CONTRIBUTIONS TO THE ANALYSES OF RECURRENT EVENTS AND COMPETING RISKS

Jung In Kim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

Jason Fine

Alan Brookhart

Feng-Chang Lin

Michael Hudgens

Anastasia Ivanova

ABSTRACT

Jung In Kim: Contributions to the Analyses of Recurrent Events and Competing Risk
(Under the direction of Jason Fine)

There is a growing interest in the analysis of recurrent events data. Recurrent events are frequently considered as an outcome when a subject could possibly experience more than one event over follow-up period. Thus, It is important to consider previous events history to explore the relationship between the effects of covariates and the correlated failure times. We extend the Cox-type model with time-varying effect depending on the number and the gap time between previous events to enhance both model fit and prediction. Parameter estimation and statistical inference can be achieved via the partial likelihood. A statistical test procedure is provided to assess the existence of the triggering effects. We demonstrate our approach via comprehensive simulation studies and chronic pseudomonas infections in young cystic fibrosis patients data. Significantly, our model provides better predictions than the existing models.

When some patients do not adhere to their assigned treatments in a randomized trial, the standard intention-to-treat analysis may not properly estimate the effect of treatment on the outcome. Also, considering only received treatment without accounting for unmeasured confounders could be biased. Therefore, it is challenging to obtain the true treatment effect, which can be observed when all subjects comply their assigned regime. Instrumental variable methods help us to consistently estimate the average causal effect of an exposure on some outcome of interest even in the presence of latent confounding. We apply Abadie's weighting scheme to estimate corresponding local average response functions in survival analysis. The method is demonstrated by simulation studies and the colorectal cancer screening data, designed and sponsored by the National Cancer Institute.

Competing risks also occur when subjects can experience one or more events which *compete* with the outcome of interest. In these cases, the competing risk inhibits to observe the event of interest or modifies the chance that this event occurs. We extend existing parametric approaches to estimate the cumulative incidence function for considering both left truncation and interval in competing risks settings. This parametric method is applied to data from the study of osteoporotic fractures to bone mineral density testing interval with age as time scale.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisors, Prof. Jason Fine. He has been leading me through the entire Ph.D program. Also, I would like to thank all my committees Prof. Feng-Chang Lin, Prof. Alan Brookhart, Prof. Michael Hudgens, and Prof. Anastasia Ivanova for all valuable comments.

I really do appreciate my family: my parents and my brother for supporting me throughout my life with strong faith and consistent love.

I would like to thank all my friends who stand by me. My 2010 classmates who have been through this long journey together Kyungsu Kim, Wansuk Choi, Hojin Yang, Yunro Chung, Byeongyeob Choi, Sooyoung Lee, other my UNC friends Eunjung Jackie Kim, Heejoon Jo, Joowhi Lee, Noorie Hyun, Minjin Ha, Sunyoung Shin, Jonathan Hibbard, Pratyaydipta Rudra, Kuan-Chieh Huang, Tara Rao, Lu Mao, Andy Ni, Sunghee Park, Hyeyoung Hwang, Minjeong Jo, Jeehae Chung, Hyoyoung Choi, Taeyeon Eom, Chanhee Yi, Bosung Kim, Sunhyung Kim, and Hyunjee Lee, Jaewon Yoon, Soohyun Baek, Sungae Moon, Dongah Won, Sinae Lee, Wonkyung Seok, Seungjun Shin, Mihye Ahn, Jaewon Jang, Eunjoo Park, Hyungseok Tak. Last but not least, I am also grateful to Prof. Yoonseo Choi, Prof. Seong-Deog Yang, Prof. Keeyoung Kim.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

1.1 Survival Analysis

1.1.1 Single Failure Time Data

The Cox proportional hazards regression model [1] is the most commonly applied for analyzing censored survival data. The model quantifies the hazard rate $\lambda(t) = \lim_{h\downarrow 0} \frac{1}{h} P(T \leq t + h | T > t)$, where T is the survival time of an individual, with covariate vector Z such as

$$\lambda(t) = \lambda_0(t) \exp(\beta' Z), \tag{1.1.1}$$

where $t \geq 0$, $\lambda_0(t)$ is an unspecified baseline hazard function and β is a p-vector of unknown regression coefficients. Let $\lambda(t|z_1)$ and $\lambda(t|z_2)$ be the hazard functions given covariate vectors z_1 and z_2 respectively. Define the hazard ratio of z_1 with respect to z_2 as $r(t|z_1, z_2) = \frac{\lambda(t|z_1)}{\lambda(t|z_2)} = \exp(\beta' z_1 - \beta' z_2)$, which does not depend on time t. The hazard ratio is interpreted as the instantaneous failure at time t of a subject with covariate vector z_1 is $r(t|z_1, z_2)$ times as likely as a subject with covariate vector z_2. This ratio is of primary interest in survival analysis.

Let T_i be the failure (survival) time, C_i denote the potential censoring time, $X_i = \min(T_i, C_i)$ denote the observed time for the subject i, and $\delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Suppose that there are n independent subjects, i.e., $i = 1, \ldots, n$. The following partial likelihood function is considered to estimate β [1, 2].

$$PL(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp(\beta' Z_i)}{\sum_{k \in R_i} \exp(\beta' Z_k)} \right\}^{\delta_i}, \tag{1.1.2}$$

where the risk set $R_i = \{k : X_k \geq t_i\}$, i.e., all individuals who have not died or been censored yet by t_i. Then, β is estimated by $\hat{\beta}$ maximizing (1.1.2).

1.1.2 Recurrent Failure Time Data

Repeated failure events are frequently considered in a longitudinal study when subjects could possibly experience more than one event during the study period. The failures can be repeated by the same type of event or caused by different natures. In this thesis, Chapter 2 and 3 are related to the former case and Chapter 4 covers the latter case.

To explore the relationship between the effects of covariates and the correlated failure times, the Cox proportional hazards model [1] was extended to a multivariate counting process model allowing for time varying covariates with assuming independent increments [3]. To formulate in a counting process form, the data for the subject i, (X_i, δ_i) is re-expressed as $\{N_i(t), Y_i(t)\}$. The right continuous $N(t)$ is referred to as the counting process defined by $N_i(t) = N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the number of events that occur during the interval $[0, t]$ and C_i is the censoring time for subject i. The left continuous $Y(t)$ is referred to as the at-risk process given by $Y_i(t) = I(X_i \geq t)$, indicating the ith subject is under observation by the value t. Thus, the β also can be estimated by maximizing the logarithm of the following partial likelihood

$$C(\beta) = \sum_{i=1}^{n} \int_0^\tau \beta' Z_i(s) dN_i(s) - \int_0^\tau \log \left\{ \sum_{i=1}^{n} Y_i(s) \exp\{\beta' Z_i(s)\} \right\} d\bar{N}(s), \tag{1.1.3}$$

where τ is the end of study time and $\bar{N} = \sum_{i=1}^{n} N_i$. The estimator $\hat{\beta}$ is defined as the solution to the score equation given by

$$U(t; \beta) = \sum_{i=1}^{n} \int_0^t Z_i(s) dN_i(s) - \int_0^t \frac{\sum_{i=1}^{n} Y_i(s) Z_i(s) \exp\{\beta' Z_i(s)\}}{\sum_{i=1}^{n} Y_i(s) \exp\{\beta' Z_i(s)\}} d\bar{N}(s). \tag{1.1.4}$$

