An estimation method of marginal treatment effects on correlated longitudinal and survival outcomes

QING PAN* AND GRACE Y. YI

This paper concerns treatment effects on correlated longitudinal and time to event processes. The marginal mean of the longitudinal outcome in the presence of event occurrence is often of interest from clinical and epidemiological perspectives. When the probability of the event is treatment-dependent, differences between treatment-specific longitudinal outcome means are usually not constant over time. In this paper, we propose a measure to quantify treatment effects using time-varying differences in longitudinal outcome means, which accounts for the constantly changing population composition due to event occurrences. Generalized linear mixed models (GLMM) and proportional hazards (PH) models are employed to construct the proposed measure. The proposed method is applied to analyze the motivating data arising from the study of weight loss in the Diabetes Prevention Program where weights after diabetes occurrence are systematically different from diabetes-free weights.

KEYWORDS AND PHRASES: Correlated longitudinal and survival processes, Generalized linear mixed model, Marginal treatment effects, Piecewise constant hazard, Proportional hazards model.

1. INTRODUCTION

Correlated longitudinal and time to event outcomes arise commonly in biomedical studies. There has been increasing interest in joint modeling of longitudinal and survival data. Typically, methods can be classified into three categories. *Selection models* postulate the marginal distribution of the longitudinal measures and the conditional distribution of the event time given the longitudinal measures (Diggle and Kenward, 1994), while *pattern-mixture models* factorize the joint distribution into the marginal distribution of the event time and the distribution of the longitudinal measures conditional on the event time (Little, 1993). *Latent models* assume some underlying latent variables, and conditional on those variables, repeated measurements and event time are assumed independent (Wu and Carroll, 1988; Wulfsohn and Tsiatis, 1997). A variety of methods have rooted from the

*Corresponding author.

three basic categories (e.g., Gruttola and Tu, 1994; Henderson et al., 2000; Xu and Zeger, 2001; Lin et al., 2002; Ratcliffe et al., 2004; Chi and Ibrahim, 2006; Elashoff et al., 2007; Ding and Wang, 2008; Liu and Ying, 2007; Zeng and Cai, 2005).

Depending on the research goal, longitudinal outcomes and event time are handled differently. If the prognostic value of the longitudinal measurements on the survival outcome is of prime interest, one may take the longitudinal measurements as covariates (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2001). If treatment effects on the longitudinal outcome under the same event occurrence process are of interest, the event time can be taken as informative censoring of the longitudinal responses (Lin et al., 2004). Different from methods such as Inverse Probability Censoring Weighting (Robins and Rotnitzky, 1992) where information on the longitudinal process is missing after censoring, we consider scenarios where longitudinal outcomes are observed both before and after the event occurrence. Our method is not aimed to evaluate the treatment effects on the longitudinal process given the same censoring distribution. Instead, we target at the marginal treatment effects on the longitudinal processes while acknowledging the treatment-specific differences on the event process.

We propose a measure of marginal treatment effects that drive both the longitudinal and time to event processes. A three-step procedure is particularly adopted to construct such a measure. In the first step, we divide the longitudinal process into two sub-processes – one before the event and the other after it to accommodate heterogeneity among the longitudinal process. GLMMs are used to characterize the longitudinal outcomes in the two sub-processes, and the PH model is called in to feature the time to event process. Secondly, the weighted average of longitudinal outcomes in each treatment group during the follow-up is calculated, where weights are included to feature the probabilities or percentages of having or not having the event. Finally, the marginal treatment effects are estimated by the differences between treatment-specific longitudinal outcome means.

The proposed method is partly motivated by the data arising from the Diabetes Prevention Program targeting at diabetes prevention through weight loss for diabetes-free subjects with impaired glucose tolerance (Diabetes Prevention Program Research Group, 2002). Weights and diabetes

occurrences are the primary outcomes. During the follow-up, a portion of subjects developed diabetes, and often subjects' weights would change after a diabetes diagnosis due to diabetes medication. Models which are used to feature pooled diabetes-free and diabetic data usually assume a constant difference between weights in the prevention and placebo groups, and they do not explicitly reflect that the composition of the population is constantly changing due to the occurrence of diabetes. Here we aim to address this issue. First, we consider three separate models – one for diabetes occurrence over time, another targeting at the weight change rate in diabetes-free subjects, and the third one for the weight change rate in diabetic subjects. Then, we use the proposed measure to quantify the treatment effects marginally. Our proposed method can also be applied to a variety of settings, such as the study of quality of life in the presence of death.

The rest of the paper is organized as follows. Section 2 describes the models and presents the measure to quantify the marginal treatment effects. Asymptotic properties of the estimator are presented in Section 3. In Section 4, we evaluate the performance and robustness of the proposed estimator through simulation studies. The proposed method is applied to the motivating example in Section 5. Concluding remarks are provided in Section 6.

2. NOTATION AND MODEL FORMULATION

Suppose N subjects are independently followed up over a time period, say, $[0, \tau)$, and longitudinal and survival outcomes are both observed for each subject. For $i = 1, \ldots, N$, event time and censoring time for the ith subject are denoted by $\tilde{T}_i$ and C_i, respectively. The observed time is $T_i = \tilde{T}_i \wedge C_i$, where $\wedge$ denotes the operation of taking the minimum of two items. Let $\Delta_i = I(\tilde{T}_i < C_i)$ be the event indicator and $S_i(t)$ be the survival probability for subject i at time t. The longitudinal process is further split into two sub-processes – $(0, \tilde{T}_i)$ and $(\tilde{T}_i, \tau)$. Change rates before and after the event are respectively denoted as $Y_i(t)$ for $t \leq \tilde{T}_i$ and $V_i(t)$ for $t > \tilde{T}_i$. Here the change rate of the longitudinal outcome is defined as the change between two adjacent measures divided by the elapsed time length. Let t_{ij} be the time points where longitudinal measures are collected, $j = 1, \ldots, n_i$.

We employ generalized linear mixed models to postulate the longitudinal change rates (Zeger and Liang, 1986)

$$(1) \qquad Y(t_{ij}) = \mu_Y(t_{ij}) + \mathbf{A}'(t_{ij})\mathbf{a}_i + \nu_i + \epsilon(t_{ij}),$$
$$(2) \qquad V(t_{ij}) = \mu_V(t_{ij}) + \mathbf{B}'(t_{ij})\mathbf{b}_i + \upsilon_i + \varsigma(t_{ij}),$$

where (ν_i, υ_i) and $(\mathbf{a}_i, \mathbf{b}_i)$ are respectively subject-specific random variables, and $\epsilon(t_{ij})$ and $\varsigma(t_{ij})$ represent additional random variation among the repeated measures of subject i at time t_{ij}. Furthermore, $\epsilon(t_{ij})$ and $\varsigma(t_{ij})$, $j = 1, \ldots, n_i$, are assumed to be independent, each following a normal distribution with mean zero and standard deviation σ,

and they are independent of (ν_i, υ_i) and $(\mathbf{a}_i, \mathbf{b}_i)$. Often, subject-specific random effects (ν_i, υ_i) and $(\mathbf{a}_i, \mathbf{b}_i)$ are assumed to have zero mean so that marginal mean of $Y(t_{ij})$ and $V(t_{ij})$ can be respectively featured by $\mu_Y(t_{ij})$ and $\mu_V(t_{ij})$. Also, we assume $\mathbf{a}_i$ and $\mathbf{b}_i$ follow multivariate normal distributions with covariance matrices $\mathbf{V}_a$ and $\mathbf{V}_b$, respectively, and they are independent of each other as well as ν_i and υ_i. The subject-level random intercepts ν_i and υ_i are assumed to follow two correlated normal distributions with covariance matrix Γ. Notice that two sets of equivalent notations are used in describing the longitudinal quantities – $Y(t_{ij}), V(t_{ij}), \mathbf{Z}(t_{ij}), \mathbf{W}(t_{ij}), \mathbf{A}(t_{ij}), \mathbf{B}(t_{ij})$ and $Y(t), V(t), \mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{A}_i(t), \mathbf{B}_i(t)$. They refer to the same processes.

