CSCanada
Progress in Applied Mathematics
Vol. 4, No. 2, 2012, pp. [128–147]
DOI: 10.3968/j.pam.1925252820120402.S0803
ISSN 1925-251X [Print]
ISSN 1925-2528 [Online]
www.cscanada.net
www.cscanada.org

Robust Inference for Incomplete Binary Longitudinal Data

Sanjoy K. Sinha[a],*

[a] School of Mathematics and Statistics, Carleton University, Canada.

* Corresponding author.
Address: School of Mathematics and Statistics, Carleton University, Ottawa, ON, K1S 5B6, Canada; E-Mail: sinha@math.carleton.ca

Supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Received: August 20, 2012/ Accepted: October 10, 2012/ Published: October 31, 2012

Abstract: Missing data occur in many longitudinal studies. When data are nonignorably missing, it is necessary to incorporate the missing data mechanism into the observed data likelihood function. A full likelihood analysis of nonignorable missing data is complicated algebraically, and often requires intensive computation, especially when there are many follow-up times. To avoid such computational difficulties, pseudo-likelihood methods have been proposed in the literature under minimal parametric assumptions. However, like the classical maximum likelihood estimators, these pseudo-likelihood estimators are also sensitive to potential outliers in the data. In this article, we propose and explore a robust method in the framework of a pseudo-likelihood function that is derived under the working assumption that the longitudinal responses are independent over time. The performance of the proposed robust method is investigated in simulations. The method is also illustrated in an example using actual data on CD4 counts from clinical trials of HIV-infected patients.

Key words: Incomplete data; Longitudinal study; Marginal models; Nonignorable missingness; Outliers; Robust estimation

Sinha, S.K. (2012). Robust Inference for Incomplete Binary Longitudinal Data. *Progress in Applied Mathematics, 4*(2), 128–147. Available from http://www.cscanada.net/index.php/pam/article/view/j.pam.1925252820120402.S0803 DOI: 10.3968/j.pam.1925252820120402.S0803

1. INTRODUCTION

In many longitudinal studies, individuals are measured repeatedly over a fixed set of assessment times. For example, longitudinal data are often collected in AIDS, cancer, and cardiovascular clinical trials as well as in observational studies. Here we focus on the case where the response over time is binary, and we are interested in modeling the marginal means of the binary responses. Methods for analyzing binary longitudinal data have been extensively studied in the literature (e.g., Le Cessie & Van Houwelingen, 1994; Liang & Zeger, 1986; Meester & MacKay, 1994; Molenberghs & Lesaffre, 1994; Prentice, 1988; and many others). In the absence of a suitable likelihood function to work with, often longitudinal data are analyzed using a multivariate analogue of the quasi-likelihood function (Wedderburn, 1974). The development of the "quasi-score equations", however, requires correct specification of the correlation matrix of the repeated responses over time. Liang and Zeger (1986) suggested a simplified "working" correlation matrix, and argued that the estimators obtained by solving their proposed "generalized estimating equations" (GEEs) are consistent even under a misspecified correlation structure.

The modeling of binary longitudinal data is often complicated by the fact that the response variable is not always observed at all assessment times. If missingness does not depend on the values of the data, missing or observed, then the data are called missing completely at random (MCAR). Any method that yields valid inferences in the absence of missing data would also yield valid inferences when data are missing completely at random and the analysis is based on the available data. A less restrictive assumption than MCAR is that missingness depends only on the observed values of the variables, but not on the values that are missing. In this case, data are called missing at random (MAR). Many authors considered extensions of the aforementioned quasi-likelihood approaches to the analysis of incomplete longitudinal data under the MAR assumption. Robins *et al.* (1995) proposed a weighted generalized estimating equation (WGEE) approach based on "inverse probability weights" (IPW) for analyzing longitudinal data under the assumption of MAR.

Note that missingness in the longitudinal data often depends on the unobserved value of the response variable at that time. In such cases, the missing data mechanism is called nonignorable (Little & Rubin, 2002). As an example, here we consider a data set from two clinical trials of HIV-infected patients, initially analyzed by Gallant *et al.* (1992) and Kahn *et al.* (1992). In this experiment, 431 patients were diagnosed with AIDS or AIDS-related complex, and the study was designed to compare two therapeutic treatments, zidovudine (Azt) and didanosine (Ddi). The response is a binary CD4 cell count variable, measured at baseline (week 0), and every week for up to 5 weeks from baseline. The interest is on the effect of treatment on changes in CD4 cell count over time. As in many longitudinal studies, here the analysis is complicated by missing response data over time. Although measurements on CD4 cell counts were taken from all 431 patients at baseline, but incomplete measurements were taken from only 383 patients (88.95%) at week 1, 345 patients (80.0%) at week 2, 324 patients (75.2%) at week 3, 306 patients (71.0%) at week 4, and only 285 patients (66.1%) at week 5. The missing data pattern is also non-monotone, that is, some patients' responses are missing at one visit, but observed at the next visit. There are 109 (25.3%) patients who missed at least one visit, but returned for a later visit. A decline in CD4 count normally indicates disease progression, and patients with low CD4 counts are more likely to make all scheduled visits, as compared to patients with normal CD4 counts. This would imply that missingness in the CD4 cell counts depends on the unobserved outcome and so is "nonignorable".

Statistical analyses with missing data based on the likelihood approach were considered by many authors (e.g., Brown, 1990; Dantan *et al.*, 2008; Diggle & Kenward, 1994; Ibrahim *et al.*, 1999, 2001; and Sinha *et al.*, 2010, 2011). Ibrahim *et al.* (1999) propose an EM algorithm for maximum likelihood estimation in generalized linear models for data with nonignorable missing covariates. Ibrahim *et al.* (2001) extended the EM method to the analysis of generalized linear mixed models with nonignorable missing responses. Recently, Sinha *et al.* (2010) investigates a multivariate logistic regression model for analyzing multiple binary outcomes with incomplete covariate data where auxiliary information is available. The auxiliary data are extraneous to the regression model of interest but predictive of the covariate with missing data.

A full likelihood analysis of longitudinal data under nonignorable missingness often requires intensive computation, especially when there are many follow-up times. To overcome this problem, a pseudo-likelihood approach was proposed by Troxel *et al.* (1998) under minimal parametric assumptions. This pseudo-likelihood approach was developed under the working assumption that the longitudinal outcomes are independent over time, and 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.

It is well-known that the ordinary maximum likelihood and maximum pseudo-likelihood estimators are sensitive to potential outliers in the data. To bound the influence of outliers, robust methods were studied by a number of authors (e.g., Cantoni & Ronchetti, 2001; Preisser & Qaqish, 1999; Sinha, 2004). Most of these robust methods focused on the estimation in generalized linear models in a complete-data setting. Sinha (2004) proposed a robust method for analyzing clustered correlated data in the framework of maximum likelihood approach for generalized linear mixed models. Also, Sinha (2008) proposed a robust method for fitting generalized linear models with nonignorable missing covariates.

In this paper, we focus on a robust analysis of longitudinal binary data with nonignorable missing responses. As mentioned earlier, a full likelihood analysis of nonignorable missing data typically involves intensive computation. Here our goal is to adopt a suitable method which is computationally feasible and also provides robust estimators in the presence of potential outliers in the data. In this note, we consider a robust method in the framework of the pseudo-likelihood of Troxel *et al.* (1998). The proposed method requires much less computation as compared to the full likelihood analysis of binary longitudinal data. Note that in the case of binary data, although no outliers are desired in the binary outcomes, outliers are still important in the residual sense; standardized residuals could be unbounded for a binary model. The proposed robust method is useful for bounding the influence of both outliers in the residuals and leverage points in the design space.

