Published in final edited form as:
Biometrics. 2011 September ; 67(3): 1119–1126. doi:10.1111/j.1541-0420.2010.01525.x.

A Bivariate Pseudo-Likelihood for Incomplete Longitudinal Binary Data with Nonignorable Non-monotone Missingness

Sanjoy K. Sinha[1,*], **Andrea B. Troxel**[2], **Stuart R. Lipsitz**[3], **Debajyoti Sinha**[4], **Garrett M. Fitzmaurice**[3], **Geert Molenberghs**[5], and **Joseph G. Ibrahim**[6]

[1]School of Mathematics and Statistics, Carleton University, Ottawa, ON, K1S 5B6 Canada

[2]University of Pennsylvania School of Medicine, Philadelphia, PA, U.S.A.

[3]Harvard Medical School, Boston, MA, U.S.A.

[4]Florida State University, Tallahassee, FL, U.S.A.

[5]Hasselt University, Diepenbeek, Belgium

[6]University of North Carolina, Chapel Hill, U.S.A.

Summary

For analyzing longitudinal binary data with nonignorable and non-monotone missing responses, a full likelihood method is complicated algebraically, and often requires intensive computation, especially when there are many follow-up times. As an alternative, a pseudo-likelihood approach has been proposed in the literature under minimal parametric assumptions. This formulation only requires specification of the marginal distributions of the responses and missing data mechanism, and uses an independence working assumption. However, this estimator can be inefficient for estimating both time-varying and time-stationary effects under moderate to strong within-subject associations among repeated responses. In this article, we propose an alternative estimator, based on a bivariate pseudo-likelihood, and demonstrate in simulations that the proposed method can be much more efficient than the previous pseudo-likelihood obtained under the assumption of independence. We illustrate the method using longitudinal data on CD4 counts from two clinical trials of HIV-infected patients.

Keywords

Logistic regression; Longitudinal data; Marginal model; Maximum likelihood; Missing data mechanism

1. Introduction

In many longitudinal studies, individuals are observed repeatedly at a fixed number of time points. For example, longitudinal data are often collected in AIDS, cardiovascular, and cancer clinical trials as well as in observational studies. Here we focus on the case where the response variable over time is binary, and we are interested in modeling the marginal means or success probabilities. There is an extensive statistical literature on methods for the analysis of longitudinal binary data (e.g., Cox, 1972; Liang & Zeger, 1986; Le Cessie & Van Houwelingen, 1994; Molenberghs & Lesaffre, 1994; Meester & MacKay, 1994; and many others). The modeling of longitudinal binary data is often complicated by the fact that the

*sinha@math.carleton.ca.

outcome variable is not always observed at all assessment times. This missingness often depends on the unobserved value of the outcome at that time. In such cases, the missing data mechanism is referred to as nonignorable (Little & Rubin, 1987). In clinical trials, an individual's response is often missing at one follow-up time but observed at the next follow-up time, resulting in a large class of "non-monotone" missingness patterns. In this article, we assume that all individuals in a study have complete data on the response obtained at the first measurement occasion; for example, this is commonly the case in many longitudinal studies that require a baseline measure of the response as an inclusion criteria.

An example of a data set with this structure comes from two similar longitudinal clinical trials of HIV-infected patients (Kahn et al., 1992; Gallant et al., 1992). The two clinical trials were randomized phase III double-blind trials, designed to compare two treatments, zidovudine (AZT) and didanosine (DDI); they have been used in several combined analyses (Finkelstein et al., 1996). The response of interest is a binary CD4 cell count variable (dichotomized at > 200 versus ≤ 200 cells per cubic millimeter), measured at baseline (week 0), and every week for up to 5 weeks from baseline. The cutoff of 200 was chosen because of its strong predictive value for development of opportunistic infections, and has been adopted as a standard threshold of clinical importance (Phair et al., 1990). In this analysis, we consider the 431 patients with AIDS at baseline. The main question of scientific interest is the effect of treatment on changes in CD4 cell count sufficiency over time. As with most longitudinal studies, missing outcome data over time complicates the analysis. Although CD4 cell counts were obtained from all 431 patients at baseline, only 383 patients (88.95 %) had measurements at week 1, only 345 patients (80.0 %) had measurements taken at week 2, only 324 patients (75.2 %) had measurements taken at week 3, only 306 patients (71.0 %) had measurements taken at week 4, and only 285 patients (66.1 %) had measurements at week 6. Even though the overall percentage with an observed response decreases over time, the missing data pattern is non-monotone, i.e., some patients' responses are missing at one occasion and observed at the next occasion. In particular, there are 109 (25.3%) patients who missed at least one measurement, but returned for a later measurement. Typically, a decline in CD4 count indicates disease progression, and patients with low CD4 counts are more likely to make all of the scheduled study visits, as compared to patients with normal CD4 counts, who may not feel the need to make planned all of the scheduled study visits. Thus, in this setting, it is quite plausible that patients with low CD4 counts are more likely make all of the scheduled study visits. This would imply that missingness depends on the unobserved outcome of interest and is "nonignorable."

Numerous approaches for analyzing incomplete binary data with nonignorable missingness have been proposed (e.g., Baker, 1995; Baker & Laird, 1988; Ibrahim et al., 2001). To define a full likelihood for nonignorable and non-monotone missing responses over time, one needs to specify a joint distribution for the repeated binary outcomes as well as a model for the missing data mechanism. Although various likelihood approaches have been proposed, e.g., models based on bivariate and higher-order correlations (Bahadur, 1961; Zhao and Prentice, 1990), models based on bivariate and higher-order odds ratios (McCullagh and Nelder, 1989; Lipsitz et al., 1991; Molenberghs and Lesaffre, 1994), none of these likelihood-based models have proven to be of real practical use unless the number of repeated measures is relatively small (i.e., less than or equal to 3). As the number of repeated measures increases, the number of parameters that need to be specified and estimated proliferates rapidly for any of these models for the joint distribution, and a solution to the likelihood equations quickly becomes intractable. Thus as an alternative to the full likelihood, a pseudo-likelihood (Gong & Samaniego, 1981; Liang & Self, 1996) approach has been proposed by Troxel et al. (1998) under minimal parametric assumptions.

