Title
Regression analysis of longitudinal data with outcome-dependent sampling and informative censoring.

Journal
Scandinavian Journal of Statistics, 46(3)

ISSN
0303-6898

Authors
Liu, Suyu
Chen, Yong
Ning, Jing
et al.

Publication Date
2019-09-01

DOI
10.1111/sjos.12373

Peer reviewed

HHS Public Access
Author manuscript
Scand Stat Theory Appl. Author manuscript; available in PMC 2020 February 17.

Published in final edited form as:
Scand Stat Theory Appl. 2019 September ; 46(3): 831–847. doi:10.1111/sjos.12373.

Regression analysis of longitudinal data with outcome-dependent sampling and informative censoring

Weining Shen[1], **Suyu Liu**[2], **Yong Chen**[3], **Jing Ning**[2]

[1]Department of Statistics, University of California, Irvine

[2]Department of Biostatistics, The University of Texas MD Anderson Cancer Center

[3]Department of Biostatistics and Epidemiology, The University of Pennsylvania

Abstract

We consider regression analysis of longitudinal data in the presence of outcome-dependent observation times and informative censoring. Existing approaches commonly require correct specification of the joint distribution of the longitudinal measurements, observation time process and informative censoring time under the joint modeling framework, and can be computationally cumbersome due to the complex form of the likelihood function. In view of these issues, we propose a semi-parametric joint regression model and construct a composite likelihood function based on a conditional order statistics argument. As a major feature of our proposed methods, the aforementioned joint distribution is not required to be specified and the random effect in the proposed joint model is treated as a nuisance parameter. Consequently, the derived composite likelihood bypasses the need to integrate over the random effect and offers the advantage of easy computation. We show that the resulting estimators are consistent and asymptotically normal. We use simulation studies to evaluate the finite-sample performance of the proposed method, and apply it to a study of weight loss data that motivated our investigation.

Keywords

Biased sampling; composite likelihood; informative censoring; joint modeling; time-varying covariate

---

1. INTRODUCTION

In medical follow-up and observational studies, longitudinal data are typically collected together with the duration of the observation window until a terminal event occurs. Two sampling issues commonly arise in analyzing the covariate effect on the longitudinal outcomes: (a) the longitudinal observation times are correlated with the longitudinal outcomes (Robins, 1995; Lipsitz et al., 2002), and (b) the terminal event is associated with the outcomes, which results in informative censoring (Wu and Carroll, 1988; Follmann and Wu, 1995; Little, 1995). Failure to take these biased sampling issues into account may result in misleading analytic results.

Our paper is motivated by a recent study of a web-based weight loss program. Obesity has become a worldwide health issue. In 2010, two-thirds of US adults were estimated to be obese or overweight (Wang et al., 2011). Among the available weight loss programs, web-based programs have gained popularity because of their flexibility and low cost. It is thus of an increasing need to evaluate the effectiveness of the web-based program on the weight loss over the time. During our motivating web-based weight loss program, a participant was able to voluntarily report his/her weight to the system at each login. A preliminary analysis revealed two interesting facts. First, there was a significant positive association between favorable weight outcomes and higher frequencies of weight entries (p-value < .01). This is not surprising because a participant who is making better progress in losing weight may be more willing to report his/her progress. Second, the length of participation in the program was strongly associated with weight loss. The average weight loss was significantly different between participants who dropped out of the study within 6 months and participants who stayed in the study longer than 6 months (two-sample t test, p-value < .001). Clearly, these two findings suggest that the self-reported weight data have the aforementioned issues of outcome-dependent sampling and informative censoring. A standard analysis such as generalized estimating equation (GEE) Liang and Zeger (1986) that ignores such biases may conclude that the program appears to be more successful than it actually is.

There is an extensive body of literature in regression analysis with outcome-dependent sampling and informative censoring. When the observation time is outcome-dependent, a common approach is based on likelihoods, where there is a need either to specify the joint distribution of the observation time process and repeated measurements (Liu, 2009; Han et al., 2007; Liang et al., 2009; Song et al., 2009) or to consider marginal or transition models (Lipsitz et al., 2002; Yi et al., 2011). Empirical likeihood approaches have also received success, see Zhou et al. (2002); Chan (2013) for examples. See Ding et al. (2017) for a review of outcome-dependent sampling design for time-to-event outcomes. Alternatively, the estimating equation approach has gained popularity by modeling the marginal mean of the response given covariates (Rotnitzky et al., 1998; Scharfstein et al., 1999; Lin et al., 2004; Sun et al., 2005, 2007). Some authors have also considered the extension to time-varying covariates (Song et al., 2012; Chen et al., 2015). When the censoring is informative, one useful approach is to first build a marginal model for the longitudinal or event variable and then consider a conditional model (Diggle and Kenward, 1994; Little, 1995). Another popular method is joint modeling, in which a shared random effect model is commonly used to characterize the relationship between repeated measurements and a time-to-event process (Self and Pawitan, 1992; Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997; Wang and Taylor, 2001; Liu and Ying, 2007). This idea has been extended to treat recurrent event time processes as well (Wang et al., 2001); see Tsiatis and Davidian (2004) for an excellent overview.

The main goal of this paper is to provide an alternative to the existing approaches for regression analysis of time-varying covariates in the situation in which the observation and censoring times are correlated with the longitudinal outcomes. The proposed method enjoys two nice properties: (a) it does not require the specification of the observation time process and repeated measurement process; and (b) it bypasses the need to model the random effect distribution and allows for easy computation as no integration over the random effect is needed. The main idea is based on a conditional approach using order statistics and the formulation of a composite likelihood function. Similar ideas have been discussed by several authors under different settings (Kalbfleisch, 1978; Liang and Qin, 2000; Chen et al., 2015). The inference proceeds by solving the maximization of the composite likelihood function, which is in a much simpler form compared with the functions used in the aforementioned methods (e.g., joint modeling). The resulting estimators are shown to be consistent and asymptotically normally distributed. The rest of this paper is organized as follows. We describe the methodology and present the asymptotic results in Section 2. We report simulation results in Section 3, and discuss an application to the weight loss program data analysis in Section 4. We provide proofs and technical details in the Appendix.

2. METHOD

We consider a longitudinal study with n subjects. For subject i, let $Y_i(t)$ be his/her response at time t and $\boldsymbol{X}_i(t)$ be a p-dimensional vector of time-varying covariates such as the usage of the weight loss program up to time t in our motivating example. We observe longitudinal outcomes at time points $t_{i1} < t_{i2} < \ldots < t_{iK_i}$ for subject i, where K_i is the total number of observations. Define the number of observations of subject i by $N_i(t) = \sum_{j=1}^{K_i} 1\!\mathrm{l}(t_{ij} \le t)$ up to time t, where 1l is the indicator function. The observation times can be viewed as realizations from an underlying counting process $N_i^*(t)$, which is censored at the end of follow-up. More specifically, let $N_i^*(t) = N_i(t \wedge C_i)$, where C_i is an informative censoring time, and $a \wedge b = \min(a, b)$. The process $Y_i(t)$ is observed when $C_i > t$ and $dN_i(t) = 1$, in which the derivative is taken with respect to the counting measure. For simplicity, denote $y_{ij} = Y_i(t_{ij})$, $N_{ij} = N_i(t_{ij})$, and $\boldsymbol{x}_{ij} = \boldsymbol{X}_i(t_{ij}) = (x_{ij1}, \ldots, x_{ijp})^T$ for $j = 1, \ldots, K_i$ and $i = 1, \ldots, n$.