From the score function (1.1.4), the following form can be derived in terms of a local mar-

tingale:

$$U(t;\beta) = \sum_{i=1}^{n} \int_0^t Z_i(s) dM_i(s) - \int_0^t \frac{\sum_{i=1}^n Y_i(s) Z_i(s) \exp\{\beta' Z_i(s)\}}{\sum_{i=1}^n Y_i(s) \exp\{\beta' Z_i(s)\}} d\bar{M}(s), \tag{1.1.5}$$

where $\bar{M} = \sum_{i=1}^n M_i$, and $M_i(t) = N_i(t) - \int_0^t \lambda_i(s) ds$. There are additional definitions such as: $S^{(r)}(t;\beta) = n^{-1} \sum_{j=1}^n Z_j(t)^{\otimes r} Y_j(t) \exp\{\beta' Z_j(t)\}$, $r = 0, 1, 2$, $E(t;\beta) = \frac{S^{(1)}(t;\beta)}{S^{(0)}(t;\beta)}$, and $V(t;\beta) = \frac{S^{(2)}(t;\beta)}{S^{(0)}(t;\beta)} - E(t;\beta)^{\otimes 2}$. Also, their corresponding limiting values were defined as $s^{(r)}$ for $r = 0, 1, 2$, $e = s^{(1)}/s^{(0)}$, and $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$, respectively. Under some regularity conditions in [3], the estimated $\hat{\beta}$ has consistency and asymptotic normality with mean β_0 and covariance matrix Σ^{-1}, where $\Sigma = \int_0^\tau v(t,\beta) s^{(0)}(t,\beta) \lambda_0(t) dt$. The Σ is estimated by the partial likelihood observed information evaluating at $\beta = \hat{\beta}$. The estimator of the cumulative baseline intensity function $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is given by

$$\hat{\Lambda}_0(t;\hat{\beta}) = \int_0^t \frac{\sum_{i=1}^n dN_i}{\sum_{i=1}^n Y_i(s) \exp\{\hat{\beta}' Z_i(s)\}}. \tag{1.1.6}$$

Prentice *et al.* [4] proposed two arbitrary baseline intensity functions. One is a function of time from the beginning of the study and the other one is a function of time from the previous failure time. Let $Z(t) = \{Z_1(t), \ldots, Z_p(t)\}$ denote p covariate processes at time t, and $N(t)$ be a counting process, i.e., the random number of failures prior up to time t. Thus, the corresponding intensity functions are following:

$$\begin{aligned} \lambda_{CP}\{t|N(t)\} &= \lambda_{0m}(t) \exp\{\beta_m' Z(t)\}, \\ \lambda_{GT}\{t|N(t)\} &= \lambda_{0m}(t - t_{N(t)}) \exp\{\beta_m' Z(t)\}, \end{aligned} \tag{1.1.7}$$

where $\lambda_{0m}(\cdot) \geq 0$, $m = N(t) + 1$, i.e., $m = 1, 2, \ldots$ and β_m is a stratum-specific regression coefficients vector. The corresponding partial likelihoods can be written as

$$PL_{CP}(\beta) = \prod_{m \geq 1} \prod_{i=1}^{d_m} \frac{\exp\{\beta_m' Z_{mi}(t_{mi})\}}{\sum_{k \in R(t_{mi}, m)} \exp\{\beta_m' Z_{mk}(t_{mi})\}}, \tag{1.1.8}$$

where t_{mi} denotes a failure time in stratum m for subject i and d_m is a total number of failures in stratum m, and

$$PL_{GT}(\beta) = \prod_{m \geq 1} \prod_{i=1}^{d_m} \frac{\exp\{\beta_m' Z_{mi}(t_{mi})\}}{\sum_{k \in R(u_{mi}, m)} \exp\{\beta_m' Z_{mk}(l_k + u_{mi})\}}, \tag{1.1.9}$$

where l_k is the last failure time of subject k prior to the entry into stratum m and u_{mi} implies that subject i fails in stratum m with a gap time u_{mi} at time t_{mi}. Kelly *et al.* [5] suggested to use the gap time model to analyze recurrent event data when within-subject events are independent because it determines whether the treatment is effective for the mth event since the time from the previous event.

Wei *et al.* [6] suggested marginal distributions for the multivariate failure times. Each of the distributions is also based on the Cox proportional hazards model [1]. There is no specific constraint for dependence among the different failure times within each subject. Similar to Prentice *et al.* [4] each event or event type is modeled as a separate stratum. Within each strata, the marginal data is used, that is, all information is ignored except the given event type. As a result, each patient normally appears in all of the strata, barring deletion due to missing values [7]. For the mth type of failure, $m = 1, \ldots, M$, of the ith subject, the hazard function is given by

$$\lambda_m(t) = \lambda_{0m}(t) \exp\{\beta_m' Z_m(t)\}, \tag{1.1.10}$$

where β_m is the failure-specific regression coefficients vector. The corresponding mth failure-specific partial likelihood is

$$PL_m(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta_m' Z_{mi}(t_{mi})\}}{\sum_{k \in R(t_{mi})} \exp\{\beta_m' Z_{mk}(t_{mi})\}} \right]^{\delta_{mi}}. \tag{1.1.11}$$

Similarly, β_m is estimated by solving the equation $\partial \log PL_m(\beta)/\partial\beta = 0$. The estimated $(\hat{\beta}_1', \ldots, \hat{\beta}_M')'$ is approximately normal with mean $(\beta_1', \ldots, \beta_M')$ and covariance matrix $\Sigma_{M\times M}$ with $\sigma_{kl}(\beta_k, \beta_l) = A_k^{-1}(\beta_k) E[w_{k1}(\beta_k) w_{l1}(\beta_l)'] A_l^{-1}(\beta_l)$ for $k, l = 1, \ldots, M$, where

$$A_k(\beta_k) = E\left[\int_0^\tau \{Z_k(t) - e_k(t; \beta_k)\}^{\otimes 2} Y_k(t) \exp\{\beta_k' Z_k(t)\} d\Lambda_{0k}(t) \right],$$

$$w_{ki}(\beta_k) = \int_0^\tau \{Z_{ki}(t) - e_k(t; \beta_k)\} dM_{ki}(t),$$

$$M_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(s) \lambda_{ki}(s) ds,$$

$$S_k^{(r)}(t;\beta) = n^{-1}\sum_{i=1}^{n} Z_{ki}^{\otimes r} Y_{ki}(t)\exp\{\beta' Z_{ki}(t)\},\ \text{for } r = 0, 1, 2,$$

$E_k(t;\beta) = S_k^{(1)}(t;\beta)/S_k^{(0)}(t;\beta)$, and $e_k(t;\beta)$ be the corresponding limit. This marginal method is a useful tool for making inferences on the population average effect of covariates on failure times. However, it cannot provide insights into the interrelationship among failure times [7, 8]. Also, Kelly *et al.* [5] suggested this method is more appropriate to data where there are different types of events from a same person than to recurrent event data. These two aforementioned approaches by Prentice *et al.* [4] and by Wei *et al.* [6] use stratified proportional hazards model with a separate stratum depending on the number of previous events. Thus, those methods could be proper for only handling a small number of recurrent events. Since there is a case with some strata that does not have enough subjects or a case where there are too many number of parameters to be estimated, those models might not get stable hazard estimates.

The existing Cox type hazards functions for analyzing recurrent event data have been reviewed. These methods are assumed that the underlying counting process is a time variant Poisson process. Lin *et al.* [9] suggested a robust procedure for treating all recurrent events within a subject as a single counting process without assumption related to the poisson process. By denoting $E[dN(t)|Z(t)] = d\mu(t)$,

$$d\mu(t) = \exp\{\beta_0' Z(t)\} d\mu_0(t),\ \text{or} \tag{1.1.12}$$

$$\mu(t) = \int_0^t \exp\{\beta_0' Z(s)\} d\mu_0(s), \tag{1.1.13}$$

where $\mu_0(\cdot)$ is an unknown function. Model (1.1.12) is referred to as the proportional rates (means) model. Model (1.1.1) implies model (1.1.12) with $d\mu_0(t) = \lambda_0(t)$, but not in reverse. The corresponding inferences are similar to arguments in [3]. [9] imposed the following regularity conditions, for $i = 1, \ldots, n$:

(i) $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ are independent and identically distributed.