Troxel et al.'s (1998) pseudo-likelihood is based on a working assumption that the longitudinal binary measurements are independent over time. Specifically, their pseudo-likelihood uses a marginal logistic regression model for the response at each time point, and assumes that the missingness probability at a given time depends only on the missing response at that time and the covariates (the covariates are assumed to be fully observed). This pseudo-likelihood approach is attractive in that it substantially eases the numerical complexities of the full likelihood approach. Further, it alleviates the need to specify and estimate many nuisance parameters involved in a full likelihood.

Although the pseudo-likelihood approach of Troxel et al. (1998) yields asymptotically unbiased estimators of the regression parameters when the marginal model for the response at each time point and the model for missingness have been correctly specified, these pseudo-likelihood estimators can be highly inefficient, especially when there is moderate to strong associations among the repeated responses over time. In this article, we propose a new pseudo-likelihood that uses a bivariate Bahadur (1961) distribution for all possible pairings of the first (or baseline) binary response with all subsequent binary responses. A limited simulation study suggests that our proposed bivariate pseudo-likelihood is much more efficient than the "working independence" pseudo-likelihood of Troxel et al. (1998).

The paper is organized as follows. In Section 2, we introduce the model and notation to define a pseudo-likelihood for incomplete binary longitudinal data. In Section 3, we review the existing pseudo-likelihood of Troxel et al. (1998), and then describe our proposed pseudo-likelihood as a robust alternative to the full likelihood. Section 4 presents an application of the proposed method using longitudinal data on CD4 counts from the two clinical trials of HIV-infected patients described earlier. In Section 5, we present results from our simulation study of efficiency, and demonstrate that our proposed method can provide much more efficient estimators than that of Troxel et al. (1998). Section 6 concludes the paper with some remarks.

2. Model and Notation

Suppose N individuals are observed at a fixed set of T time points, $t = 1,\dots, T$. For the i^{th} individual ($i = 1,\dots, N$), we can form a $T \times 1$ vector, $\mathbf{Y}_i = (Y_{i1},\dots, Y_{iT})^t$ of binary random variables, Y_{it}. Each individual also has a $J \times 1$ vector of covariates, $\mathbf{x}_{it}$, and we assume that all the covariates are fully observed. The marginal distribution of Y_{it} is assumed to be Bernoulli with the probability of success

$$p_{it}=E(Y_{it}|\mathbf{x}_{it},\beta)=P(Y_{it}=1|\mathbf{x}_{it},\beta)=\frac{\exp(\mathbf{x}_{it}^t\beta)}{1+\exp(\mathbf{x}_{it}^t\beta)}. \tag{1}$$

Here the goal is to draw inferences about the regression parameters $\boldsymbol{\beta}$, whereas the within-subject association among the responses is regarded as a nuisance characteristic of the data. The association between a pair of binary outcomes is typically measured in terms of marginal odds ratios or marginal correlations. Marginal odds ratios can be used to derive a multivariate Plackett (1965) distribution. Also, marginal correlations can be used to derive a multivariate Bahadur (1961) model.

In a typical longitudinal study, individuals are not observed at all T occasions on account of some stochastic missing data mechanism. We assume that all individuals are observed at baseline ($t = 1$). However, subjects can be missing at any post-baseline follow-up time. We introduce $(T - 1)$ binary random variables, R_{it}, ($t = 2,\dots,T$), with R_{it} equal to 1 if Y_{it} is observed, and 0 if Y_{it} is missing. The pseudo-likelihood of Troxel et al. (1998) assumes that the marginal distribution of the binary random variable R_{it} is Bernoulli, with the probability of being observed,

$$\pi_{it}=P(R_{it}=1|y_{it},\mathbf{x}_{it},\gamma)=\frac{\exp(\gamma_0+\gamma_1 y_{it}+\gamma_2^t\mathbf{x}_{it})}{1+\exp(\gamma_0+\gamma_1 y_{it}+\gamma_2^t\mathbf{x}_{it})}. \tag{2}$$

Note that if $\gamma_1 \neq 0$, then the missing data mechanism is non-ignorable since the probability of missingness depends on possibly unobserved data Y_{it}. In the next section, we discuss the pseudo-likelihood of Troxel et al. (1998), and then describe our proposed bivariate pseudo-likelihood.

3. Estimators

3.1 Independent Pseudo-Likelihood

In this section, we briefly review the pseudo-likelihood approach of Troxel et al. (1998) who use a working assumption that the repeated responses are independent over time. To describe this pseudo-likelihood, let $f(y_{it}, r_{it}|\mathbf{x}_{it}, \beta, \gamma)$ be the marginal distribution of (Y_{it}, R_{it}) at time t, which can be expressed as

$$f(y_{it},r_{it}|\mathbf{x}_{it},\beta,\gamma)=f(y_{it}|\mathbf{x}_{it},\beta)f(r_{it}|y_{it},\mathbf{x}_{it},\gamma),$$

where $f(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})$ is Bernoulli with probability of success p_{it} as given in (1), and $f(r_{it}|y_{it}, \mathbf{x}_{it}, \boldsymbol{\gamma})$ is Bernoulli with probability of being observed as given in (2). Troxel et al.'s (1998) pseudo-likelihood, which treats the observations at different times as independent, is defined as