The parameters of interest are the effects of the time-varying covariates $\boldsymbol{X}_i(t)$ on the response variable $Y_i(t)$. We extend the semiparametric proportional likelihood ratio model (Luo and Tsai, 2012) to a random effect model by allowing for a subject-specific effect ξ_i. Specifically, the density of Y_{ij} given the covariate $\boldsymbol{x}_{ij}$ and random effect ξ_i is

$$f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i) = \exp(\alpha_{ij} + y_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) f_0(y_{ij}/\xi_i), \tag{2.1}$$

where $f_0(\cdot)$ is an unspecified baseline density function with covariates $\boldsymbol{x}_{ij} = 0$, $\boldsymbol{\beta}$ is the parameter of interest that quantifies the subject-specific effects of the time-varying covariates, and α_{ij} is the normalizing constant, which is defined as

$$\alpha_{ij} = -\log \int \exp(y\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) f_0(y/\xi_i) dy .$$

The conditional density of the *observed* Y_{ij} given $C_i \ge t_{ij}$, $dN_{ij}^* = 1$, covariates $\boldsymbol{x}_{ij}$ and the random effect ξ_i is

$$f\left\{y_{ij} \mid dN_{ij}^* = 1, \boldsymbol{x}_{ij}, \xi_i, C_i > t_{ij}\right\} = \frac{f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)\Pr(dN_{ij}^* = 1 \mid y_{ij}, \boldsymbol{x}_{ij}, \xi_i, C_i \ge t_{ij})}{\Pr(\mathrm{d}N_{ij}^* = 1, C_i \ge t_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)},$$

where the denominator $\Pr(dN_{ij}^{*} = 1,\ C_i \geq t_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)$ can be calculated as $\int f(y \mid \boldsymbol{x}_{ij}, \xi_i)\Pr(dN_{ij}^{*} = 1,\ C_i \geq t_{ij} \mid y, \boldsymbol{x}_{ij}, \xi_i)dy$. In general, valid inference on the covariate effect $\boldsymbol{\beta}$ requires correct specification of the joint distribution of the observation time process and informative censoring and $\Pr(dN_{ij}^{*} = 1,\ C_i \geq t_{ij} \mid y_{ij}, \boldsymbol{x}_{ij}, \xi_i)$ and the distribution of the random effect $g(\xi_i)$. These specifications are subject to potential model misspecification, which may lead to biased inference (Neuhaus et al., 1992).

Here, we propose an alternative method that does not require the specification of the observation time process, the informative censoring and the random effect's distribution. Our strategy is to identify the observed conditional density that is a functional form of the density of interest, $f_{ij}(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)$. The following assumptions are needed.

(A1) For each subject i, assume that responses $y_{i1}, y_{i2}, \ldots, y_{iK_i}$ are independent given their covariates and the unobserved random effect ξ_i.

(A2) Conditional on $\boldsymbol{X}(\cdot)$ and ξ, C is independent of $(N(\cdot),\ Y(\cdot))$.

(A3) We assume that the probability of observing the response given the response variable, the covariates and the random effect is

$$\Pr(dN_{ij}^{*} = 1 \mid y_{ij}, \boldsymbol{x}_{ij}, \xi_i) = a_1(y_{ij})a_2(\boldsymbol{x}_{ij})a_3(\xi_i), \tag{2.2}$$

where $a_1(\cdot)$, $a_2(\cdot)$ and $a_3(\cdot)$ are completely unspecified nonnegative functions.

Assumption (A1) is standard in the literature. Assumption (A2) is commonly used in the literature to handle informative censoring; see Section 2 of Wang et al. (2001) for an example. Under (A2), the censoring time is allowed to depend on the latent variable, and such a conditional independence assumption substantially relaxes the usual non-informative censoring assumption. Although (A3) specifies a structure through three unspecified functions for the sampling mechanism, it is generally enough to handle the situation with outcome-dependent sampling. For example, when $a_2(\cdot) = a_3(\cdot) = 1$, it implies that the probability of observing the response at a particular time point depends on the response only. This is common in practice, e.g., publication bias in meta-analyses, where the probability of a study being published or not depends on the p-value (significant or not) from that study. When $a_1(\cdot) = a_2(\cdot) = 1$, it implies that the probability of observing the response at a particular time point depends on the subject-specific random effect only.

Chen et al. (2015) developed a pairwise likelihood to handle outcome-dependent sampling for longitudinal data. The fundamental step in the construction of the pairwise likelihood in Chen et al. (2015) is to consider two observations from a pair of independent subjects. However, the two observations in such a pair are not comparable due to the informative censoring; hence the pairwise likelihood requires proper adjustment. To bypass the challenge of dealing with incomparable pairs, we consider two observations from the same individual: say the jth and kth observations of individual i. Under assumptions (A2) and (A3), the conditional density of observing responses at the jth and kth time points for individual i, (y_{ij}, y_{ik}), given their order statistic $(y^{(1)}, y^{(2)})$, covariates $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{ik}$, and random effect ξ_i is

$$
\begin{aligned}
& f(y_{ij}, y_{ik} \mid y^{(1)}, y^{(2)}, dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i, C_i \geq t_{ij}, C_i \geq t_{ik}) \\
& = \frac{f(y_{ij}, y_{ik} \mid dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i)}{f(y_{ij} \mid dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i) + f(y_{ik}, y_{ij} \mid dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i)} \\
& = \frac{f(y_{ij} \mid dN_{ij}^{*} = 1, \boldsymbol{x}_{ij}, \xi_i) f(y_{ik} \mid dN_{ik}^{*} = 1, \boldsymbol{x}_{ik}, \xi_i)}{f(y_{ij} \mid dN_{ij}^{*} = 1, \boldsymbol{x}_{ij}, \xi_i) f(y_{ik} \mid dN_{ik}^{*} = 1, \boldsymbol{x}_{ik}, \xi_i) + f(y_{ij} \mid dN_{ik}^{*} = 1, \boldsymbol{x}_{ik}, \xi_i) f(y_{ik} \mid dN_{ij}^{*} = 1, \boldsymbol{x}_{ij}, \xi_i)}
\end{aligned} \tag{2.3}
$$

.

By (A3) and Bayes formula,

$$
f(y_{ij} \mid dN_{ij}^{*} = 1, \boldsymbol{x}_{ij}, \xi_i) = \frac{\Pr(dN_{ij}^{*} = 1 \mid y_{ij}, \boldsymbol{x}_{ij}, \xi_i) f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)}{\Pr(dN_{ij}^{*} = 1 \mid \boldsymbol{x}_{ij}, \xi_i)} = \frac{a_1(y_{ij}) a_2(\boldsymbol{x}_{ij}) a_3(\xi_i) f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i)}{\Pr(dN_{ij}^{*} = 1 \mid \boldsymbol{x}_{ij}, \xi_i)}
$$

Hence equation (2.3) can be further simplified by canceling out the common factors $a_1(\cdot)$, $a_2(\cdot)$, $a_3(\cdot)$ and $\Pr(dN_i(\cdot) = 1 \mid \boldsymbol{x}_i(\cdot), \xi_i)$,

$$
\begin{aligned}
& f(y_{ij}, y_{ik} \mid y^{(1)}, y^{(2)}, dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i) \\
& = \frac{f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i) f(y_{ik} \mid \boldsymbol{x}_{ik}, \xi_i)}{f(y_{ij} \mid \boldsymbol{x}_{ij}, \xi_i) f(y_{ik} \mid \boldsymbol{x}_{ik}, \xi_i) + f(y_{ij} \mid \boldsymbol{x}_{ik}, \xi_i) f(y_{ik} \mid \boldsymbol{x}_{ij}, \xi_i)} .
\end{aligned} \tag{2.4}
$$

Equation (2.4) implies that the conditional density of the observed data (biased sample) given the order statistic is a function of the density functions of the target population. By model (2.1), equation (2.4) becomes

$$
\begin{aligned}
& f(y_{ij}, y_{ik} \mid y^{(1)}, y^{(2)}, dN_{ij}^{*} = 1, dN_{ik}^{*} = 1, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \xi_i) \\
& = \frac{\exp\left\{y_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}\right\}}{\exp\left\{y_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}\right\} + \exp\left\{y_{ij}\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right\}} .
\end{aligned} \tag{2.5}
$$

The above conditional density does not involve the random effect ξ_i. Thus, the specification of $g(\xi_i)$ and integration over the random effect are not required. In addition, both baseline density function $f_0(\cdot \mid \xi_i)$ and the normalizing constants $(\alpha_{ij}, \alpha_{ik})$ are eliminated by the above conditioning procedure. Therefore, the specification of the baseline density function is not needed either.