The paper is organized as follows. Section 2 introduces the model and notation for analyzing longitudinal data. Section 3 introduces the proposed robust method and studies the asymptotic properties of the robust estimators. Section 4 provides an illustrative example to describe the computational issues of the robust estimation. Section 5 investigates the empirical properties of the robust estimators based on a simulation study. Section 6 presents an application of the proposed method using the CD4 cell count data introduced earlier. Section 7 concludes the paper with some discussion.

2. MODEL AND NOTATION

Suppose N individuals, $i = 1, \ldots, N$, are observed at a fixed set of T time-points, $t = 1, \ldots, T$. Let $\mathbf{y}_i$ represent a $T \times 1$ vector of repeated responses, $(y_{i1}, \ldots, y_{iT})$, for the ith individual. Also, let $\mathbf{x}_{it}$ represent a $p \times 1$ vector of covariates associated with the response y_{it} for individual i at time t. 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

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

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

We focus on the case where individuals in a longitudinal study are not observed at all T follow-up times on account of some stochastic missing data mechanism. We introduce T binary random variables, v_{it}, $(t = 1, \ldots, T)$, with v_{it} equal to 1 if the response y_{it} is observed, and 0 if y_{it} is missing. We assume that the marginal distribution of the binary random variable v_{it} is Bernoulli, with the probability of being observed,

$$\pi_{it} = P(v_{it} = 1|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau}) = \frac{\exp(\tau_0 + \boldsymbol{\tau}_1^t \mathbf{x}_{it} + \tau_2 y_{it})}{1 + \exp(\tau_0 + \boldsymbol{\tau}_1^t \mathbf{x}_{it} + \tau_2 y_{it})}. \tag{2}$$

Note that if $\tau_2 \neq 0$, then the missing data mechanism is nonignorable since the probability of missingness depends on possibly unobserved data y_{it}. In the next section, we briefly review the pseudo-likelihood approach of Troxel *et al.* (1998) for analyzing incomplete binary longitudinal data, and then introduce our proposed robust method in the framework of this pseudo-likelihood function.

3. ESTIMATORS

3.1. Independent Pseudo-Likelihood (IPL) Estimator

Troxel *et al.* (1998) proposed a pseudo-likelihood under the working assumption that the repeated responses are independent over time. To describe this, let $f_{y,v}(y_{it}, v_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\tau})$ be the marginal distribution of (y_{it}, v_{it}) at time t, which can be expressed as

$$f_{y,v}(y_{it}, v_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\tau}) = f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}) f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau}), \tag{3}$$

where $f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})$ is Bernoulli with probability of success p_{it} as given in (1), and $f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})$ is Bernoulli with probability of being observed as given in (2). Then treating the repeated observations at different follow-up times as independent, Troxel *et al.* (1998) defined the pseudo-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \boldsymbol{\tau}^t)^t$ by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{t=1}^{T}\{f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})\}^{v_{it}}\left\{\sum_{y_{it}=0}^{1} f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})\right\}^{1-v_{it}} \tag{4}$$

The logarithm of the pseudo-likelihood function may be obtained as

$$\begin{aligned}\log L(\boldsymbol{\theta}) = &\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}\left\{\log f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}) + \log f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})\right\}\\ &+\sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})\log\left\{\sum_{y_{it}=0}^{1} f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})f_v(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})\right\}.\end{aligned} \tag{5}$$

An estimator of $\boldsymbol{\theta}$ may be obtained by maximizing the pseudo-likelihood function (4) or the log-pseudo-likelihood function (5).

From (5), the pseudo-score equations for the regression parameters $\boldsymbol{\beta}$ take the form

$$\begin{aligned}\mathbf{0} &= \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \equiv S_\beta(\boldsymbol{\theta}) \equiv \sum_{i=1}^{N} S_{i,\beta}(\boldsymbol{\theta})\\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}\frac{\partial \log f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\left\{\left.\frac{\partial \log f_y(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right| v_{it}\right\}\\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}(y_{it}-p_{it})\mathbf{x}_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\{(y_{it}-p_{it})\mathbf{x}_{it}|v_{it}\},\end{aligned} \tag{6}$$

where $E_{y|v}$ represents the expectation with respect to the conditional distribution of the response y_{it} given the value of the missing data indicator v_{it}.

Similarly, the pseudo-score equations for the parameters $\boldsymbol{\tau}$ of the missing data model (2) may be obtained as

$$\begin{aligned}\mathbf{0} =&\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \equiv S_\tau(\boldsymbol{\theta}) \equiv \sum_{i=1}^{N} S_{i,\tau}(\boldsymbol{\theta})\\ =&\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}\frac{\partial \log f_{v|y}(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\\ &+\sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\left\{\left.\frac{\partial \log f_{v|y}(v_{it}|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right| v_{it}\right\}\\ =&\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}(v_{it}-\pi_{it})\mathbf{z}_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\{(v_{it}-\pi_{it})\mathbf{z}_{it}|v_{it}\},\end{aligned} \tag{7}$$

where $\mathbf{z}_{it} = (\mathbf{x}_{it}^t, y_{it})^t$.

Equations (6) and (7) are solved simultaneously for the "independent pseudo-likelihood" (IPL) estimators $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}^t, \tilde{\boldsymbol{\tau}}^t)^t$ of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \boldsymbol{\tau}^t)^t$.

3.1.1. Standard Errors of IPL Estimators

We can write the pseudo-score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ in matrix form as

$$\frac{\partial}{\partial \boldsymbol{\theta}} S(\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ S_\beta(\boldsymbol{\theta})^t, S_\tau(\boldsymbol{\theta})^t \right\}^t = \begin{pmatrix} \dfrac{\partial S_\beta(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} & \dfrac{\partial S_\tau(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ \dfrac{\partial S_\beta(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} & \dfrac{\partial S_\tau(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \end{pmatrix},$$

where

$$\begin{aligned} \frac{\partial S_\beta(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= -\sum_{i=1}^{N} \sum_{t=1}^{T} p_{it}(1 - p_{it}) \mathbf{x}_{it} \mathbf{x}_{it}^t \\ &\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - v_{it}) \left[E_{y|v}\{(y_{it} - p_{it})^2 | v_{it}\} - \{E_{y|v}[(y_{it} - p_{it}) | v_{it}]\}^2 \right] \mathbf{x}_{it} \mathbf{x}_{it}^t, \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial S_\tau(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} &= -\sum_{i=1}^{N} \sum_{t=1}^{T} \pi_{it}(1 - \pi_{it}) \mathbf{z}_{it} \mathbf{z}_{it}^t \\ &\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - v_{it}) E_{y|v}\{(v_{it} - \pi_{it})^2 \mathbf{z}_{it} \mathbf{z}_{it}^t | v_{it}\} \\ &\quad - \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - v_{it}) E_{y|v}\{(v_{it} - \pi_{it}) \mathbf{z}_{it} | v_{it}\} E_{y|v}\{(v_{it} - \pi_{it}) \mathbf{z}_{it} | v_{it}\}^t, \end{aligned} \tag{9}$$

and

$$\begin{aligned} \frac{\partial S_\beta(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} &= \frac{\partial S_\tau(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^t} \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - v_{it}) E_{y|v}\{(y_{it} - p_{it})(v_{it} - \pi_{it}) \mathbf{x}_{it} \mathbf{z}_{it}^t | v_{it}\} \\ &\quad - \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - v_{it}) E_{y|v}\{(y_{it} - p_{it}) \mathbf{x}_{it} | v_{it}\} E_{y|v}\{(v_{it} - \pi_{it}) \mathbf{z}_{it} | v_{it}\}^t. \end{aligned} \tag{10}$$

To estimate the asymptotic variance-covariance matrix of the IPL estimators, we can use a sandwich-type variance-covariance matrix in the form

$$\text{Var}(\tilde{\boldsymbol{\theta}}) \approx \left\{ \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}^{-1} \left\{ \sum_{i=1}^{N} S_i(\boldsymbol{\theta}) S_i^t(\boldsymbol{\theta}) \right\} \left\{ \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}^{-1}, \tag{11}$$

where

$$S_i(\boldsymbol{\theta}) = \left\{ S_{i,\beta}(\boldsymbol{\theta})^t, S_{i,\tau}(\boldsymbol{\theta})^t \right\}^t.$$

An estimate of the variance of $\tilde{\boldsymbol{\theta}}$ is obtained by evaluating the right-hand side of (11) at the IPL estimate $\tilde{\boldsymbol{\theta}}$.