$$\begin{aligned}\mathscr{L}_{ind}(\beta,\gamma) &= \prod_{i=1}^{N}\prod_{t=1}^{T}\{f(y_{it}|\mathbf{x}_{it},\beta)f(r_{it}|y_{it},\mathbf{x}_{it},\gamma)\}^{r_{it}}\left\{\sum_{y_{it}=0}^{1}f(y_{it}|\mathbf{x}_{it},\beta)f(r_{it}|y_{it},\mathbf{x}_{it},\gamma)\right\}^{(1-r_{it})}\\ &= \prod_{i=1}^{N}\prod_{t=1}^{T}\{f(y_{it}|\mathbf{x}_{it},\beta)\pi_{it}\}^{r_{it}}\left\{\sum_{y_{it}=0}^{1}f(y_{it}|\mathbf{x}_{it},\beta)(1-\pi_{it})\right\}^{(1-r_{it})}.\end{aligned}$$

The pseudo-likelihood estimator of Troxel et al. (1998) is obtained by maximizing this pseudo-likelihood function, derived under a "working independence" assumption.

Although the pseudo-likelihood estimator of Troxel et al. (1998) is consistent, it can be highly inefficient, especially when there is strong within-subject associations among the repeated responses. In the next section, we propose a new pseudo-likelihood approach, which generally provides more efficient estimators than that of Troxel et al. (1998).

3.2 Proposed Bivariate Pseudo-Likelihood

Recall that we assume Y_{i1} is observed for all subjects. The marginal distribution of Y_{i1} ($i = 1, \ldots, N$) for all subjects is the product of Bernoulli distributions over N subjects, denoted by

$$\prod_{i=1}^{N}f(y_{i1}|\mathbf{x}_{it},\beta)=\prod_{i=1}^{N}p_{i1}^{y_{i1}}(1-p_{i1})^{(1-y_{i1})}.$$

Since Y_{i1} is observed for all subjects, for all $t > 1$ we can consider the conditional distribution of $f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, where $\boldsymbol{\alpha}$ denotes a vector of pairwise association parameters between Y_{i1} and Y_{it} for all $t > 1$. We can write this conditional distribution as

$$f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \beta, \alpha, \gamma) = f(y_{it}|y_{i1}, \mathbf{x}_{it}, \beta, \alpha) f(r_{it}|y_{i1}, y_{it}, \mathbf{x}_{it}, \gamma), \tag{3}$$

where $f(y_{it}|y_{i1}, \mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ can be easily obtained by first specifying the bivariate distribution $f(y_{i1}, y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha})$. There are many potential candidates for the bivariate distributions, $f(y_{i1}, y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha})$, such that the marginal logistic model for y_{it} holds. When the pairwise associations between Y_{i1} and Y_{it} are specified in terms of marginal odds ratios, one could adopt the bivariate Plackett (1965) distribution. When the pairwise associations between Y_{i1} and Y_{it} are specified in terms of marginal correlations, one could adopt the bivariate Bahadur (1961) model. Specifically, if (Y_{i1}, Y_{it}) follows the bivariate Bahadur (1961) distribution, then

$$f(y_{i1}, y_{it}|\mathbf{x}_{it}, \beta, \alpha) = p_{i1}^{y_{i1}}(1 - p_{i1})^{(1-y_{i1})} p_{it}^{y_{it}}(1 - p_{it})^{(1-y_{it})} \left\{1 + \alpha_{i1t} \frac{(y_{i1} - p_{i1})(y_{it} - p_{it})}{\sqrt{p_{i1}(1 - p_{i1})p_{it}(1 - p_{it})}}\right\}, \tag{4}$$

where $\alpha_{i1t} = \mathrm{Corr}(Y_{i1}, Y_{it}|\mathbf{x}_{it})$. For the bivariate distribution in (4), it can be shown that the conditional distribution of Y_{it} given Y_{i1} is Bernoulli, with success probability

$$p_{itI} = P(Y_{it} = 1|y_{i1}, \mathbf{x}_{it}, \beta, \alpha) = p_{it} \left\{1 + \alpha_{i1t} \frac{(1 - p_{it})(y_{i1} - p_{i1})}{\sqrt{p_{it}(1 - p_{it})p_{i1}(1 - p_{i1})}}\right\},$$

so that

$$f(y_{it}|y_{i1}, \mathbf{x}_{it}, \beta, \alpha) = p_{itI}^{y_{it}}(1 - p_{itI})^{(1-y_{it})}.$$

The density of the missing data mechanism $f(r_{it}|y_{i1}, y_{it}, \mathbf{x}_{it}, \boldsymbol{\gamma})$ can be specified by using a logistic regression model similar to (2), except that y_{i1} is included as an additional ‘covariate’ in the model.

To obtain estimators of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ we propose maximizing a pseudo-likelihood formed by treating the density $f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ as independent over the t’s:

$$\mathscr{L}_{biv}(\beta, \alpha, \gamma) = \prod_{i=1}^{N} \mathscr{L}_{biv,i}(\beta, \alpha, \gamma) = \prod_{i=1}^{N} f(y_{i1}|\mathbf{x}_{it}, \beta) \prod_{t=2}^{T} \{f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \beta, \alpha, \gamma))\}^{r_{it}} \left\{\sum_{y_{it}=0}^{1} f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \beta, \alpha, \gamma))\right\}^{(1-r_{it})} \tag{5}$$