Equation (2.5) represents the contribution of one pair of observations (j, k) from the ith individual to the proposed likelihood. We can consider all possible pairs of observations from the same individual and derive their contributions according to equation (2.5). By taking the product of all possible conditional densities while leaving out their correlation, we obtain a composite likelihood function as follows,

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{i<k}\frac{\exp\left\{y_{ij}\boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ik}^{T}\boldsymbol{\beta}\right\}}{\exp\left\{y_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ik}^{T}\boldsymbol{\beta}\right\} + \exp\left\{y_{ij}\boldsymbol{x}_{ik}^{T}\boldsymbol{\beta} + y_{ik}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right\}}$$
$$= \prod_{i=1}^{n}\prod_{j<k}\left[1 + \exp\left\{-(y_{ij} - y_{ik})(\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik})^{T}\boldsymbol{\beta}\right\}\right]^{-1}. \tag{2.6}$$

We will show that the covariate effects $\boldsymbol{\beta}$ can be consistently estimated by maximizing $L_p(\boldsymbol{\beta})$, or equivalently maximizing the following log-likelihood function in the counting process notation, without estimating the nonparametric component $f_0(\cdot)$ and specifying the distribution assumption for the unobservable latent variable ξ_i,

$$\log L_p(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau}\left[-\log\left\{1 + \exp\left\{-(Y_i(s) - Y_i(t))(\boldsymbol{X}_i(s) - \boldsymbol{X}_i(t))^{T}\boldsymbol{\beta}\right\}\right\}\right]dN_i(s)dN_i(t).$$

It is worth mentioning that our construction of the likelihood in (2.5) and (2.6) is different from that by Chen et al. (2015). Their paper considered the comparison of outcomes from different individuals, Y_{ij} and $Y_{i'j'}$. In contrast, we are comparing the outcomes from the same individual, Y_{ij} and $Y_{ij'}$, which allows our method to take care of the informative censoring.

By using the asymptotic results of the proposed composite likelihood (Lindsay, 1988; Cox and Reid, 2004), we show in the following statements that the resulting estimating equation produces consistent and asymptotically normally distributed estimators. The proof is provided in the Appendix.

Theorem 1. *Under the conditions (a)–(c) listed in the* Appendix, *the maximizer of* $\log L_p(\boldsymbol{\beta})$, *denoted by* $\hat{\boldsymbol{\beta}}$, *converges to the true* $\boldsymbol{\beta}_0$ *with probability tending to one as* $n \to \infty$. *Moreover,* $\hat{\boldsymbol{\beta}}$ *is asymptotically normal with mean* $\boldsymbol{\beta}_0$ *and covariance matrix* $\boldsymbol{V} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$, where

$$\sum\nolimits_1 = -E\left\{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{T}}A_i(\boldsymbol{\beta});\boldsymbol{\beta}_0\right\}, \quad \sum\nolimits_2 = cov\left\{\frac{\partial}{\partial\boldsymbol{\beta}}A_i(\boldsymbol{\beta});\boldsymbol{\beta}_0\right\}, \quad and$$

$$A_i(\boldsymbol{\beta}) = -\sum_{j<k}\log\left[1 + \exp\left\{-(y_{ij} - y_{ik})(\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik})^{T}\boldsymbol{\beta}\right\}\right].$$

The covariance matrix V can be empirically estimated by $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$, where

$$\hat{\Sigma}_1 = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{T}}A_i(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\right\}, \quad \text{and} \quad \hat{\Sigma}_2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\boldsymbol{\beta}}A_i(\boldsymbol{\beta})\right\}^{\otimes 2}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.$$

It is worth mentioning that the constructed likelihood function in (2.6) cannot be directly used for estimating the regression effects of constant covariates. The reason is that the parameter $\boldsymbol{\beta}$ is associated with the product $(\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik})^T\boldsymbol{\beta}$. Suppose the l-th covariate in $\boldsymbol{x}$ is a constant, then $x_{ijl} = x_{ikl}$ for every j, k, which makes β_l not estimable since the product is always 0. Inference for constant covariate effects may require additional modeling assumption such as the distribution of random effects and the observation time process.

3. SIMULATION

We evaluate the numerical performance of the proposed method via simulation studies. We compare the proposed estimator with estimators obtained from three existing methods: the GEE method, joint modeling (JM) of longitudinal outcomes and informative censoring, and the pairwise-likelihood (PL) method (Chen et al., 2015), in which the authors did not consider informative censoring.

We consider a cohort of 200 study subjects and generate three time-varying covariates for each subject,

$$\boldsymbol{X} = (X_1, X_2, X_3)^T,\ X_1 \sim N(.5 + .5\sqrt{t}, 0.04),\ X_2 = t,\ X_3 = t^2, \tag{3.1}$$

where t is the time after the study enrollment and takes values in a set of grid points {0.1, 0.2, …, 9.9, 10}.

In the first scenario, we generate responses Y_{ij} from a normal distribution, with mean $\boldsymbol{X}_i^T\boldsymbol{\beta} + \xi_i$ and variance .09, where the true value of $\boldsymbol{\beta}$ is (1, 2, −.1), and $\xi_i \sim N(0, \sigma_1^2)$ are i.i.d. random effects with two noise levels $\sigma_1 = .2$ and $\sigma_1 = .4$. The probability of observing response $Y = y_{ij}$ is generated from a logistic model,

$$\text{logit}\left\{\Pr(dN_{ij} = 1 \mid y_{ij}; \boldsymbol{\gamma})\right\} = \gamma_0 + \gamma_1 y_{ij} + \gamma_2 y_{ij}^2, \tag{3.2}$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^T$ controls the level of association between the sampling probability and covariates. Specifically, we let $\boldsymbol{\gamma}$ take three sets of different values that correspond to different levels of association: (1) independence (“indep”), where $\boldsymbol{\gamma} = (2, 0, 0)^T$, (2) weak association (“weak”), where $\boldsymbol{\gamma} = (.1, 2, -.2)^T$, and (3) strong association (“strong”), where $\boldsymbol{\gamma} = (-1.5, 4.5, 1)^T$. The censoring time C is generated from a normal distribution $N(7 + 5\xi^2, .5^2)$.

The GEE and JM methods are implemented in R Console using “geepack” and “JM” packages. We summarize the results based on 1000 replications in Table 1. For each method, we present its estimation bias (Bias), estimated standard error (SE), empirical standard error (SD) and coverage probability (CP %) of the 95% confidence interval. The proposed method provides desired estimation accuracy and yields reasonable CP for the 95% confidence intervals for all simulation settings. Since both the GEE method and PL method ignore the informative censoring, they work poorly under the simulation setting. For example, when $\sigma_1 = .4$, the CPs of the PL method range from 0 to 25%, far below the nominal value of 95%. The joint model implemented in the R package can appropriately handle the informative censoring, but cannot deal with the outcome-dependent sampling. When assuming “independent” sampling, JM works well as expected; its CPs are close to 95% and its standard errors are smaller than those of the proposed method since the distribution information of the informative censoring is used in the likelihood function under the JM. The proposed method shows great advantages when the sampling is outcome-dependent. For example, when $\sigma_1 = 0.2$ and a "weak" association between the outcome and sampling exists, the proposed method can still provide a small estimation bias, and its CPs for all parameters are greater than 94%. However, the CPs for all other methods are much lower than those obtained from the proposed method, and the bias can be as high as 0.8. When there is "strong" association between the sampling and the outcome, this advantage for the proposed method can be even more obvious.