3.2. Proposed Robust Pseudo-Likelihood (RPL) Estimator

Note that in the case of binary longitudinal data, as the response y is binary, outliers can arise in the data only through the covariates $\mathbf{x}$. We focus on the case where the covariates are continuous, but the method can be generalized for discrete or a mixture of discrete and continuous measurements.

From (6) and (7), it is clear that the pseudo-score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are proportional to the covariates $\mathbf{x}$, so the influence of an outlier on the ordinary IPL estimators is unbounded. In other words, the IPL estimators are not robust against outliers. To obtain robust estimators of the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, we propose to solve the estimating equations

$$\sum_{i=1}^{N} \Psi_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = \mathbf{0}, \tag{12}$$

$$\sum_{i=1}^{N} \Psi_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = \mathbf{0}, \tag{13}$$

where

$$\begin{aligned}\Psi_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = &\sum_{t=1}^{T} v_{it}\left\{\psi_c(r_{it}) - E_y\{\psi_c(r_{it})\}\right\}\sigma_{it}w_{it}\mathbf{x}_{it} \\ &+ \sum_{t=1}^{T}(1-v_{it})E_{y|v}\left[\left\{\psi_c(r_{it}) - E_y\{\psi_c(r_{it})\}\right\}|v_{it}\right]\sigma_{it}w_{it}\mathbf{x}_{it}\end{aligned} \tag{14}$$

and

$$\begin{aligned}\Psi_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = &\sum_{t=1}^{T} v_{it}\left\{\psi_c(r^*_{it}) - E_{v|y}\{\psi_c(r^*_{it})\}\right\}\sigma^*_{it}w^*_{it}\mathbf{z}_{it} \\ &+ \sum_{t=1}^{T}(1-v_{it})E_{y|v}\left[\left\{\psi_c(r^*_{it}) - E_{v|y}\{\psi_c(r^*_{it})\}\right\}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}|v_{it}\right]\end{aligned} \tag{15}$$

with $r_{it} = (y_{it} - p_{it})/\sigma_{it}$, $\sigma^2_{it} = \mathrm{var}(y_{it})$, $r^*_{it} = (v_{it} - \pi_{it})/\sigma^*_{it}$, and $\sigma^{*2}_{it} = \mathrm{var}(v_{it})$. The function ψ_c is considered as the Huber's psi function, $\psi_c(r) = \max\{-c, \min(r, c)\}$, which is used to bound the influence of any outliers in the residuals when estimating the model parameters.

The weights $w_{it} = w(\mathbf{x}_{it})$ and $w^*_{it} = w(\mathbf{z}_{it})$ are used to downweight any leverage points in the design space. Here we choose the weight function w as a function of the Mahalanobis distance d_x in the form

$$w(\mathbf{x}) = w(\mathbf{x}, \mathbf{c}_x, \mathbf{S}_x) = \min\left\{1, \left(\frac{b_0}{d_x}\right)^{\gamma_0/2}\right\} \tag{16}$$

where $d_x = (\mathbf{x} - \mathbf{c}_x)^T\mathbf{S}_x^{-1}(\mathbf{x} - \mathbf{c}_x)$, the tuning constants $\gamma_0 \geq 1$ and b_0 is chosen as the 95th percentile of the chi-squared distribution with degrees of freedom equal to the dimension of $\mathbf{x}$, and $\mathbf{c}_x$ and $\mathbf{S}_x$ are some robust estimators of the location and scale parameters for the distribution of $\mathbf{x}$, such as the minimum volume ellipsoid (MVE) estimators of Rousseeuw and van Zomeren (1990). Note that the choice $\psi_c(r) = r$ and $w(\mathbf{x}) = 1$ leads to the ordinary pseudo-likelihood (IPL) estimators of the model parameters.

3.2.1. Newton-Raphson Method for RPL Estimators

The proposed RPL estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are obtained by solving the estimating equations (12) and (13) simultaneously using an iterative method. We focus on the iterative Newton-Raphson method and the scoring technique for solving these equations. The iterative equations for the RPL estimators of $\boldsymbol{\beta}$ may be expressed in the form

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left\{\sum_{i=1}^{N} \dot{\Psi}_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}^{(m)}, \boldsymbol{\tau})\right\}^{-1} \sum_{i=1}^{N} \Psi_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}^{(m)}, \boldsymbol{\tau}), \quad (17)$$

for $m = 0, 1, 2, \ldots$, where

$$\begin{aligned}\dot{\Psi}_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = & \sum_{t=1}^{T} v_{it} \left[\frac{\partial}{\partial \eta_{it}} \{\psi_c(r_{it}) - E_y\{\psi_c(r_{it})\}\}\right] \sigma_{it} w_{it} \mathbf{x}_{it} \mathbf{x}_{it}^t \\ & + \sum_{t=1}^{T} (1 - v_{it}) E_{y|v} \left[\frac{\partial}{\partial \eta_{it}} \{\psi_c(r_{it}) - E_y\{\psi_c(r_{it})\}\}\middle| v_{it}\right] \sigma_{it} w_{it} \mathbf{x}_{it} \mathbf{x}_{it}^t.\end{aligned}$$

Similarly, the iterative equations for the RPL estimators of $\boldsymbol{\tau}$ can be expressed in the form

$$\boldsymbol{\tau}^{(m+1)} = \boldsymbol{\tau}^{(m)} - \left\{\sum_{i=1}^{N} \dot{\Psi}_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}^{(m)})\right\}^{-1} \sum_{i=1}^{N} \Psi_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}^{(m)}), \quad (18)$$

for $m = 0, 1, 2, \ldots$, where

$$\begin{aligned}\dot{\Psi}_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau}) = & \sum_{t=1}^{T} v_{it} \left[\frac{\partial}{\partial \eta_{it}} \left\{\psi_c(r_{it}^*) - E_{v|y}\{\psi_c(r_{it}^*)\}\right\}\right] \sigma_{it}^* w_{it}^* \mathbf{z}_{it} \mathbf{z}_{it}^t \\ & + \sum_{t=1}^{T} (1 - v_{it}) E_{y|v} \left[\frac{\partial}{\partial \eta_{it}} \left\{\psi_c(r_{it}^*) - E_{v|y}\{\psi_c(r_{it}^*)\}\right\} \sigma_{it}^* w_{it}^* \mathbf{z}_{it} \mathbf{z}_{it}^t \middle| v_{it}\right].\end{aligned}$$

The complete algorithm for these robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ is described as follows:

1. Choose initial values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\tau}^{(0)}$. These initial values can be chosen as the ordinary IPL estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. Set $m = 0$.
2. (a) Calculate $\boldsymbol{\beta}^{(m+1)}$ and $\boldsymbol{\tau}^{(m+1)}$ from the iterative equations (17) and (18), respectively.
   (b) Set $m = m + 1$.
3. Continue step 2 until a convergence is achieved. Declare the estimates at convergence to be the RPL estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\tau}}$.