Note that this pseudo-likelihood approach uses bivariate distributions between y_{it} and y_{i1}, and is expected to be more efficient than the independent pseudo-likelihood approach of Troxel et al. (1998). The bivariate pseudo-likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})$ can be obtained as the solution to $S(\hat{\boldsymbol{\theta}}) = 0$, where $S(\boldsymbol{\theta})$ is the pseudo-score function

$$S(\theta) = \sum_{i=1}^{N} S_i(\theta) = \sum_{i=1}^{N} \frac{\partial \log \mathscr{L}_{biv,i}(\theta)}{\partial \theta}.$$

Heuristically, using method of moments ideas, the bivariate pseudo-likelihood estimator $\hat{\boldsymbol{\theta}}$ is consistent since it is the solution to $S(\hat{\boldsymbol{\theta}}) = 0$, and it can be shown that $E[S(\boldsymbol{\theta})] = 0$ at the true $\boldsymbol{\theta}$ as long as (3) is correctly specified, even though the terms $f(y_{it}, r_{it}|y_{i1}, \mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ in (5) are not truly independent. The the bivariate pseudo-likelihood estimator $\hat{\boldsymbol{\theta}}$ is also asymptotically normal, with asymptotic variance described below. General proofs of the consistency and asymptotic normality of pseudo-likelihood estimators are given in Gong & Samaniego (1981), White(1982), and Lindsay (1988). The proof of the consistency of the independent pseudo-likelihood estimator is given in the Appendix of Troxel et al. (1998), and the proof of our pseudo-likelihood estimate is similar.

To estimate the asymptotic variance of the proposed bivariate maximum pseudo-likelihood estimator, we use a sandwich-type variance-covariance matrix (White, 1982) of the form

$$\mathrm{Var}(\hat{\theta}) \approx \left\{\frac{\partial S(\theta)}{\partial \theta}\right\}^{-1} \left\{\sum_{i=1}^{N} S_i(\theta) S_i^t(\theta)\right\} \left\{\frac{\partial S(\theta)}{\partial \theta}\right\}^{-1}. \tag{6}$$

An estimate of the asymptotic variance of $\hat{\boldsymbol{\theta}}$ is obtained by evaluating the right-hand side of (6) at the bivariate pseudo-likelihood estimator $\hat{\boldsymbol{\theta}}$.

4. Application: Analysis of AIDS Data

We present an analysis of the CD4 count data from the AIDS clinical trials described in the Introduction. The parameters are estimated using the proposed bivariate pseudo-likelihood, and Troxel et al.’s (1998) independent pseudo-likelihood under the assumption of nonignorable missingness. The AIDS clinical trials are randomised phase III double-blind trials, designed to compare two therapeutic treatments: zidovudine (AZT) and didanosine (DDI). Our study contains 431 patients who were diagnosed with AIDS or AIDS-related complex. The response of interest is normal CD4 cell count (> 200 cells per cubic millimeter) versus abnormal CD4 cell count (≤ 200) measured at baseline (week 0), and every week for up to 5 weeks from baseline; the outcome is defined as $Y_{it} = 1$ if the CD4 count exceeds 200 and 0 otherwise. As discuss in the Introduction, the cutoff of 200 cells per cubic millimeter was initially chosen because of its strong predictive value for development of opportunistic infections, and has been adopted as a standard threshold of clinical importance. The main question of scientific interest is the effect of treatment on changes in CD4 cell count sufficiency over time. As is common in many longitudinal clinical trials, missing responses over time complicate the analysis. The percent of missingness in CD4 counts ranges from 11% to 44% at the five follow-up occasions.

The probability that CD4 count > 200 at a given time, $p_{it} = pr(Y_{it} = 1|x_{it})$, is modeled, using a logistic regression model, as a function of treatment, time and baseline age. Treatment is defined by the following indicator variable

$$\mathrm{AZT}_i = \begin{cases} 1 \text{ if the } i^{th} \text{ subject is randomized to AZT,} \\ 0 \text{ if the } i^{th} \text{ subject is randomized to DDI.} \end{cases}$$

Because of the stratified randomization, to control for baseline age, we define the indicator variable

$$\mathrm{AGE}_i = \begin{cases} 1 \text{ if the } i^{th} \text{ subject has baseline age } \geq 35, \\ 0 \text{ otherwise,} \end{cases}$$

Specifically, the following logistic regression model was fit to the data,

$$\text{logit}(p_{it})=\beta_0+\beta_1\text{AGE}_i+\beta_3 t+\beta_4 t * \text{AZT}_i,$$

for $t = 0, 1,\ldots,5$. Note the exclusion of a main effect of treatment (AZT_i). In a model with a treatment (AZT_i) by time (t) interaction, the main effect of AZT corresponds to the treatment effect at baseline ($t = 0$). However, due to randomization, it is assumed that there is no treatment effect at baseline, i.e., the main effect of AZT equals 0.

Recall that the bivariate pseudo-likelihood requires specification of the correlations, α_{1t}. We estimated the parameters under both banded and exchangeable correlations; the results were so similar that for simplicity, we present results under an exchangeable assumption. Further, in terms of goodness of fit, the exchangeable correlation had the largest composite likelihood (Lindsay, 1988) information criterion (similar to the AIC) proposed for composite likelihoods by Varin and Vidoni (2005).