(ii) $Pr(Y_i(\tau) = 1) > 0$, where τ is a predetermined study end time.

(iii) $N_i(\tau)$ are bounded.

(iv) $Z_i(\cdot)$ are bounded total variations, i.e., $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| \leq K$ for all $j = 1, \ldots, p$, where Z_{ij} is the jth element of Z_i and K is a constant.

(v) $A \equiv E[\int_0^\tau \{Z(t) - E(\beta_0, t)\}^{\otimes 2} Y(t) \exp\{\beta_0' Z(t)\} d\mu_0(t)]$ is positive definite.

Then, $n^{-1/2}U(t;\beta_0)$, $0 \leq t \leq \tau$ converges weakly to a continuous zero mean Gaussian process with covariance matrix

$$\Sigma(s,t) = E\left[\int_0^s \{Z_1(u) - E(u;\beta_0)\} dM_1(u) \int_0^t \{Z_1(v) - E(v;\beta_0)\} dM_1(v)\right],$$

$0 \leq s, t \leq \tau$ between time points s and t. Also, they proved that $\hat{\beta}$ has asymptotic normality with mean β_0 and covariance matrix $\Gamma = A^{-1}\Sigma A^{-1}$, where $\Sigma = \Sigma(\tau, \tau)$. The covariance matrix Γ can be consistently estimated with these subsequent quantities:

$$\hat{\mu}_0(t) = \int_0^t \frac{d\bar{N}(s)}{nS^0(s;\hat{\beta})},$$

$$\hat{A} = -n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(s) - E(s;\hat{\beta})\}^{\otimes 2} Y_i(s) \exp\{\hat{\beta}' Z_i(s)\} d\hat{\mu}_0(s),$$

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(u) - E(u;\hat{\beta})\} d\hat{M}_i(u) \int_0^\tau \{Z_i(v) - E(v;\hat{\beta})\} d\hat{M}_i(v),$$

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp\{\hat{\beta}' Z_i(s)\} d\hat{\mu}_0(s).$$

Additionally, a random weight function $Q(t;\hat{\beta})$ is incorporated into $U(\tau,\beta)$, then the following weighted estimating functions for β_0 are obtained:

$$U_Q(\tau;\beta) = \sum_{i=1}^n \int_0^\tau Q(s;\hat{\beta})\{Z_i(s) - E(s;\beta)\} dN_i(s). \tag{1.1.14}$$

They assumed that the weight function is non-negative, bounded and monotone in t, and converges almost surely to a continuous deterministic function $q(t)$ in $t \in [0,\tau]$. By solving $U_Q(\tau;\beta) = 0$, β_0 can be estimated by $\hat{\beta}_Q$. Also, they showed that $n^{1/2}(\hat{\beta}_Q - \beta_0)$ is

asymptotically zero mean normal with covariance matrix $A_Q^{-1}\Sigma_Q A_Q^{-1}$, where

$$A_Q = E\left[\int_0^\tau q(t)\{Z_1(t) - E(t;\beta_0)\}^{\otimes 2} Y_1(t)\exp\{\beta_0' Z_1(t)\}d\mu_0(t)\right],$$

$$\Sigma_Q = E\left[\int_0^\tau q(u)\{Z_i(u) - E(u;\beta_0)\}dM_1(u)\int_0^\tau q(v)\{Z_i(v) - E(v;\beta_0)\}dM_1(v)\right].$$

Note that the weight function $Q(t;\hat{\beta})$ does not relate to β.

1.2 Instrumental Variable Methods

Instrumental variable (IV) analysis is commonly applied to estimating the exposure effect for the data with unmeasured confounders. For example, in a randomized trial, subjects often do not comply with their assigned treatment protocols. Hence, a subject's actual treatment may differ from his or her assigned treatment, i.e., compliance might not be random. To estimate the true causal effect of an exposure on an outcome, we may use IV. The valid IV should have the following properties such as i) affect treatment (endogenous covariate) or be associated with treatment by sharing a common cause; ii) be a factor that is as good as randomly assigned and iii) related to the outcome only through the treatment [10]. Even though recurrent failure event time data are frequently obtained in many randomized clinical trials or observational studies, there is only a handful of studies that analyze it by dealing with non-compliance problem. Thus, existing IV methods mostly for time-to-event data, i.e., univariate failure time data, are reviewed.

An intent-to-treat analysis is more commonly used without considering the presence of non-compliance. For example, if Z is denoted as a randomization indicator, i.e., $Z = 1$ for individuals randomized to receive an active treatment and $Z = 0$ otherwise, then we can model the intent-to-treat treatment effect as $\lambda(t|Z) = \lambda_0(t)\exp(\beta_0 Z)$. The parameter β_0 is not to measure the true causal effect of treatment, but rather a mixture of the effect on the compliers with the absence of effect on the non-compliers due to their non-compliance.

Thus, the measure of treatment effect under intent-to-treat analysis would diminish as non-compliance increased.

Several IV methods for right-censored event time outcome data have been proposed. The following models could plausibly be used: the rank preserving structural failure time model (RPSFTM) [11], the marginal structural Cox proportional hazards model [12–14] and the method of inverse probability of censoring weighted log-rank tests [15].

The causal parameter of interest will often be a function of the survival differences that would have been observed, contrary to fact that all subjects remained on protocol. Robins and Tsiatis [11] proposed the RPSFTM using a class of rank estimators to estimate the survival differences with correcting non-compliance issue. This model is a strong version of the accelerated failure time (AFT) model with time-dependent covariates proposed by [16]. The ranking preserving implies that given any two subjects i and j, if subject i fails before subject j when both followed a particular treatment regime, then subject i would fail before subject j when both followed any other regime. Let U_i denote the survival time of the subject i when he or she is never to receive treatment, i.e., $D_i(t) = 0$ for all t. In the absence of censoring, the observable random variables are $\{T_i, H_i(T_i), Zi\}$ are independently and identically distributed, where T_i is the observed survival time of ith subject, $H_i(T_i) = (D_i(s); 0 < s \leq T_i)$ is the observed treatment history, and Z_i is a randomization group indicator. Simply, the RPSFTM assumes that the baseline latent lifetime variable, U_i, is related to these observable random variables by the relationship

$$U_i = \int_0^{T_i} \exp\{\beta_0 D_i(s)\} ds, \tag{1.2.1}$$

where β_0 is an unknown parameter. More generally, the RPSFTM links the variable U_i to $\{T_i, H_i(T_i)\}$ by assuming $U_i = \psi(T_i, H_i(T_i), \beta_0)$, where $\beta_0 \in \mathbb{R}^p$ is an unknown parameter and $\psi(\cdot)$ is a known smooth function. The following properties are satisfied by:

(i) **smoothness**: $\psi_1(t, h(t), \beta)$, $\psi_{3,p}(t, h(t), \beta)$, and $\psi_{13,p}(t, h(t), \beta)$ are continuous for all β

and almost all t, where $\psi_1(t, h(t), \beta) \equiv \partial\psi(t, h(t), \beta)/\partial t$, $\psi_{3,p}(t, h(t), \beta) \equiv \partial\psi(t, h(t), \beta)/\partial\beta_p$, $\psi_{13,p}(t, h(t), \beta) \equiv \partial\psi_1(t, h(t), \beta)/\partial\beta_p$.

(ii) **Monotonicity**: $\psi(t, h(t), \beta) > \psi(s, h(s), \beta)$ if $t > s$.

(iii) **Identity**: $\psi(t, h(t), 0) = t$

(iv) **Independence and Identification**: There exists a unique β_0 such that $U(\beta_0) \perp\!\!\!\perp Z$, where $U(\beta) = \psi(T, H(T), \beta)$.

Thus, β_0 can be considered as the true value of β with a casual interpretation. Standard nonparametric methods can be applied to test the null hypothesis $\beta_0 = 0$.