3.2.2. Asymptotics for RPL Estimators

A sketch of the development of the asymptotic distributions of the RPL estimators is given here. Recall the estimating equations (12) and (13) for the robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. These equations can be reexpressed in the form

$$\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \Psi_i(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\theta}) = \mathbf{0}, \tag{19}$$

where $\Psi_i(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\theta}) = \{\Psi_{\beta,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau})^t, \Psi_{\tau,i}(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\beta}, \boldsymbol{\tau})^t\}^t$. Let $\hat{\boldsymbol{\theta}}_N = (\hat{\boldsymbol{\beta}}^t_N, \hat{\boldsymbol{\tau}}^t_N)^t$ denote the robust estimators obtained by solving (19). Assume that the "true" values $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$ are obtained by solving the equations

$$\bar{\mathbf{G}}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} E\left\{\Psi_i(\mathbf{y}_i, \mathbf{v}_i|\boldsymbol{\theta})\right\} = \mathbf{0} \tag{20}$$

with respect to $\boldsymbol{\theta}$. By the Mean Value Theorem, we can write

$$\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\hat{\boldsymbol{\theta}}_N) = \mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}_0) + \mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0), \tag{21}$$

where vector $\tilde{\boldsymbol{\theta}}$ lies on the segment connecting $\hat{\boldsymbol{\theta}}_N$ and $\boldsymbol{\theta}_0$, and $\mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\tilde{\boldsymbol{\theta}})$ is the derivative of $\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ evaluated at $\tilde{\boldsymbol{\theta}}$. From (21), we can write

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) = \left\{-\mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\tilde{\boldsymbol{\theta}})\right\}^{-1}\left\{\sqrt{N}\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}_0)\right\}. \tag{22}$$

Now, the asymptotic normality of $\hat{\boldsymbol{\theta}}_N$ will follow if $\mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\tilde{\boldsymbol{\theta}})$ in (22) converges appropriately, and if the vector $\sqrt{N}\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}_0)$ has the central limit property. Under appropriate regularity conditions as described in Sinha (2004) (see also White, 1982), it can be shown that $\hat{\boldsymbol{\theta}}_N \to \boldsymbol{\theta}_0$ a.s. and $\| \mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{G}}'_N(\boldsymbol{\theta}_0) \| \to 0$ a.s. as $N \to \infty$. Also, we can show that $\sqrt{N}\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}_0)$ is asymptotically $N\{\mathbf{0}, \mathbf{Q}_N(\boldsymbol{\theta}_0)\}$, where $\mathbf{Q}_N(\boldsymbol{\theta}_0) = \text{var}\left\{\sqrt{N}\mathbf{G}_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta}_0)\right\}$. Then from (22), we can argue that

$$\sqrt{N}\mathbf{Q}_N(\boldsymbol{\theta}_0)^{-1/2}\mathbf{M}_N(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \stackrel{.}{\sim} N(\mathbf{0}, \mathbf{I}), \tag{23}$$

where $\mathbf{M}_N(\boldsymbol{\theta}) = -\bar{\mathbf{G}}_N(\boldsymbol{\theta}) = -E\{\mathbf{G}'_N(\mathbf{y}, \mathbf{v}|\boldsymbol{\theta})\}$. The asymptotic variance-covariance matrix of the robust estimators $\hat{\boldsymbol{\theta}}_N$ may be obtained from

$$\mathbf{V}_N(\boldsymbol{\theta}_0) = \mathbf{M}_N^{-1}(\boldsymbol{\theta}_0)\,\mathbf{Q}_N(\boldsymbol{\theta}_0)\,\mathbf{M}_N^{-1}(\boldsymbol{\theta}_0). \tag{24}$$

The computational aspects of this variance-covariance matrix are discussed in an illustrative example in the next section.

4. ILLUSTRATIVE EXAMPLE

Here we describe the computational issues of the proposed robust estimator using a simple example. Suppose in a clinical study, y_{it} represents a binary response from individual i at time t. To describe the success probability p_{it} as a function of time t and a baseline covariate x_i for individual i, consider a marginal logistic regression model in the form

$$\begin{aligned} y_{it} &\sim \text{Bernoulli}(p_{it}),\ i = 1, \ldots, N,\ t = 1, \ldots, T, \\ \eta_{it} &= \text{logit}(p_{it}) = \beta_0 + \beta_1 x_i + \beta_2(t-1). \end{aligned} \tag{25}$$

Define $v_{it} = 1$ if the response y_{it} is observed and $v_{it} = 0$ if y_{it} is missing. Assume that the marginal distribution of the binary random variable v_{it} is Bernoulli, with the probability of being observed,

$$\pi_{it} = P(v_{it} = 1 | y_{it}, x_i, \boldsymbol{\tau}) = \frac{\exp(\tau_0 + \tau_1 x_i + \tau_2 y_{it})}{1 + \exp(\tau_0 + \tau_1 x_i + \tau_2 y_{it})} \tag{26}$$

Here the parameters of interest are the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^t$, with $\boldsymbol{\tau} = (\tau_0, \tau_1, \tau_2)^t$ being considered as nuisance parameters of the missing data model.

For the robust estimators of the regression parameters $\boldsymbol{\beta}$, the iterative equations (17) can be expressed in the form

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left(\sum_{i=1}^{N} \mathbf{X}_i^t \mathbf{W}_i \mathbf{D}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^t \mathbf{W}_i \mathbf{d}_i, \tag{27}$$

where the second term on the right side is evaluated at $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\tau}^{(m)}$, $\mathbf{X}_i$ is the design matrix for subject i, and $\mathbf{W}_i$ is a diagonal matrix with diagonal elements $\sigma_{it} w_{it}, t = 1, \ldots, T$. Also, $\mathbf{d}_i$ is a $T \times 1$ vector with elements

$$d_{it} = v_{it} \tilde{d}_{it} + (1 - v_{it}) E_{y|v}(\tilde{d}_{it} | v_{it}), \tag{28}$$

where $\tilde{d}_{it} = \psi_c(r_{it}) - E_y\{\psi_c(r_{it})\}$ and $\mathbf{D}_i$ is a diagonal matrix with diagonal elements

$$D_{it} = -v_{it} \frac{\partial \tilde{d}_{it}}{\partial \eta_{it}} - (1 - v_{it}) E_{y|v} \left(\left. \frac{\partial \tilde{d}_{it}}{\partial \eta_{it}} \right| v_{it} \right), \tag{29}$$

for $t = 1, \ldots, T$. To calculate D_{it} in (29), it can be shown that

$$\frac{\partial}{\partial \eta_{it}} \psi_c(r_{it}) = -\sqrt{p_{it}(1 - p_{it})} \psi_c'(r_{it}) - (1/2 - p_{it}) \psi_c'(r_{it}) r_{it}.$$

Also, it can be shown that

$$\begin{aligned} \frac{\partial}{\partial \eta_{it}} E_y\{\psi_c(r_{it})\} &= E_y \left\{ \frac{\partial}{\partial \eta_{it}} \psi_c(r_{it}) \right\} + \sqrt{p_{it}(1 - p_{it})} E_y\{r_{it} \psi_c(r_{it})\} \\ &= -\sqrt{p_{it}(1 - p_{it})} E_y\{\psi_c'(r_{it})\} - (1/2 - p_{it}) E_y\{\psi_c'(r_{it}) r_{it}\} \\ &\quad + \sqrt{p_{it}(1 - p_{it})} E_y\{r_{it} \psi_c(r_{it})\}. \end{aligned}$$