For the proposed bivariate pseudo-likelihood, recall that we must also model the probability of being observed at each time point, given the outcome at baseline. It was conjectured that CD4 count is nonignorably missing since sicker patients may be more likely to come in for a further GP visit, e.g., sicker patients may have been hospitalized. We considered the following missing data mechanism:

$$\text{logit}(\pi_{it})=\text{logit}[P(R_{it}=1|y_{i1}, y_{it}, x_{it}, \gamma)]=\gamma_0+\gamma_1 y_{i1}+\gamma_2 y_{it}+\gamma_3\text{AZT}_i+\gamma_4\text{AGE}_i+\gamma_5 t+\gamma_6 y_{i1}*\text{AZT}_i+\gamma_7 y_{i1}*t+\gamma_8 y_{it}*\text{AZT}_i+\gamma_9 y_{it}*t$$

for $t > 0$. Note that to choose a suitable mode for the missing-data mechanism, we again used a model selection criterion following Varin and Vidoni (2005). The above model produced the maximum “information criterion” among a number of models considered in the analysis. In general, the non-ignorable models suggest that subjects with normal CD4 counts and on AZT are less likely to be seen over time. For Troxel et al. (1998)’s pseudo-likelihood, we also used the above model except that y_{i1} (and any interaction term that included y_{i1}) was excluded as a “covariate” from the missingness model, since π_{it} can only be a function of variables at time t.

Table 1 displays estimates and standard errors for the parameters $\boldsymbol{\beta}$ for all models and methods, as well as the estimates of the missing data model. From Table 1, we see that the estimates from the non-ignorable independent and bivariate pseudo-likelihoods are similar, but the bivariate pseudo-likelihood yields smaller estimated standard errors. For example, for the time-stationary age main effect, the estimated relative efficiency (ratio of estimated variances) is 77% for independent versus bivariate pseudo-likelihood. For the AZT*TIME interaction, the parameter of primary scientific interest in a clinical trial comparing two treatments, the estimated relative efficiency (ratio of estimated variances) is 6.3% for the independent versus bivariate pseudo-likelihood. The estimated exchangeable correlation is 0.60, indicating high correlation among the repeated responses. This highlights the potential gains in efficiency from use of the bivariate pseudo-likelihood, in particular when the correlation among repeated measures is relatively high. To examine the finite sample efficiency of these approaches, in the next section we conduct a a simulation study that compares their finite sample properties.

5. Simulation Study

We conducted a simulation study to compare Troxel et al.’s (1998) pseudo-likelihood estimator under independence, the ML estimator under a correctly specified model, and the proposed bivariate pseudo-likelihood estimator with $T = 5$ time points. Even though we perform simulations with $T = 5$ times points, we have found in these simulations reported below that the relative bias is so high for ML that we would advise against using it unless the sample size is extremely large (say, over $N = 1000$); thus, as discussed in the Introduction, ML is only really feasible with $T \leq 3$. Therefore, to compare the pseudo-likelihood approaches to ML in the simulation study, we restricted the number of occasions to $T = 5$. The binary outcomes, $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5})$, for the i^{th} individual are assumed to follow a Bahadur (1961) model, with joint probabilities

$$P\{(Y_{i1}=y_{i1}),(Y_{i2}=y_{i2}),\ldots,(Y_{i5}=y_{i5})|x_i,\beta,\alpha\}=\left\{\prod_{t=1}^{5}p_{it}^{y_{it}}(1-p_{it})^{1-y_{it}}\right\}\times\left\{1+\sum_{st}\alpha_{st}z_{is}z_{it}+\sum_{stu}\alpha_{stu}z_{is}z_{it}z_{iu}+\sum_{stuv}\alpha_{stuv}z_{is}z_{it}z_{iu}z_{iv}+\alpha_{12345}z_{i1}z_{i2}z_{i3}z_{i4}z_{i5}\right\}$$

where $Z_{it}=(Y_{it}-p_{it})/\sqrt{p_{it}(1-p_{it})}$; $\alpha_{st} = \mathrm{Corr}(Y_{is}, Y_{it}) = E[Z_{is}Z_{it}|x_i]$; $\alpha_{stu} = E[Z_{is}Z_{it}Z_{iu}|x_i]$; $\alpha_{stuv} = E[Z_{is}Z_{it}Z_{iu}Z_{iv}|x_i]$; $\alpha_{12345} = E[Z_{i1}Z_{i2}Z_{i3}Z_{i4}Z_{i5}|x_i]$; and $\mathrm{logit}(p_{it}) = \beta_0 + \beta_x x_i + \beta_\tau(t - 1)$, for $t = 1, 2,\ldots,5$. We consider $\boldsymbol{\alpha} = (\{\alpha_{st}\}, \{\alpha_{stu}\}, \{\alpha_{stuv}\}, \alpha_{12345})^t$ as the vector of association parameters. We chose $\beta_0 = -0.2$, $\beta_x = 0.6$, and $\beta_\tau = -0.2$. The values of the covariate x were assumed to follow a Uniform(0, 2) distribution. A variety of different correlation structures were investigated. We present results from two different correlation structures: 1) an exchangeable correlation structure with $\alpha_{st} = \alpha$, and 2) a banded correlation structure, with $\alpha_{st} = \alpha_{t-s}$, where $t > s$ and $t - s = 1, 2, 3, 4$.

We assumed the following true non-ignorable missing data mechanism holds

$$\prod_{t=2}^{5}\pi_{it}^{r_{it}}(1-\pi_{it})^{(1-r_{it})}; \qquad \mathrm{logit}(\pi_{it})=\mathrm{logit}[P(R_{it}=1|y_{it},\gamma)]=\gamma_0+\gamma_1 y_{it}+\gamma_2 x_i, \tag{7}$$

for $t > 1$. The true model parameters in (7) were chosen as $\gamma_0 = -0.2$, $\gamma_1 = 1.0$, and $\gamma_2 = -0.5$. In this mechanism, non-monotone missingness can occur in that an outcome can be missing at time s ($R_{is} = 0$), but observed at time $t > s$ ($R_{it} = 1$). Each simulation run was based on 2500 replications, with $N = 120$ and $N = 240$ subjects.

