Regression analysis of longitudinal data with irregular and informative observation times

Yong Chen

Division of Biostatistics,

The University of Texas, Health Science Center at Houston,

School of Public Health, Houston, Texas 77030, USA

yong.chen@uth.tmc.edu

Phone: 713-500-9569,

Jing Ning

Department of Biostatistics,

The University of Texas M. D. Anderson Cancer Center,

Houston, Texas 77030, USA

Chunyan Cai

Department of Biostatistics,

The University of Texas M. D. Anderson Cancer Center,

Houston, Texas 77030, USA

Authors' Footnote: Yong Chen is Assistant Professor, Division of Biostatistics, The University of Texas, Health Science Center at Houston, TX 77030 (yong.chen@uth.tmc.edu); Jing Ning is Assistant Professor, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, TX 77030 (jing.ning@uth.tmc.edu); and Chunyan Cai is Ph.D. candidate, Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX 77030 (Chunyan.Cai@uth.tmc.edu).

Yong Chen and Jing Ning were partially supported by start-up funds from the University of Texas School of Public Health and the University of Texas MD Anderson Cancer Center respectively.

SUMMARY

In longitudinal data analyses, the observation times are often assumed to be independent of the outcomes. Such assumption can be violated in some applications. In these applications, the standard inferential approach using a generalized estimating equation may lead to biased inference. The current methods require the correction specification of either the observation time process or the repeated measures process with correct covariance structure. In this article, we construct a novel pairwise pseudolikelihood method for longitudinal data that allows dependence between observation times and outcomes. This method incorporates both time-dependent and time-independent covariates, while leaving the observation time process unspecified. The novelty of this method is that neither specification of the observation time process nor the covariance structure of repeated measures process is required. Large sample properties of the regression coefficient estimates and pseudolikelihood-ratio test statistic are established. Simulation studies demonstrate that the proposed method performs well in finite samples and is robust to the true form of the observation time process. An analysis of weight loss data from a web-based program is presented to illustrate the proposed method.

Key words: Informative observation times; Irregular observation times; Longitudinal data analysis; Pairwise likelihood

1. INTRODUCTION

In many clinical trials and observational studies, subjects are followed over a period of time and are often scheduled to be assessed at a common set of pre-specified visit times. In practice, however, the actual observation times may deviate considerably from the schedule and possibly vary from subject to subject. As a result, the observation times can be highly irregular. Furthermore, the observation times are possibly outcome dependent. For example, in a recent study for evaluating the effectiveness of a web-based weight loss program, the self-reported times are subject-specific and the data structure is highly imbalanced. Moreover, the self-reported times of a member may be correlated with the weight of that member. For instance, the overweight/obese participants of the weight loss program are more likely to lose weight and are more willing to report their weight to the web-based program than normal weight participants. This outcome dependent self-report mechanism can cause the web-based weight loss program to appear more successful than it really is. In general, when the observation times are correlated with the responses, the standard inferential approach using a generalized estimating equation (GEE) of Liang and Zeger (1986) may lead to biased inferences (Lin et al., 2004). Therefore, careful consideration is needed to enable valid inference of covariate effects on the longitudinal responses. A great deal of attention has been devoted to this problem.

In this regard, at least two general statistical approaches have been suggested in the literature. One is the likelihood-based approach. In general, this approach requires full parametric specification for the joint distribution of the repeated measure process and the observation time process. The dependence between these two processes are often characterized by shared random effects (Liu et al., 2008; Liu and Huang, 2009; Liu, 2009) or latent class variables (Han et al., 2007). The inference on the regression coefficients is then based on the likelihood of observed data. One exception that full specification for joint distribution of two processes is not needed is the models proposed by Lipsitz et al. (2002) and Fitzmaurice et al. (2006). The authors took a transition modeling strategy by specifying the whole repeated measure process and its covariance structure. The mathematical relationship between the parameters in marginal models and that in transition models is then established case by case, for example, for multivariate Gaussian data (Lipsitz et al., 2002) and repeated binary data (Fitzmaurice et al., 2006). The second general approach is the estimating equation approach. Instead of specifying the distribution of repeated measure or observation time process, the estimating equation approach directly models the marginal mean of response variable conditional on covariates. The dependence between two possibly correlated processes is accounted for either by shared latent variables (Sun et al., 2005, 2007; Liang et al., 2009; Zhu et al., 2011) or by incorporating an inverse intensity-of-visit weights in the estimating equation (Rotnitzky and Robins, 1995; Robins, 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999; Lin et al., 2004).

These approaches have their own strength given particular data structure and the appropriateness of the underlying assumptions. However, all of the forementioned methods require either modeling the observation time process, parametrically or semiparametrically, or having the correct specification of the repeated measures process with covariance structure. In Section 2, we propose an alternative estimating procedure for regression coefficients in which neither specification of the observation time process nor the covariance structure of repeated measures process is required. The asymptotic properties of the proposed procedure are established in Section 3. In Section 4 we discuss extension to density ratio models. Section 5 presents its finite sample performance through simulation studies followed by a real data example in Section 6. We provide brief discussion in Section 7.

2. Method

Consider a longitudinal study with a random sample of n subjects. For the ith subject, let $Y_i(t)$ be the response variable at time t and let $\mathbf{X_i(t)}$ be a p-dimensional vector of possibly time-dependent covariates. The response variable $Y_i(t)$ is evaluated at sequential time points $t_{i1} < t_{i2} < \ldots < t_{iK_i}$, where K_i is the total number of observations on the ith subject. The number of observations from the ith subject up to time t is $N_i(t) = \sum_{j=1}^{K_i} I(t_{ij} \leq t)$, where $I(A)$ is the indicator function and equals 1 if A is true and 0 otherwise. The observation times are regarded as realizations from an underlying counting process $N_i^*(t)$ that is censored at the end of follow-up. Specifically, $N_i(t) = N_i^*(t \wedge C_i)$, where C_i is the noninformative censoring time, and $a \wedge b = \min(a, b)$. The process $Y_i(t)$ is observed only at the jump points of $N_i(t)$, i.e., $\mathrm{d}N_i(t) = 1$ where the derivative is taken respect to the counting measure. For simplicity of notation, denote $y_{ij} = Y_i(t_{ij})$ and $\mathbf{x_{ij}} = \mathbf{X_i}(\mathbf{t_{ij}})$ for $j = 1, 2, \ldots, K_i$ and $i = 1, 2, \ldots, n$. The observed data are denoted by $\{y_{i1}, y_{i2}, \ldots, y_{iK_i}, \mathbf{x_{i1}}, \mathbf{x_{i2}}, \ldots, \mathbf{x_{iK_i}}, t_{i1}, t_{i2}, \ldots, t_{iK_i}, C_i\}$ for $i = 1, 2, \ldots, n$.