We also assess the performance of the proposed method with other distributed outcomes such as the exponential distribution. We let $X_1 \sim N(.5 + .5\sqrt{t},\ 0.01)$ and generate X_2 and X_3 in the same way as in the previous simulation. We generate responses Y_{ij} from an exponential distribution with mean $(\boldsymbol{X}_{\boldsymbol{i}}^T\boldsymbol{\beta} + \xi_i)^{-1}$, where $\boldsymbol{\beta} = (2, .5, .1)^T$, and the random effects ξ_i are generated i.i.d from a uniform distribution on $(-\sigma_2, \sigma_2)$ with two different noise levels $\sigma_2 = .1$ and .5. The probability of observing Y follows a logistic model as in (3.2), where $\boldsymbol{\gamma}$ is chosen from three sets of values: (1) independence, where $\boldsymbol{\gamma} = (2, 0, 0)^T$, (2) weak association, $\boldsymbol{\gamma} = (.3, 1, 0)^T$, and (3) strong association, $\boldsymbol{\gamma} = (-1, 10, 2)^T$.

We generate the censoring time from a normal distribution $C \sim N(8 + 4\xi^2, .4^2)$. Note that among the four aforementioned methods, only our proposed method and the GEE method do not need to specify the distribution of the longitudinal outcomes. Hence for this scenario, we only evaluate the proposed method and the GEE method on 1000 replicated data sets; the results are listed in Table 2. The 95% confidence intervals produced by the proposed method have coverage probabilities between 93.3% and 95.2%, indicating accurate estimation under all scenarios although the standard error tends to be underestimated. In contrast, $\hat{\boldsymbol{\beta}}_{\mathrm{GEE}}$ has a larger bias, even when $\boldsymbol{\gamma}$ is independent (e.g., when $\sigma_2 = .5$). When there is a strong association between the sampling and the outcomes, the coverage probability of GEE is always below 40%.

To evaluate the robustness of the proposed method, we conduct sensitivity analysis for which the condition (2.2) in Assumption (A3) is violated. We consider two scenarios for generating the probability of observing response $Y = y_{ij}$,

$$\mathrm{logit}\left\{\Pr(dN_{ij} = 1 \mid y_{ij}, \boldsymbol{x}_{ij})\right\} = -1 + y_{ij} + x_1(t_j), \tag{3.3}$$

and

$$\mathrm{logit}\left\{\Pr(dN_{ij} = 1 \mid y_{ij}, \boldsymbol{x}_{ij})\right\} = y_{ij} + .2t_j + x_1(t_j) * y_{ij}/3\,. \tag{3.4}$$

Those two scenarios can be viewed as moderate derivations from the "weak" dependence situation under (A3). The corresponding results are summarized in "mis-specified 1" and "mis-specified 2" panels in Tables 1 and 2, respectively. It can be seen that the coverage of the regression coefficients are still quite close to the nominal level for most cases, which confirms that the proposed method has some degree of robustness to the violation of (A3).

We also consider a simulation example that compares the performance of our method with that of the partial likelihood (PL) approach (Chen et al. 2015) when the terminal event is non-informative. Here, we generate the data in the same way with the first simulation scenario (linear case), except where the censoring time is generated independently of the random effect following either (i) a “weak” censoring rate from a uniform distribution from 8 to 10 (average censoring rate is 10%) or (ii) a “strong” censoring rate from an exponential distribution $C = 4.3+\mathrm{Exp}(1/2)$, with an average censoring rate of 39%. We summarize the results in Table 3. As expected, both methods manage to estimate the coefficients consistently under all scenarios. PL has better efficiency than our method (SE is smaller by an average of 20%) for most cases, although our method is more computationally efficient (computational complexity is $O(n^3)$ compared with $O(n^4)$ for PL).

4. APPLICATION TO WEIGHT LOSS DATA STUDY

We have applied the proposed method to evaluate the effectiveness of the web-based weight control program by using the self reported longitudinal weight data. Web-based programs have become popular among many weight loss programs because of their flexibility and low cost. The effectiveness of such programs has been recently evaluated, and research has shown findings of a positive association between weight loss and the use of web-based programs (Neve et al., 2011). However, most of the published analyses were based on the evaluation of the weight difference, i.e., the difference in a participant’s body weights when entering the study compared to that when leaving the study. Inevitably, ignoring the information of weight and other predictors in the middle of the study will result in misleading interpretations. For example, patients may experience a rebound in weight loss, i.e., lose 10 lbs in the first 3 months and then gain 8 lbs back in the next 2 months. An examination of the overall weight difference will not be able to reveal such patterns. Hence, it is necessary to use weight data over time and study the effect of web-based programs during the entire study period.

The main objective of our study was to conduct regression analysis to evaluate the effect of a web-based weight loss program and weight trend over time after enrollment in the program. We considered data collected from a web-based weight loss program, including records collected in 2008 from 5477 participants. Each participant voluntarily reported his/her weight and the web system also recorded his/her number of logins to the system up to any time t, which can be used as a measure of the usage of the program up to time t. The number of self-reported data entries for each participant varied from 2 to 168, with a mean of 6.2. The study duration had a mean of 8.2 months and a maximum of 27.6 months. As expected, we found that the length of enrollment was strongly associated with weight loss. The average weight loss was 3.31 lbs for participants who stayed in the study less than 8 months, and 2.00 lbs for participants who stayed longer than 8 months (two-sample t test, p-value < .001).

We applied the proposed method to the aforementioned data set and considered several time-varying covariates: the number of logins (login), months in the study (month) and its square (month^2) up to each time of self-report. The estimated coefficients (Est.), associated standard errors (SE) and p-values are summarized in Table 4. We also present the results using the method of Chen et al. (2015) and the GEE method. We found that all three covariates had a significant effect on weight. Usage of the web-based program (measured by the number of logins) was positively associated with weight loss. We also found that the enrollment duration had a positive effect (negative sign of month effect) in helping participants lose weight, although there may have been a long-term rebound effect suggested by the positive sign of the month^2 effect. These results confirm the findings of Chen et al. (2015), although our estimated covariate effects were smaller in magnitude than those of the other two methods. This is not surprising: both dependent censoring and biased sampling tend to over-estimate the covariate effects. In contrast, the results obtained using the method of Chen et al. (2015) were based on an independent censoring assumption and the GEE did not account for informative censoring or biased sampling.

We also conducted an analysis for the "active" participants who had at least 10 self-reported records (790 participants). We included an additional covariate called gap, which we defined as the time gap since the last report (equal to 0 if it was the first report). The results, shown in Table 5, demonstrate that the estimates are very close among the "general" and "active" sub-populations (report records ≥ 2 and ≥ 10). The positive sign of the "gap" confirms the conjecture that the patients who are making progress in weight loss are more likely to report more frequently than those who are not. This justifies the use of the proposed method on this data set because it does not require additional assumptions/modeling on the outcome sampling scheme.