To reflect the dependence on covariates, we invoke regression models

$$g_Y(\mu_Y(t_{ij})) = \mathbf{Z}'(t_{ij})\theta,$$
$$g_V(\mu_V(t_{ij})) = \mathbf{W}'(t_{ij})\gamma,$$

where $\mathbf{Z}(t_{ij})$ and $\mathbf{W}(t_{ij})$ are covariates that could be time-varying, and θ and γ are the corresponding regression coefficient vectors. Link functions g_Y and g_V are assumed known, commonly taking forms such as identity, logarithm, logit or probit. For ease of exposition, a common link function g is assumed for $Y(t_{ij})$ and $V(t_{ij})$. Extensions to accommodating distinct links are straightforward.

The survival outcome is modeled by the proportional hazards model (e.g., Kalbfleisch and Prentice, 2002)

$$(3) \qquad \lambda_i(t) = \lambda_0(t) \exp(\mathbf{X}_i'(t)\beta),$$

where $\mathbf{X}_i(t)$ denotes the covariate vector and β is the corresponding vector of regression coefficients. The baseline hazard function $\lambda_0(t)$ can be modeled parametrically or nonparametrically. Here we adopt a weakly parametric approach, the piecewise constant method, which is widely used in survival analysis (e.g., Lawless and Zhan, 1998; He and Lawless, 2003). To be specific, the cumulative baseline hazard $\Lambda_0(t)$ can be written as:

$$\Lambda_0(t) = \sum_{k=1}^{K} \lambda_{0k} u_k(t),$$

where we pre-specify K time points, $0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_K$, dividing the follow-up time period into K pieces, λ_{0k} denotes the constant hazard rate on the kth interval $[\alpha_{k-1}, \alpha_k)$, $k = 1, \ldots, K$, and $u_k(t) = \max(0, \min(\alpha_k, t) - \alpha_{k-1})$ is the length of the intersection of the kth interval with $(0, t]$. In our simulation studies, we select five cut points between 0 and τ so that approximately equal number of events falls in each subinterval.

The five sets of covariates $\mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{A}_i(t), \mathbf{B}_i(t)$ and $\mathbf{X}_i(t)$ may or may not overlap. Note that although we allow time-varying covariates, the adjustment covariates are

restricted to defined and ancillary ones. That is, time varying covariates whose paths are fixed ahead or decided by external factors, contrary to internal time-varying covariates that are generated by the subjects during follow-up (Kalbfleisch and Prentice, 2002). If the covariates are affected by the treatment or exposure under study, inclusion of internal time-varying covariates would attentuate the treatment effects and require methods such as Inverse Probability Treatment Weighting (Robins et al., 2000).

To feature the marginal cumulative change in the longitudinal outcomes, we propose the following measure

$$
(4) \quad \eta(t) \equiv E\bigg(\int_0^t [E\{Y_i(s)|\mathbf{Z}_i(s)\}E\{S_i(s)|\mathbf{X}_i(s)\} \\
+ E\{V_i(s)|\mathbf{W}_i(s)\}E\{1 - S_i(s)|\mathbf{X}_i(s)\}]ds \bigg),
$$

where $E\{Y_i(s)|\mathbf{Z}_i(s)\}$ and $E\{V_i(s)|\mathbf{W}_i(s)\}$ are the expected change rates, determined by models (1) and (2), respectively. They are weighted by two probabilities related to the event, $E\{S_i(s)|\mathbf{X}_i(s)\}$ and $E\{1 - S_i(s)|\mathbf{X}_i(s)\}$ respectively, which are determined by model (3). Here the inner expectations are conditional expectations of the change rates and the survival probabilities given covariates $\{\mathbf{Z}_i(s), \mathbf{W}_i(s), \mathbf{X}_i(s)\}$, and the outer expectation is taken with respect to the distribution of $\{\mathbf{Z}_i(s), \mathbf{W}_i(s), \mathbf{X}_i(s)\}$ in the target population. Given the model setup of (1), (2) and (3), and assuming that

$$
(5) \quad E\{Y_i(t)|\mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{X}_i(t); \theta\} \\
= E\{Y_i(t)|\mathbf{Z}_i(t); \theta\},
$$

$$
(6) \quad E\{V_i(t)|\mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{X}_i(t); \gamma\} \\
= E\{V_i(t)|\mathbf{W}_i(t); \gamma\}, \text{ and}
$$

$$
(7) \quad E\{S_i(t)|\mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{X}_i(t); \Lambda_0(t), \beta\} \\
= E\{S_i(t)|\mathbf{X}_i(t); \Lambda_0(t), \beta\},
$$

we can express the cumulative change $\eta(t)$, defined by (4), as

(8)

$$
\eta(t; \theta, \gamma, \lambda_0(t), \beta) \\
= E\bigg(\int_0^t \bigg[g^{-1}(\mathbf{Z}_i'(s)\theta) \exp\bigg\{ -\int_0^s \lambda_0(u)e^{\mathbf{X}_i'(u)\beta}du \bigg\} \\
+ g^{-1}(\mathbf{W}_i'(t)\gamma)\bigg\{ 1 - \exp\bigg(-\int_0^s \lambda_0(u)e^{\mathbf{X}_i'(u)\beta}du \bigg) \bigg\} \bigg] ds \bigg).
$$

To explicitly spell out the dependence of $\eta(t)$ on the treatment indicator, we rewrite $\mathbf{Z}_i'(t) = (Z_{i1}(t)\ \mathbf{Z}_{i2}'(t))'$, $\mathbf{W}_i'(t) = (W_{i1}(t)\ \mathbf{W}_{i2}'(t))'$ and $\mathbf{X}_i'(t) = (X_{i1}(t)\ \mathbf{X}_{i2}'(t))'$, where $Z_{i1}(t)$, $W_{i1}(t)$ and $X_{i1}(t)$ are the treatment indicator for the two longitudinal sub-processes and the event process, respectively, and $\mathbf{Z}_{i2}(t)$, $\mathbf{W}_{i2}(t)$ and $\mathbf{X}_{i2}(t)$ denote vectors

of other covariates. Correspondingly, the regression coefficients can be divided into the treatment coefficients (θ_1, γ_1, β_1) and the vectors of other coefficients (θ_2, γ_2, β_2). Denote $\mathbf{Z}_i^{(r)}(t) = (r, \mathbf{Z}_{i2}'(t))'$, $\mathbf{W}_i^{(r)}(t) = (r, \mathbf{W}_{i2}'(t))'$ and $\mathbf{X}_i^{(r)}(t) = (r, \mathbf{X}_{i2}'(t))'$, $r = 1, 0$. Here $\mathbf{Z}_i^{(1)}(t)$, $\mathbf{W}_i^{(1)}(t)$ and $\mathbf{X}_i^{(1)}(t)$ are hypothetical covariate values if subject i took the treatment, and $\mathbf{Z}_i^{(0)}(t)$, $\mathbf{W}_i^{(0)}(t)$ and $\mathbf{X}_i^{(0)}(t)$ denote hypothetical covariate values if subject i took placebo. In reality, each subject often takes either the treatment or placebo but not both, so we would not be able to observe both sets of values.