We are interested in the association between covariates of interest $\mathbf{X_i}(\mathbf{t})$ and the response variable $Y_i(t)$. This association can be studied by regression analysis through modeling the conditional distribution of y_{ij} given $\mathbf{x_{ij}}$, i.e., $f(y_{ij}|\mathbf{x_{ij}}; \boldsymbol{\beta})$. However, in practice, such conditional approach may not be feasible when the observation times depend on the response variables. For example, in the web-based weight loss study, the probability of conducting self-report at a certain time for a subject may depend on that subject's weight at that time. Specifically, at time t, the probability of $\mathrm{d}N_i(t) = 1$ may depend on $Y_i(t)$. In this case, the conditional probability of observing y_{ij} given $\mathrm{d}N_i(t_{ij}) = 1$ and $\mathbf{x_{ij}}$ is

$$\Pr(y_{ij}|\mathrm{d}N_i(t_{ij}) = 1, \mathbf{x_{ij}}) = \frac{f(y_{ij}|\mathbf{x_{ij}}; \boldsymbol{\beta})\Pr(\mathrm{d}N_i(t_{ij}) = 1|y_{ij}; \boldsymbol{\gamma})}{\Pr(\mathrm{d}N_i(t_{ij}) = 1|\mathbf{x_{ij}}; \boldsymbol{\beta}, \boldsymbol{\gamma})}, \tag{1}$$

where $\Pr(\mathrm{d}N_i(t_{ij}) = 1|\mathbf{x_{ij}}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \int f(y|\mathbf{x_{ij}}; \boldsymbol{\beta})\Pr(\mathrm{d}N_i(t_{ij}) = 1|y; \boldsymbol{\gamma})\,\mathrm{d}y$. This conditional probability depends on both parameters of interest, $\boldsymbol{\beta}$, and nuisance parameter, $\boldsymbol{\gamma}$, which specifies the dependence of the observation time process on the response process. Therefore, unless the observation time process is assumed to be independent of response measurement process conditional on covariates, i.e. $\{\mathrm{d}N_i(t) = 1\} \perp Y_i(t)|\mathbf{X_i}(\mathbf{t})$, in which the above conditional probability reduces to $f(y_{ij}|\mathbf{x_{ij}}; \boldsymbol{\beta})$, valid inference based on the above conditional probability, in general, requires the correct specification of the observation time process, i.e. $\Pr(\mathrm{d}N_i(t_{ij}) = 1|y_{ij}; \boldsymbol{\gamma})$.

Now we propose an alternative estimating procedure for the parameter of interest $\boldsymbol{\beta}$ in which no correct specification of observation time process is required. The strategy is to use pairwise conditional probability, which was first introduced by Kalbfleisch (1978) in the setting of nonparametric tests and later studied by Liang and Qin (2000) in regression analysis of cross-sectional data. Denote $a(y_{ij}; \boldsymbol{\beta}, \boldsymbol{\gamma}) =$

$\Pr(\mathrm{d}N_i(t_{ij}) = 1|y_{ij})$ and $b(\mathbf{x_{ij}};\boldsymbol{\gamma}) = \{\Pr(\mathrm{d}N_i(t_{ij}) = 1|\mathbf{x_{ij}})\}^{-1}$, the equation (1) can be written as, $\Pr(y_{ij}|\mathrm{d}N_i(t_{ij}) = 1, \mathbf{x_{ij}}) = a(y_{ij};\boldsymbol{\beta},\boldsymbol{\gamma})b(\mathbf{x_{ij}};\boldsymbol{\gamma})f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})$.

Consider two observations from a pair of subjects: the jth observation of subject i and the j'th observation of subject i'. The conditional probability of the responses at the jth time point for subject i and at the j'th time point for subject i', $(y_{ij}, y_{i'j'})$, given their order statistic $(y^{(1)}, y^{(2)})$, is calculated as,

$$\begin{aligned}
& \Pr(y_{ij}, y_{i'j'}|y^{(1)}, y^{(2)}, dN_i(t_{ij}) = 1, dN_{i'}(t_{i'j'}) = 1, \mathbf{x_{ij}}, \mathbf{x_{i'j'}}) \\
= & \frac{a(y_{ij};\boldsymbol{\beta},\boldsymbol{\gamma})a(y_{i'j'};\boldsymbol{\beta},\boldsymbol{\gamma})b(\mathbf{x_{ij}};\boldsymbol{\gamma})b(\mathbf{x_{i'j'}};\boldsymbol{\gamma})f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{i'j'}};\boldsymbol{\beta})}{a(y^{(1)};\boldsymbol{\beta},\boldsymbol{\gamma})a(y^{(2)};\boldsymbol{\beta},\boldsymbol{\gamma})b(\mathbf{x_{ij}};\boldsymbol{\gamma})b(\mathbf{x_{i'j'}};\boldsymbol{\gamma})\{f(y^{(1)}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y^{(2)}|\mathbf{x_{i'j'}};\boldsymbol{\beta}) + f(y^{(1)}|\mathbf{x_{i'j'}};\boldsymbol{\beta})f(y^{(2)}|\mathbf{x_{ij}};} \\
= & \frac{f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{i'j'}};\boldsymbol{\beta})}{f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{i'j'}};\boldsymbol{\beta}) + f(y_{ij}|\mathbf{x_{i'j'}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{ij}};\boldsymbol{\beta})} \\
= & \Big\{1 + R(y_{ij}, \mathbf{x_{ij}}, y_{i'j'}, \mathbf{x_{i'j'}};\boldsymbol{\beta})\Big\}^{-1},
\end{aligned}$$

where

$$R(y_{ij}, \mathbf{x_{ij}}, y_{i'j'}, \mathbf{x_{i'j'}};\boldsymbol{\beta}) = \frac{f(y_{ij}|\mathbf{x_{i'j'}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{ij}};\boldsymbol{\beta})}{f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{i'j'}};\boldsymbol{\beta})}.$$

Notice that both $a(y_{ij};\boldsymbol{\beta},\boldsymbol{\gamma})$ and $b(\mathbf{x_{ij}};\boldsymbol{\gamma})$ are eliminated through conditioning on the order statistic.