5. DISCUSSION

Achieving sufficiently fast computation and avoiding the issue of non-convergence remain challenging problems for the joint modeling approach. In the presence of informative censoring and outcome-dependent observation times, multiple integrations and failure of convergence of the optimization are commonly encountered in practice. Some existing efforts have considered approximation (Sweeting and Thompson, 2011) or Bayesian MCMC methods (Faucett and Thomas, 1996; Brown and Ibrahim, 2003), but they only apply to specific models. In the simulation examples, the implementation of the JM package in R occasionally encountered convergence issues (roughly 2 out of 100 times) when solving the optimization using the quasi-Newton method ("BFGS" option in R). Our method did not experience any convergence issues. The average running time for a single iteration was about 10 seconds for the proposed method, 15 seconds for joint modeling, and 445 seconds for the pairwise likelihood on a cluster machine with Dual 2.2GHz Single Core AMD operation 252 CPUs, 16GB RAM, and 64-bit CentOS Linux System. This is expected since the proposed composite likelihood has a simpler form than the likelihood of joint modeling, and provides a more computationally efficient alternative to the pairwise likelihood. The robustness of the proposed composite likelihood approach due to the requirement of fewer model/distribution assumptions is another advantage.

The idea of conditioning on the order statistics and canceling out the distribution of nuisance variables such as the latent variables has been previously used by Kalbfleisch (1978) for nonparametric testing and by Liang and Qin (2000) for handling missing data under regression analysis. Different from the conditional event used in the literature, we particularly consider a pair of observations from the same subject to handle the informative censoring issue.

Our work is motivated by evaluating how a web-based weight loss program affects the participant's weight, and the parameter on which we focus is the effect of the time-varying intervention. One limitation of the proposed composite likelihood approach is that the effects of constant covariates are not estimable, and it seems that there is no easy way to extend the current methodology without making additional modeling assumptions on the distribution of the outcomes, random effects, correlation between the informative censoring and the outcome, and the observation time process. Hence, we leave this important topic for future research. The methodology we have developed is applicable to weight loss program studies and other studies that share a similar interest in analyzing the effects of covariates that are changing over the study period. Another challenge of our methodology is the model specification for the likelihood ratio. Note that the nonparametric components such as random effects and the baseline density function of the outcome are not estimated in the proposed estimating procedure. Standard diagnostic tools, such as residual-based methods, are not applicable and cannot handle informative censoring. Developing rigorous test procedures for the modeling assumptions such as a proportional likelihood ratio will be a very interesting and important future research direction. Although the sensitivity analysis suggests some degree of robustness of our method when assumption (A3) is violated, caution should be taken when that assumption is not met. Developing statistical approaches for testing this assumption will be of interest for future research.

ACKNOWLEDGMENTS

Shen's research is partially supported by the Simons Foundation (Award 512620) and the National Science Foundation (NSF DMS 1509023). Ning's research is partially supported by grants from the National Cancer Institute (R01CA193878 and P30CA016672) and the Andrew Sabin Family Fellowship. The authors thank the editor, the associate editor, and two reviewers for their constructive comments that have greatly improved the initial version of this article. Shen thanks D.D for the inspiration and encouragement.

APPENDIX

Regularity conditions

We first state a set of regularity conditions needed to establish asymptotic results. For any fixed time point $t \in [0, \tau]$, we assume that $z(t) = \{Y_i(t), X_i(t), N_i(t)\}$, $i = 1, 2, \ldots, n$, are independent, identically distributed with a joint density function $h\{z(t); \boldsymbol{\beta}\}$, which satisfies the conditions $\mathscr{R}$ in Chernoff (1954). The set of conditions is listed as follows.

a. There exists a neighborhood $\mathcal{N}_{\boldsymbol{\beta}_0}$ of $\boldsymbol{\beta}_0$ such that for almost all z and every $\boldsymbol{\beta} \in \mathcal{N}_{\beta_0}$, the following derivatives exist

$$\frac{\partial \log h\{z(t); \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}}, \frac{\partial^2 \log h\{z(t); \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}, \frac{\partial^3 \log h\{z(t); \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}}.$$

b. For any $\boldsymbol{\beta} \in \mathcal{N}_{\beta_0}$,

$$\left|\frac{\partial h\{z(t);\boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}}\right| < H\{z(t)\}, \quad \left|\frac{\partial^2 h\{z(t);\boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right| < H\{z(t)\}, \quad \left|\frac{\partial^3 h\{z(t);\boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T\partial \boldsymbol{\beta}}\right| < H\{z(t)\},$$

for some finitely integrable function H and $E(H\{\mathbf{z}(t)\}) < M$ with M independent of $\boldsymbol{\beta}$ and t.

c. We assume the matrices $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ defined in Theorem 1 to be positive definite and finite.

Assumption (c) is commonly used in the literature, and holds for most situations except some extreme cases. For example, we can show that both $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ are always non-negative definite. The positive definiteness assumption only rules out the extreme situation where the likelihood function is flat in $\boldsymbol{\beta}$ (e.g., the data do not contain adequate information to estimate one of regression efficients, and further model check is needed to revise the model).

Proof of Theorem 1

Let $\boldsymbol{\beta_0}$ be the true value of $\boldsymbol{\beta}$. We first show the consistency result. Note that for any $\delta > 0$, the intersection of the parameter space of $\boldsymbol{\beta}$ and the closure of a δ-neighborhood of $\boldsymbol{\beta_0}$ is closed. Therefore log $L_p(\boldsymbol{\beta})$ has a local maximum on the intersection. Hence, it suffices to show that the maximum of log $L_p(\boldsymbol{\beta})$ has a l_2-distance less than δ from $\boldsymbol{\beta_0}$, with the probability going to 1. Consider the Taylor expansion of log $L_p(\boldsymbol{\beta})$ around $\boldsymbol{\beta_0}$,

$$\begin{aligned}\frac{1}{n}\left\{\log L_p(\boldsymbol{\beta}) - \log L_p(\boldsymbol{\beta}_0)\right\} &= \frac{1}{n}\frac{\partial L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &+ \frac{1}{2}\frac{1}{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\frac{\partial^2 L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + n^{-1}o_p(\left\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\right\|_2^2).\end{aligned} \quad (A.1)$$

As $n \to \infty$, since $\boldsymbol{\beta_0}$ is the unique maximum in its neighborhood, we have

$$\frac{1}{n}\frac{\partial L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \to \mathbf{0}, \quad \frac{1}{n}\frac{\partial^2 L_p(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} \to \boldsymbol{A},$$

for some negative-definite matrix $\boldsymbol{A}$. Note that the first term in (A.1) is $o_p(1)$, and the third term is negligible compared to the second term. The second term in (A.1) is negative. Hence log $L_p(\boldsymbol{\beta}) <$ log $L_p(\boldsymbol{\beta_0})$ with probability 1 for every $\boldsymbol{\beta}$ that satisfies $\|\boldsymbol{\beta} - \boldsymbol{\beta_0}\|_2 < \delta$. The consistency of $\hat{\boldsymbol{\beta}}$ holds.