Let

$$
\eta^{(1)}(t; \theta, \gamma, \lambda_0(t), \beta) \\
= E\bigg[\int_0^t \{E(Y_i(s)|\mathbf{Z}_i^{(1)}(s; \theta))E(S_i(s)|\mathbf{X}_i^{(1)}(s; \Lambda_0(s), \beta)) \\
+ E(V_i(s)|\mathbf{W}_i^{(1)}(s; \gamma))E(1 - S_i(s)|\mathbf{X}_i^{(1)}(s; \Lambda_0(s), \beta))\}ds \bigg]
$$

be the expected marginal cumulative change if the subject took the treatment, and it is calculated by replacing $(Z_{i1}(s), W_{i1}(s), X_{i1}(s))$ in (8) with one. Similarly, the expected marginal cumulative change given the control is

$$
\eta^{(0)}(t; \theta, \gamma, \lambda_0(t), \beta) \\
= E\bigg[\int_0^t \{E(Y_i(s)|\mathbf{Z}_i^{(0)}(s))E(S_i(s)|\mathbf{X}_i^{(0)}(s)) \\
+ E(V_i(s)|\mathbf{W}_i^{(0)}(s))E(1 - S_i(s)|\mathbf{X}_i^{(0)}(s))\}ds \bigg],
$$

and it is calculated by replacing $Z_{i1}(s)$, $W_{i1}(s)$ and $X_{i1}(s)$ in (8) with zero. Here, we are estimating the longitudinal outcome mean changes if the same cohort went through treatment or placebo. As a result, the population-level marginal treatment effects up to time t can be measured by

$$
\eta_d(t) = \eta^{(1)}(t; \theta_0, \gamma_0, \lambda_0(t), \beta_0) - \eta^{(0)}(t; \theta_0, \gamma_0, \lambda_0(t), \beta_0).
$$

3. MAXIMUM LIKELIHOOD ESTIMATOR AND ASYMPTOTIC PROPERTIES

In this section, we discuss estimation procedures and establish the asymptotic results. Without exception, these results require the following standard regularity conditions

(a) $\{Y_i(t), V_i(t), \tilde{T}_i, C_i, \mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{X}_i(t), \mathbf{A}_i(t), \mathbf{B}_i(t)\}$ are independent and identically distributed.
(b) The censoring time C_i is conditionally independent of the event time $\tilde{T}_i$ given covariates, $\lim_{\xi \to 0} \frac{1}{\xi}Pr\{t \le \tilde{T}_i < t + \xi|\tilde{T}_i > t, C_i > t, \mathbf{X}_i(t)\} = \lim_{\xi \to 0} \frac{1}{\xi}Pr\{t \le \tilde{T}_i < t + \xi|\tilde{T}_i > t, \mathbf{X}_i(t)\}$.
(c) C_i and $(Y_i(t), V_i(t))$ are independent.
(d) Elements of $\mathbf{Z}_{i2}(t), \mathbf{W}_{i2}(t), \mathbf{A}_i(t), \mathbf{B}_i(t)$ and $\mathbf{X}_{i2}(t)$ are bounded almost surely.

(e) The information matrix for the maximum likelihood estimators $(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \ldots, \widehat{\lambda}_{0K})$ (to be discussed) is positive-definite.

(f) $Pr(T_i \geq \tau) > 0$.

(g) $\int_0^\tau Y_i(t)dt < \infty$ and $\int_0^\tau V_i(t)dt < \infty$.

Condition (a) is a regular condition required in applying the Central Limit Theorem. Condition (b) corresponds to independent censorship for the event process. Condition (c) guarantees censoring is also independent of the longitudinal processes, which ensures that the longitudinal outcomes can be consistently estimated after censoring. Condition (d), bounded covariates, is a common assumption. Condition (e) excludes the cases that $\mathbf{Z}_i(t)$, $\mathbf{W}_i(t)$ or $\mathbf{X}_i(t)$ has any linear dependency. Condition (f) means that the event process is identifiable up to time point τ, and condition (g) ensures finite longitudinal outcome changes over interval $(0, \tau)$.

With the distributions of $Y_i(t)$, $V_i(t)$ and $\tilde{T}_i$ fully parameterized, the consistency and asymptotic covariance matrix of the parameter estimators, $(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K})$, can be established by adapting the arguments of Guo and Carlin (2004). Specifically, we start with the joint likelihood of $(Y_i(t), V_i(t), T_i, \Delta_i, \nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i)$, given by

$$L(Y_i(t), V_i(t), T_i, \Delta_i, \nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i|\theta, \lambda_0(t), \beta, \sigma, \boldsymbol{\Gamma}, \mathbf{V}_a, \mathbf{V}_b)$$
$$= \prod_{i=1}^N L_i(Y_i(t)|\nu_i, \mathbf{a}_i, \theta, \sigma) L_i(V_i(t)|v_i, \mathbf{b}_i, \gamma, \sigma)$$
$$\quad L_i(T_i, \Delta_i|\lambda_0(t), \beta) L_i(\nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i|\boldsymbol{\Gamma}, \mathbf{V}_a, \mathbf{V}_b).$$

Integrating over the distribution of the latent variables $(\nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i)$, the marginal distribution of $(Y_i(t), V_i(t), T_i, \Delta_i)$ goes as

(9)

$$L(Y_i(t), V_i(t), T_i, \Delta_i|\theta, \lambda_0(t), \beta, \sigma, \boldsymbol{\Gamma}, \mathbf{V}_a, \mathbf{V}_b)$$
$$= \prod_{i=1}^N \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty L_i(Y_i(t)|\nu_i, \mathbf{a}_i, \theta, \sigma) L_i(V_i(t)|v_i, \mathbf{b}_i, \gamma, \sigma)$$
$$\quad L_i(T_i, \Delta_i|\lambda_0(t), \beta) L_i(\nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i|\boldsymbol{\Gamma}, \mathbf{V}_a, \mathbf{V}_b)$$
$$\quad d\nu_i dv_i d\mathbf{a}_{i1} \ldots d\mathbf{a}_{ip_a} d\mathbf{b}_{i1} \ldots d\mathbf{b}_{ip_b},$$

where p_a and p_b denote the dimension of $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively. The four components composing the likelihood (9) are respectively given by

$$L_i(Y_i(t)|\nu_i, \mathbf{a}_i, \theta, \sigma)$$
$$= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(Y_{ij} - g^{-1}(\mathbf{Z}'_{ij}\theta) - \nu_i - \mathbf{A}'_{ij}\mathbf{a}_i)^2}{2\sigma^2}\right\},$$

$$L_i(V_i(t)|v_i, \mathbf{b}_i, \gamma, \sigma)$$
$$= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(V_{ij} - g^{-1}(\mathbf{W}'_{ij}\gamma) - v_i - \mathbf{B}'_{ij}\mathbf{b}_i)^2}{2\sigma^2}\right\},$$

$$L_i(T_i, \Delta_i|\lambda_0(t), \beta)$$
$$= \exp\left\{-\int_0^{T_i} \lambda_0(t)\exp(\mathbf{X}'_i(t)\beta)dt\right\}$$
$$\quad \{\lambda_0(T_i)\exp(\mathbf{X}'_i(T_i)\beta)\}^{\Delta_i},$$

$$L_i(\nu_i, v_i, \mathbf{a}_i, \mathbf{b}_i|\boldsymbol{\Gamma}, \mathbf{V}_a, \mathbf{V}_b)$$
$$= \prod_{i=1}^N \frac{\exp\{-\frac{1}{2}(\nu_i, v_i)'\Gamma^{-1}(\nu_i, v_i) - \frac{1}{2}\mathbf{a}'_i V_a^{-1}\mathbf{a}_i - \frac{1}{2}\mathbf{b}'_i V_b^{-1}\mathbf{b}_i\}}{(2\pi)^{\frac{2+p_a+p_b}{2}}|\boldsymbol{\Gamma}|^{.5}|\mathbf{V_a}|^{.5}|\mathbf{V_b}|^{.5}}.$$

The integration in formula (9) can be achieved through numeric methods such as adaptive Gauss-Quadrature. Then $(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \hat{\lambda}_{02}, \ldots, \hat{\lambda}_{0K})$, the maximum likelihood estimators of $(\theta, \gamma, \beta, \lambda_{01}, \lambda_{02}, \ldots, \lambda_{0K})$, are solutions to the corresponding score equations. The asymptotic covariance matrix, $\boldsymbol{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \hat{\lambda}_{02}, \ldots, \hat{\lambda}_{0K})$, is obtained by taking the inverse of the information matrix.