In the marginal structural Cox proportional hazards model [12–14], the Cox proportional hazards model [1, 16], which is a standard semi-parametric method, was adopted rather than the AFT model. The parameter of interest in the marginal structural Cox proportional hazard model is the counterfactual hazard ratio rather than survival time itself. Loeys and Goetghebeur [14] proposed the Complier PROPortional Hazards Effect of Treatment (C-PROPHET). Suppose that n independent individuals were randomized over experimental exposure ($Z_i = 1$) or control ($Z_i = 0$). Either (D_{i1}, T_{i1}) or (D_{i0}, T_{i0}) is observed, where $D_{i1} = 1$ denotes that subject i received the treatment and T_{i1} is the corresponding right censored survival time when $Z_i = 1$. The following assumptions are required:

(i) $(D_{i1}, T_{i1}, D_{i0}, T_{i0}, Z_i)$ are independent and identically distributed random variables, implying that potential outcomes for each subject are unrelated to the treatment or outcome experienced by other individuals.

(ii) **randomization**: $(D_{i0}, T_{i0}, D_{i1}, T_{i1}) \perp\!\!\!\perp Z_i$.

(iii) No access to treatment on the control arm. That is, $D_{i0} = 0$ for all subjects, and T_{i0} represents the treatment-free survival time outcome, when randomized to control.

(iv) **exclusion restriction**: $Pr(T_{i1} > t|D_{i1} = 0) = Pr(T_{i0} > t|D_{i1} = 0)$.

Under the randomized treatment assignment Z_i, the treatment actually received can be written as $D_i = (1 - Z_i)D_{i0} + Z_i D_{i1}$. The observation time is defined as $X_i = \min(T_i, C_i)$, with $T_i = (1 - Z_i)T_{i0} + Z_i T_{i1}$, the observed survival time, possibly censored at time C_i. The corresponding censoring indicator is denoted by $\delta_i = I(T_i \leq C_i)$. Additionally, they assumed

(v) **independent censoring**: $(T_i, D_i, Z_i) \perp\!\!\!\perp C_i$ or weaker assumption that censoring is non-informative for the control arm as a whole, while in the experimental arm, censoring is a non-informative conditional on treatment exposure.

To estimate C-PROPHET, $Pr(T_{i0} > t|D_{i1} = 1) = Pr(T_i > t|Z_i = 0, D_{i1} = 1)$ and $Pr(T_{i1} > t|D_{i1} = 1) = Pr(T_i > t|Z_i = 1, D_{i1} = 1)$ are compared by using a proportional hazards model. That is, C-PROPHET implies that the proportional hazards of treatment in the subpopulation that has received the treatment. Within this subgroup $D_{1i} = 1$, the two hazard rate functions have a relationship as follows:

$$\lambda(t|Z_i = 1, D_{i1} = 1) = \lambda(t|Z_i = 0, D_{i1} = 1) \exp(\psi_0), \tag{1.2.2}$$

where $\exp(\psi_0)$ denotes the causal proportional hazards effect within the treatable subpopulation. However, the estimator proposed by Loeys and Goetghebeur [14] is limited to the setting of all-or-nothing compliance exposure. To overcome this limitation, Loeys *et al.* [17] extended it to more general causal proportional hazards models that allow for time-constant discrete and continuous exposure levels. Suppose that there are n independent subjects are randomly assigned to treatment ($Z_i = 1$) or control ($Z_i = 0$). Let X_i denote a covariate vector measured prior to randomization. Subjects randomized to treatment arm may receive control therapy, but subjects on the control arm have no access to treatment and thus adhere to their assigned treatment. Let U_i denote that subject i's potential exposure to the treatment if the subject were randomized to treatment and $D_i = Z_i U_i$ indicated the observed exposure for all subjects. That is, U_i cannot be observed in control arm but $D_i = 0$. Under non-informative censoring, $X_i = \min(T_i, C_i)$, and corresponding censoring indicator

$\delta_i = I(T_i \leq C_i)$ are defined. To contrast the observed hazard in the treated group versus the unobserved subpopulation-specific hazards in the control group, the following causal proportional hazards model is considered:

$$\lambda(t|Z_i = 1, U_i = u, X_i = x_i) = \lambda(t|Z_i = 0, U_i = u, X_i = x_i) \exp(\psi_0 u), \tag{1.2.3}$$

where ψ_0 is the unknown parameter of interest. The fundamental problem of estimating ψ_0 is that all subjects in the control group ($Z_i = 0$) have latent U_i. Randomization is the key for the estimation procedure. The Eq.(1.2.3) is re-written in terms of survival distributions such as

$$S(t|Z_i = 1, U_i = u, X_i = x_i) = S(t|Z_i = 0, U_i = u, X_i = x_i)^{\exp(\psi_0 u)}. \tag{1.2.4}$$

The survival probability in the control group are defined as a mixture of unobserved compliance-specific probabilities given by

$$S(t|Z_i = 0, X_i = x_i) = \sum_u S(t|Z_i = 0, U_i = u, X_i = x_i) Pr(U_i = u|Z_i = 0, X_i = x_i),$$

where U_i is discrete. Also, if Eq.(1.2.4) holds, then $S(t|Z_i = 0, X_i = x_i)$ can be expressed as $\sum_u S(t|Z_i = 1, U_i = u, X_i = x_i)^{\exp(-\psi_0 u)} Pr(U_i = u|Z_i = 1, X_i = x_i)$ since $Pr(U_i = u|Z_i = 0, X_i = x_i) = Pr(U_i = u|Z_i = 1, X_i = x_i)$ by definition of U_i and randomization. They defined $\hat{S}_{1\to 0}(t|X_i; \psi)$ as

$$\sum_u \hat{S}(t|Z_i = 1, U_i = u, X_i = x_i)^{\exp(-\psi u)} \hat{Pr}(U_i = u|Z_i = 1, X_i = x_i).$$

The unknown parameter ψ_0 is estimated by the value of ψ that minimizes the distance between the $\hat{S}_{1\to 0}(t|X_i; \psi)$ and the fitted treatment-free survival distribution in the control arm conditional on X_i.

Robins and Finkelstein [15] proposed the inverse probability of censoring weighted (IPCW) log-rank test for estimating the effect of treatment with non-compliance and a informative censoring scheme. They assumed that there is no unmeasured confounders of censoring, that is, the cause-specific intensity rate of censoring time $\lambda_C(t)$, where $C_i = t$, does not depend on

the possibly unobserved failure time T_i conditional on the assigned treatment Z_i and recorded history $H_i(t)$ of all associated time variant covariates $X_i(t)$, where $H_i(t) = \{H_i(s); 0 \leq s \leq t\}$. This assumption can be expressed as $\lambda_C(t|H(t), Z, T, T > t) = \lambda_C(t|H(t), Z, T > t)$ which is different from the usual independent censoring assumption $\lambda_T(t|Z, C \geq t) = \lambda_T(t|Z)$, where $\lambda_T(t|Z, C \geq t)$ denotes the cause-specific hazard of failure time $T_i = t$. Under the assumption, the following model is considered

$$\lambda_C(t|H(t), Z, T > t) = \lambda_{0Z}(t) \exp\{\beta'_Z H(t)\}. \tag{1.2.5}$$