For each possible pair of subjects (i, i'), and for any two respective jumps of $N_i(t)$ and $N_{i'}(t)$, we can calculate the above conditional probability. Multiplying these probabilities together, we obtain the following pairwise pseudolikelihood function for all observations,

$$L_p(\boldsymbol{\beta}) = \prod_{i<i'}\prod_{j=1}^{K_i}\prod_{j'=1}^{K_{i'}} \left[\left\{1 + \frac{f(y_{ij}|\mathbf{x_{i'j'}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{ij}};\boldsymbol{\beta})}{f(y_{ij}|\mathbf{x_{ij}};\boldsymbol{\beta})f(y_{i'j'}|\mathbf{x_{i'j'}};\boldsymbol{\beta})}\right\}^{-1}\right].$$

Equivalently, the corresponding log pairwise pseudolikelihood can be represented in stochastic process notation as,

$$\log L_p(\boldsymbol{\beta}) = \sum_{i<i'}\int_0^\tau\int_0^\tau \left[-\log\left\{1+\frac{f(y_i(s)|\mathbf{x_{i'}}(t);\boldsymbol{\beta})f(y_{i'}(t)|\mathbf{x_i}(s);\boldsymbol{\beta})}{f(y_i(s)|\mathbf{x_i}(s);\boldsymbol{\beta})f(y_{i'}(t)|\mathbf{x_{i'}}(t);\boldsymbol{\beta})}\right\}\right]\mathrm{d}N_i(s)\mathrm{d}N_{i'}(t),$$

where τ is the maximum follow-up time. To obtain parameter estimates, we maximize the log pairwise pseudolikelihood or solve the score equations by setting the first derivative of $\log L_p(\boldsymbol{\beta})$ to zero. Notice that the above likelihood function $L_p(\boldsymbol{\beta})$ is not a true likelihood function because the probabilities in equation (2) are multiplied together without accounting for the possible correlations between the repeated observations of the same subject. Nevertheless, since each term in $L_p(\boldsymbol{\beta})$ is a legitimate conditional probability, the corresponding score equation is an unbiased estimating equation for $\boldsymbol{\beta}$.

3. Asymptotic behavior

As mentioned in last section, pairwise pseudolikelihoods cannot be treated the same as likelihood functions for the purpose of making inferences about parameters. Instead, the asymptotic behavior of $\hat{\boldsymbol{\beta}}$, i.e., a solution of $\partial \log L_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = 0$, must be investigated. Now we establish the asymptotic results of $\hat{\boldsymbol{\beta}}$.

Lemma 1 *Assuming that* $\{y_i(t), \mathbf{x_i}(t), N_i(t)\}$, $i = 1, 2, \ldots, n$, *are independent and identically distributed, then under the regularity conditions* $\Re$ *in Chernoff (1954), with probability tending to* 1 *as* $n \to \infty$ *there exists a sequence of points,* $\hat{\boldsymbol{\beta}}$, *at which local maxima of* $\log L_p(\boldsymbol{\beta})$ *occur, and that converges to* $\boldsymbol{\beta}_0$, *the true* $\boldsymbol{\beta}$*-value, in probability. Moreover,* $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1)$.

A sketch of the proof is given in Appendix. Theorem 1 states the large sample distribution of $\hat{\boldsymbol{\beta}}$.

Theorem 1 *Assuming that* $\{y_i(t), \mathbf{x_i}(t), N_i(t)\}$, $i = 1, 2, \ldots, n$, *are independent and identically distributed, then under the regularity conditions* $\Re$ *in Chernoff (1954),* $\hat{\boldsymbol{\beta}}$ *is asymptotically normal with mean* $\boldsymbol{\beta}_0$ *and covariance* $\mathbf{V} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$, *where*

$$\boldsymbol{\Sigma}_1 = -E\Big(- \frac{\partial \boldsymbol{\psi}_{12}}{\partial \boldsymbol{\beta}}; \boldsymbol{\beta}_0 \Big),$$

$$\boldsymbol{\Sigma}_2 = 4cov(\boldsymbol{\psi}_{12}, \boldsymbol{\psi}_{13}; \boldsymbol{\beta}_0),$$

and

$$\boldsymbol{\psi}_{ii'} = -\int_0^\tau \int_0^\tau \frac{\partial \log\{1 + R(y_i(s), \mathbf{x}_i(s), y_{i'}(t), \mathbf{x}_{i'}(t); \boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}} \, dN_i(s) dN_{i'}(t).$$

A sketch of the proof is given in Appendix. The variance covariance matrix $\mathbf{V}$ can be estimated by $\hat{\boldsymbol{\Sigma}}_{\mathbf{1}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{2}}\hat{\boldsymbol{\Sigma}}_1^{-1}$, where

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{1}} = \binom{n}{2}^{-1} \sum_{i<i'} \frac{\partial \boldsymbol{\psi}_{ii'}}{\partial \beta}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}},$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{2}} = \frac{4}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{\substack{i',i''\neq i \\ i'\neq i''}} \boldsymbol{\psi}_{ii'} \boldsymbol{\psi}_{ii''}{}^{\mathrm{T}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.$$

The above results on asymptotic distribution of $\hat{\boldsymbol{\beta}}$ are useful for Wald-based inference. However, it has been well documented in the literature that such inference could be ill-behaved, and that this concern can be alleviated by using likelihood-ratio-based inference (Hauck and Donner, 1977). To this end, we establish the results for the pairwise pseudolikelihood ratio based inference. In practice, we are often interested in testing for one coefficient of interest or a subset of the parameters $\boldsymbol{\beta}$. In general, consider the partition of $\boldsymbol{\beta}$ into two parts: the parameters to be tested, $\boldsymbol{\theta}$, of dimension p_1 and the remaining parameters, $\boldsymbol{\eta}$, of dimension $p-p_1$. The null hypothesis of interest is specified as $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Let T be the pairwise pseudolikelihood ratio test statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e. $T = -2\log\{L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0))/L_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})\}$ where $\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) = \arg\max_{\boldsymbol{\eta}} L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta})$ and $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\eta}} L_p(\boldsymbol{\theta}, \boldsymbol{\eta})$. Theorem 2 states the large sample distribution of T under H_0.

Theorem 2 *If the regularity conditions* $\Re$ *in Chernoff (1954) hold, then*

1. *the matrix* $\left\{\boldsymbol{\Sigma}_1^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{1,\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \end{pmatrix}\right\}\boldsymbol{\Sigma}_{\mathbf{2}}$ *has* p_1 *positive eigenvalues,* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{p_1} > 0$*, and* $p - p_1$ *zero eigenvalues;*

2. *the pairwise pseudolikelihood ratio,* T*, is distributed as* $U = \sum_{j=1}^{p_1} \lambda_j U_j$ *where* U_j*s are independent* χ_1^2 *variables.*

where

$$\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}\boldsymbol{\eta}} = -E\Big(-\frac{\partial \boldsymbol{\phi}_{12}}{\partial \boldsymbol{\eta}}; \boldsymbol{\beta}_0\Big)$$

and

$$\boldsymbol{\phi}_{ii'} = -\int_0^{\tau}\int_0^{\tau} \frac{\partial \log\{1 + R(y_i(s), \mathbf{x}_i(s), y_{i'}(t), \mathbf{x}_{i'}(t); \boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\eta}} \, dN_i(s) dN_{i'}(t).$$

A sketch of the proof is given in Appendix. For the most common situation where we are testing for one coefficient of interest, i.e. $p_1 = 1$, the asymptotic distribution of T is a weighted χ_1^2.