Now we turn attention to estimating the proposed measure $\eta_d(t)$, which is of primary interest here. First, replace parameters $(\theta_0, \gamma_0, \lambda_0(t), \beta_0)$ in $E\{Y_i(t)|\mathbf{Z}_i^{(1)}\}$, $E\{V_i(t)|\mathbf{W}_i^{(1)}\}$, $E\{S_i(t)|\mathbf{X}_i^{(1)}\}$ and $E\{Y_i(t)|\mathbf{Z}_i^{(0)}\}$, $E\{V_i(t)|\mathbf{W}_i^{(0)}\}$, $E\{S_i(t)|\mathbf{X}_i^{(0)}\}$ in formulas (5), (6) and (7) with their empirical counterparts, $(\hat{\theta}, \hat{\gamma}, \hat{\lambda}_0(t), \hat{\beta})$. Then, we estimate the expectation in (8) with the sample average by plugging in covariates $\mathbf{Z}_{i2}(t)$, $\mathbf{W}_{i2}(t)$ and $\mathbf{X}_{i2}(t)$, where $i = 1, \ldots, N$, in the representative sample and take an average. Therefore, we obtain

$$\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t))$$
$$= \hat{\eta}^{(1)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\Lambda}_0(t)) - \hat{\eta}^{(0)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\Lambda}_0(t))$$
$$= \frac{1}{N} \sum_{i=1}^N \int_0^t \left[g^{-1}\{\mathbf{Z}_i^{(1)}(s)'\hat{\theta}\}e^{-\int_0^s \hat{\lambda}_0(u)e^{\mathbf{X}_i^{(1)}(u)'\hat{\beta}}du}\right.$$
$$\left. + g^{-1}\{\mathbf{W}_i^{(1)}(s)'\hat{\gamma}\}\{1 - e^{-\int_0^s \hat{\lambda}_0(u)e^{\mathbf{X}_i^{(1)}(u)'\hat{\beta}}du}\}\right]ds$$
$$- \frac{1}{N} \sum_{i=1}^N \int_0^t \left[g^{-1}\{\mathbf{Z}_i^{(0)}(s)'\hat{\theta}\}e^{-\int_0^s \hat{\lambda}_0(u)e^{\mathbf{X}_i^{(0)}(u)'\hat{\beta}}du}\right.$$
$$\left. + g^{-1}\{\mathbf{W}_i^{(0)}(s)'\hat{\gamma}\}\{1 - e^{-\int_0^s \hat{\lambda}_0(u)e^{\mathbf{X}_i^{(0)}(u)'\hat{\beta}}du}\}\right]ds.$$

Finally, we discuss the asymptotic distribution of the estimator $\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t))$.

Theorem 3.1. *Under the preceding regularity conditions, we have*

(1) $\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0)$ *is a strong consistent estimator for* $\eta_d(t)$. *That is,* $\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0) \xrightarrow{a.s.} \eta_d(t)$ *as* $n \to \infty$.

(2) $\sqrt{n}\{\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0) - \eta_d(t)\}$ *has an asymptotic Gaussian distribution with mean zero and covariance that can be consistently estimated by* $V(\hat{\eta}_d(t); \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0)$, *whose detailed expression is given in the Appendix.*

The consistency comes as a result of the consistency of $\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0$ by the Slutsky Theorem (Sen and Singer, 1993).

The asymptotic distribution is derived from the empirical Central Limit Theorem (van der Vaart and Wellner, 1996; van der Vaart, 2000) and Delta method. Details are reported in the Appendix.

4. SIMULATION STUDIES

4.1 Performance of the proposed method

We examine a special but important type of events – the survival outcome is a terminating event, such as death. Notice that the longitudinal outcomes are not missing after the terminating event, death. In many clinical measures, for example the health related quality of life derived from SF-36 (McHorney et al., 1994), the score for dead people is well defined, which is zero, representing the worst status. Therefore, we take the special case where the longitudinal outcome after the event occurrence degenerates to constant zero. And only the longitudinal sub-process before the event occurrence and the survival process need be modeled. In the simulation, each configuration contains correlated longitudinal and survival outcomes for 300 independent and identically distributed subjects. The longitudinal outcomes are generated as follows. Two repeated measures, $Y(t_{i1})$ and $Y(t_{i2})$, are generated for each subject from the linear mixed model

$$Y(t_{ij}) = \theta_1 Z_{i1} + \theta_2 Z_{i2} + \nu_i + \epsilon(t_{ij}),$$

for $i = 1, \ldots, N$ and $j = 1, 2$. Notice we assume random intercepts but no random slopes. The two observation times are $t_{i1} = 50$ and $t_{i2} = 100$. Covariate Z_{i1}, representing the treatment indicator, is generated from Bernoulli distribution with equal probabilities to be 0 or 1. Covariate Z_{i2} is generated from a normal distribution with mean zero and standard deviation 0.1, and the regression coefficient θ_2 is fixed at 1. Random effects ν_i and ϵ_{ij} follow two independent mean zero normal distributions with variances σ_l^2 and σ^2, respectively. The survival outcome, time to event, is generated from the exponential distribution with hazard rate $0.02 \exp(\beta_1 X_{i1} + \beta_2 X_{i2})$. Let $X_{i1} = Z_{i1}$ to indicate the treatment. Covariate X_{i2} follows normal$(0, 0.01)$ with coefficient $\beta_2 = 1$. Censoring time is generated from uniform distribution on $(0, 100]$, creating censoring percentages from 35% to 55%.

We examine the performance of the proposed estimator $\hat{\eta}_d(t)$ under various scenarios. First, the coefficients for the treatment indicator in the mixed model and in the proportional hazards model, (θ_1, β_1), are set as $(0, 0)$, $(1, 1)$ or $(1, -1)$ to examine the cases of no treatment effects, treatment effects in the same direction and those in opposite directions. Second, the variances of (ν_i, ϵ_{ij}), (σ_l^2, σ^2), equal either $(0.5, 0.1)$ representing the cases when variation between subjects is the main source of variation or $(0.1, 0.5)$ when variation within subjects dominates.

True values of the cumulative differences between the treatment and control groups up to time point t, $\eta_d(t)$, are obtained in three steps. First we calculate $E(Y_i(t)|\mathbf{Z}_i^{(1)}(t); \theta) E(S_i(t)|\mathbf{X}_i^{(1)}(t); \lambda_0(t), \beta) - E(Y_i(t)|\mathbf{Z}_i^{(0)}(t); \theta) E(S_i(t)|\mathbf{X}_i^{(0)}(t); \lambda_0(t), \beta)$ where all parameters take their true values. Then expectations over the distribution of adjustment covariates Z_{i2} and X_{i2} are taken through the Gauss-Hermit method. Finally, the expectations are integrated over time using numerical integration.

Table 1 reports the simulation results on the performance of the proposed estimator $\hat{\eta}_d(t)$ where $\tau = 100$, as well as the treatment regression coefficient on the longitudinal outcomes, $\hat{\theta}_1$, and that on the survival process, $\hat{\beta}_1$. Each row corresponds to a unique combination of true treatment coefficients value (θ_1, β_1) and variance level (σ_l^2, σ^2). The true value of $\eta_d(100)$ is also listed before its estimator, $\hat{\eta}_d(100)$. We particularly report on the biases, asymptotic standard errors (ASE), empirical standard deviations (ESD) and coverage probabilities (CP). Our analysis yields consistent estimates with small finite sample biases. Model-based standard errors agree well with empirical standard deviations. In all the scenarios, coverage probabilities are close to the nominal level 0.95.