Next, we show the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}$. Using the Taylor expansion and the regularity conditions, we obtain for some $C_1 > 0$ that,

$$\frac{1}{n}\left\{\log L_p(\hat{\boldsymbol{\beta}}) - \log L_p(\boldsymbol{\beta_0})\right\} = \frac{1}{n}\frac{\partial \log L_p(\boldsymbol{\beta_0})}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) + \frac{1}{2}\frac{1}{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})^T\frac{\partial^2 \log L_p(\boldsymbol{\beta_0})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) + C_1 n^{-1}\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}\right\|_2^3.$$

Since $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_{\mathbf{0}}$, and $n^{-1}\frac{\partial^2 \log L_p(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \xrightarrow{a.s.} -\boldsymbol{\Sigma}_1$ by strong law of large numbers, then for every $\epsilon > 0$, there exists a sequence of positive numbers $c_{n\epsilon} \to 0$, such that

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}\right\|_2 < c_{n\epsilon}, \quad \left\|\frac{1}{n}\frac{\partial^2 \log L_p(\boldsymbol{\beta}_{\mathbf{0}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} + \boldsymbol{\Sigma}_{\mathbf{1}}\right\|_2 < c_{n\epsilon}.$$

Also, since $n^{-1} \log L_p(\boldsymbol{\beta})$ is an empirical average based on i.i.d. observations from each study individual, it is of the order of $O_p(n^{-1/2})$ at $\boldsymbol{\beta}_{\mathbf{0}}$, i.e., there exists $K_\epsilon > C_1$ such that

$$\left\|\frac{1}{n}\frac{\partial \log L_p(\boldsymbol{\beta}_{\mathbf{0}})}{\partial\boldsymbol{\beta}}\right\|_2 < n^{-1/2}K_\epsilon,$$

Therefore,

$$0 \le \frac{1}{n}\left\{\log L_p(\hat{\boldsymbol{\beta}}) - \log L_p(\boldsymbol{\beta}_{\mathbf{0}})\right\} \le n^{-1/2}K_\epsilon\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}\right\|_2 - \frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}})^T\,\Sigma_1\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}) + C_1c_{n\epsilon}\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}\right\|_2^2.$$

Since $\boldsymbol{\Sigma}_{\mathbf{1}}$ is positive definite, $n^{1/2}\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_2 = O_p(1)$. The root-n consistency holds.

For asymptotic normality result, define

$$R_i(s, t; \boldsymbol{\beta}) = \exp\left\{-(Y_i(s) - Y_i(t))(\boldsymbol{X_i}(s) - \boldsymbol{X_i}(t))^T\boldsymbol{\beta}\right\}.$$

Applying the Taylor expansion of $\partial \log L_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ at $\boldsymbol{\beta}_{\mathbf{0}}$, we obtain

$$\begin{aligned}\mathbf{0} &= \sum_{i=1}^{n}\int_0^\tau\int_0^\tau\left[-\frac{\partial \log\left\{1 + R_i(s, t; \hat{\boldsymbol{\beta}})\right\}}{\partial\boldsymbol{\beta}}\right]\mathrm{d}N_i(s)\mathrm{d}N_i(t)\\ &= \sum_{i=1}^{n}\int_0^\tau\int_0^\tau\left[-\frac{\partial \log\left\{1 + R_i(s, t; \boldsymbol{\beta}_{\mathbf{0}})\right\}}{\partial\boldsymbol{\beta}}\right]\mathrm{d}N_i(s)\mathrm{d}N_i(t)\\ &+ \left(\sum_{i=1}^{n}\int_0^\tau\int_0^\tau\left[-\frac{\partial^2 \log\left\{1 + R_i(s, t; \boldsymbol{\beta}_{\mathbf{0}})\right\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]\mathrm{d}N_i(s)\mathrm{d}N_i(t)\right)^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}) + O_p(n^{-1}),\end{aligned}$$

where the last term $O_p(n^{-1})$ is due to the fact that $\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_2^2 = O_p(n^{-1})$. Therefore, we can multiply both sides by $n^{1/2}$, and obtain

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathbf{0}}) = \boldsymbol{A}_{\boldsymbol{n}}^{-1}n^{1/2}\boldsymbol{B}_{\boldsymbol{n}} + O_p(n^{-1/2}),$$

where

$$A_n = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau}\left[-\frac{\partial^2 \log\{1 + R_i(s,t;\boldsymbol{\beta_0})\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]\mathrm{d}N_i(s)dN_i(t),$$

$$B_n = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau}\left[-\frac{\partial \log\{1 + R_i(s,t;\boldsymbol{\beta_0})\}}{\partial\boldsymbol{\beta}}\right]\mathrm{d}N_i(s)\mathrm{d}N_i(t).$$

Since $\boldsymbol{A_n}$ and $\boldsymbol{B_n}$ are both empirical average of functionals based on i.i.d. observations, we have $\boldsymbol{A_n} \xrightarrow{p} \boldsymbol{\Sigma}_1$ and $n^{1/2}\boldsymbol{B_n} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_2)$ as $n \to \infty$ by the law of large numbers and the CLT. This completes the proof.

REFERENCES

Brown ER and Ibrahim JG (2003), “A Bayesian semiparametric joint hierarchical model for longitudinal and survival data,” Biometrics, 59, 221–228. [PubMed: 12926706]

Chan KCG (2013), “Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation.” Biometrika, 100, 269–276.

Chen Y, Ning J, and Cai C (2015), “Regression analysis of longitudinal data with irregular and informative observation times,” Biostatistics, 16, 727–739. [PubMed: 25813646]

Chernoff H (1954), “On the distribution of the likelihood ratio,” The Annals of Mathematical Statistics, 25, 573–578.

Cox D and Reid N (2004), “A note on pseudolikelihood constructed from marginal densities,” Biometrika, 91, 729–737.

Diggle P and Kenward MG (1994), “Informative Drop-Out in Longitudinal Data Analysis,” Journal of the Royal Statistical Society. Series C (Applied Statistics), 43, 49–93.

Ding J, Lu TS, Cai J, and Zhou H (2017), “Recent progresses in outcome-dependent sampling with failure time data,” Lifetime Data Analysis, 23, 57–82. [PubMed: 26759313]

Faucett CL and Thomas DC (1996), “Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach,” Statistics in Medicine, 15, 1663–1685. [PubMed: 8858789]

Follmann D and Wu M (1995), “An approximate generalized linear model with random effects for informative missing data.” Biometrics, 51, 151–168. [PubMed: 7766771]

Han J, Slate E, and Peña E (2007), “Parametric latent class joint model for a longitudinal biomarker and recurrent events,” Statistics in Medicine, 26, 5285–5302. [PubMed: 17542002]

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

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

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

Liang Y, Lu W, and Ying Z (2009), “Joint modeling and analysis of longitudinal data with informative observation times,” Biometrics, 65, 377–384. [PubMed: 18759841]

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

Lindsay B (1988), “Composite likelihood methods,” Contemporary Mathematics, 80, 221–39.

Lipsitz S, Fitzmaurice G, Ibrahim J, Gelber R, and Lipshultz S (2002), "Parameter estimation in longitudinal studies with outcome-dependent follow-up," Biometrics, 58, 621–630. [PubMed: 12229997]

Little R (1995), "Modeling the drop-out mechanism in repeated-measures studies," Journal of the American Statistical Association, 90, 1112–1121.

Liu L (2009), "Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data," Statistics in Medicine, 28, 972–986. [PubMed: 19040210]

Liu M and Ying Z (2007), "Joint analysis of longitudinal data with informative right censoring," Biometrics, 63, 363–371. [PubMed: 17425632]

Luo X and Tsai W (2012), "A proportional likelihood ratio model," Biometrika, 99, 211–222.

Neuhaus JM, Hauck WW, and Kalbfleisch JD (1992), "The effects of mixture distribution misspecification when fitting mixed-effects logistic models," Biometrika, 79, 755–762.

Neve M, Morgan PJ, and Collin C. E. s. (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. [PubMed: 21993231]

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, Robins J, and Scharfstein D (1998), "Semi parametric regression for repeated outcomes with nonignorable nonresponse," J. Am. Stat. Assoc, 93, 1321–1339.

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

Self S and Pawitan Y (1992), "Modeling a marker of disease progression and onset of disease," in AIDS Epidemiology, Birkhauser Boston, pp. 231–255.