4. Extension to density ratio models

In this section, we assume a class of more flexible semiparametric models instead of parametric models described above to evaluate the association between covariates $\mathbf{X_i(t)}$ and the response variable $Y_i(t)$. Specifically, we consider the density ratio models (Qin, 1998; Gilbert et al., 1999; Wang et al., 2011; Luo and Tsai, 2011), which uses a log-linear (i.e., exponential multiplier) model for the density ratio, leaving the baseline density unspecified.

The density ratio model assumes that

$$f(y|\mathbf{x}) = \frac{f_0(y)\exp(y\boldsymbol{\beta}^T\mathbf{x})}{\int f_0(y)\exp(y\boldsymbol{\beta}^T\mathbf{x})\ dy}, \tag{3}$$

where $f_0(\cdot)$ is the unspecified density function of the subgroup with covariates $\mathbf{x} = \mathbf{0}$. Under model (3), for two individuals with outcome and covariate values of $(y_i, \mathbf{x_i})$ and $(y_{i'}, \mathbf{x_{i'}})$, we have

$$\frac{f(y_i|\mathbf{x_i})/f(y_{i'}|\mathbf{x_i})}{f(y_i|\mathbf{x_{i'}})/f(y_{i'}|\mathbf{x_{i'}})} = \exp\{-(y_{i'} - y_i)\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x_{i'}} - \mathbf{x_i})\}.$$

This implies that the regression coefficients $\boldsymbol{\beta}$ characterize the effect of covariates on the "generalized odds" of y through its density function (Liang and Qin, 2000).

Following the arguments in Section 2, we can adopt the pairwise pseudolikelihood method to estimate the regression coefficients $\boldsymbol{\beta}$ in model (3). Recall that by equation (2), the conditional probability of the responses at the jth time point for subject i and at the j'th time point for subject i', $(y_{ij}, y_{i'j'})$, given their order statistic $(y^{(1)}, y^{(2)})$, is

$$\begin{aligned}
& \Pr(y_{ij}, y_{i'j'}|y^{(1)}, y^{(2)}, dN_i(t_{ij}) = 1, dN_{i'}(t_{i'j'}) = 1, \mathbf{x_{ij}}, \mathbf{x_{i'j'}}) \\
= & \left\{1 + \frac{f(y_{ij}|\mathbf{x_{i'j'}})f(y_{i'j'}|\mathbf{x_{ij}})}{f(y_{ij}|\mathbf{x_{ij}})f(y_{i'j'}|\mathbf{x_{i'j'}})}\right\}^{-1} \\
= & \Big[1 + \exp\{(y_{ij} - y_{i'j'})\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x_{ij}} - \mathbf{x_{i'j'}})\}\Big]^{-1}.
\end{aligned} \tag{4}$$

This conditional probability depends on the regression coefficients $\boldsymbol{\beta}$ but not the nonparametric component $f_0(\cdot)$. The corresponding pairwise pseudolikelihood for $\boldsymbol{\beta}$ can be constructed by the product of all possible conditional probabilities. As a result, the regression coefficients $\boldsymbol{\beta}$ can be estimated by maximizing the pairwise pseudolikelihood, without estimating the nonparametric component $f_0(\cdot)$.

5. Simulation

We conduct simulation studies to evaluate the performance of the proposed method and compare it with the GEE method. For all simulation settings, we generate 1000 samples, each with 200 or 400 independent subjects. We consider regression models with four covariates: continuous variable X_1 generated from uniform distribution on $(0, 10)$, Bernoulli variable X_2 with probability 0.6 and time-dependent variables t and t^2. Two types of outcomes are considered: continuous and binary. The probability of observing y_i at time t is set as $\text{logit}\{\Pr(\text{d}N_i(t_{ij}) = 1|y_{ij}; \boldsymbol{\gamma})\} = \gamma_0 + \gamma_1 y_{ij} + \gamma_2 y_{ij}^2$, where $\text{logit}(a) = \log\{a/(1-a)\}$. For each type of outcome, we consider a relatively "weak" dependence and a "strong" dependence between the observation times and outcomes, described by the parameters $\boldsymbol{\gamma}$. For continuous outcomes, we generate the correlated outcomes from a random effect model with a normally distributed random effects. For binary outcomes, we generate the multivariate correlated binary outcomes from the Bahadur model (Bahadur, 1961; Emrich and Piedmonte, 1991).

We compare the results from the proposed method to the GEE method. The empirical bias, average of asymptotic standard error estimates, empirical standard error and the coverage probability of the confidence intervals based on 1000 samples of continuous outcomes and binary outcomes are shown in Tables 1 and 2 respectively. When the sample size is 200, the estimates of regression coefficients β from the proposed method perform well in that the biases of the estimates of β are small and the coverage probabilities are close to the nominal level of 95%. On the other hand, the GEE method, which requires independence between the observation times and outcomes, results in biased estimation as well as poor coverage probabilities. As the strength of the dependence between the observation times and outcomes increases from "weak" to "strong", the biases of the proposed method remain small whereas the biases of the GEE method increase and the coverage properties of the confidence intervals from the GEE method deteriorate. As the sample size increases to 400, the biases of the proposed method become smaller and the coverage probabilities of the proposed method are closer to the nominal level, whereas the biases of the GEE method remain large, leading to worse coverage probabilities. In summary, the simulation studies suggest that under all settings considered here, the proposed method performs well and is robust to the true form of the observation time process, whereas the GEE method can have substantial bias, leading to poor coverage probabilities.

6. Application

Over the past decades, the dramatic increase in the incidence of obesity becomes a worldwide health issue, which contributes significantly to the incidence of many diseases. As of 2010, around one-third of U.S. adults are obese, while another one-third is overweight (Wang et al., 2011). People seeking assistance with weight control can choose from a wide array of weight loss programs including the free web-based weight loss programs. Although currently there is no universally effective method of weight management that assures long-term maintenance of lost weight (Levy et al., 2007), many individuals who wish to avoid weight gain participate in web-based weight loss programs such as Sparkpeople (`http://sparkpeople.com`). Recently, the effectiveness of such web-based weight loss programs has been intensively evaluated, and it was claimed that greater weight loss is likely to be associated with increased use of web-based program features (Kelders et al., 2011; Fiona and Jane, 2011; Nijland et al., 2011; Neve et al., 2011). However, the aforementioned evaluation is focused on the weight loss, defined as the difference between the last and first self-reported weight and ignores other intermediate information on weight over time. To make full use of the collected data, our objective is to evaluate the effect of the usage of the weight loss program on weight by including longitudinal data and use the proposed pairwise pseudo-likelihood for inference to account for the possible outcome dependent self-report mechanism.