Since both the baseline hazards $\lambda_{0Z}(t)$ and β_Z may depend on the treatment arm, this model (1.2.5) can be separated by treatment-arm-specific models for censoring and then they are fit to data from the two arms separately. Let $X = \min(T, C)$, $Y(t) = I(X \geq t)$ be the at-risk indicator, and let $\delta = I(T \leq C)$ be the censoring indicator. A consistent estimate of the conditional probability that a subject i is uncensored through time t given $(H(T), Z, T)$ is provided by the following time-dependent extension of the Kaplan-Meier (K-M) product limit estimator for censoring:

$$\hat{K}_i^H(t) = \prod_{\{j:X_j<t, \delta_j=0, Z_j=Z_i\}} [1 - \hat{\lambda}_{Z_i}(X_j) \exp\{\beta_{Z_i} H_i(X_j)\}], \tag{1.2.6}$$

where

$$\hat{\lambda}_Z(X_j) = \frac{1 - \delta_j}{\sum_{i=1}^n Y_i(X_j) \exp\{\beta_Z H_i(X_j)\} I(Z_i = Z)}$$

is the baseline hazard estimator for censoring λ_{0Z} in arm Z. Let $\hat{K}_i^0(t)$ be the usual treatment-arm-specific Kaplan-Meier estimator of the probability of being uncensored by time t in treatment arm Z_i, then the subject specific weight is defined $\hat{W}_i(t) = \hat{K}_i^0(t)/\hat{K}_i^H(t)$ so that $\hat{W}_i(t)$ will be converged to one for all t if and only if there is no dependent censoring within each treatment arm. Thus, IPCW K-M estimate of the treatment arm specific marginal probability of remaining alive through time t is

$$\hat{S}_T(t|z) = \prod_{\{i:X_i<t\}} \frac{1 - \delta_i \hat{W}_i(X_i) I(Z_i = z)}{\sum_{k=1}^n Y_k(X_i) \hat{W}_k(X_i) I(Z_k = z)}. \tag{1.2.7}$$

Yu *et al.* [18] extended semiparametric linear transformation models that include the proportional hazards model that was considered by Cuzick *et al.* [19] to estimate the distribution of survival times in the treatment and control groups, conditionally not only on observed covariates, but on the latent compliance type. Under the two-arm randomized trials with all-or-nothing compliance and time-to-event outcomes, let X_i be a p-dimensional covariate vector, and let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ denote the n-vector of treatment assignments with $0 < Pr(Z_i = 1) < 1$, for $i = 1, \ldots, n$. For all possible assignment vectors $\mathbf{z} = (z_1, \ldots, z_n) \in \{0, 1\}^n$, let $D_{i\mathbf{z}} = 1$ denote that a subject i received the treatment under the assignment $\mathbf{z}$ and let $T_{i\mathbf{z}}$ and $C_{i\mathbf{z}}$ similarly denote the potential event time and potential right censoring time for subject i under assignment $\mathbf{z}$, respectively. Let $U_i = k$ denote subject i's latent class membership: U_i equals 1 if i is an always-taker ($D_{i0} = D_{i1} = 1$), 2 if i is a complier ($D_{i0} = 0$; $D_{i1} = 1$), 3 if i is a never-taker ($D_{i0} = D_{i1} = 0$) and 4 if i is a defier ($D_{i0} = 1$; $D_{i1} = 0$). The following assumptions are required:

(i) **stable unit treatment value assumption**: For any assignments $\mathbf{z}$ and $\mathbf{z}'$, if $z_i = z'_i$, then $D_{i\mathbf{z}} = D_{i\mathbf{z}'}$, $T_{i\mathbf{z}} = T_{i\mathbf{z}'}$ and $C_{i\mathbf{z}} = C_{i\mathbf{z}'}$ for $i = 1, \ldots, n$. Thus, $D_{i\mathbf{z}} \equiv D_{iz_i} \equiv D_{iz}$, where $z = 0$ if $z_i = 0$, and $z = 1$ otherwise; similarly, $T_{i\mathbf{z}} \equiv T_{iz_i} \equiv T_{iz}$ and $C_{i\mathbf{z}} \equiv C_{iz_i} \equiv C_{iz}$.

(ii) **random sampling**: $(D_{iz}, T_{iz}, C_{iz}, X_i, Z_i)$, $i = 1, \ldots, n$, are independent and identically distributed from the distribution of a random vector (D_z, T_z, C_z, X, Z).

(iii) **random assignment conditional on covariates**: $D_0, T_0, C_0, D_1, T_1, C_1 \perp\!\!\!\perp Z|X$.

(iv) **conditional non-null compliance class**: $Pr(U = 2|X = x) > 0$.

(v) **conditional monotonicity**: $Pr(D_1 \geq D_0|X = x) = 1$.

(vi) **exclusion restriction**: For $k = 1, 3$ and for all t, $Pr(T_0 \leq t|U = k, X = x) = Pr(T_1 \leq t|U = k, X = x)$.

They considered three estimands such as the (conditional) complier average causal effect

(CACE), the time t effect on the (conditional) complier survival probability (CESP), and the (conditional) complier quantile causal effect (CQCE):

$$CACE(x) = E[T_1 - T_0|U = 2, X = x],$$

$$CESP(t; x) = Pr(T_1 > t|U = 2, X = x) - Pr(T_0 > t|U = 2, X = x),$$

$$CQCE(q; x) = \sup\{t : Pr(T_1 \leq t|U = 2, X = x) \leq q\} - \sup\{t : Pr(T_0 \leq t|U = 2, X = x) \leq q\}.$$

The causal linear transformation model for the potential event time distributions cannot be directly used for estimation by using the observed data. Since only $T_i = Z_i T_{i1} + (1Z_i)T_{i0}$ is observed. Thus, they proposed the following model with only using the parameters that are related to the T_i:

$$\log\{H(T_i)\} = -\sum_{k=1}^{3}(\beta_{0k} + \beta_k Z_i + \eta_k' X_i)I(U_i = k) + \epsilon_i, \tag{1.2.8}$$

where $H(\cdot)$ is an unspecified continuously differentiable increasing function with $H(0) = 0$, the random errors ϵ_i are independent, identically distributed, and independent of U_i, X_i, Z_i. Additionally they assumed

(vii) **(independent censoring of potential outcomes)** For $z = 0, 1$, $C_z \perp\!\!\!\perp \epsilon_z$, $C_z \perp\!\!\!\perp U|X$.

Thus, the observed data are $(Y_i, \Delta_i, X_i, D_i, Z_i)$ with realized values $(y_i, \delta_i, x_i, d_i, z_i)$ for $i = 1, \ldots, n$. The log likelihood of the observed data is a mixture of distributions depending on the compliance type probabilities $p_{ik} = Pr(U_i = k|X_i = x_i)$ since compliance class cannot be fully observed. The p_{ik} can be calculated with the multinomial logit model with $\log\{p_{ik}(\theta)/p_{i2}(\theta)\} = \theta_k' x_i$, for $k = 1, 3$, and $\theta = (\theta_1', \theta_3')'$. The corresponding likelihood function of $(y_i, \delta_i, x_i, d_i, z_i)$ for $i = 1, \ldots, n$ is given by

$$\prod_{i=1}^{n}\{(p_{i1}g_{i1} + p_{i2}g_{i2})I(d_i = z_i = 1) + p_{i3}g_{i3}I(d_i = 0, z_i = 1) + p_{i1}g_{i1}I(d_i = 1, z_i = 0)$$
$$+ (p_{i2}g_{i2} + p_{i3}g_{i3})I(di = zi = 0)\}\lambda_{C|X,Z}^{1-\delta_i}(y_i|x_i, z_i)S_{C|X,Z}(y_i|x_i, z_i)f_{X,Z}(x_i, z_i), \tag{1.2.9}$$

where

$$g_{ik} = \left\{ \lambda_\epsilon[\log H(y_i) + \beta_{0k} + \beta_{k_{zi}} + \eta_k' x_i] \frac{h(y_i)}{H(y_i)} \right\}^{\delta_i} \exp\left\{ \Lambda_\epsilon[\log H(y_i) + \beta_{0k} + \beta_{kz_i} + \eta_k' x_i] \right\},$$

$\lambda_{C|X,Z}$ and $S_{C|X,Z}$ are the conditional hazard and survivor functions of C given X, Z and $f_{X,Z}$ is the density of X, Z. Since the distributions of the censoring time and covariates do not depend on the parameters of interest, the likelihood (1.2.9) can be simplified as

$$\prod_{d_i=z_i=1} (p_{i1}g_{i1} + p_{i2}g_{i2}) \prod_{d_i=0,z_i=1} p_{i3}g_{i3} \prod_{d_i=1,z_i=0} p_{i1}g_{i1} \prod_{d_i=z_i=0} (p_{i2}g_{i2} + p_{i3}g_{i3}). \tag{1.2.10}$$

The maximum likelihood estimator (MLE) can be obtained by maximizing the likelihood (1.2.10) with respect to $(\theta, \beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3, h1, \ldots, h_{n_u})$ subject to the constraints in which $h_1, \ldots, h_{n_u}$ are non-negative. They also showed that the MLE is consistent and has asymptotic normality under the regularity conditions.