For the exchangeable correlation model, Table 2 presents the empirical percent relative biases, $100 \times (\hat{\beta} - \beta)/\beta$ mean squared errors, and coverage probabilities of the regression estimators obtained from ML, the independent pseudo-likelihood of Troxel et al. (1998), and the proposed bivariate pseudo-likelihood. The coverage probabilities were obtained for 95% confidence intervals, $\hat{\beta} \pm 1.96\, s.e.(\hat{\beta})$, where the standard errors, $s.e.(\hat{\beta})$'s, of the pseudo-likelihood estimators, $\hat{\beta}$'s, were obtained from the sandwich-type variance estimator given in (6).

It is clear from Table 2 that all of the two pseudo-likelihood methods provide approximately unbiased estimators of the regression parameters under all simulation configurations; however, the ML estimate can give appreciable bias (in simulations not shown, this bias becomes negligible as the sample size increases). However, we are mainly interested in investigating the efficiency gains of the proposed bivariate pseudo-likelihood (BPL) estimators over the independent pseudo-likelihood (IPL) estimators of Troxel et al. (1998). In the following, we assume that any bias makes a negligible contribution to the MSEs. The BPL estimators appear to provide considerable gains in efficiency over the IPL estimators for both β_τ for correlations greater than 0.10, and for both sample sizes. For all configurations in Table 2, the MSEs of the BPL estimators are smaller than the MSEs of the corresponding IPL estimators. In general, for the time effect, the BPL estimator is substantially more efficient than the IPL estimator when the correlation is moderate to high. For example, when $N = 240$ and $\alpha = 0.25$, for the time effect, the IPL estimator is only 67% as efficient as the BPL estimator. Also, when $N = 240$ and $\alpha = 0.40$, for the time effect, the IPL estimator is only 64% as efficient as the BPL estimator.

For weak correlations ($\alpha = 0.10$, for example), both BPL and IPL estimators are nearly as efficient as the ML estimators, as might be expected. But for stronger correlations, unlike the IPL estimators, the proposed BPL estimators appear to be competitive to the ML estimators for the time trend. For example, when $N = 120$ and $\alpha = 0.40$, for the time effect, the IPL estimator is only 47% as efficient as the ML estimator, whereas the BPL estimator is 86% as efficient as the ML estimator. When comparing the BPL to the ML, we observe that the BPL has at least 85% efficiency for any simulation configuration. For this configuration, we found the sample size needs to be at least $N = 3000$ for the ML to have bias under 5% for all parameters.

We next describe the results of the simulation for the banded correlation structure, i.e.,

$$\mathrm{Corr}(Y_i, Y_{it}|\mathrm{x}_{it}) = \alpha_{|t-s|},$$

and $|t - s| = 1, 2, 3, 4$. We specified three sets of α's. In the first set, we have $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.20, 0.14, 0.06, 0.00)$; $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.35, 0.29, 0.21, 0.15)$; and $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.50, 0.43, 0.37, 0.30)$. The constraints on the Bahadur are such that we could not specify a range of more than 0.2 for any of the configurations. In the first set, we have, on average, weak correlation; in the second set, we have, on average, moderate correlation; in the last set, we have approximately the highest correlation possible for the Bahadur.

The results for Table 3 are similar to those in Table 2. First, it is striking the relative bias in the MLE, much higher than in the exchangeable simulation. This is due to the constraints on the parameter space of the Bahadur model for a banded correlation, i.e., the joint probabilities for the Bahadur probabilities must be between 0 and 1, but the possible values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are highly constrained in order for these probabilities to be in (0,1). We used the Newton-Raphson algorithm to find the maximum, and we found it always converged to a stationary point, but we also found that there were typically many different sets of parameter values that gave similar values of the likelihood as the MLE. We advise against using the likelihood approach here, unless the sample size is large (in simulations not shown, we found that the sample size should be at least $N = 700{,}000$ for the MLE to have at most 5% bias in these simulations. Thus, for these simulations, we will only compare BPL to IPL. For weak correlations on average $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.20, 0.14, 0.06, 0.00)$, IPL is at least 98% efficient versus BPL, as might be expected. For moderate correlations $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.35, 0.29, 0.21, 0.15)$, IPL is at least 83% efficient versus BPL. For strong correlations $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.50, 0.43, 0.37, 0.30)$, IPL is approximately 69% efficient versus BPL. Thus, for this type (banded) of correlation, IPL is more efficient than it was for an exchangeable correlation; however, as the correlation increases, it again loses efficiency.

We also performed another simulation with five time points to explore the robustness with respect to bias of the BPL approach with a mis-specified correlation model. In particular, we let the true model be banded as above, but we fit an exchangeable correlation model when using BPL. We used the same marginal model as before with $N = 120, 240$, and let $\alpha = 0.01, 0.25, 0.40$. Each simulation run was based on 2500 replications. The IPL approach will be asymptotically unbiased for any correlation structure, and ML will be asymptotically unbiased when the correlation model is correctly specified. In the simulations, we found that IPL and ML have a relative bias very similar to that in Table 2. We found the relative bias of BPL to be slightly higher than that in Table 2, but never larger than 5%. This simulation, as well as other simulations not shown, suggests that the BPL estimates of the marginal regression parameters have minimal bias when the correlation is mis-specified.

6. Conclusions

The purpose of this research was to provide a better alternative to the independent pseudo-likelihood approach for analyzing longitudinal binary data with possible nonignorable and non-monotone missingness. For the proposed pseudo-likelihood, we need to model only the bivariate density of (y_{i1}, y_{it}), for $t > 1$. Unlike the full likelihood approach, our proposed bivariate pseudo-likelihood is computationally less expensive. Also, when compared to the independent pseudo-likelihood of Troxel et al. (1998), we have observed in the simulations that the proposed bivariate pseudo-likelihood can be much more efficient under moderate to strong within-subject association among the repeated responses. We advise against using the likelihood approach for more than just 3 time points, unless the sample size is at least 3000. In particular, in our simulations, even for the simplest correlation structure (exchangeable), we found that the sample size should be at least $N = 3000$ for the MLE to have at most 5% bias.