Sparkpeople is an online free weight loss program with over 8 million registered members. We consider a subset of SparkPeople data by including records collected in 2008 from participants who enrolled in 2008 and have at leat two self-reported weight. As a result, a total of $26,937$ self-reported weight from $2,153$ participants were included in our analysis. In summary, participants in the analyzed data have a mean (SD) age of 35.5(11.3) years, a mean (SD) BMI of 32.0(7.6) and are predominately female (92.0%). The number of self-report weight for each participant has a mean (SD) of 12.5(15.8), ranging from 2 to 168. It is not surprising to observe that the self-reported times of a participant depend on the weight of that participant. Specifically, the subgroup analysis and associated t-tests suggest that the overweight/obese participants are more likely to lose weight and have more frequent self-reported weight than participants with normal weight.

A longitudinal linear regression model is used to evaluate the effect of usage of the program, measured by the number of logins up to each self-report time, on the self-reported weight. To control for potential confounders, additional covariates such as gender, BMI at enrollment, age at enrollment, and months in the program are included in the model. The effect sizes with corresponding standard errors and p-values from the proposed pairwise pseudolikelihood method are listed in Table 3. Interestingly, the results in Table 3 suggest that the Sparkpeople program may produce weight loss over the first 4.2 months after the enrollment, suggested by the negative effect of the months in the program adjusting for the usage of the Sparkpeople program and the baseline participant characteristics. However, the positive coefficient of time square implicates such weight loss could diminish over time and the long-term maintenance of weight loss is a hard task. More importantly, the results in Table 3 suggest that the usage of the Sparkpeople program, measured by the login number, can significantly reduce the weight (p-value = 0.018). After controlling for the time effects, participants with 10 more logins are expected to lose 0.5 more pounds compared to participants with similar baseline characteristics. For comparison, we also present the estimated regression coefficients by the GEE method which ignores the outcome dependent self-report mechanism. As expected, the GEE method is likely to overestimate the effect of the usage of the Sparkpeople program by ignoring the informative self-reported processes.

7. Discussion

In this paper, we proposed a pseudolikelihood based method for regression analysis of longitudinal data when the observation times are possibly outcome dependent. The proposed method has the advantage that no correct specification of observation time process is required. Our simulation studies confirmed its robustness as the resulting estimates for the regression coefficients are unbiased regardless the true form of the observation time process. However, it is worth to point out that if there is no or weak dependence between the observation times and outcomes, the standard inferential approach using a generalized estimating equation (GEE) could be more efficient than the proposed method. It will be important to develop statistical tests for such dependence, which can be an interesting area for future research.

Appendix

Proof of Lemma 1

Proof: We first establish the consistency result and then the $n^{1/2}$-consistency result. For simplicity of notation, let $\boldsymbol{\beta}_0 = 0$. For any $\delta > 0$, since the intersection of the parameter space of $\boldsymbol{\beta}$ and the closure of a δ-neighborhood of origin is closed, $\log L_p(\boldsymbol{\beta})$ has a local maximum on this set. If we can show the maximum lies in the parameter space of $\boldsymbol{\beta}$ at a distance from origin less than δ with probability tending to 1, then the existence and consistency of $\hat{\boldsymbol{\beta}}$ will follow immediately. This can be shown by proving $\log L_p(\boldsymbol{\beta}) < \log L_p(0)$ with probability tending toward 1 for all $\boldsymbol{\beta}$ in its parameter space that are at a distance δ from the origin. Such fact can be established by Taylor expansion of $\log L_p(\boldsymbol{\beta})$ around 0. For simplicity, assume $\boldsymbol{\beta}$ is one dimensional, we have,

$$\binom{n}{2}^{-1}\{\log L_p(\delta) - \log L_p(0)\} = \binom{n}{2}^{-1}\frac{\partial L_p(0)}{\partial \beta}\delta + \frac{1}{2}\binom{n}{2}^{-1}\frac{\partial^2 L_p(\delta^*)}{\partial \beta^2}\delta^2,$$

where δ^* is a point between origin and δ.

Notice that as n goes to infinity,

$$\binom{n}{2}^{-1}\frac{\partial L_p(0)}{\partial \beta} \to 0 \quad \text{and} \quad \binom{n}{2}^{-1}\frac{\partial^2 L_p(\delta^*)}{\partial \beta^2} \to a < 0.$$

We have $\binom{n}{2}^{-1}\{\log L_p(\delta) - \log L_p(0)\} < 0$ with probability tending to 1.

To establish the $n^{1/2}$-consistency, we only need to show for any $\epsilon > 0$, there exists $N_\epsilon, M_\epsilon > 0$ such that under H_0, $Pr(|n^{1/2}\hat{\beta}| < M_\epsilon) > 1 - \epsilon$ for $n > N_\epsilon$. Notice that under the regularity conditions $\Re$ in Chernoff (1954), we have,

$$
\begin{aligned}
& \binom{n}{2}^{-1}\{\log L_p(\hat{\beta}) - \log L_p(0)\} \\
= & \binom{n}{2}^{-1}\Big\{\frac{\partial \log L_p(0)}{\partial \beta}\Big\}\hat{\beta} + \frac{1}{2}\Big\{\binom{n}{2}^{-1}\frac{\partial^2 \log L_p(0)}{\partial \beta^2}\Big\}^{-1}\hat{\beta}^2 + O_p(1)|\hat{\beta}^3|.
\end{aligned}
$$

For each $\epsilon > 0$, there is a sequence $c_{n\epsilon} \to 0$ and a $K_\epsilon > 0$ such that with probability greater than $1-\epsilon$,

$$
|\hat{\beta}| < c_{n\epsilon}, \quad \Big|\binom{n}{2}^{-1}\frac{\partial \log L_p(0)}{\partial \beta}\Big| < n^{-1/2}K_\epsilon, \quad \Big|\binom{n}{2}^{-1}\frac{\partial^2 \log L_p(0)}{\partial \beta^2} - \Sigma_1\Big| < c_{n\epsilon},
$$

and the term represented by $O_p(1)|\hat{\beta}^3|$ is less than $K_\epsilon|\hat{\beta}^3|$. When these inequalities are satisfied there is a K_ϵ^* such that

$$
\begin{aligned}
0 & < \binom{n}{2}^{-1}\{\log L_p(\hat{\beta}) - \log L_p(0)\} \\
& = \binom{n}{2}^{-1}\Big\{\frac{\partial \log L_p(0)}{\partial \beta}\Big\}\hat{\beta} + \frac{1}{2}\Big\{\binom{n}{2}^{-1}\frac{\partial^2 \log L_p(0)}{\partial \beta^2}\Big\}^{-1}\hat{\beta}^2 + O_p(1)|\hat{\beta}^3| \\
& < -\frac{1}{2}\Sigma_1\hat{\beta}^2 + K_\epsilon^*(|n^{-1/2}\hat{\beta}| + c_{n\epsilon}|\hat{\beta}^2|).
\end{aligned}
$$

Therefore, there is a K_ϵ^{**} such that $\hat{\beta} < K_\epsilon^{**}/n^{1/2}$ by positivity of the matrix Σ_1. The lemma follows.