1.3 Competing Risks Analysis

So far, we have considered that there is only one event type of interest. However, in many contexts it is likely that we can have several different types of failure. The occurrence of one type of failure may prevent us from observing the other types of failures. The causes of failure compete to occur or to be observed so it is referred to as *competing* risks. We restrict ourselves to competing risk events where the follow-up duration of a patient ends at the onset of the first event, and do not focus on multiple or recurrent events occurring in a patient. For example, in cardiovascular studies, deaths from other causes are considered as competing risks. There are two different ways to deal with competing risks setting such as a latent failure time formulation and an approach of bivariate random variables with the event time and the type of event. The former has the identifiability problem so the latter is commonly used in modern competing risks analyses.

Specifically, one possible access for characterizing competing risks data is the latent failure time. Let $\tilde{T}_j$ is the latent time variable due to cause j for $j = 1, \ldots, n_J$, i.e, n_J is the number

of failure types. We observe $T = \min(\tilde{T}_1, \ldots, \tilde{T}_J)$ in the absence of censoring. Functions corresponding to the latent failure times are called marginal. The marginal hazard function is defined as

$$\tilde{\lambda}_j(t) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leq \tilde{T}_j < t + h | \tilde{T}_j > t). \tag{1.3.1}$$

Also, the marginal survival function is $\tilde{S}_j(t) = P(\tilde{T}_j > t) = \exp\{-\tilde{\Lambda}_j(t)\}$, where $\tilde{\Lambda}_j(t) = \int_0^t \tilde{\lambda}(s)ds$. Note that the marginal functions for cause j do not consider other causes which may not be practically relevant. The marginal functions are only estimable when $\tilde{T}_1 \perp\!\!\!\perp \tilde{T}_2 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \tilde{T}_J$. However, this assumption is not verifiable [20].

In addition to the marginal functions, we can consider a bivariate random variable (T, J), where T is a random variable for the event time and J is a random variable for the event type. Thus, it leads to crude functions such as cause-specific hazard functions [21] and cumulative incidence functions (subdistribution functions). The corresponding analysis can be performed without identifiability problems and all measures can be estimated from observable data. The cause-specific hazard for failure type j is given by

$$\lambda_j^*(t) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leq T < t + h, J = j | T > t) \tag{1.3.2}$$

for $j = 1, \ldots, J$, and is interpreted as cause j failure hazard at time t, among subjects alive at time t with acknowledging the existence of other causes by treating failures from other causes as censored. Even though the cause-specific hazard function has a clear interpretation, the corresponding function $S_j^*(t) = \exp\{-\Lambda_j^*(t)\}$, where $\Lambda_j^*(t) = \int_0^t \lambda_j^*(s)ds$, does not. It implies that when cause j is the only cause of failure then $S_j^*(t)$ would be equal to the survival function. Another commonly used crude function is the cumulative incidence function (CIF) defined as

$$F_j(t) = P(T \leq t, J = j). \tag{1.3.3}$$

It means the probability of observing the event of interest from cause j, acknowledging that the subject may experience the event due to other causes first. It is not the same quantity with $P(\tilde{T}_j < t)$, because $\tilde{T}_j$ cannot be observed when $\tilde{T}_k < \tilde{T}_j$, $k \neq j$. The name

subdistribution function is driven by the fact that $F_j(t)$ does not converge to one as t is going to infinity. It represents the overall probability of the type j event such that

$$\lim_{t\to\infty} F_j(t) = P(J = j),$$

so it is not a proper probability distribution. Corresponding subsurvival function is $Q_j(t) = P(T > t, J = j)$ and $Q_j(t) + F_j(t) = P(J = j)$ by the law of total probability. These CIF and subsurvival function are estimable without assuming independence between the causes of the failure. The CIF can be expressed in terms of the cause-specific hazards:

$$\begin{aligned} F_j(t) &= \int_0^t S(u)\lambda_j^*(u)du = \int_0^t \exp\{-\Lambda(u)\}\lambda_j^*(u)du \\ &= \int_0^t \exp\left\{-\sum_{j=1}^{J}\int_0^u \lambda_j^*(v)dv\right\}\lambda_j^*(u)du, \end{aligned} \tag{1.3.4}$$

where $S(t)$ is the overall survival function, $S(t) = 1 - \sum_{j=1}^{J} F_j(t)$. This is often used for the estimation of $F_j(t)$. We can also define a cause-specific density at time t, say

$$f_j(t) = \lim_{h\downarrow 0}\frac{1}{h}P(t \le T_i < t + h, J = j) = \lambda_j^*(t)S(t). \tag{1.3.5}$$

By the law of total probability, we have $\lambda(t) = \sum_{j=1}^{J}\lambda_j^*(t)$, because failure must be due to one of the J causes, and similarly $F(t) = \sum_{j=1}^{J} F_j(t)$ and $f(t) = \sum_{j=1}^{J} f_j(t)$.

Various nonparametric and semiparametric methods have been developed for modeling the cumulative incidence function $F_j(t)$. A subdistribution hazard function is considered directly from the cumulative incidence function for cause j by Gray [22] in order to compare the cumulative incidence of a particular type of failure amongst K different groups. The corresponding function is defined by

$$\begin{aligned} \lambda_j(t) &= \frac{dF_j(t)}{\{1 - F_j(t)\}} \\ &= \lim_{h\downarrow 0}\frac{1}{h}P\{t \le T_i < t + h, J = j | T \ge t \text{ or } (T \le t \text{ and } J \ne j)\}. \end{aligned} \tag{1.3.6}$$

The conditional expression includes two different scenarios: 1) the event has not occurred at time t, and 2) the event has occurred from a different cause before t. The difference between

λ_j and $\lambda*_j$ is due to the risk set. The former includes subjects who have failed from other causes, but the latter excludes subjects who have failed from other causes up to time t.

Fine and Gray [23] suggested semiparametric proportional sub-distribution by adopting the Cox proportional hazards model, which is following:

$$\lambda_j(t|Z) = \lambda_{j0} \exp(\beta_j' Z), \tag{1.3.7}$$

where λ_{j0} is an unspecified, nonnegative baseline subdistribution hazard and β_j is a p vector of unknown regression parameters. Also, the covariates are linear on a complementary log-log transformed cumulative incidence function. They showed that the partial likelihood approach is still valid for the estimation in complete data and censoring complete data and proposed the weighted score function for the right censored data. Fine [24] generalized the Fine and Gray model [23] by using a transformation of the cumulative incidence function to have flexibility. Given the assumption that $g(.)$ is a known and differentiable function, the following model is considered:

$$g\{F_j(t|Z)\} = h_j(t) - \beta_j' Z, \tag{1.3.8}$$

where $h_j(t)$ is the baseline failure probability, unspecified, invertible and strictly increasing in t. By choosing $g(x) = \log\{-\log(1-x)\}$, it gives the proportional subdistribution model. However, the proposed estimation method is less efficient under this model. Also, the proportional odds model with $g(x) = logit(x)$ is specified.

