-participation and death in longitudinal aging studies. *Statistics in Medicine*. 2010; 29:2260–2268. [PubMed: 20645281]
5. Shardell M, Hicks GE, Miller RR, Magaziner J. Semiparametric regression models for repeated measures of mortal cohorts with non-monotone missing outcomes and time-dependent covariates. *Statistics in Medicine*. 2010; 29:2282–2296. [PubMed: 20564729]
6. Fitzmaurice GM, Laird NM, Rotnitzky A. Regression models for discrete longitudinal responses. *Statistical Science*. 1993; 8:284–299.

7. Yi GY, Cook RJ. Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association*. 2002; 97:1071–1080.
8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
9. Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46:673–687. [PubMed: 2242409]
10. Goldstein, H. *Multilevel Statistical Models*. 3. Edward Arnold Publishers, Ltd; London: 2003.
11. Lin H, Liu D, Zhou XH. A correlated random-effects model for normal longitudinal data with nonignorable missingness. *Statistics in Medicine*. 2010; 29:236–247. [PubMed: 19941316]
12. Shen C, Gao S. A mixed-effects model for cognitive decline with non-monotone non-response from a two-phase longitudinal study of dementia. *Statistics in Medicine*. 2007; 26:409–425. [PubMed: 16345034]
13. Qu Y, Williams GW, Beck GJ, Medendorp SV. Latent variable models for clustered Dichotomous data with multiple subclusters. *Biometrics*. 1992; 48:1095–1102.
14. Tan M, Qu Y, Rao JS. Sensitivity of the latent variable model for correlated binary data. *Biometrics*. 1999; 55:25–31.
15. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*. 1995; 90:106–121.
16. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*. 1998; 93:1321–1339.
17. Shardell M, Miller R. Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine*. 2008; 27:1008–1025. [PubMed: 17579923]
18. Horton HJ, Laird NM. Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research*. 1998; 8:37–50. [PubMed: 10347859]
19. Ibrahim JG, Chen MH, Lipsitz SR. Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*. 2001; 88:551–564.
20. Ibrahim JG, Lipsitz SR, Horton N. Using auxiliary data for parameter estimation with nonignorable missing outcomes. *Applied Statistics*. 2001; 50:361–373.
21. Fitzmaurice GM, Molenberghs G, Lipsitz SR. Regression models for longitudinal binary data responses with informative drop-outs. *Journal of the Royal Statistical Society, Series B*. 1995; 57:691–704.
22. Chen Q, Ibrahim JG, Chen M, Senchaudhuri P. Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis*. 2008; 99:1302–1331. [PubMed: 19169388]
23. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2. John Wiley and Sons, Inc; New York: 2002.
24. Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, Hubbard JL, Koepsell TD, Morris JC, Kukull WA. NIA ADCs. The National Alzheimers Coordinating Center (NACC) database: the uniform data set. *Alzheimer Disease and Related Disorders*. 2007; 21:249–258.
25. Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation*. 1994; 23:939–951.
26. Harville DA, Mee RW. A mixed model procedure for analyzing ordered categorical data. *Biometrics*. 1984; 40:393–408.
27. Robins JM, Gill R. Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine*. 1997; 16:39–56. [PubMed: 9004382]
28. Newey, WK.; Mcfadden, D. Large sample estimation and hypothesis testing. In: McFadden, D.; Engler, R., editors. *Handbook of Econometrics*. Vol. 4. Elsevier; Amsterdam, North-Holland: 1994.
29. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons; New York: 1987.

30. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Statistical Methods in Medical Research*. 2007; 16:199–218. [PubMed: 17621468]
31. Carpenter J, Goldstein H. Multiple imputation in MLwiN. *Multilevel Modelling Newsletters*. 2004; 16:9–18.
32. Chen B, Yi GY, Cook RJ. Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *Journal of the American Statistical Association*. 2010; 105:336–353.
33. Chen B, Zhou XH. Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*. 2011; 67:830–842. [PubMed: 21281272]

Table I

Empirical biases, standard errors, and coverage probabilities for the three approaches to estimation and inference with incomplete multi-level clustering data.