Proof of Theorem 1

Proof: For given $\boldsymbol{\beta}$ and two possibly different time points s and t, define $R_{ii'}(s,t;\boldsymbol{\beta})$ as a function of a pair of $(p+1)$-dimensional random vectors $(\mathbf{x_i}(s), y_i(s))$ and $(\mathbf{x_{i'}}(t), y_{i'}(t))$ as follows,

$$
R_{ii'}(s,t;\boldsymbol{\beta}) = \frac{f(y_i(s)|\mathbf{x_{i'}}(t);\boldsymbol{\beta})f(y_{i'}(t)|\mathbf{x_i}(s);\boldsymbol{\beta})}{f(y_i(s)|\mathbf{x_i}(s);\boldsymbol{\beta})f(y_{i'}(t)|\mathbf{x_{i'}}(t);\boldsymbol{\beta})}.
$$

The log pseudolikelihood can be written as

$$
\log L_p(\boldsymbol{\beta}) = \sum_{i<i'}\int_0^\tau\int_0^\tau -\log\{1 + R_{ii'}(s,t;\boldsymbol{\beta})\}\,\mathrm{d}N_i(s)\mathrm{d}N_{i'}(t).
$$

Notice that $\hat{\boldsymbol{\beta}}$ is a solution of $\partial \log L_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}$. Apply the Taylor expansion of

$\partial \log L_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ around $\boldsymbol{\beta}_0$:

$$\begin{aligned}
\mathbf{0} &= \sum_{i<i'} \int_0^\tau \int_0^\tau \Big[-\frac{\partial \log\{1+R_{ii'}(s,t;\hat{\boldsymbol{\beta}})\}}{\partial \boldsymbol{\beta}} \Big] \mathrm{d}N_i(s)\mathrm{d}N_{i'}(t) \\
&= \sum_{i<i'} \int_0^\tau \int_0^\tau \Big[-\frac{\partial \log\{1+R_{ii'}(s,t;\boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}} \Big] \mathrm{d}N_i(s)\mathrm{d}N_{i'}(t) \\
&\quad + \Big(\sum_{i<i'} \int_0^\tau \int_0^\tau \Big[-\frac{\partial^2 \log\{1+R_{ii'}(s,t;\boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}^2} \Big] \mathrm{d}N_i(s)\mathrm{d}N_{i'}(t) \Big)^{\mathrm{T}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(n).
\end{aligned}$$

Therefore, we have

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{A_n}^{-1} n^{1/2} \mathbf{B_n} + O_p(n^{-1/2}), \tag{5}$$

where

$$\mathbf{A_n} = -\binom{n}{2}^{-1} \sum_{i<i'} \int_0^\tau \int_0^\tau \Big[-\frac{\partial^2 \log\{1+R_{ii'}(s,t;\boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}^2} \Big] \mathrm{d}N_i(s)\mathrm{d}N_{i'}(t),$$

and

$$\mathbf{B_n} = \binom{n}{2}^{-1} \sum_{i<i'} \int_0^\tau \int_0^\tau \Big[-\frac{\partial \log\{1+R_{ii'}(s,t;\boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}} \Big] \mathrm{d}N_i(s)\mathrm{d}N_{i'}(t).$$

Notice that $\mathbf{A_n}$ and $\mathbf{B_n}$ are U-statistics, and

$$\begin{aligned}
& E\Big[-\frac{\partial \log\{1+R_{ii'}(s,t;\boldsymbol{\beta}_0)\}}{\partial \boldsymbol{\beta}} \Big] \\
= \;& E\Big\{ \frac{\partial}{\partial \boldsymbol{\beta}} \log \Pr(y_i(s), y_{i'}(t) | y^{(1)}, y^{(2)}, \mathrm{d}N_i(s)=1, \mathrm{d}N_{i'}(t)=1, \mathbf{x_i}(s), \mathbf{x_{i'}}(t); \boldsymbol{\beta}_0) \Big\} \\
= \;& E\Big[E\Big\{ \frac{\partial}{\partial \boldsymbol{\beta}} \log \Pr(y_i(s), y_{i'}(t) | y^{(1)}, y^{(2)}, \mathrm{d}N_i(s)=1, \mathrm{d}N_{i'}(t)=1, \mathbf{x_i}(s), \mathbf{x_{i'}}(t); \boldsymbol{\beta}_0) | y^{(1)}, y^{(2)}, \\
& \qquad \mathrm{d}N_i(s)=1, \mathrm{d}N_{i'}(t)=1, \mathbf{x_i}(s), \mathbf{x_{i'}}(t) \Big\} \Big] = 0.
\end{aligned}$$

By theorem 10 of Lehmann and D'abrera (1975), $\mathbf{A_n} \to \Sigma_1$ in probability and $n^{1/2}\mathbf{B_n} \to N(0, \Sigma_2)$ as $n \to \infty$. The proof is then completed by recalling equation (5).

Proof of Theorem 2

We restate the first result in Theorem 2 as a general result in linear algebra.

Lemma 2 *Suppose the symmetric positive definite matrix* $\mathbf{A}$ *can be partitioned to* $\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^{\mathrm{T}} & \mathbf{A}_{22} \end{pmatrix}_{p\times p}$, *where* $\mathbf{A}_{11}$ *is a* $p_1 \times p_1$ *matrix and* $\mathbf{A}_{22}$ *is a* $p_2 \times p_2$ *matrix* $(p_1 + p_2 = p)$. *Denote* $\mathbf{B}$ *the* $p \times p$ *matrix defined by* $\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix}$. *Then*

1. *There exists a non-singular matrix* $\mathbf{P}$ *such that* $\mathbf{A}^{-1} = \mathbf{P}^{\mathrm{T}}\mathbf{P}$ *and* $\mathbf{B} = \mathbf{P}^{\mathrm{T}}\mathbf{D}\mathbf{P}$, *where* $\mathbf{D} = diag(\mathbf{0}_{p_1}, \mathbf{1}_{p_2})$;
2. *The matrix* $\mathbf{A}^{-1} - \mathbf{B}$ *has* p_1 *positive eigenvalues and* $p - p_1$ *zero eigenvalues;*
3. *For any positive definite matrix* $\mathbf{G}$, *the matrix* $(\mathbf{A}^{-1} - \mathbf{B})\mathbf{G}$ *has* p_1 *positive eigenvalues and* $p - p_1$ *zero eigenvalues.*