CHAPTER 2: SELF-TRIGGERING COX MODEL FOR RECURRENT EVENT DATA

2.1 Introduction

Recurrent event data frequently arise when subjects may experience more than one event during the observation period. For example, cystic fibrosis patients have have repeated *Pseudomonas aeruginosa* infections, which are prognostic for progressive lung disease and are highly associated with the mortality and morbidity in early life. The most common approaches to modeling recurrent event data are the Andersen and Gill (AG) model [3], the Prentice, Williams and Peterson total time (PWP-CP) and gap time (PWP-GT) models [4], and the Wei, Lin and Weissfeld model (WLW) model [6]. Other models, such as Lee, Wei and Amato model [25] based on the marginal Cox model or frailty models [26] may also be considered. While these model specifications target different endpoints and may have different interpretations, the methods have been compared in simulation experiments and real data analyses in terms of model fit [5, 7, 8, 27, 28]. The application of such methods generally ignores previous event history when modeling the risk of future events, with the focus on the effects of baseline covariates, like treatment, on the recurrent event trajectory. The incorporation of time-dependent covariates capturing this history may improve a model fit and prediction, but may also obscure the effects of time-independent covariates. Additionally, it is difficult to deal with a large number of events when either of the PWP or WLW approaches is applied.

There is limited literature discussing the potential use of event history for modeling the occurrence of recurrent events. Recently, Chen and Chen [29] proposed an m-memory Cox-

type self-triggering intensity model to account for the correlation among the occurrence of events by utilizing a time-dependent decay function that describes the effects of previous events. However, this work neither provides a formal statistical procedure to test if such self-triggering effect exists, nor demonstrates the improvements that such effects may have on model fit and the prediction of future event occurrences. This Chapter addresses these issues.

The remaining sections are organized as follows. In Section 2.2, we briefly review the Cox-type model with self-triggering scheme and discuss the use of partial likelihood for parameter estimation and inference. In addition, we propose a hypothesis testing procedure for the existence of the triggering scheme. Evaluation of model prediction is also discussed. In Section 2.3, we will report simulation studies conducted to examine the feasibility of the partial likelihood estimation and hypothesis testing procedures. The cystic fibrosis registry data is analyzed in Section 2.4 with comparison between different modeling approaches in terms of prediction accuracy.

2.2 Inference

2.2.1 Notations, Model, and Estimation

Let T_{ij} denote the event occurrence time for the jth event of subject i, and let t_{ij} be the observed realization of T_{ij}. Letting C_i denote the censoring time, one can define $N_i(t) = N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the number of events that are observed during the interval $[0, t]$ for subject i, and $Y_{ij}(t) = I(C_i \geq t)$, where $I(A)$ is an indicator function of event A. Also, let $Z_i = (Z_{i1}, \ldots, Z_{ip})'$ denote the p-vector covariate with corresponding regression parameters $\gamma = (\gamma_1, \ldots, \gamma_p)'$ in a Cox-type regression model that also includes a non-increasing function which accommodates the self-triggering effects of previous events. The intensity function is

given by

$$\lambda_i(t) = \lambda_0(t) \exp\left\{\gamma' Z_i + \int_0^t \rho(t-s) dN_i(s)\right\}, \tag{2.2.1}$$

where ρ is the self-triggering function that describes the decaying effects of previous events. Assuming that $\rho(x) = \alpha \exp(-\beta x)$ with unknown parameters $\alpha \geq 0$ and $\beta \geq 0$, the model (2.2.1) can then be written as

$$\lambda_i(t) = \lambda_0(t) \exp\left[\gamma' Z_i + \sum_{j=1}^{N_i(t-)} \alpha \exp\{-\beta(t-t_{ij})\} I\{N_i(t-) > 0\}\right], \tag{2.2.2}$$

where the parameter α controls the magnitude of the cumulative effects from previous events and parameter β controls for the decay rate of the function. When $\alpha = 0$, model (2.2.2) assumes no self-triggering effects from previous events. The parameter values $\alpha \neq 0$ and $\beta > 0$ imply that more recent events have stronger effects than more distant events, while $\alpha \neq 0$ and $\beta = 0$ indicate non-differential effects from previous events and lead to a regular Cox-type model with the total number of events, $N(t-)$, as a time-varying covariate.

It is important to recognize that model (2.2.2) can be explosive. The intensity may become arbitrarily large when either α is large or β is close to 0, and the process may become non-stationary under these conditions. To address this issue, one may restrict to bounded ρ by assuming there is no effect of previous events after either a certain time lag or after a certain number of previous events. Chen and Chen [29] considered an m-lag model, which is defined by

$$\lambda_i(t) = \lambda_0(t) \exp\left[\gamma' Z_i + \sum_{j \in \ell(t,m)} \alpha \exp\{-\beta(t-t_{ij})\}\right], \tag{2.2.3}$$

where $\ell(t,m) = \{j : \max(N(t-) - m + 1, 1) \leq j \leq N(t-), m \in \mathbb{N}^+, N(t-) > 0\}$. That is, the occurrence rate is influenced by the nearest m events in the history. Although m can be predetermined by a researcher based on prior knowledge, it can also be determined from the observed data based on cross-validation or likelihood based information measures, like the Akaike information criterion (AIC). Of course, model (2.2.3) is equivalent to model (2.2.2) as $m \to \infty$.

Let n_i denote the total number of observed events of subject i by $C_i, i = 1, \ldots, n$. Assuming C_i is independent of each T_{ij}, conditionally on Z_i, Chen and Chen [29] proposed a partial likelihood function for $\theta = (\gamma, \alpha, \beta)'$, which is defined as

$$L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left[\frac{\exp\{\phi_i(t_{ij}, m; \theta)\}}{\sum_{k \in R(t_{ij})} \exp\{\phi_k(t_{ij}, m; \theta)\}} \right]^{dN_i(t_{ij})}, \tag{2.2.4}$$

where $\phi_i(t, m; \theta) = \gamma' Z_i + \sum_{j \in \ell(t,m)} \alpha \exp\{-\beta(t - t_{ij})\}$ is the aggregated effects and $R(t_{ij}) = \{k : C_k \geq t_{ij}\} = \sum_i Y_{ij}(t_{ij})$ is the set of subjects who are at risk at time t_{ij}. The estimator $\hat{\theta}$ can be defined as a solution of $\partial \log\{L(\theta)\}/\partial\theta = 0$. The large sample properties of $\hat{\theta}$ follow from standard partial likelihood theory given that the self-triggering function satisfies certain regularity conditions. Thus, $n^{1/2}(\hat{\theta} - \theta)$ is consistent and asymptotically zero mean with covariance matrix Σ^{-1}, where

$$\Sigma = E\left[\int_0^\tau \left\{ \partial_\theta \phi(t, m; \theta) - \frac{E[\{\partial_\theta \phi(t, m; \theta)\} Y(t) \exp\{\phi(t, m; \theta)\}]}{E[Y(t) \exp\{\phi(t, m; \theta)\}]} \right\}^{\otimes 2} Y(t) \lambda(t) dt \right]. \tag{2.2.5}$$

The covariance matrix Σ can be consistently estimated with

$$\begin{aligned} \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \int_0^\tau & \left\{ \partial_\theta \phi(t, m; \hat{\theta}) - \frac{\sum_{j=1}^n \{\partial_\theta \phi_j(t, m; \hat{\theta})\} Y_j(t) \exp\{\phi_j(t, m; \hat{\theta})\}}{\sum_{j=1}^n Y_j(t) \exp\{\phi_j(t, m; \hat{\theta})\}} \right\}^{\otimes 2} \\ & \times Y_i(t) \exp\{\phi_i(t, m; \hat{\theta})\} d\hat{\Lambda}_0(t) dt, \end{aligned} \tag{2.2.6}$$

where $\hat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(s)}{\sum_{j=1}^n Y_j(s) \exp\{\phi_j(s, m; \hat{\theta})\}}$, and $\bar{N}(s) = \sum_{i=1}^n N_i(s)$.