4.2 Evaluation of robustness: Unmeasured covariates

The marginal survival model (3) is fitted separately from the GLMMs, essentially assuming no additional correlation between the longitudinal and survival outcomes beyond the covariates $\mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{X}_i(t)$. This could be true because the two longitudinal sub-processes are conditional on whether the subject has or has not experienced the event. And we can try to include all relevant information in the covariates. However, there might be cases with unmeasured covariates for the survival process. If any unmeasured covariate is correlated with $\mathbf{X}_i(t)$, it becomes a confounder. Even when all unmeasured covariates are independent of the adjusted covariates, they could be correlated with the random variables in the GLMMs such as ν_i and υ_i. To verify the robustness of our method, we generate time to event data assuming a random variable in the PH model as follows

$$\lambda_i(t) = \lambda_0(t) \exp(\mathbf{X}_i'(t)\beta + \omega_i),$$

random variable ω_i represents an unmeasured covariate. The normally distributed ω_i is correlated with the random intercepts in the GLMMs, ν_i, υ_i. We assume that conditional on $(\nu_i, \upsilon_i, \omega_i, \mathbf{a}_i, \mathbf{b}_i, \mathbf{Z}_i(t), \mathbf{W}_i(t), \mathbf{A}_i(t), \mathbf{B}_i(t), \mathbf{X}_i(t))$, $(\tilde{T}_i, Y_i(t), V_i(t))$ are mutually independent. Parameters are set up in the same way as Table 1. The only difference is the addition of ω_i, which follows a normal distribution with mean zero and variance σ_s^2. Set $\sigma_s^2 = \sigma_l^2$ and the correlation coefficient ρ between ν_i and ω_i as 0.5.

We notice that Table 2 demonstrates satisfactory point estimates and coverage probabilities obtained from the

Table 1. Simulation results: Performance of the proposed method

(σ_l^2, σ^2)	θ_1	$\hat{\theta}_1$ bias	ESD	ASE	CP	β_1	$\hat{\beta}_1$ bias	ESD	ASE	CP	$\eta_d(100)$	$\hat{\eta}_d(100)$ bias	ESD	ASE	CP
(.1,.5)	1	−.00	.07	.07	.95	1	.06	.21	.21	.95	18.4	−.31	3.09	3.10	.95
(.1,.5)	0	−.00	.07	.07	.94	0	.02	.19	.18	.94	0	−.09	2.95	2.91	.96
(.1,.5)	1	−.00	.07	.07	.93	−1	−.09	.27	.27	.94	70.7	.34	5.87	5.85	.96
(.5,.1)	1	−.00	.09	.09	.93	1	.02	.24	.24	.94	18.4	0.17	4.14	3.96	.93
(.5,.1)	0	.00	.09	.09	.95	0	.03	.20	.19	.94	0	.03	3.79	3.67	.94
(.5,.1)	1	.00	.08	.09	.94	−1	−.03	.29	.28	.94	70.7	−.67	6.47	6.79	.95

ESD = Empirical Standard Deviation;   ASE = average Asymptotic Standard Error;
CP = Coverage Probability for 95% confidence intervals; $\eta_d(100)$ = true value of the proposed cumulative difference at $t = 100$.

Table 2. Sensitivity analysis: Unmeasured covariates

(σ_l^2, σ^2)	θ_1	$\hat{\theta}_1$ bias	ESD	ASE	CP	β_1	$\hat{\beta}_1$ bias	ESD	ASE	CP	$\eta_d(100)$	$\hat{\eta}_d(100)$ bias	ESD	ASE	CP
(.1,.5)	1	.00	.07	.07	.95	1	.07	.22	.22	.95	18.4	−.22	3.02	3.13	.95
(.1,.5)	0	−.01	.07	.07	.95	0	.03	.18	.18	.95	0	−.22	2.95	2.90	.95
(.1,.5)	1	.01	.07	.07	.95	−1	−.09	.27	.27	.94	70.7	−.30	5.54	5.88	.96
(.5,.1)	1	.00	.09	.09	.95	1	.04	.25	.24	.92	18.4	.16	3.86	3.96	.95
(.5,.1)	0	.01	.08	.09	.94	0	.02	.19	.19	.92	0	.28	3.59	3.67	.95
(.5,.1)	1	.00	.09	.09	.94	−1	−.02	.27	.28	.94	70.7	−.59	6.55	6.77	.97

marginal models. When unmeasured covariates are of concern, any modeling efforts trying to adjust for the correlated latent variables make assumptions about the distribution of the latent variables and the form of the correlation, which are hard to validate if possible at all. The marginal method has the advantage of robustness over the method requiring specification of the joint models, as the former method needs less model assumption and is less sensitive to model misidentification. The coefficient estimates from marginal models can be considered as covariate effects averaged over the distribution of the unmeasured variables. What's more, the marginal method has computational advantage as it is often much easier and quicker to fit separate marginal models.

4.3 Evaluation of robustness: Piecewise exponential assumption

Piecewise exponential distribution is widely used to model event time distribution in cases of unknown underlying distributions. The piecewise constant hazard rate assumption usually approximates the true distribution well as long as there are at least 4–6 cut points and the sample size within each interval is not too small (Yi and Lawless, 2007). In order to examine the robustness of the piecewise exponential assumption to different underlying distributions, we consider two scenarios where the piecewise exponential assumption is violated. Specifically, data from Weibull and lognormal distributions are generated. The hazard rate for the Weibull distribution is set up as $\{0.02 \exp(X_{i1}+X_{i2})\}^{1.5}$. The logarithm of the lognormal distributed event time follows a normal distribution with mean $4 + X_{i1} + X_{i2}$ and standard deviation 1.5. In both cases, covariates X_{i1} and

X_{i2} are set up in the same way as those in Table 1. The random effects, ν_i, and the longitudinal outcomes, $Y_i(t)$, are generated using the parameter values in the first row of Table 1. Censoring time is uniformly distributed over (0,100], leading to approximately 30% of censoring in the Weibull data and 45% in the lognormal data.

Three sets of models are fitted to each underlying distribution – the accelerated failure time (AFT) model assuming a Weibull distribution, the AFT model assuming a lognormal distribution and the AFT model assuming a piecewise exponential distribution with 6 pieces. Although we are proposing a proportional hazards model in Section 2 and using AFT models here, in the special case of piecewise exponential or exponential distributions the coefficients from an AFT model and those from a proportional hazards model have the same absolute values and only differ by a sign. Therefore, when both models assume the same piecewise exponential distribution, the bias, ASE and ESD of the proposed proportional hazards model are the same as those from an AFT model. Bias, ASE and ESD for $\hat{\beta}_1$ and $\hat{\eta}_d(100)$ are reported in Table 3.

The variance estimators turn out to be robust and the proposed method yields ASE close to ESD, even under misspecified models. However, the biases for $\hat{\beta}_1$ and $\hat{\eta}_d(100)$ differ quite a lot when different distributions are assumed. The biases are smallest for the models assuming the true distribution. The biases inflate under the piecewise exponential model with coverage probabilities between 0.90 and 0.95, lower than the nominal level. Under the models assuming distribution other than the true underlying distribution or piecewise exponential, the biases are largest. In summary, the piecewise exponential model doesn't perform as well as

Table 3. Sensitivity analysis: Piecewise exponential assumption

Distribution	Model	β_1	$\hat{\beta}_1$ bias	ESD	ASE	$\eta_d(100)$	$\hat{\eta}_d(100)$ bias	ESD	ASE	CP
Weibull	Weibull	1	$-.02$	.11	.10	16.7	0.77	2.57	2.58	0.94
Weibull	Piecewise	1	$-.03$	.13	.14	16.7	0.43	2.67	2.66	0.95
Weibull	Lognormal	1	.04	.13	.12	16.7	3.08	2.66	2.86	0.84
Lognormal	Weibull	1	$-.13$	.20	.19	34.9	$-.87$	4.61	4.51	0.94
Lognormal	Piecewise	1	$-.02$	.25	.26	34.9	-3.40	5.10	5.20	0.90
Lognormal	Lognormal	1	$-.00$	.20	.20	34.9	$-.06$	4.37	4.49	0.96

the model correctly specifying the underlying distribution, however, it is generally better than models with misspecified distributions. In cases when the underlying model is unknown, piecewise exponential distribution is a reasonable and robust choice.