Proof: Let $\mathbf{Q} = \begin{pmatrix} \mathbf{I}_{(p_1\times p_1)} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{(p_2\times p_2)} \end{pmatrix}$, we have

$$\mathbf{A}^{-1} = \mathbf{Q}^{\mathrm{T}} \begin{pmatrix} (\mathbf{A}_{11}^{*})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix} \mathbf{Q}$$

and

$$\mathbf{B} = \mathbf{Q}^{\mathrm{T}} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix} \mathbf{Q}$$

where $\mathbf{A}_{11}^{*} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^{\mathrm{T}}$ is positive definite. Let $\mathbf{U}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{U}$ and $\mathbf{O}^{\mathrm{T}}\boldsymbol{\eta}\mathbf{O}$ be the eigendecomposition of $(\mathbf{A}_{11}^{*})^{-1}$ and $\mathbf{A}_{22}^{-1}$. Denote $\mathbf{S} = \mathbf{\Lambda}^{1/2}\mathbf{U}$ and $\mathbf{T} = \boldsymbol{\eta}^{1/2}\mathbf{O}$ and $\mathbf{P} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{pmatrix} \mathbf{Q} = \begin{pmatrix} \mathbf{S} & -\mathbf{S}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{T} \end{pmatrix}$. It is easy to show $\mathbf{A}^{-1} = \mathbf{P}^{\mathrm{T}}\mathbf{P}$ and $\mathbf{B} = \mathbf{P}^{\mathrm{T}}\mathbf{D}\mathbf{P}$, where $\mathbf{D} = diag(\mathbf{0}_{p_1}, \mathbf{1}_{p_2})$. Now we prove the second result in Theorem 2. **Proof:**

$$\begin{aligned} T &= 2\{\log L_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) - \log L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0))\} \\ &= 2\{\log L_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) - \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\} - 2\{\log L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)) - \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\}. \quad (6) \end{aligned}$$

To calculate the first term of eq. (6), note that

$$\begin{aligned} & 2\{\log L_p(\hat{\boldsymbol{\beta}}) - \log L_p(\boldsymbol{\beta}_0)\} \\ = \;& 2\Big\{\frac{\partial \log L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\Big\{\frac{\partial \log L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\Big\}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1) \\ = \;& (n^{1/2}\mathbf{B}_n)^{\mathrm{T}}\mathbf{A}_n^{-1}(n^{1/2}\mathbf{B}_n) + o_p(1), \end{aligned}$$

where the last step is due to equation (5). To calculate the second term of eq. (6), we expand the equation

$$\frac{\partial \log L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\eta}} = 0$$

about $\boldsymbol{\eta}_0$ and get

$$\sqrt{n}\Big\{\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}_0\Big\} = \Big\{-n^{-1}\frac{\partial^2 \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^2}\Big\}^{-1}\Big\{n^{-1/2}\frac{\partial \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\Big\} + o_p(1). \quad (7)$$

Expanding $\log L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0))$ around $\boldsymbol{\eta}_0$ and using eq. (7) yield

$$\begin{aligned} & 2\{\log L_p(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0)) - \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\} \\ = \ & 2\Big\{\frac{\partial \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\Big\}^{\mathrm{T}}\Big\{\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}_0\Big\} + \Big\{\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}_0\Big\}^{\mathrm{T}}\Big\{\frac{\partial^2 \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^2}\Big\}\Big\{\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}_0\Big\} + o_p(1) \\ = \ & \Big\{n^{-1/2}\frac{\partial \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\Big\}^{\mathrm{T}}\Big\{-n^{-1}\frac{\partial^2 \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^2}\Big\}^{-1}\Big\{n^{-1/2}\frac{\partial \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}\Big\} + o_p(1). \end{aligned}$$

Therefore T is asymptotically equivalent to $(n^{1/2}\mathbf{B}_n)^{\mathrm{T}}\mathbf{D}(n^{1/2}\mathbf{B}_n)$ where $\mathbf{D} = \mathbf{A}_n^{-1} - \mathbf{C}_n$ and $\mathbf{C}_n = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Big\{-n^{-1}\frac{\partial^2 \log L_p(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}^2}\Big\}^{-1} \end{pmatrix}$.

As n goes to infinity, $\mathbf{A}_n \to \Sigma_1$ in probability and $n^{1/2}\mathbf{B}_n \to N(0, \Sigma_2)$. The proof is completed following theorem 4.4.4 of Graybill (1976).

References

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

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist 25*(3), 573–578.

Emrich, L. and M. Piedmonte (1991). A method for generating high-dimensional multivariate binary variates. *The American statistician 45*(4), 302–304.

Fiona, J. and W. Jane (2011). The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss program: a retrospective analysis. *International Journal of Behavioral Nutrition and Physical Activity 8*, e83.

Fitzmaurice, G., S. Lipsitz, J. Ibrahim, R. Gelber, and S. Lipshultz (2006). Estimation in regression models for longitudinal binary data with outcome-dependent follow-up. *Biostatistics 7*(3), 469–485.

Gilbert, P., S. Lele, , and Y. Vardi (1999). Maximum likelihood estimation in semiparametric selection bias models with application to aids vaccine trials. *Biometrika 86*(1), 27–43.

Graybill, F. A. (1976). *Theory and application of the linear model*. Duxbury Press North Scituate, Massachusetts.

Han, J., E. Slate, and E. Peña (2007). Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in medicine 26*(29), 5285–5302.

Hauck, W. W. and A. Donner (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association 72*(360), 851–853.

Kalbfleisch, J. (1978). Likelihood methods and nonparametric tests. *Journal of the American Statistical Association 73*(361), 167–170.

Kelders, S. M., J. E. W. C. Van Gemert-Pijnen, A. Werkman, N. Nijland, and E. R. Seydel (2011). Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. *Journal of Medical Internet Research 13*, e32.

Lehmann, E. and H. D'abrera (1975). *Nonparametrics: statistical methods based on ranks*, Volume 351. Holden-Day San Francisco.

Levy, R., E. Finch, M. Crowell, N. Talley, and R. Jeffery (2007). Behavioral intervention for the treatment of obesity: strategies and effectiveness data. *Am J Gastroenterol 102*, 2314–2321.

Liang, K. and J. Qin (2000). Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(4), 773–786.

Liang, K. and S. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Liang, Y., W. Lu, and Z. Ying (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics 65*(2), 377–384.

Lin, H., D. Scharfstein, and R. Rosenheck (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(3), 791–813.

Lipsitz, S., G. Fitzmaurice, J. Ibrahim, R. Gelber, and S. Lipshultz (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics 58*(3), 621–630.

Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in medicine 28*(6), 972–986.

Liu, L. and X. Huang (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 58*(1), 65–81.

Liu, L., X. Huang, and J. O'Quigley (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics 64*(3), 950–958.

Luo, X. and W. Tsai (2011). A proportional likelihood ratio model. *Biometrika in press*.