$\tilde{\sigma}_1^2$	α_2	Parameter	Proposed method			GEE-I			CC		
			Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
$(\sigma_1^2, \sigma_2^2)=(0.3, 0.6)$											
0.3	-1	β_0	-1.414	0.571	94.4	-1.818	0.612	94.8	31.361	0.179	76.4
0.3	-1	β_1	0.464	0.251	95.1	1.798	0.289	94.8	-0.531	0.189	95.0
0.3	0.5	β_0	-0.800	0.363	94.2	-2.224	0.370	94.1	-16.937	0.163	83.2
0.3	0.5	β_1	1.094	0.176	94.5	1.103	0.177	94.3	-1.286	0.180	93.2
0.3	1.0	β_0	0.931	0.309	94.8	1.144	0.318	94.4	-31.801	0.168	66.6
0.3	1.0	β_1	1.843	0.155	94.6	1.712	0.168	94.5	-2.160	0.181	93.6
0.6	-1	β_0	-3.957	0.577	94.3	1.847	0.618	94.6	31.815	0.180	76.0
0.6	-1	β_1	-2.049	0.272	94.8	5.323	0.293	94.6	-1.383	0.198	92.8
0.6	0.5	β_0	1.532	0.324	94.4	1.866	0.334	94.3	-14.515	0.166	81.8
0.6	0.5	β_1	0.977	0.175	95.1	2.003	0.177	94.2	0.667	0.169	96.2
0.6	1.0	β_0	1.578	0.300	95.0	0.847	0.309	94.7	-30.503	0.165	67.0
0.6	1.0	β_1	0.703	0.147	94.5	1.463	0.158	94.4	-2.997	0.181	93.6
$(\sigma_1^2, \sigma_2^2)=(0.3, 0)$											
0.3	-1	β_0	-1.416	0.376	94.6	1.325	0.409	94.9	14.414	0.166	88.6
0.3	-1	β_1	-1.137	0.272	94.6	1.219	0.273	94.8	1.811	0.194	93.6
0.3	0.5	β_0	-2.214	0.248	94.3	1.460	0.250	94.6	-5.959	0.150	90.0
0.3	0.5	β_1	-0.760	0.179	94.9	1.219	0.184	94.4	-1.331	0.167	96.6
0.3	1.0	β_0	0.920	0.229	94.7	0.475	0.229	94.8	-9.440	0.146	88.0
0.3	1.0	β_1	1.173	0.175	94.6	0.448	0.184	94.3	-1.538	0.171	94.0
0.6	-1	β_0	-0.263	0.388	94.5	0.603	0.399	94.3	9.944	0.172	87.6
0.6	-1	β_1	0.216	0.263	94.0	1.210	0.270	94.6	-0.028	0.195	93.0
0.6	0.5	β_0	-0.233	0.217	95.3	2.210	0.226	95.2	-4.661	0.157	88.4

σ_1^2	α_2	Parameter	Proposed method			GEE-I			CC		
			Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
0.6	0.5	β_1	0.258	0.178	95.2	-0.110	0.188	95.3	-0.799	0.176	94.8
0.6	1.0	β_0	-1.686	0.218	94.6	1.500	0.224	94.5	-10.320	0.148	87.2
0.6	1.0	β_1	-1.003	0.176	95.2	1.673	0.182	94.2	-1.232	0.167	94.8

Table II

Empirical biases, standard errors, and coverage probabilities for the three approaches to estimation and inference with incomplete multi-level clustering data.

$\tilde{\sigma}_1^2$	α_2	Parameter	Proposed method			GEE-I			CC		
			Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
$(\sigma_1^2, \sigma_2^2)=(0, 0.6)$											
0.3	-1	β_0	-3.173	0.316	94.0	4.627	0.358	94.5	19.379	0.149	91.0
0.3	-1	β_1	-0.940	0.271	95.2	2.302	0.288	95.0	0.138	0.189	96.0
0.3	0.5	β_0	0.313	0.191	95.1	1.555	0.193	94.2	-10.970	0.136	91.2
0.3	0.5	β_1	0.822	0.167	94.8	1.409	0.175	94.3	-0.881	0.170	97.0
0.3	1.0	β_0	-1.558	0.181	94.4	-0.697	0.186	94.7	-21.469	0.134	82.4
0.3	1.0	β_1	-0.142	0.158	94.7	-0.344	0.169	94.5	-2.404	0.168	94.6
0.6	-1	β_0	1.849	0.384	95.0	-3.598	0.389	95.6	20.143	0.157	88.0
0.6	-1	β_1	1.530	0.265	94.4	0.484	0.275	94.9	-0.227	0.194	94.4
0.6	0.5	β_0	0.354	0.189	94.9	1.145	0.198	94.5	-10.110	0.136	90.0
0.6	0.5	β_1	0.704	0.168	94.5	0.351	0.178	94.7	0.225	0.171	94.8
0.6	1.0	β_0	-1.257	0.169	95.2	-1.774	0.170	94.4	-21.938	0.136	83.0
0.6	1.0	β_1	0.235	0.170	94.8	0.394	0.173	94.6	-1.732	0.169	95.0
$(\sigma_1^2, \sigma_2^2)=(0, 0)$											
0.3	-1	β_0	1.505	0.244	94.8	-1.938	0.248	94.4	0.567	0.144	95.0
0.3	-1	β_1	2.185	0.292	94.8	1.542	0.299	94.5	0.537	0.193	95.0
0.3	0.5	β_0	0.353	0.159	94.9	1.695	0.159	94.4	2.842	0.126	95.4
0.3	0.5	β_1	0.892	0.203	94.6	0.791	0.210	94.7	2.944	0.170	96.0
0.3	1.0	β_0	-0.531	0.144	94.8	2.162	0.144	94.3	0.295	0.126	95.0
0.3	1.0	β_1	0.356	0.191	94.3	1.867	0.193	94.2	0.632	0.168	95.2
0.6	-1	β_0	1.889	0.221	94.6	-3.762	0.220	94.6	0.021	0.144	94.4
0.6	-1	β_1	2.606	0.359	94.7	-2.255	0.266	94.3	-0.866	0.192	94.8
0.6	0.5	β_0	-1.253	0.142	94.3	0.820	0.142	94.9	0.108	0.130	94.8

σ_1^2	α_2	Parameter	Proposed method			GEE-I			CC		
			Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
0.6	0.5	β_1	-1.370	0.188	94.4	0.405	0.194	94.7	0.536	0.175	93.8
0.6	1.0	β_0	1.162	0.131	95.3	-0.099	0.144	95.2	1.052	0.127	94.2
0.6	1.0	β_1	-0.163	0.187	94.6	0.734	0.190	94.8	2.148	0.171	94.4

Table III

Parameter estimates for the National Alzheimer's Coordinating Center Uniform Data Set: response models.