Acknowledgments

We are grateful for the support provided by grants from the U.S. National Institutes of Health, and the Natural Sciences and Engineering Research Council of Canada.

References

Bahadur, RR. A representation of the joint distribution of responses to *n* dichotomous items. In: Solomon, H., editor. Studies in Item Analysis and Prediction. Stanford University Press; 1961. p. 158-168.

Baker SG. Marginal regression for repeated binary data with outcomes subject to nonignorable nonresponse. Biometrics. 1995; 51:1042–1052. [PubMed: 7548689]

Baker SG, Laird NM. Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. Journal of the American Statistical Association. 1988; 83:62–69.

Cox DR. The analysis of multivariate binary data. Applied Statistics. 1972; 21:113–120.

Finkelstein DM, Williams PL, Molenberghs G, Feinberg J, Powderly W, Kahn J, Dolins R, Cotton D. Patterns of opportunistic infections in patients with HIV infection. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology. 1996; 12:38–45. [PubMed: 8624759]

Gallant JE, Moore RD, Richman DD, Keruly J, Chaisson RE. Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. Journal of Infectious Diseases. 1992; 166:1223–1227. [PubMed: 1358986]

Gong G, Samaniego F. Pseudo maximum likelihood estimation: theory and applications. Annals of Statistics. 1981; 9:861–869.

Ibrahim JG, Chen MH, Lipsitz SR. Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. Biometrika. 2001; 88:551–564.

Kahn JO, Lagakos SW, Richman DD. A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. New England Journal of Medicine. 1992; 327:581–587. [PubMed: 1353607]

Le Cessie, Van Houwelingen JC. Logistic regression for correlated binary data. Applied Statistics. 1994; 43:95–108.

Liang K-Y, Self SG. On the asymptotic behavior of the pseudolikelihood ratio test statistic. Journal of the Royal Statistical Society, Series B. 1996; 58:785–796.

Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

Lindsay, BG. Composite likelihood methods. In: Prabhu, NU., editor. Statistical Inference from Stochastic Processes. Providence: American Mathematical Society; 1988. p. 221-239.

Lipsitz SR, Laird NM, Harrington DP. Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. Biometrika. 1991; 78:153–160.

Little, RJ.; Rubin, DB. Statistical Analysis with Missing Data. New York: Wiley & Sons; 1987.

McCullagh, P.; Nelder, JA. Generalized Linear Models. 2nd Edition. London: Chapman and Hall; 1989.

Meester SG, MacKay J. A parametric model for cluster correlated categorical data. Biometrics. 1994; 50:954–963. [PubMed: 7787008]

Molenberghs G, Lesaffre E. Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. Journal of the American Statistical Association. 1994; 89:633–644.

Phair J, Munoz A, Detels R, Kaslow R, Rinaldo C, Saah A. The risk of *Pneumocystis carinii* pneumonia among men infection with human immunodeficiency virus type 1. New England Journal of Medicine. 1990; 332:161–165. [PubMed: 1967190]

Plackett RM. A class of bivariate distributions. Journal of the American Statistical Association. 1965; 60:516–522.

Troxel AB, Lipsitz SR, Harrington DP. Marginal models for the analysis of longitudinal measurements subject to nonignorable non-monotone missing data. Biometrika. 1998; 85:661–672.

Varin C, Vidoni P. A note on composite likelihood inference and model selection. Biometrika. 2005; 92:519–528.

White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50:1–26.

Zhao LP, Prentice RL. Correlated binary regression using a quadratic exponential model. Biometrika. 1990; 77:642–648.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Longitudinal analysis of CD4 count data

	Proposed pseudo-likelihood			Troxel's pseudo-likelihood		
Variable	**Estimate**	**Std error**	***z* value**	**Estimate**	**Std error**	***z* value**
INTERCEPT	−2.6083	0.2213	−11.79	−2.7558	0.2451	−11.24
AGE	0.9403	0.2728	3.45	1.1209	0.2881	3.89
TIME	−0.1324	0.0355	−3.73	−0.1173	0.0405	−2.90
TIME * AZT	0.1665	0.0605	2.75	0.2407	0.1504	1.60
Exchangeable corr:						
ρ	0.6028	0.0516	11.68			
Missing data model:						
INTERCEPT	2.1236	0.1754	12.11	2.0622	0.1915	10.77
y_1	−0.5165	0.7326	−0.71			
y_t	4.9929	1.4788	3.38	3.8444	0.9632	3.99
AZT	0.0185	0.1804	0.10	0.2540	0.3706	0.69
AGE	−0.1866	0.1669	−1.12	−0.1092	0.1930	−0.57
TIME	−0.3111	0.0352	−8.84	−0.2955	0.0385	−7.68
y_1 * AZT	−0.3555	0.6211	−0.57			
y_1 * TIME	0.3530	0.1318	2.68			
y_t * AZT	−2.7838	1.0211	−2.73	−4.4208	1.3952	−3.17
y_t * TIME	−0.5016	0.2757	−1.82	−0.1483	0.1652	−0.90

Table 2

Empirical percentage relative biases, mean squared errors, and coverage probabilities of Troxel's independent pseudo-likelihood (IPL), proposed bivariate pseudo-likelihood (BPL), and ML estimators for exchangeable correlation. (True parameter values: $\beta_0 = -0.2$, $\beta_\tau = -0.2$, $\beta_x = 0.6$).