Song R, Zhou H, and Kosorok MR (2009), "A note on semiparametric efficient inference for two-stage outcome-dependent sampling with a continuous outcome," Biometrika, 96, 221–228. [PubMed: 20107493]

Song X, Mu X, and Sun L (2012), "Regression Analysis of Longitudinal Data with Time-Dependent Covariates and Informative Observation Times," Scandinavian Journal of Statistics, 39, 248–258.

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

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

Sweeting MJ and Thompson SG (2011), "Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture," Biometrical Journal, 53, 750–763. [PubMed: 21834127]

Tsiatis AA and Davidian M (2004), "Joint modeling of longitudinal and time-to-event data: an overview." Statistica Sinica, 14, 809–834.

Tsiatis AA, De Gruttola V, and Wulfsohn MS (1995), "Modelling the relationship of survival to longitudinal data measured with error – application to survival and CD4 counts in patients with AIDS," Journal of the American Statistical Association, 90, 27–37.

Wang MC, Qin J, and Chiang CT (2001), "Analyzing Recurrent Event Data with Informative Censoring," Journal of the American Statistical Association, 96, 1057–1065.

Wang Y and Taylor JMG (2001), "Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome," Journal of the American Statistical Association, 96, 895–905.

Wang YC, McPherson K, Marsh T, Gortmaker SL, and Brown M (2011), "Health and economic burden of the projected obesity trends in the USA and the UK," Lancet, 378, 815–825. [PubMed: 21872750]

Wu M and Carroll R (1988), "Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process." Biometrics, 44, 175–188.

Wulfsohn MS and Tsiatis AA (1997), “A joint model for survival and longitudinal data measured with error,” Biometrics, 53, 330–339. [PubMed: 9147598]

Yi G, Zeng L, and Cook R (2011), “A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters,” Canadian Journal of Statistics, 39, 34–51.

Zhou H, Weaver MA, Qin J, Longnecker MP, and Wang MC (2002), “A Semiparametric Empirical Likelihood Method for Data from an Outcome-Dependent Sampling Scheme with a Continuous Outcome,” Biometrics, 58, 413–421. [PubMed: 12071415]

Author Manuscript
Author Manuscript
Author Manuscript

Table 1:

Simulation scenario 1, comparing the performance of the proposed estimator $\widehat{\beta}$ with those of the GEE ($\widehat{\beta}_{\mathrm{GEE}}$), joint modeling ($\widehat{\beta}_{\mathrm{JM}}$) and pairwise-likelihood ($\widehat{\beta}_{\mathrm{PL}}$) estimators

			$\widehat{\beta}$				$\widehat{\beta}_{\mathrm{GEE}}$				$\widehat{\beta}_{\mathrm{JM}}$				$\widehat{\beta}_{\mathrm{PL}}$			
σ_1	γ	β	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**
0.2	‘indep”	β_1	.012	.191	.197	93.2	−.054	.202	.208	93.8	−.002	.144	.146	94.4	−.323	.152	.157	42.8
		β_2	−.001	.099	.099	94.7	.011	.121	.128	94.1	−.009	.080	.082	94.3	−.572	.095	.099	0
		β_3	.000	.012	.012	94.8	−.010	.016	.017	89.5	.001	.009	.010	94.3	.018	.012	.013	63.6
	“weak”	β_1	.014	.216	.220	94.1	.875	.198	.202	1.0	.373	.151	.153	31.5	−.320	.171	.175	52.3
		β_2	−.002	.114	.112	94.7	−.193	.126	.131	64.7	−.152	.086	.087	58.6	−.578	.108	.113	.5
		β_3	.000	.013	.013	94.0	−.009	.016	.017	89.0	.004	.010	.010	93.1	.019	.013	.014	67.0
	“strong”	β_1	.013	.237	.237	94.0	2.23	.187	.193	0.0	.958	.159	.157	0.0	−.327	.192	.192	59.5
		β_2	.002	.131	.128	95.3	−.512	.122	.128	1.5	−.434	.085	.085	0.0	−.589	.127	.129	.7
		β_3	.000	.014	.015	94.8	−.013	.015	.016	82.1	.010	.010	.010	80.4	.021	.015	.016	68.4
	“mis-specified 1”	β_1	−.165	.217	.218	88.1	.435	.209	.214	43.2	.133	.156	.155	87.0	−.508	.170	.167	16.3
		β_2	−.013	.115	.112	95.3	−.048	.133	.141	91.5	−.059	.090	.090	90.0	−.577	.109	.114	.3
		β_3	.001	.013	.013	95.5	−.017	.017	.019	81.5	−.001	.010	.010	94.9	.019	.014	.015	67.5
	“mis-specified 2”	β_1	.037	.206	.211	93.8	.628	.203	.210	14.0	.292	.150	.154	51.8	−.300	.163	.163	54.7
		β_2	−.060	.110	.109	90.5	−.149	.128	.136	77.1	−.133	.086	.087	65.5	−.623	.106	.111	.1
		β_3	.003	.012	.012	93.6	−.011	.017	.018	88.5	.003	.010	.010	93.1	.02	.013	.014	63.2
0.4	“indep”	β_1	.012	.189	.198	93.7	−.450	.382	.390	75.7	.004	.161	.164	93.8	−.693	.107	.110	0
		β_2	−.001	.089	.090	94.9	.289	.207	.217	65.1	−.010	.071	.072	94.7	−1.129	.084	.084	0
		β_3	.000	.009	.009	94.5	−.073	.028	.030	29.0	.001	.007	.008	95.2	.031	.009	.010	8.4
	“weak”	β_1	.012	.216	.226	93.8	1.341	.365	.370	2.8	.153	.171	.177	84.9	−.694	.125	.131	0

σ_1	γ	β	$\hat{\beta}$				$\hat{\beta}_{GEE}$				$\hat{\beta}_{JM}$				$\hat{\beta}_{PL}$			
			Bias	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**
		β_2	−.002	.105	.106	95.1	−.172	.224	.232	90.3	−.144	.078	.078	55.2	−1.145	.101	.103	0
		β_3	.000	.011	.011	93.8	−.053	.030	.032	52.3	.006	.008	.008	88.5	.033	.012	.012	18.1
	"strong"	β_1	−.003	.240	.239	94.8	3.39	.350	.349	0.0	.401	.171	.168	34.8	−.685	.148	.150	.1
		β_2	.001	.121	.123	94.6	−.732	.223	.233	4.8	−.401	.079	.079	0.0	−1.191	.127	.138	0
		β_3	.000	.012	.013	94.5	−.033	.028	.031	67.3	.017	.008	.008	48.8	.039	.014	.016	24.5
	"mis-specified 1"	β_1	−.166	.215	.218	87.5	.481	.386	.396	79.0	.034	.175	.182	92.9	−.875	.121	.121	0
		β_2	−.009	.103	.102	95.3	.141	.232	.240	85.1	−.075	.082	.143	81.2	−1.13	.097	.099	0
		β_3	.001	.010	.011	94.8	−.076	.030	.032	34.3	.002	.009	.010	94.5	.031	.011	.011	18.3
	"mis-specified 2"	β_1	.040	.204	.209	93.3	.818	.374	.379	40.1	.209	.167	.235	71.2	−.659	.116	.119	0
		β_2	−.051	.098	.098	91.5	−.016	.224	.231	92.9	−.133	.077	.139	60.7	−1.17	.097	.099	0
		β_3	.001	.010	.010	93.7	−.070	.029	.031	36.5	.003	.008	.010	87.4	.032	.011	.011	17.5

Table 2:

Simulation scenario 2, comparing the performance of the proposed estimator $\hat{\boldsymbol{\beta}}$ with that of the GEE ($\hat{\boldsymbol{\beta}}_{\text{GEE}}$).