One exception in Table 3 is worth noting. When the underlying distribution is lognormal, fitting a Weibull model results in smaller bias for $\hat{\eta}_d(100)$ ($-.87$) compared to the bias from a piecewise exponential model (-3.40). Notice that the bias in $\hat{\beta}_1$ from fitting a piecewise exponential model ($-.02$) is much smaller than that from a Weibull model ($-.13$). The reversed pattern in the biases of $\hat{\eta}_d(100)$ might be explained as follows – in the Weibull model, the bias for $\hat{\beta}_1$ and the bias for $\hat{\lambda}_0(t)$ are probably in different directions and cancel out in calculating $\hat{S}(t)$. Hence, smaller bias in $\hat{\eta}_d(100)$ does not necessarily mean better fit.

5. APPLICATION

Type 2 diabetes, also called non-insulin-dependent diabetes, affects about 8% of adults in the United States. Since treatments are costly and only slow the progression of the disease without a complete cure, prevention is preferred for people at risk. The Diabetes Prevention Program (DPP) (Diabetes Prevention Program Research Group, 2002) is a large-scale, multi-center, randomized clinical trial targeting at diabetic-free adults with elevated fasting and post-load blood glucose levels. From 1996 to 1997, qualified adults were recruited at 27 clinic centers across US. Among them, 1,082 were randomized to the placebo group, and 1,079 to a lifestyle modification program. The goals of the lifestyle modification program were for subjects to lose weight through at least 150 minutes of physical activity per week and controlled calories intake. A recent data collection ended on October 31, 2008. At that time, subjects were followed up to 4,419 days. During the follow-up, 39% (417) subjects in the lifestyle group developed diabetes. In contrast, 49% (527) diabetes cases occurred in the placebo group. Variables collected at randomization included treatment assignment, age, race, gender, life habits such as smoking, marriage and employment status, and previous disease history. Laboratory measures such as diabetes status and weight were updated quarterly. Weight is one of the most

important and sensitive indicators of general health and diabetes risk. A prevention effect on weight is a main research topic in DPP. The research question of interest here is the difference in average weight changes between subjects receiving the lifestyle program and placebo.

Linear mixed models are used to postulate the repeated measures of weight change rates. Adjustment covariates include length of follow-up, demographic characteristics, smoking status, cardiovascular disease history and high blood cholesterol history, defined as the low density lipid (LDL) cholesterol concentration higher than $130mmol/L$. Subject-specific random slopes over follow-up length and random intercepts are assumed. Notice that the only time-varying covariate, follow-up length, is not affected by the treatment and all other covariates are measured at baseline. Therefore, no internal covariate is of concern here. After a subject is diagnosed with diabetes, he or she receives appropriate medication to prevent further degradation. The diabetes medication often affects weights. Patients' weights after diabetes occurrence differ dramatically from diabetes-free weights when subjects are not taking the medications. Therefore, two separate models are employed to model the weight change rates before and after diabetes occurrence respectively. Their coefficients estimates are reported side by side in Table 4.

In the diabetes-free subjects, the lifestyle modification program reduces weights significantly more than the placebo group (-0.39 kilogram/year). However, the difference diminishes in diabetic subjects (p-value $= 0.35$). It would be inappropriate to fit a single model pooling both diabetes-free and diabetic weights, which can not distinguish different coefficient values in these two groups. The proposed method is flexible to accommodate nonhomogeneous samples because diabetic-free and diabetic weight change rates are estimated separately.

Next, we estimate the probability of being diabetes-free or diabetic. The output from the proportional hazards model assuming a piecewise exponential distribution for diabetes occurrences are in Table 5. Constant hazard rates are assumed within the first year and every two years afterwards. The baseline hazard rate in each time period can be estimated by exponentiating the corresponding coefficient. Lifestyle modification program recipients have 68%

Table 4. Application to weight change rates (kg/yr) in DPP: Estimated regression parameters from mixed models

Covariate, $Z_{ik} = W_{ik}$	Pre-Diabetes			Post-Diabetes		
	Est.	SE	p-value	Est.	SE	p-value
Lifestyle	-0.39	0.10	$< .01$	0.15	0.16	0.35
Female	0.07	0.11	0.51	-0.30	0.17	0.08
Age ≤ 45	0.66	0.15	$< .01$	-0.02	0.25	0.94
Age 45–59	0.46	0.13	$< .01$	0.10	0.22	0.64
African American	0.21	0.13	0.11	0.02	0.19	0.90
Hispanic	-0.01	0.14	0.95	0.16	0.22	0.48
Asian	0.09	0.23	0.70	0.46	0.35	0.19
Smoke	-0.15	0.20	0.46	0.13	0.30	0.66
Cardiovascular disease history	-0.18	0.15	0.25	0.11	0.22	0.62
History of High LDL cholesterol	-0.08	0.10	0.45	0.00	0.16	0.96

Table 5. Analysis of diabetes risk in DPP: Estimated regression parameters from proportional hazards model with piecewise constant hazards

Covariate, X_{ik}	Est.	SE	p-value	BHR/HR
Follow-up: ≤ 1 year	-9.04	0.14	$< .01$	0.00
Follow-up: 1–3 years	-9.23	0.13	$< .01$	0.00
Follow-up: 3–5 years	-9.16	0.13	$< .01$	0.00
Follow-up: 5–7 years	-9.44	0.14	$< .01$	0.00
Follow-up: 7–9 years	-9.49	0.15	$< .01$	0.00
Follow-up: 9–11 years	-8.77	0.11	$< .01$	0.00
Lifestyle	-0.38	0.07	$< .01$	0.68
African American	0.34	0.08	$< .01$	1.40
Hispanic	0.07	0.09	0.46	1.07
Asian	0.06	0.15	0.70	1.06
Female	-0.13	0.07	0.07	0.88
Age ≤ 45	0.51	0.10	$< .01$	1.67
Age 45–59	0.27	0.10	$< .01$	1.31
Cardiovascular disease history	0.28	0.10	$< .01$	1.32
History of High LDL cholesterol	-0.08	0.07	0.22	0.92
History of stroke	0.51	0.28	0.07	1.66

BHR/HR: Baseline Hazards Rate or Hazards Ratio.

Table 6. Analysis of DPP effects on weights: Differences in mean weight changes (kg) between prevention and placebo recipients

t	Weight Change in Lifestyle		Weight Change in Placebo		$\hat{\eta}_d(t)$	$\hat{SE}\{\hat{\eta}_d(t)\}$	p-value
	$\hat{\eta}^{(1)}(t)$	$\hat{SE}\{\hat{\eta}^{(1)}(t)\}$	$\hat{\eta}^{(0)}(t)$	$\hat{SE}\{\hat{\eta}^{(0)}(t)\}$			
1 year	-0.25	0.07	0.28	0.07	-0.53	0.10	$< .0001$
2 year	-0.50	0.13	0.54	0.15	-1.04	0.20	$< .0001$
3 year	-0.75	0.20	0.80	0.22	-1.55	0.30	$< .0001$
4 year	-1.00	0.27	1.07	0.29	-2.07	0.40	$< .0001$
5 year	-1.25	0.34	1.33	0.37	-2.59	0.50	$< .0001$
6 year	-1.51	0.40	1.57	0.44	-3.08	0.60	$< .0001$
7 year	-1.76	0.47	1.81	0.51	-3.56	0.69	$< .0001$
8 year	-2.01	0.54	2.04	0.59	-4.05	0.79	$< .0001$
9 year	-2.26	0.60	2.27	0.66	-4.53	0.89	$< .0001$
10 year	-2.51	0.67	2.42	0.73	-4.92	0.99	$< .0001$
11 year	-2.76	0.74	2.56	0.80	-5.32	1.09	$< .0001$