Then we have,

$$\begin{aligned} &\frac{\partial \tilde{d}_{it}}{\partial \eta_{it}} = \frac{\partial}{\partial \eta_{it}} \{\psi_c(r_{it}) - E_y[\psi_c(r_{it})]\} \\ &= -\sqrt{p_{it}(1 - p_{it})} \{\psi_c'(r_{it}) - E_y[\psi_c'(r_{it})]\} - (1/2 - p_{it}) \{r_{it} \psi_c'(r_{it}) - E_y[r_{it} \psi_c'(r_{it})]\} \\ &\quad - \sqrt{p_{it}(1 - p_{it})} E_y[r_{it} \psi_c(r_{it})]. \end{aligned} \tag{30}$$

For the robust estimators of the nuisance parameters $\boldsymbol{\tau}$, the iterative equations (18) can be expressed in the form

$$\boldsymbol{\tau}^{(m+1)} \quad = \quad \boldsymbol{\tau}^{(m)} + \left[\sum_{i=1}^{N} \{ \mathbf{Z}_i^t \mathbf{V}_i \mathbf{W}_i^* \mathbf{D}_i^* \mathbf{Z}_i + E_{y|v} (\mathbf{Z}_i^t \bar{\mathbf{V}}_i \mathbf{W}_i^* \mathbf{D}_i^* \mathbf{Z}_i) \} \right]^{-1}$$

$$\times \sum_{i=1}^{N} \{\mathbf{Z}_i^t \mathbf{V}_i \mathbf{W}_i^* \mathbf{d}_i^* + E_{y|v}(\mathbf{Z}_i^t \bar{\mathbf{V}}_i \mathbf{W}_i^* \mathbf{d}_i^*)\}, \tag{31}$$

where $\mathbf{Z}_i$ is a design matrix with rows $\mathbf{z}_{it}$, $\mathbf{V}_i$ and $\bar{\mathbf{V}}_i$ are diagonal matrices with diagonal elements v_{it} and $1-v_{it}$ respectively, $\mathbf{W}_i^*$ is a diagonal matrix with diagonal elements $\sigma_{it}^* w_{it}^*$, $\mathbf{d}_i^*$ is a vector with elements $d_{it}^* = \psi_c(r_{it}^*) - E_{v|y}\{\psi_c(r_{it}^*)\}$ and $\mathbf{D}_i^*$ is a diagonal matrix with diagonal elements $D_{it}^* = -\partial d_{it}^* / \partial \eta_{it}^*$ with $\eta_{it}^* = \mathbf{z}_{it}^t \boldsymbol{\tau}$. The derivatives in D_{it}^* can be obtained using similar arguments as shown in (28)–(30).

Equations (27) and (31) are solved iteratively until convergence for the RPL estimators $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^t, \hat{\boldsymbol{\tau}}^t)^t$. An approximate variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ may be obtained from (24) as

$$\text{Var}(\hat{\boldsymbol{\theta}}) \approx \mathbf{M}^{-1}(\hat{\boldsymbol{\theta}})\, \mathbf{Q}(\hat{\boldsymbol{\theta}})\, \mathbf{M}^{-1}(\hat{\boldsymbol{\theta}}), \tag{32}$$

where the matrices $\mathbf{M}(\boldsymbol{\theta})$ and $\mathbf{Q}(\boldsymbol{\theta})$ may be partitioned as:

$$\mathbf{M}(\boldsymbol{\theta}) = \begin{pmatrix} M_{\beta\beta} & M_{\beta\tau} \\ M_{\tau\beta} & M_{\tau\tau} \end{pmatrix},$$

and

$$\mathbf{Q}(\boldsymbol{\theta}) = \begin{pmatrix} Q_{\beta\beta} & Q_{\beta\tau} \\ Q_{\tau\beta} & Q_{\tau\tau} \end{pmatrix},$$

with

$$\begin{aligned} M_{\beta\beta} &= -\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}\sigma_{it} w_{it} \mathbf{x}_{it}\, \mathbf{x}_{it}^t \\ &\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it}) E_{y|v}\{\tilde{d}_{it}(y_{it}-p_{it})|v_{it}\}\sigma_{it} w_{it} \mathbf{x}_{it}\, \mathbf{x}_{it}^t \\ &\quad - \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it}) E_{y|v}(\tilde{d}_{it}|v_{it}) E_{y|v}\{(y_{it}-p_{it})|v_{it}\}\sigma_{it} w_{it} \mathbf{x}_{it}\, \mathbf{x}_{it}^t, \end{aligned}$$

$$\begin{aligned} M_{\beta\tau} &= \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it}) E_{y|v}\{d_{it}^*\sigma_{it}^* w_{it}^* (y_{it}-p_{it})\mathbf{z}_{it}|v_{it}\}\, \mathbf{x}_{it}^t \\ &\quad - \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it}) E_{y|v}(d_{it}^*\sigma_{it}^* w_{it}^* \mathbf{z}_{it}|v_{it}) E_{y|v}\{(y_{it}-p_{it})|v_{it}\}\, \mathbf{x}_{it}^t, \end{aligned}$$

$$\begin{aligned} M_{\tau\beta} &= \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it})\sigma_{it} w_{it}\, \mathbf{x}_{it} E_{y|v}\{\tilde{d}_{it}(v_{it}-\pi_{it})\mathbf{z}_{it}^t|v_{it}\} \\ &\quad - \sum_{i=1}^{N}\sum_{t=1}^{T} (1-v_{it})\sigma_{it} w_{it}\, \mathbf{x}_{it} E_{y|v}\{\tilde{d}_{it}|v_{it}\} E_{y|v}\{(v_{it}-\pi_{it})\mathbf{z}_{it}^t|v_{it}\}, \end{aligned}$$

$$M_{\tau\tau} = -\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it} D_{it}^* \sigma_{it}^* w_{it}^* \mathbf{z}_{it}\, \mathbf{z}_{it}^t$$

$$
\begin{aligned}
&-\sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}(D^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}\,\mathbf{z}^t_{it}|v_{it})\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\{d^*_{it}(v_{it}-\pi_{it})\sigma^*_{it}w^*_{it}\mathbf{z}_{it}\,\mathbf{z}^t_{it}|v_{it}\}\\
&-\sum_{i=1}^{N}\sum_{t=1}^{T}(1-v_{it})E_{y|v}\{d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}|v_{it}\}E_{y|v}\{(v_{it}-\pi_{it})\mathbf{z}^t_{it}|v_{it}\}.
\end{aligned}
$$

Also,

$$
Q_{\beta\beta}=\sum_{i=1}^{N}\left(\sum_{t=1}^{T}d_{it}w_{it}\sigma_{it}\mathbf{x}_{it}\right)\left(\sum_{t=1}^{T}d_{it}w_{it}\sigma_{it}\mathbf{x}^t_{it}\right),
$$

$$
Q_{\beta\tau}=\sum_{i=1}^{N}\left(\sum_{t=1}^{T}d_{it}w_{it}\sigma_{it}\mathbf{x}_{it}\right)\left[\sum_{t=1}^{T}\{v_{it}d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}+(1-v_{it})E_{y|v}(d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}|v_{it})\}\right]^t,
$$

$$
\begin{aligned}
Q_{\tau\tau} \;=\; &\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\{v_{it}d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}+(1-v_{it})E_{y|v}(d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}|v_{it})\}\right]\\
&\times\left[\sum_{t=1}^{T}\{v_{it}d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}+(1-v_{it})E_{y|v}(d^*_{it}\sigma^*_{it}w^*_{it}\mathbf{z}_{it}|v_{it})\}\right]^t,
\end{aligned}
$$

and $Q_{\tau\beta}=Q^t_{\beta\tau}$.

We carried out a simulation study to investigate the empirical properties of the RPL estimators based on the above example. The simulation results are discussed in the next section.