			$\hat{\beta}$				$\hat{\beta}_{\text{GEE}}$			
random effect	γ	β	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**
low	"indep"	β_1	.001	.431	.445	93.7	−.003	.085	.084	95.9
$\sigma_2 = .1$	(2, 0, 0)	β_2	.001	.146	.146	94.0	.001	.084	.082	96.0
		β_3	.000	.015	.015	95.0	.000	.012	.012	95.0
	"weak"	β_1	.007	.460	.484	93.6	−.259	.084	.085	14.7
	(.3, 1, 0)	β_2	.003	.159	.162	94.6	−.001	.087	.088	94.0
		β_3	.000	.017	.017	94.9	.002	.013	.013	93.8
	"strong"	β_1	−.011	.432	.427	95.2	−.212	.066	.066	11.6
	(−1, 10, 2)	β_2	.006	.151	.151	94.7	−.261	.065	.065	1.8
		β_3	.000	.017	.017	93.8	−.023	.010	.010	33.8
	"mis-specified 1"	β_1	.083	.453	.470	93.6	−.234	.082	.082	20.5
	(−1, 10, 2)	β_2	.006	.154	.153	94.9	.023	.084	.083	94.0
		β_3	.000	.016	.016	95.3	.001	.012	.012	94.3
	"mis-specified 2"	β_1	−.095	.499	.513	93.0	−.224	.093	.092	32.4
	(−1, 10, 2)	β_2	.027	.166	.167	94.8	.034	.90	.089	94.2
		β_3	.000	.017	.017	95.2	.001	.013	.012	95.2
high	"indep"	β_1	.004	.42.6	.43.7	94.0	−.087	.088	.087	83.8
$\sigma_2 = .5$	(2, 0, 0)	β_2	.002	.142	.142	94.7	.042	.082	.080	92.0
		β_3	.000	.014	.015	94.8	−.003	.011	.011	93.6
	"weak"	β_1	.004	.451	.471	93.3	−.349	.086	.086	3.3
	(.3, 1, 0)	β_2	.005	.153	.155	94.7	.044	.084	.085	91.5
		β_3	−.001	.016	.016	94.6	−.002	.012	.012	93.0
	"strong"	β_1	−.008	.423	.416	95.0	−.283	.069	.069	2.7
	(−1, 10, 2)	β_2	.005	.145	.146	94.9	−.229	.064	.064	5.3
		β_3	.000	.016	.016	94.2	−.025	.009	.009	22.2
	"mis-specified 1"	β_1	.121	.446	.439	94.9	−.318	.085	.084	4.2
	(−1, 10, 2)	β_2	−.006	.149	.143	96.2	.063	.081	.081	86.9
		β_3	.000	.015	.015	95.5	−.003	.011	.011	94.1
	"mis-specified 2"	β_1	−.06	.494	.485	94.3	−.314	.095	.094	10.8
	(−1, 10, 2)	β_2	.020	.162	.156	95.6	.079	.087	.087	85.1
		β_3	.000	.016	.016	95.7	−.003	.012	.012	94.2

Table 3:

Simulation scenario 3, comparing the performance of the proposed estimator $\hat{\boldsymbol{\beta}}$ with those of the pairwise-likelihood ($\hat{\boldsymbol{\beta}}_{\mathrm{PL}}$) estimators under non-informative censoring.

				$\hat{\boldsymbol{\beta}}$				$\hat{\boldsymbol{\beta}}_{\mathrm{PL}}$			
censor	$\boldsymbol{\sigma_1}$	$\boldsymbol{\gamma}$	$\boldsymbol{\beta}$	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**
“weak”	0.2	“indep”	β_1	.004	.214	.216	94.9	.03	.168	.180	92.4
			β_2	.002	.096	.099	94.6	.020	.095	.101	93.2
			β_3	.000	.009	.009	94.1	−.002	.008	.009	92.3
		“weak”	β_1	.003	.238	.238	95.0	.033	.187	.196	93.3
			β_2	.002	.109	.111	94.9	.015	.103	.109	93.1
			β_3	.000	.010	.010	94.6	−.002	.009	.009	92.2
		“strong”	β_1	.008	.256	.256	94.0	.035	.206	.209	95.3
			β_2	−.001	.121	.124	93.7	.006	.114	.119	93.5
			β_3	.000	.011	.011	93.8	−.001	.010	.011	92.4
	0.4	“indep”	β_1	.002	.214	.217	95.0	.044	.123	.133	91.4
			β_2	.001	.096	.099	94.2	−.006	.084	.091	91.9
			β_3	.000	.009	.009	94.8	−.002	.007	.008	91.8
		“weak”	β_1	.003	.244	.243	94.5	.044	.145	.151	93.0
			β_2	.000	.112	.114	94.8	−.019	.095	.101	92.9
			β_3	.000	.010	.010	94.3	−.001	.008	.009	92.6
		“strong”	β_1	.004	.265	.260	94.6	.051	.168	.172	93.5
			β_2	.006	.125	.127	94.0	−.041	.114	.121	91.5
			β_3	−.001	.011	.011	93.5	.000	.010	.011	94.4
“strong”	0.2	“indep”	β_1	−.011	.243	.251	93.6	.035	.191	.192	94.7
			β_2	.006	.113	.118	93.3	.014	.114	.122	92.4
			β_3	.000	.012	.014	91.2	−.001	.014	.016	92.1
		“weak”	β_1	−.005	.278	.289	93.6	.035	.218	.211	95.0
			β_2	.007	.128	.134	93.5	.012	.125	.131	93.9
			β_3	.000	.014	.015	90.8	−.001	.015	.016	91.4
		“strong”	β_1	−.007	.313	.313	95.5	.027	.249	.243	94.7
			β_2	.005	.145	.156	91.8	.005	.141	.148	93.4
			β_3	.000	.015	.017	89.4	−.000	.017	.018	92.0
	0.4	“indep”	β_1	−.012	.244	.252	94.7	.047	.139	.142	92.7
			β_2	.007	.113	.118	93.0	−.020	.102	.110	92.0

				$\widehat{\beta}$				$\widehat{\beta}_{PL}$			
censor	σ_1	γ	β	**Bias**	**SE**	**SD**	**CP**	**Bias**	**SE**	**SD**	**CP**
			β_3	.000	.012	.013	91.7	.001	.013	.015	90.1
		"weak"	β_1	−.010	.285	.287	95.3	.044	.167	.167	93.9
			β_2	.008	.132	.136	93.2	−.027	.115	.123	91.6
			β_3	.000	.014	.015	91.4	.001	.015	.016	91.5
		"strong"	β_1	.003	.265	.262	94.5	.056	.201	.204	93.3
			β_2	.004	.125	.141	93.9	−.041	.140	.151	90.9
			β_3	−.001	.011	.012	93.5	.000	.017	.019	91.3

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Weight loss data analysis using the proposed method, the method of Chen et al. (2015) and the GEE.

	Proposed			Chen et al. (2015)			GEE		
Covariate	**Est**	**SE**	**p-value**	**Est**	**SE**	**p-value**	**Est**	**SE**	**p-value**
Login	−.010	.003	< .001	−.049	.007	< .001	−.044	.002	< .001
Month	−.190	.061	.002	−.261	.247	.292	−.472	.063	< .001
Month^2	.008	.002	< .001	.025	.009	.005	.025	.002	< .001

Table 5:

Weight loss data analysis using the proposed method

	Report records ≥ 2			Report records ≥ 10		
Covariate	**Est**	**SE**	**p-value**	**Est**	**SE**	**p-value**
Login	−.008	.003	.003	−.008	.003	.003
Month	−.252	.066	< .001	−.246	.068	< .001
$Month^2$	.009	.002	< .001	.009	.002	< .001
Gap	.180	.026	< .001	.215	.031	< .001