$\hat{\eta}^{(1)}(t)$: average weight change in the lifestyle group at the end of 1 to 11 years;
$\hat{\eta}^{(0)}(t)$: average weight change in the placebo group at the end of 1 to 11 years;
$\hat{\eta}_d(t)$: differences in mean weight changes between lifestyle group and placebo group at the end of 1 to 11 years;
p-values: two-sided P-value for $H_0 : \eta_d = 0$ vs. $H_a : \eta_d \neq 0$.
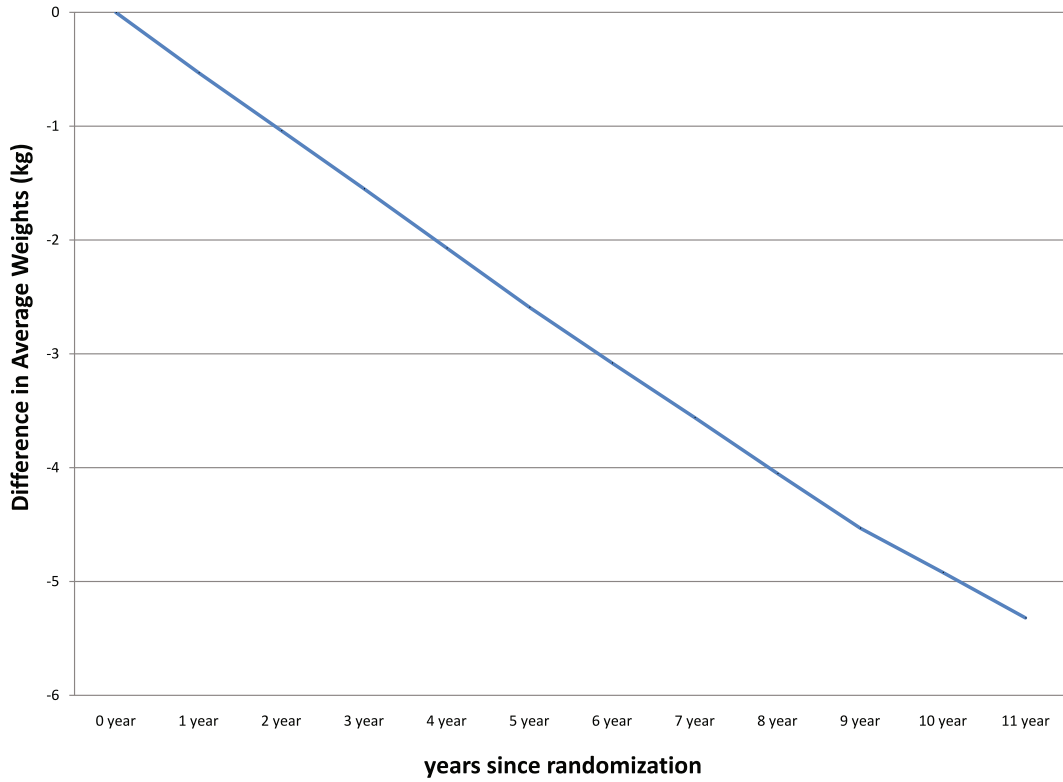
Figure 1. Difference in mean weight changes between the DPP lifestyle and placebo groups over follow-up.

of the diabetes risk compared to those receiving placebo (p-value < 0.01).

Table 6 lists the proposed estimators $\hat{\eta}^{(1)}(t)$, $\hat{\eta}^{(0)}(t)$ and $\hat{\eta}_d(t)$. Lifestyle prevention recipients on average lose 0.25 kilograms at the end of year one, while subjects in the placebo group gain an average of 0.28 kilograms one year after enrollment. The loss of weight in the lifestyle group increases with time and reaches 2.76 kilograms at the end of 11 years. As illustrated in Figure 1, the difference between the mean weight changes in the lifestyle and placebo recipients reaches 5.32 kilograms after 11 years. At each time point, the average weight loss in the lifestyle group is significantly different from zero. And the differences between the two groups over time are significantly nonzero with two-sided p-values less than 0.0001. In summary, taking the lifestyle modification program leads to significant weight loss regardless whether diabetes occur or not in the follow-up.

6. DISCUSSION

We propose a measure combining treatment effects on correlated longitudinal and time to event outcomes. The proposed method involves estimating event probabilities and longitudinal outcomes among subjects with or without the event in three separate models – a proportional hazards model and two generalized linear mixed models. Both theoretical and numerical studies are provided to evaluate the performance of the proposed method. In the analysis of DPP data, a challenge is present due to the fact that the composition of the cohorts changes constantly as diabetes occur and the occurrence rates differ in the Lifestyle and Placebo groups. The proposed method provides a method to assess the lifestyle prevention effects on mean weight changes. It is found that lifestyle prevention recipients consistently lose more weight than subjects in the placebo group, and the statistically significant difference increases with time.

The proposed method provides estimation of subject-level outcome-specific treatment effects β_1, γ_1 and θ_1. Moreover, it enables us to provide population-level marginal treatment effect measure $\eta_d(t)$ which combines the three effects. Together they provide a comprehensive view of the treatment impacts.

The proposed method is flexible in several aspects. First, models other than the GLMM and the piecewise exponential PH model can be employed to describe other types of longitudinal outcomes and survival probabilities. Secondly, one may include time-varying covariates $\mathbf{Z}_i(t)$, $\mathbf{W}_i(t)$ and $\mathbf{X}_i(t)$ in the models to feature temporal effects explicitly. Thirdly, as the marginal treatment effects $\eta_d(t)$ is characterized by the integrals over the time period $(0, t]$, the proposed method can accommodate varying visit times across subjects. Finally, the proposed estimator can be combined with other methods to handle more complicated data structure such as different degrees of compliance or hierarchical random effects.

An estimation method of marginal treatment effects on correlated longitudinal and survival outcomes 507

The proposed measure $\eta_d(t)$ has potential applications in broad areas. For example, if $Y_i(t)$ and $V_i(t)$ represent the spending rates of medical costs, $\eta_d(t)$ can be interpreted as the mean difference in total costs up to time t; if $Y_i(t)$ and $V_i(t)$ are quality of life score change rates, $\eta_d(t)$ can be used to measure the average gained/lost quality of life scores over time period $(0, t]$.

APPENDIX A. ASYMPTOTIC VARIANCE OF $\hat{\eta}_D(T)$

We derive the asymptotic variance of $\hat{\eta}(t)$ through the covariance matrix of $(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K})$ by the Delta method. The derivatives of $\hat{\eta}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K})$ over the parameters are as follows

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\theta}} = N^{-1} \sum_{i=1}^{N} \int_0^t (g^{-1})' \mathbf{Z}_i(s) E(S_i(s)) ds,$$

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\gamma}} = N^{-1} \sum_{i=1}^{N} \int_0^t (g^{-1})' \mathbf{W}_i(s)\{1 - E(S_i(s))\} ds,$$

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\beta}} = N^{-1} \sum_{i=1}^{N} \int_0^t [-g^{-1}\{\mathbf{Z}_i^{(1)}(s)'\hat{\theta}\} + g^{-1}\{\mathbf{W}_i^{(1)}(s)'\hat{\gamma}\}]$$
$$E(S_i(s)) \int_0^s \hat{\lambda}_0(u) e^{\mathbf{X}_i'(u)\hat{\beta}} \mathbf{X}_i'(u) du ds,$$

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\lambda}_{01}} = N^{-1} \sum_{i=1}^{N} \int_0^t [-g^{-1}\{\mathbf{Z}_i^{(1)}(s)'\hat{\theta}\} + g^{-1}\{\mathbf{W}_i^{(1)}(s)'\hat{\gamma}\}]$$
$$E(S_i(s)) \int_0^{t_1 \wedge t} e^{\mathbf{X}_i'(u)\hat{\beta}} du ds,$$