Parameter	Proposed method			GEE-I			CC		
	Est.	SE	p	Est.	SE	p	Est.	SE	p
Fixed effects									
DEMENTIA									
(Intercept)	1.4293	0.1188	<0.0001	1.5026	0.1200	<0.0001	0.6200	0.1584	0.0001
SEX(F)	-0.1182	0.0252	<0.0001	-0.1174	0.0254	<0.0001	-0.0071	0.0366	0.8453
CVCHF	0.0469	0.0667	0.4823	0.0486	0.0671	0.4672	-0.0113	0.0916	0.9022
MMSE	-0.0713	0.0020	<0.0001	-0.0730	0.0022	<0.0001	-0.0215	0.0012	<0.0001
FHDEM	0.0634	0.0275	0.0211	0.0633	0.0278	0.0215	-0.0671	0.0402	0.0952
DIABETE	-0.1775	0.0375	<0.0001	-0.1792	0.0377	<0.0001	-0.1669	0.0538	0.0019
HYPERT	-0.1772	0.0260	<0.0001	-0.1781	0.0260	<0.0001	-0.2005	0.0384	<0.0001
EDUC	0.0008	0.0015	0.6157	0.0011	0.0016	0.4829	0.0019	0.0019	0.3356
AGE	0.0096	0.0014	<0.0001	0.0091	0.0015	<0.0001	0.0101	0.0020	<0.0001
DEPRESSION									
(Intercept)	0.5732	0.1157	<0.0001	0.5667	0.1160	<0.0001	1.4954	0.1543	<0.0001
SEX(F)	-0.0268	0.0261	0.3044	-0.0272	0.0261	0.2976	0.1564	0.0355	<0.0001
CVCHF	0.1919	0.0665	0.0039	0.2022	0.0668	0.0023	0.0843	0.0897	0.3472
MMSE	-0.0097	0.0010	<0.0001	-0.0091	0.0012	<0.0001	-0.0064	0.0012	<0.0001
FHDEM	0.0260	0.0286	0.3641	0.0249	0.0292	0.3839	-0.0162	0.0391	0.6787
DIABETE	0.1474	0.0384	0.0001	0.1480	0.0388	0.0001	0.0039	0.0532	0.9415
HYPERT	0.1410	0.0271	<0.0001	0.1421	0.0277	<0.0001	0.1085	0.0371	0.0034
EDUC	-0.0071	0.0016	<0.0001	-0.0066	0.0017	<0.0001	-0.0038	0.0019	0.0447
AGE	-0.0175	0.0014	<0.0001	-0.0177	0.0015	<0.0001	-0.0264	0.0019	<0.0001
Random effects									
σ_1^2	0.0599	0.1004	0.5504	0.0713	0.1006	0.4603	0.0728	0.1349	0.5897
σ_2^2	0.5545	0.1152	<0.0001	0.5719	0.1246	<0.0001	0.3667	0.2064	0.0757
σ_3^2	0.1648	0.0688	0.0166	0.1396	0.0665	0.0356	0.2096	0.1265	0.0977

Table IV
Parameter estimates for the National Alzheimer's Coordinating Center Uniform Data Set: missing models

Parameter	DEMENTIA			DEPRESSION		
	Est.	SE	p	Est.	SE	p
Fixed effects						
(Intercept)	-0.1390	0.1133	0.2198	3.4460	0.1470	<0.0001
SEX(F)	-0.4685	0.0256	<0.0001	-0.2231	0.0333	<0.0001
CVCHF	0.1602	0.0691	0.0203	0.0590	0.0866	0.4957
MMSE	-0.0264	0.0010	<0.0001	-0.0240	0.0008	<0.0001
FHDEM	0.1562	0.0279	<0.0001	-0.2118	0.0360	<0.0001
DIABETE	0.1619	0.0389	<0.0001	-0.0881	0.0497	0.0765
HYPERT	0.0240	0.0262	0.3597	0.0992	0.0339	<0.0034
EDUC	-0.0096	0.0015	<0.0001	-0.0007	0.0018	0.7020
AGE	0.0143	0.0014	<0.0001	-0.0134	0.0018	<0.0001
Pre. obs. response	1.0373	0.0700	<0.0001	0.4408	0.0356	<0.0001
Random effects						
σ_1^2	0.0569	0.1082	0.5990			
σ_2^2	0.4971	0.1302	0.0001			
σ_3^2	0.1672	0.0867	0.0537			