5. SIMULATION STUDY

To explore the performance of the proposed RPL estimators, we ran two sets of simulations. In the first set, the estimators were studied for the case when no outliers were considered in the data. In the second set, they were studied in the presence of design outliers.

For each simulation run, we generated a series of 1000 data sets, each of size $N=200$, using a Bahadur (1961) type multivariate binary model for three longitudinal outcomes, y_{i1}, y_{i2} and y_{i3}, with joint probabilities

$$
P(y_{i1},y_{i2},y_{i3}|x_i,\boldsymbol{\beta},\boldsymbol{\alpha})=\left\{\prod_{t=1}^{3}p_{it}^{y_{it}}(1-p_{it})^{1-y_{it}}\right\}\left\{1+\sum_{st}\alpha_{st}z_{is}z_{it}+\alpha_{123}z_{i1}z_{i2}z_{i3}\right\} \tag{33}
$$

where $z_{it}=(y_{it}-p_{it})/\sqrt{p_{it}(1-p_{it})}$; $\alpha_{st}=\mathrm{corr}(y_{is},y_{it})=E[z_{is}z_{it}|x_i]$; $\alpha_{123}=E[z_{i1}z_{i2}z_{i3}|x_i]$; and $\mathrm{logit}(p_{it})=\beta_0+\beta_1x_i+\beta_2(t-1)$, for $t=1,2,3$. We consider $\boldsymbol{\alpha}=(\alpha_{12},\alpha_{13},\alpha_{23},\alpha_{123})^t$ as the vector of association parameters. The regression coefficients were fixed at $\beta_0=0.5$, $\beta_1=0.25$ and $\beta_2=-0.25$. The values of the covariate x were assumed to follow a $N(2,1)$ distribution. The data were generated under three correlation structures: 1) Uncorrelated; 2) an exchangeable correlation structure, with $\alpha_{st} = \alpha$ and $\alpha_{123} = 0$; and 3) a serial correlation structure, with $\alpha_{st} = \alpha^{|t-s|}$ and $\alpha_{123} = 0$, for $t, s = 1, 2, 3$. Note that here we use the Bahadur (1961) model just to generate data with correlated binary responses. The Bahadur representation is attractive in that the marginal means of the binary responses can be easily derived from the multivariate binary distribution. However, a serious drawback is that the correlations among the binary responses are constrained by the marginal means in a complicated manner. In our robust approach, we estimate the model parameters by treating the observations independent.

Without loss of generality, we assumed that all individuals were observed at the first time-point, but incomplete data were obtained at the second and third time-points according to the probability of being observed,

$$\pi_{it} = P(v_{it} = 1|x_i, y_{it}, \tau_0, \tau_1, \tau_2) = \frac{\exp(\tau_0 + \tau_1 x_i + \tau_2 y_{it})}{1 + \exp(\tau_0 + \tau_1 x_i + \tau_2 y_{it})}, \tag{34}$$

for $t = 2, 3$. The parameters of this missing data model were fixed at $\tau_0 = -2$, $\tau_1 = 1$ and $\tau_2 = 1$ for which roughly 38% missing data occurred at each of the second and third time-points.

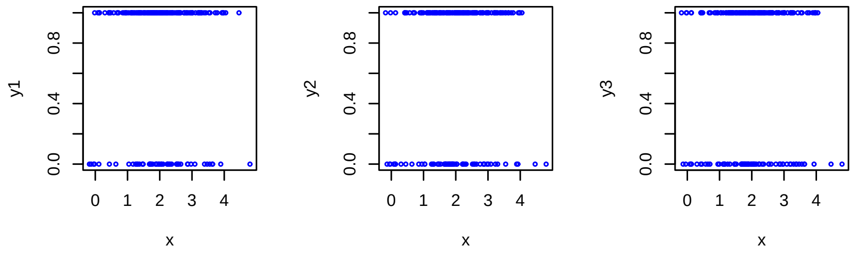

Figure 1
Scatter Plots of Longitudinal Binary Outcomes, y_1, y_2 and y_3, Against Covariate x, with no Outliers

Figure 1 exhibits representative scatter plots of the longitudinal binary outcomes, y_1, y_2 and y_3, over three observation times ($t = 1, 2, 3$), against the covariate x when the data are generated from the multivariate binary model (33) under an exchangeable correlation structure with $\alpha = 0.4$. As expected, the missing data occurred at the second and third observation times for lower values of x and y.

The proposed RPL estimates of the regression parameters $(\beta_0, \beta_1, \beta_2)$ and the nuisance parameters (τ_0, τ_1, τ_2) of the missing data model were obtained from the iterative equations (27) and (31) using the tuning constants $c = 1.2$ for the Huber's psi function $\psi_c(r)$, and $\gamma_0 = 2$ for the weight function w as defined in (16). Typically, these tuning constants are chosen so as to provide a certain level efficiency at the underlying distributions. However, it is difficult to obtain an optimal set of values for the tuning constants analytically due to the complex nature of the robust estimators under incomplete binary data. We explore different sets of tuning constants for the RPL estimators and finally chose the values $c = 1.2$ and $\gamma_0 = 2$ for which the RPL estimators of the regression parameters provided roughly 95% efficiency over the ordinary pseudo-likelihood (IPL) estimators under correctly specified models. Note that for the choice $c = \infty$ and $\gamma_0 = 0$, the RPL estimators lead to the non-robust IPL estimators.

Table 1
Empirical Biases of RPL and IPL Estimators for Data with no Outliers

Parameter	True value	Independent	Exchangeable $\alpha = 0.2$	Exchangeable $\alpha = 0.4$	Serial $\alpha = 0.2$	Serial $\alpha = 0.4$
Robust method (RPL)						
β_0	0.50	0.0184	0.0095	0.0004	–0.0057	0.0048
β_1	0.25	–0.0037	0.0012	0.0063	0.0079	0.0007
β_2	–0.25	0.0060	0.0036	–0.0029	–0.0121	0.0088
Classical method (IPL)						
β_0	0.50	0.0194	0.0093	0.0017	0.0027	0.0008
β_1	0.25	–0.0035	0.0014	0.0059	0.0044	0.0033
β_2	–0.25	0.0070	0.0026	–0.0017	–0.0093	0.0061

Table 1 presents the empirical biases of both the RPL and IPL estimators under different correlation structures. It is clear from the table that both methods provide roughly unbiased estimators of the regression parameters β_0, β_1 and β_2. Tables 2 presents the empirical mean squared errors of the RPL and IPL estimators. The RPL method appears to lose small efficiency in terms of slightly larger mean squared errors of the regression estimators. For example, under the correctly specified "independent" model, the RPL estimator of β_1 is 92.2% as efficient as the corresponding IPL estimator; under the exchangeable correlation with $\alpha = 0.2$, the efficiency of the RPL estimator of β_1 is 96.6%. Also, under the serial correlation with $\alpha = 0.4$, the efficiency of the RPL estimator of β_1 is 93.4%. We would expect to lose such small efficiencies from the robust method when there are, in fact, no outliers in the data. However, our focus is on the robust analysis of data in the presence of outliers. In the next step, we investigate the performance of the estimators under outliers.

We ran the second set of simulations by contaminating the data with a small proportion of outliers. As before, a series of 1000 data sets, each of size $N = 200$, were generated from the multivariate binary model (33). The values of the binary random variable v_{it} were generated from the missing data model (34). Note that in the case of binary responses, outliers in the data can arise only through the design points x. After generating each data set, we created a few outliers by randomly moving a small proportion of the design points from the bulk of the data. Specifically, to create these outliers, we randomly replaced 5% of the x values in the original data by $x + 5$. This type of contamination generally produces mean-shift outliers in the data. Figure 2 exhibits representative scatter plots of the longitudinal binary outcomes, y_1, y_2 and y_3, over three observation times ($t = 1, 2, 3$), against the covariate x when data are contaminated with design outliers. In these plots, the outliers are indicated by the large values of the covariate x for which the values of y are 0.