	Relative bias (%)			Mean squared error			Coverage prob.		
Parameter	**IPL**	**BPL**	**ML**	**IPL**	**BPL**	**ML**	**IPL**	**BPL**	**ML**
N = **120**									
ρ = **0.10**									
β_0	−3.044	−2.414	−13.745	0.04450	0.04418	0.03541	94.6	94.3	96.2
β_τ	−2.726	−2.691	−15.789	0.01090	0.01054	0.00927	93.2	93.4	93.6
β_x	−1.046	−0.954	−10.341	0.06153	0.06127	0.05657	94.3	94.4	94.9
ρ = **0.25**									
β_0	−5.515	−3.005	−8.711	0.04767	0.04630	0.04335	95.3	95.5	95.6
β_τ	−1.972	0.103	−0.852	0.01135	0.00783	0.00662	92.6	93.4	94.8
β_x	0.581	0.601	−4.462	0.07442	0.07375	0.07100	94.6	95.0	95.1
ρ = **0.40**									
β_0	−2.807	0.793	−23.247	0.05430	0.05300	0.03914	94.6	94.7	97.0
β_τ	−2.401	1.479	2.750	0.00992	0.00535	0.00462	93.0	94.2	94.6
β_x	−0.524	−0.088	−21.610	0.08935	0.08802	0.07733	95.0	95.3	94.0
N = **240**									
ρ = **0.10**									
β_0	−0.078	0.087	−5.768	0.02116	0.02093	0.01870	95.4	95.5	96.0
β_τ	−0.840	−0.908	−7.927	0.00501	0.00485	0.00411	94.3	94.6	93.6
β_x	0.099	0.203	−3.913	0.02902	0.02869	0.02787	95.0	95.2	94.6
ρ = **0.25**									
β_0	−1.663	−0.402	−1.376	0.02326	0.02291	0.02202	95.1	95.2	95.6
β_τ	−1.897	−0.380	1.212	0.00452	0.00303	0.00295	94.8	94.9	95.0
β_x	0.453	0.435	−0.509	0.03592	0.03542	0.03254	95.4	95.7	96.3
ρ = **0.40**									
β_0	0.129	1.292	−10.865	0.02611	0.02522	0.02047	95.0	95.2	96.7
β_τ	−0.936	0.520	4.644	0.00404	0.00258	0.00244	94.8	95.0	92.3

Parameter	Relative bias (%)			Mean squared error			Coverage prob.		
	IPL	BPL	ML	IPL	BPL	ML	IPL	BPL	ML
β_x	−0.576	−0.335	−11.220	0.04332	0.04291	0.03803	95.2	95.6	93.2

Table 3

Empirical percentage relative biases, mean squared errors, and coverage probabilities of Troxel's independent pseudo-likelihood (IPL), proposed bivariate pseudo-likelihood (BPL), and ML estimators for a banded correlation. (True parameter values: $\beta_0 = -0.2$, $\beta_\tau = -0.2$, $\beta_x = 0.6$).

	Relative bias (%)			Mean squared error			Coverage prob.		
Parameter	**IPL**	**BPL**	**ML**	**IPL**	**BPL**	**ML**	**IPL**	**BPL**	**ML**
N = **120**									
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.20, 0.14, 0.06, 0.00**)									
β_0	−2.441	−3.046	−21.003	0.04510	0.04501	0.01924	94.8	95.2	96.2
β_τ	−2.440	−3.024	−30.730	0.01177	0.01157	0.01004	93.0	93.5	93.6
β_x	−0.877	−0.684	−27.217	0.06301	0.06277	0.05544	94.5	94.7	94.9
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.35, 0.29, 0.21, 0.15**)									
β_0	−3.725	−2.605	−15.868	0.05163	0.05106	0.01639	94.9	94.9	98.7
β_τ	−3.359	−1.723	−25.668	0.01084	0.00904	0.00797	93.2	94.0	94.2
β_x	0.478	0.769	−34.286	0.07934	0.07827	0.06641	94.3	94.4	92.9
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.50, 0.43, 0.37, 0.30**)									
β_0	−6.573	−3.055	−35.958	0.05773	0.05674	0.02278	94.8	95.0	98.4
β_τ	−1.104	1.876	−17.690	0.00946	0.00657	0.00605	93.6	93.9	94.3
β_x	−0.245	0.228	−43.879	0.09534	0.09475	0.08912	94.6	94.9	94.1
N = **240**									
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.20, 0.14, 0.06, 0.00**)									
β_0	1.302	1.502	−8.489	0.02154	0.02134	0.01334	95.3	95.2	96.8
β_τ	−1.925	−1.639	−18.059	0.00546	0.00540	0.00509	94.7	95.0	92.7
β_x	0.762	0.812	−12.956	0.03065	0.03053	0.02877	95.0	94.8	92.4
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.35, 0.29, 0.21, 0.15**)									
β_0	0.955	1.442	−3.253	0.02403	0.02312	0.01054	95.2	95.6	99.1
β_τ	−0.961	0.174	−16.178	0.00509	0.00419	0.00385	95.0	95.1	94.6
β_x	0.686	0.791	−22.896	0.03958	0.03918	0.03824	94.4	94.6	92.9
$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ = (**0.50, 0.43, 0.37, 0.30**)									
β_0	−0.184	0.503	−27.114	0.02746	0.02726	0.01421	95.4	95.3	99.1
β_τ	−0.475	0.826	−4.336	0.00448	0.00309	0.00278	93.8	94.5	96.5

Parameter	Relative bias (%)			Mean squared error			Coverage prob.		
	IPL	**BPL**	**ML**	**IPL**	**BPL**	**ML**	**IPL**	**BPL**	**ML**
β_x	0.801	0.679	−33.995	0.04612	0.04582	0.05775	95.4	95.6	94.6