Neve, M., P. J. Morgan, and C. E. s. Collin (2011). Weight change in a commercial web-based weight loss program and its association with website use: cohort study. *Journal of Medical Internet Research 13*, e83.

Nijland, N., J. E. W. C. van Gemert-Pijnen, S. M. Kelders, B. J. Brandenburg, and E. R. Seydel (2011). Factors Influencing the Use of a Web-Based Application for Supporting the Self-Care of Patients with Type 2 Diabetes: A Longitudinal Study. *Journal of Medical Internet Research 13*, e71.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika 85*(3), 619–630.

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

Rotnitzky, A. and J. Robins (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika 82*(4), 805–20.

Rotnitzky, A., J. Robins, and D. Scharfstein (1998). Semi parametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association 93*(444), 1321–1339.

Scharfstein, D., A. Rotnitzky, and J. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association 94*(448), 1096–1120.

Sun, J., D. Park, L. Sun, and X. Zhao (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association 100*(471), 882–889.

Sun, J., L. Sun, and D. Liu (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association 102*(480), 1397–1406.

Wang, C., Z. Tan, and T. A. Louis (2011). Exponential tilt models for two-group comparison with censored data. *Journal of Statistical Planning and Inference 141*, 1102–1117.

Wang, Y. C., K. McPherson, T. Marsh, S. L. Gortmaker, and M. Brown (2011). Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet 378*, 815–825.

Zhu, L., J. Sun, X. Tong, and S. Pounds (2011). Regression analysis of longitudinal data with informative observation times and application to medical cost data. *Statistics in medicine 30*(12), 1429–1440.

Table 1: Summary of 1000 simulations with data generated from normal distribution: estimated bias, standard errors and coverage probabilities of parameter estimates times 100.

			$\hat{\boldsymbol{\beta}}$				$\hat{\boldsymbol{\beta}}_{\text{GEE}}$			
n	$\boldsymbol{\gamma}$	$\boldsymbol{\beta}$	Bias	ASE	ESE	CP	Bias	ASE	ESE	CP
200	"weak"	β_1	-0.6	3.6	3.8	93.6	3.2	1.6	1.6	47.4
	(-1.40, 0.15, -0.03)	β_2	1.4	13.3	14.2	93.2	-2.6	9.9	10.4	92.8
		β_3	1.7	7.3	7.8	93.1	-1.6	4.4	4.3	93.5
		β_4	-0.3	1.3	1.4	92.7	0.7	0.5	0.5	70.7
	"strong"	β_1	-0.7	4.1	4.4	93.1	8.2	1.8	1.9	1.4
	(-1.00, 0.48, -0.08)	β_2	2.1	15.5	16.8	92.5	-5.3	10.7	11.2	90.4
		β_3	2.1	8.4	9.5	90.5	-0.7	5.4	5.5	93.9
		β_4	-0.4	1.6	1.8	90.1	1.2	0.7	0.7	61.7
400	"weak"	β_1	-0.2	2.5	2.6	95.1	3.2	1.1	1.1	18.6
	(-1.40, 0.15, -0.03)	β_2	0.5	9.2	9.5	95.1	-2.7	6.9	7.1	94.1
		β_3	0.6	5.2	5.3	95.3	-1.7	3.1	3.1	90.5
		β_4	-0.1	1.0	1.0	94.7	0.7	0.4	0.4	47.8
	"strong"	β_1	-0.2	2.9	2.9	95.0	8.2	1.3	1.3	0.0
	(-1.00, 0.48, -0.08)	β_2	0.5	11.0	11.2	94.6	-5.6	7.7	7.7	89.4
		β_3	0.8	5.9	6.1	94.9	-0.6	3.8	3.9	95.3
		β_4	-0.2	1.1	1.1	94.1	1.2	0.5	0.5	39.0

Bias, empirical bias; ASE, average of asymptotic standard error estimates; ESE, empirical standard error; CP, coverage probability.

Table 2: Summary of 1000 simulations with data generated from bernoulli distribution: estimated bias, standard errors and coverage probabilities of parameter estimates times 100.

			$\hat{\boldsymbol{\beta}}$				$\hat{\boldsymbol{\beta}}_{\text{GEE}}$			
n	$\boldsymbol{\gamma}$	$\boldsymbol{\beta}$	Bias	ASE	ESE	CP	Bias	ASE	ESE	CP
200	"weak"	β_1	0.1	2.8	2.9	94.7	-1.6	2.5	2.5	90.2
	(0.5, -1.0, 0)	β_2	-0.2	16.8	16.6	95.4	-6.7	16.0	15.6	93.4
		β_3	0.8	10.5	10.7	94.4	-14.1	7.6	7.7	52.0
		β_4	-0.1	0.9	0.9	93.8	1.1	0.7	0.7	65.4
	"strong"	β_1	0.0	2.9	2.9	95.1	-2.6	2.6	2.6	83.1
	(0.8, -1.6, 0)	β_2	-0.1	17.1	17.1	95.1	-10.1	16.2	15.9	91.4
		β_3	0.8	11.1	11.5	95.2	-22.9	7.7	8.0	17.3
		β_4	-0.1	1.0	1.0	94.7	1.8	0.7	0.8	32.7
400	"weak"	β_1	0.0	2.0	2.0	94.8	-1.6	1.8	1.8	86.7
	(0.5, -1.0, 0)	β_2	-0.3	11.9	11.7	95.9	-6.7	11.3	11.1	90.7
		β_3	0.2	7.5	7.5	94.0	-14.2	5.4	5.3	24.6
		β_4	0.0	0.7	0.7	94.2	1.1	0.5	0.5	38.8
	"strong"	β_1	0.1	2.1	2.0	95.6	-2.6	1.8	1.8	73.5
	(0.8, -1.6, 0)	β_2	-0.3	12.2	12.1	95.6	-10.4	11.5	11.5	85.9
		β_3	0.4	7.8	7.9	95.1	-23.0	5.5	5.4	2.3
		β_4	0.0	0.7	0.7	95.4	1.8	0.5	0.5	7.8

Bias, empirical bias; ASE, average of asymptotic standard error estimates; ESE, empirical standard error; CP, coverage probability.

Table 3: Summary of weight data analysis

	$\hat{\boldsymbol{\beta}}$			$\hat{\boldsymbol{\beta}}_{\text{GEE}}$		
Covariates	Est.	SE	p-value	Est.	SE	p-value
Login number	-0.052	0.017	0.002	-0.081	0.004	<0.001
Gender	-23.907	7.171	0.001	-13.468	1.085	<0.001
BMI	4.398	0.853	<0.001	5.909	0.032	<0.001
Age	-0.067	0.096	0.486	0.158	0.026	<0.001
Month	-0.258	0.385	0.502	1.198	0.164	<0.001
Month^2	0.031	0.014	0.027	-0.010	0.006	0.095