Table 2
Empirical Mean Squared Errors of RPL and IPL Estimators for Data with no Outliers

Parameter	True value	Independent	Exchangeable $\alpha = 0.2$	Exchangeable $\alpha = 0.4$	Serial $\alpha = 0.2$	Serial $\alpha = 0.4$
Robust method (RPL)						
β_0	0.50	0.1045	0.1071	0.1200	0.1075	0.1282
β_1	0.25	0.0180	0.0205	0.0246	0.0210	0.0244
β_2	–0.25	0.0431	0.0367	0.0319	0.0428	0.0387
Classical method (IPL)						
β_0	0.50	0.0967	0.1036	0.1127	0.1020	0.1179
β_1	0.25	0.0166	0.0198	0.0229	0.0199	0.0228
β_2	–0.25	0.0405	0.0338	0.0299	0.0410	0.0362

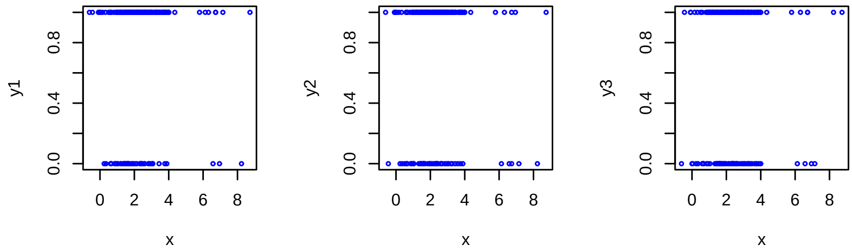

Figure 2
Scatter Plots of Longitudinal Binary Outcomes, y_1, y_2 and y_3, Against Covariate x for Data Contaminated with Design Outliers

Table 3
Empirical Biases of RPL and IPL Estimators for Data with Outliers

Parameter	True value	Independent	Exchangeable $\alpha = 0.2$	Exchangeable $\alpha = 0.4$	Serial $\alpha = 0.2$	Serial $\alpha = 0.4$
Robust method (RPL)						
β_0	0.50	0.0948	0.0910	0.0948	0.0873	0.0839
β_1	0.25	–0.0436	–0.0429	–0.0446	–0.0395	–0.0389
β_2	–0.25	0.0404	0.0407	0.0253	0.0440	0.0387
Classical method (IPL)						
β_0	0.50	0.2725	0.2746	0.2676	0.2678	0.2639
β_1	0.25	–0.1401	–0.1417	–0.1376	–0.1371	–0.1353
β_2	–0.25	0.0775	0.0781	0.0668	0.0804	0.0734

Table 3 presents the empirical biases of the RPL and IPL estimators for data under design outliers. It is clear from the table that the ordinary IPL estimators of the regression parameters generally produce much larger biases, as compared to the RPL estimators. For example, under independence, the RPL estimator of β_1 produces a small bias of -0.0436, whereas the IPL estimator of β_1 produces a much larger bias of -0.1401. Also, as clear from Table 4, the empirical mean squared errors of the IPL estimators are almost uniformly larger than those of the RPL estimators. For example, under independence, the RPL estimator of β_1 has a mean squared error of 0.0195, whereas the IPL estimator of β_1 has a larger mean squared error of 0.0288. This demonstrates the usefulness of the proposed robust method in bounding the influence of potential outliers in the data.

Table 4
Empirical Mean Squared Errors of RPL and IPL Estimators for Data with Outliers

Parameter	True value	Independent	Exchangeable		Serial	
			$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.2$	$\alpha = 0.4$
Robust method (RPL)						
β_0	0.50	0.1109	0.1216	0.1348	0.1196	0.1227
β_1	0.25	0.0195	0.0220	0.0254	0.0214	0.0224
β_2	–0.25	0.0437	0.0350	0.0290	0.0451	0.0393
Classical method (IPL)						
β_0	0.50	0.1468	0.1552	0.1608	0.1500	0.1524
β_1	0.25	0.0288	0.0311	0.0314	0.0295	0.0305
β_2	–0.25	0.0416	0.0346	0.0284	0.0430	0.0356

6. APPLICATION: ANALYSIS OF AIDS DATA

Here we present an analysis of the CD4 count data from the AIDS clinical trials described in the Introduction. The parameters are estimated by using the proposed robust approach, and Troxel *et al.*'s (1998) independent pseudo-likelihood approach under the assumption of nonignorable missingness.

Our study involves 431 patients who were diagnosed with AIDS or AIDS-related complex. All patients were observed at the baseline period $t = 0$. Among the predictors considered in the study, "Azt" is defined to be 1 if a patient is randomized to treatment Azt, and 0 if he/she is randomized to treatment Ddi; "Age" is defined to be 1 if the patient is 35 or older at baseline period, and 0 otherwise; and "BaseCD4" is defined as BaseCD4 $= \sqrt{\text{baseline CD4 count}}/10$. Note that since the covariates Azt and Age are both binary, any outliers in the data can arise only through the covariate BaseCD4. Figure 3 displays the histogram of the observed values of BaseCD4 and the corresponding scatter plot of weights w used in our robust analysis. The right tail area of the histogram indicates some extreme values in BaseCD4, which are downweighted by the weight function w, as shown in the scatter plot on the right panel.

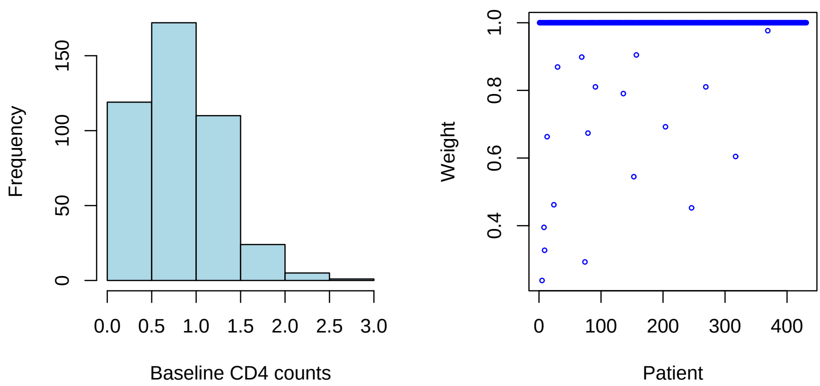

Figure 3
Histogram of Baseline CD4 Counts (BaseCD4) and Corresponding Scatter Plot of Weights w Used in Robust Analysis. Design Outliers Are Indicated by Small Weights

In this study, the response of interest is normal CD4 cell count ($>$ 200 cells per cubic millimeter) versus abnormal CD4 cell count ($\leq$ 200) measured at every week for up to 5 weeks from baseline (t=1, ..., 5); the outcome is defined as $y_{it} = 1$ if the CD4 count exceeds 200 and 0 otherwise. The cutoff of 200 cells per cubic millimeter was chosen because of its strong predictive value for the 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.