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\lambda}_{02}} = N^{-1} \sum_{i=1}^{N} \int_0^t [-g^{-1}\{\mathbf{Z}_i^{(1)}(s)'\hat{\theta}\} + g^{-1}\{\mathbf{W}_i^{(1)}(s)'\hat{\gamma}\}]$$
$$E(S_i(s)) \int_{t_1 \wedge t}^{t_2 \wedge t} e^{\mathbf{X}_i'(u)\hat{\beta}} du ds,$$
$$\vdots$$

$$\frac{\partial \hat{\eta}(t)}{\partial \hat{\lambda}_{0K}} = N^{-1} \sum_{i=1}^{N} \int_0^t [-g^{-1}\{\mathbf{Z}_i^{(1)}(s)'\hat{\theta}\} + g^{-1}\{\mathbf{W}_i^{(1)}(s)'\hat{\gamma}\}]$$
$$E(S_i(s)) \int_{t_{K-1} \wedge t}^{t_K \wedge t} e^{\mathbf{X}_i'(u)\hat{\beta}} du ds,$$

where $(g^{-1})'$ represents the derivative of g^{-1} function over $\mathbf{Z}_i^{(1)}(s)'\hat{\theta}$ or $\mathbf{W}_i^{(1)}(s)'\hat{\gamma}$, and $E(S_i(s))$ is obtained from the proportional hazards model with parameters replaced by their empirical counterparts $\hat{\beta}, \hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K}$.

Let $\hat{\lambda}_0(t) = (\hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K})$ and $\lambda_0(t) = (\lambda_{01}, \ldots, \lambda_{0K})$. Once the derivatives are obtained, the variance of $\hat{\eta}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_{01}, \ldots, \hat{\lambda}_{0K})$ goes as follows

$$V(\hat{\eta}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t))) = \mathbf{\Upsilon}'(t) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)) \mathbf{\Upsilon}(t),$$

where $\mathbf{\Upsilon}(t) = (\frac{\partial \hat{\eta}(t)}{\partial \hat{\theta}}, \frac{\partial \hat{\eta}(t)}{\partial \hat{\gamma}}, \frac{\partial \hat{\eta}(t)}{\partial \hat{\beta}}, \frac{\partial \hat{\eta}(t)}{\partial \hat{\lambda}_{01}}, \ldots, \frac{\partial \hat{\eta}(t)}{\partial \hat{\lambda}_{0K}})$. Therefore,

(10)
$$V(\hat{\eta}^{(1)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(t)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}),$$
$$V(\hat{\eta}^{(0)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(t)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}),$$
$$COV(\hat{\eta}^{(1)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)), \hat{\eta}^{(0)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(t)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}),$$
$$COV(\hat{\eta}^{(1)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)), \hat{\eta}^{(1)}(s; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(s)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}),$$
$$COV(\hat{\eta}^{(0)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)), \hat{\eta}^{(0)}(s; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(s)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}),$$
$$COV(\hat{\eta}^{(1)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)), \hat{\eta}^{(0)}(s; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(s)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}),$$
$$COV(\hat{\eta}^{(0)}(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)), \hat{\eta}^{(1)}(s; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)))$$
$$= (\mathbf{\Upsilon}'(t)|\mathbf{Z}_i^{(0)}, \mathbf{W}_i^{(0)}, \mathbf{X}_i^{(0)}) \mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t)|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{X}_i)$$
$$(\mathbf{\Upsilon}(s)|\mathbf{Z}_i^{(1)}, \mathbf{W}_i^{(1)}, \mathbf{X}_i^{(1)}).$$

In the formulas above, treatment indicators Z_{i1}, W_{i1}, X_{i1} are uniformly replaced by 1 or 0 in $\mathbf{\Upsilon}$. However, the calculation of $\mathbf{\Sigma}(\hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t))$ still takes the observed Z_{i1}, W_{i1}, X_{i1} values. Finally, given $\hat{\eta}_d(t) = \hat{\eta}^{(1)}(t) - \hat{\eta}^{(0)}(t)$, we calculate the covariance of $\hat{\eta}_d(t; \hat{\theta}, \hat{\gamma}, \hat{\beta}, \hat{\lambda}_0(t))$,

$$V(\hat{\eta}_d(t)) = V(\hat{\eta}^{(1)}(t)) + V(\hat{\eta}^{(0)}(t)) - 2COV(\hat{\eta}^{(1)}(t), \hat{\eta}^{(0)}(t)),$$

where

$$COV(\hat{\eta}_d^{(1)}(t), \hat{\eta}_d^{(0)}(s))$$
$$= COV(\hat{\eta}^{(1)}(t), \hat{\eta}^{(1)}(s)) + COV(\hat{\eta}^{(0)}(t), \hat{\eta}^{(0)}(s))$$
$$- COV(\hat{\eta}^{(1)}(t), \hat{\eta}^{(0)}(s)) - COV(\hat{\eta}^{(0)}(t), \hat{\eta}^{(1)}(s)).$$

All components are estimated in (10).

ACKNOWLEDGEMENTS

REFERENCES

CHI, Y. Y. and IBRAHIM, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62** 432–445. MR2227491

DIABETES PREVENTION PROGRAM RESEARCH GROUP (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine* **346** 393–403.

DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43** 49–93.

DING, J. and WANG, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64** 546–556. MR2432425

ELASHOFF, R. M., LI, G. and LI, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure type. *Biometrics* **64** 762–771. MR2526626

GRUTTOLA, V. D. and TU, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50** 1003–1014.

GUO, X. and CARLIN, B. P. (2004). Separate and joint modelling of longitudinal and event time data using standard computer packages. *The American Statistician* **58** 16–24. MR2055507

HE, W. and LAWLESS, J. F. (2003). Flexible maximum likelihood methods for bivariate proportional hazards models. *Biometrics* **59** 837–848. MR2025107

HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1** 465–480.

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. Wiley, Hoboken, NJ. MR1924807

LAWLESS, J. F. and ZHAN, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* **26** 549–565.

LIN, H., SCHARFSTEIN, D. O. and ROSENHECK, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society B* **66** 791–813. MR2088782

LIN, H., TURNBULL, B. W., MCCULLOCH, C. E. and SLATE, E. H. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* **97** 53–65. MR1947272

LITTLE, R. (1993). Statistical analysis of masked data. *Journal of Official Statistics* **90** 407–426.

LIU, M. and YING, Z. (2007). Joint analysis of longitudinal data with informative right censoring. *Biometrics* **63** 363–371. MR2370794

MCHORNEY, C. A., WARE, J. E., LU, J. F. R. and SHERBOURNE, C. D. (1994). The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* **32** 40–66.

RATCLIFFE, S. J., GUO, W. and TEN HAVE, T. R. (2004). Joint modeling of longitudinal and survival data via common frailty. *Biometrics* **60** 892–899. MR2133541

ROBINS, J. M., HERNAN, M. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

ROBINS, J. M. and ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology, Methodological Issues.* Birkhäuser, Boston.

SEN, P. K. and SINGER, J. M. (1993). *Large Sample Methods in Statistics.* Chapman and Hall, New York. MR1293125

TSIATIS, A. and DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88** 447–458. MR1844844

VAN DER VAART, A. W. (2000). *Asymptotic Statistics.* Cambridge, Melbourne.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York. MR1385671

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. MR0931633

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186

XU, J. and ZEGER, S. (2001). Joint analysis of longitudinal data comprising repeated measures and time to events. *Applied Statistics* **50** 375–387. MR1856332

YI, G. Y. and LAWLESS, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference* **137** 1816–1828. MR2323865

ZENG, D. and CAI, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measures and survival time. *The Annals of Statistics* **33** 2132–2163. MR2211082

ZEGER, S. L. and LIANG, K. Y. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

Qing Pan
Department of Statistics
George Washington University
Washington, DC 20052
USA
E-mail address: qpan@gwu.edu

Grace Y. Yi
Department of Statistics & Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
E-mail address: yyi@uwaterloo.ca