We model the "success probability", $p_{it} = P(y_{it} = 1)$, as a function of the covariates using the logistic regression:

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \text{BaseCD4}_i + \beta_2 \text{Age}_i + \beta_3 \text{Azt}_i + \beta_4 t,$$

for $t = 1, ..., 5$. We also model the probability of being observed at each follow-up time assuming 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 consider the missing data model:

$$\begin{aligned}\text{logit}(\pi_{it}) &= \text{logit}\{P(v_{it} = 1|\mathbf{x}_{it}, y_{it}, \boldsymbol{\tau})\} \\ &= \tau_0 + \tau_1 \text{BaseCD4}_i + \tau_2 \text{Age}_i + \tau_3 \text{Azt}_i + \tau_4 t + \tau_5 y_{it},\end{aligned}$$

for $t = 1, ..., 5$. Table 5 reports the RPL and IPL estimates of the model parameters with their approximate standard errors and corresponding z values. The RPL and IPL estimates appear to be generally close to each other for the regression parameters. However, the standard errors of the two sets of estimators are different to some extent. For example, the RPL estimator of the intercept term β_0 has a standard error of 1.2048, whereas the IPL estimator has a standard error of 0.9014; for β_1, the RPL and MPL methods provide standard errors of 0.7714 and 0.6073, respectively. The smaller standard errors from the IPL method may be justified by the fact that when there are outliers in the data in one direction (that is, when the outlying residuals are either all positives or all negatives), an ordinary non-robust method generally underestimates the standard errors.

From the robust analysis of the data, the CD4 counts appear to decrease over time. But there is no evidence of treatment effects, which indicates that the effects of the two treatments Azt and Ddi are not significantly different. The estimates of the parameters of the missing data model from the two methods are similar. The probability of a response decreases over time under both methods. This probability, however, increases when possibly unobserved response y increases from 0 to 1.

Table 5
Analysis of AIDS Data

Variable	RPL Estimate	RPL Std error	RPL z value	IPL Estimate	IPL Std error	IPL z value
Regression model						
Intercept (β_0)	–10.3243	1.2048	–8.570	–9.6419	0.9014	–10.696
BaseCD4 (β_1)	6.5015	0.7714	8.428	6.2027	0.6073	10.213
Age (β_2)	1.4716	0.4028	3.654	1.0772	0.3663	2.941
Azt (β_3)	0.1351	0.4128	0.327	0.4751	0.3491	1.361
Time (β_4)	–0.2254	0.0794	–2.840	–0.2330	0.0749	–3.110
Missing data model						
Intercept (τ_0)	1.4484	0.2493	5.810	1.4415	0.2487	5.795
BaseCD4 (τ_1)	1.0552	0.3095	3.409	1.0796	0.3140	3.438
Age (τ_2)	–0.2596	0.1685	–1.541	–0.1835	0.1638	–1.120
Azt (τ_3)	–0.0960	0.1737	–0.553	–0.0956	0.1692	–0.565
Time (τ_4)	–0.3107	0.0345	–8.995	–0.3196	0.0352	–9.075
y (τ_5)	1.3475	1.6650	0.809	0.3581	0.8467	0.423

We also investigate the residuals from the robust fit to identify any outliers in the data. Figure 4 displays scatter plots of the standardized residuals $(y_{it} - \hat{p}_{it})/\sqrt{\hat{p}_{it}(1-\hat{p}_{it})}$ for the observed data at the time-points $t = 1, \ldots, 5$. These plots show some large residuals at each time-point, which may have influenced the ordinary IPL estimates and their standard errors.

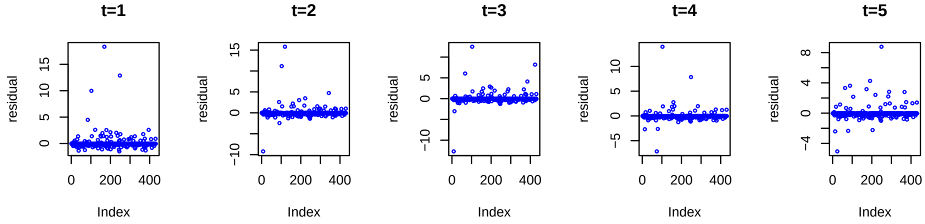

Figure 4
Standardized Residuals from the Robust Analysis of AIDS Data

7. DISCUSSION

We have developed the proposed RPL method to provide protection against outliers in the data. We have briefly studied the asymptotic properties of the RPL estimators. The empirical properties of these estimators were studied in simulations. The simulation results indicate that the RPL method is almost as efficient as the IPL method for data with no outliers. But the gain in efficiency from this robust method is generally large when data are contaminated with outliers.

In the case of binary longitudinal data, no outliers are desired in the binary responses. But still outliers can arise in the standardized residuals as these can be arbitrarily large even for a binary model. The proposed robust method can be used to bound the influence of both outliers in the residuals and leverage points in the design space.

Note that for data with nonignorable missing responses, we assumed an independent binary logistic model to describe the missing data mechanism. It is, however, not clear how the robust estimators would behave under a misspecified missing data model. A sensitivity analysis could be performed to investigate the effects model misspecification. Work remains to be done in this direction.

ACKNOWLEDGMENTS

The author is grateful for the support provided by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Bahadur, R.R. (1961). A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (Ed.), *Studies in item analysis and prediction* (pp. 158–68). Stanford University Press.

[2] Brown, C.H. (1990). Protecting against nonrandomly missing data in longitudinal studies. *Biometrics, 46*, 143–157.

[3] Cantoni, E., & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association, 96*, 1022–1030.

[4] Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics, 4*, article 14.

[5] Diggle, P., & Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics, 43*, 49–94.

[6] Finkelstein, D.M., Williams, P.L., Molenberghs, G., Feinberg, J., Powderly, W., Kahn, J., Dolins, R., & Cotton, D. (1996). Patterns of opportunistic infections in patients with HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 12*, 38–45.

[7] Gallant, J.E., Moore, R.D., Richman, D.D., Keruly, J., & Chaisson, R.E. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. *Journal of Infectious Diseases, 166*, 1223–1227.

[8] Ibrahim, J.G., Chen, M.H., & Lipsitz, S.R. (2001). Missing responses in gener-

alized linear mixed models when the missing data mechanism is nonignorable. *Biometrika, 88*, 551–564.

[9] Ibrahim, J.G., Lipsitz, S. R., & Chen, M.H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society (Ser. B), 61*, 173–190.

[10] Kahn, J.O., Lagakos, S.W., & Richman, D.D. (1992). A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. *New England Journal of Medicine, 327*, 581–587.

[11] Le Cessie, & Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Applied Statistics, 43*, 95–108.

[12] Liang, K.-Y., & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

[13] Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd ed.). New Jersey: John Wiley & Sons.

[14] Meester, S.G., & MacKay, J. (1994). A parametric model for cluster correlated categorical data. *Biometrics, 50*, 954–963.

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

[16] Plackett, R.M. (1965). A class of bivariate distributions. *Journal of the American Statistical Association, 60*, 516–522.

[17] Preisser, J.S., & Qaqish, B.F. (1999). Robust regression to clustered data with application to binary responses. *Biometrics, 55*, 574–579.

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

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

[20] Rousseeuw, P.J., & van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association, 85*, 633–639.

[21] Sinha, S.K. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association, 99*, 451–460.

[22] Sinha, S.K. (2008). Robust methods for generalized linear models with nonignorable missing covariates. *The Canadian Journal of Statistics, 36*(2), 277–299.

[23] Sinha, S.K., Laird, N.M., & Fitzmaurice, G.M. (2010). Multivariate logistic regression with incomplete covariate and auxiliary information. *Journal of Multivariate Analysis, 101*, 2389–2397.

[24] Sinha, S.K., Troxel, A. B., Lipsitz, S.R., Sinha, D., Fitzmaurice, G.M., Molenberghs, G., & Ibrahim, J.G. (2011). A bivariate pseudo-likelihood for incomplete longitudinal binary data with nonignorable non-monotone missingness. *Biometrics, 67*, 1119–1126.

[25] Troxel, A.B., Lipsitz, S.R., & Harrington, D.P. (1998). Marginal models for the analysis of longitudinal measurements subject to nonignorable non-monotone missing data. *Biometrika, 85*, 661–672.

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