2.2.2 Testing for Self-Triggering Effects

As mentioned previously, whether the event occurrence rate is independent of past events is determined by the parameter α. Testing such assumption is of practical interest in understanding the natural history of disease. To test the assumption, one tests if α equals zero. Under the null hypothesis $\alpha = 0$, one may naively use likelihood-based methods such as likelihood ratio test, score test, or Wald-type test. However, when $\alpha = 0$, the parameter β is not identifiable in model (2.2.2). The classical large sample properties of these likelihood-based

tests may not be valid under such non-identifiability [27, 30–32]. To avoid such issues, we propose the following testing procedure.

Given a sequence of fixed $\beta^{(k)}$, $k = 1, \ldots, K < \infty$, we estimate γ and α by maximizing $L(\theta^{(k)})$ in (2.2.4), where $\theta^{(k)} = (\gamma, \alpha, \beta^{(k)})'$ when $\beta = \beta^{(k)}$. Let $\hat{\gamma}^{(k)}$ and $\hat{\alpha}^{(k)}$ denote the maximizer of the function for γ and α, respectively. A Wald-type test statistic can be defined by $\hat{\alpha}^{(k)}/se(\hat{\alpha}^{(k)})$, where $se(\hat{\alpha}^{(k)})$ denotes the standard error of $\hat{\alpha}^{(k)}$. Under the null hypothesis that α equals zero, the test statistic asymptotically follows a standard normal distribution. However, to make a joint inference for α given different values of β, one requires an adjustment for multiple testing. For simplicity, we adopt the Bonferroni correction to account for multiple comparisons in these dependent hypothesis tests. If at least one p-value is less than the predetermined size divided by K, the null hypothesis is rejected so that the family-wise type I error probability can be controlled under the predetermined size. In the simulations below, we study the performance of each individual testing result and compare it with the overall procedure which adjusts for the multiple comparisons.

2.2.3 Prediction

Including event history as covariates is of importance for model specification, as well as for model prediction for future events. Let $p_{ij}(w; t)$ denote the probability of at least one event occurring within a window $(t, t + w]$, given that j events have occurred before t. Here, w can be considered as a period of time when the occurrence of an event is of interest after time t. One can show that the probability can be written as

$$\begin{aligned} p_{ij}(w; t) &= P(T_{i(j+1)} \in (t, t + w] | T_{ij} \leq t, T_{i(j+1)} > t) \\ &= 1 - S_{i(j+1)}(t + w)/S_{i(j+1)}(t), \end{aligned} \tag{2.2.7}$$

where $S_{i(j+1)}(t) = \exp\{-\int_{t_{ij}}^{t} \lambda_i(s)ds\}$ is the conditional survival function of the $(j + 1)$th event given that the jth event occurred at t_{ij}. If no event occurred before t, i.e., $j = 0$,

we would let $t_{i0} = 0$ and $S_{i1}(t) = \exp\{-\int_0^t \lambda_i(s)ds\}$ is the survival function of the time to the first event. One may simplify $p_{ij}(w;t)$ as $p_{ij}(w;t) = 1 - \exp\{\Lambda_i(t+w;t)\}$, where $\Lambda_i(t+w;t) = \int_t^{t+w} \lambda_i(s)ds$ is the cumulative hazard function in the window $(t,\, t+w]$. One can estimate $p_{ij}(w;t)$ by

$$\hat{p}_{ij}(w;t) = 1 - \exp\{\hat{\Lambda}_i(t+w;t,\hat{\theta})\},$$

where $\hat{\Lambda}_i(t+w;t,\hat{\theta}) = \int_t^{t+w} d\hat{\Lambda}_0(s)\exp\{\phi_i(s,m,\hat{\theta})\}ds$ with $\hat{\theta}$ and $\hat{\Lambda}_0$ defined in Subsection 2.2.1.

To evaluate the performance of the prediction, one may apply the receiver operating characteristic curve (ROC) for binary classification. This curve is created by plotting true positive rate (sensitivity) against false positive rate (1-specificity) at various threshold probabilities. Given a threshold probability q, one predicts at least one event would occur in $(t, t+w]$ if $\hat{p}_{ij}(w;t) > q$. Let $\delta_i = I\{\hat{p}_{ij}(w;t) > q\}$ denote the indicator of the predicted event occurrence. The true positive rate under q is estimated by $n_\Delta^{-1}\sum_{i=1}^n I(\delta_i = \Delta_i)$, where $\Delta_i = I\{N_i(t+w) - N_i(t) > 0\}$ and $n_\Delta = \sum_{i=1}^n \Delta_i$, while the false positive rate is estimated by $n_\Delta^{-1}\sum_{i=1}^n I(\delta_i = 1-\Delta_i)$. To summarize the accuracy of a model, an area under the curve (AUC) may be calculated. However, when two ROC curves are compared, one may not be interested in the entire range of the false positive rate. As an alternative, the partial AUC that considers the area under only a portion of the ROC curve may be calculated and may be more clinically relevant [33].

2.3 Simulation Studies

Extensive simulation experiments were conducted to demonstrate the feasibility of our proposed method. We considered sample sizes $n = 100, 400, 1000$ with different α, β, and m. For subject i, the recurrent event process was generated under the intensity function (2.2.3) with $\lambda_0(t) = 1$. We generated the first event time t_{i1} by solving the equation $t_{i1}\exp(\gamma z_i) +$

$\log(u_{i1}) = 0$, with subsequent event times obtained recursively by solving the equation

$$\int_{t_{i(\ell-1)}}^{t_{i\ell}} \exp\left[\gamma' z_i + \sum_{j\in\ell(t,m)} \alpha \exp\{-\beta(t - t_{ij})\}\right] dt + \log(u_{i\ell}) = 0,$$

for $t_{i\ell}$, $\ell = 2, \ldots, n_i$, until $t_{i(n_i+1)}$ is larger than the censoring time C_i, with $u_{i\ell}$ independently drawn from a uniform distribution on the interval (0,1), and C_i randomly drawn from a uniform distribution on the interval (0,4). The covariate $z_i = 1$ if i is even, and 0 otherwise.

Table 2.1 shows the simulation results when $\gamma = -0.5$, $\alpha = 0.5$, and $\beta = 0.5$, 1, with each scenario replicated $1,000$ times. We report mean (MEAN), median (MED), empirical standard deviation (ESD), defined by the sample standard deviation of the replicated estimates, average of the replicated standard deviation estimates (ASD), and empirical coverage probability (CP) at a 0.95 nominal level. In Table 2.1, the estimated γ and α have biases close to 0 but the bias of β is relatively large when $m = 1$ and $n = 100$. However, the median of the repeated estimates for β is close to the true value under this small sample size. The bias decreases as the sample size increases and the number of lags increases. The variance estimation is generally close to the empirical variance, with the empirical coverage probability close to the nominal level.

Table 2.2 shows the simulation results for testing $\alpha = 0$ when the true $\alpha = 0, 0.1, 0.2$, with $\gamma = -0.5$, $\beta = 0.5$ and $m = 2$. We implemented our hypothesis testing procedure that addresses the potential non-identifiability issues. The sample size n ranges from 100 to 1,000. We choose five different β values and estimate γ and α with β fixed at one of those five values. We report the average estimated values for γ and α from 1,000 repeated data generations. The empirical size of the test (*size*), defined as the percentage of rejections when the null hypothesis is true, i.e., $\alpha = 0$, and the empirical power (*power*), defined as the percentage of rejections when the null hypothesis is false, i.e., $\alpha \neq 0$, are shown in Table 2.2. Regardless of the true value of β, the mean estimates of α increases as the given β increases. The bias of $\hat{\alpha}$ is the smallest when β equals the true value except when $\alpha = 0$. As

Table 2.1: Simulation Results when $\gamma = -0.5$, $\alpha = 0.5$, and $\beta = 0.5$ (first half), $\beta = 1$ (second half)

	m	1			2			3			4		
	n	100	400	1000	100	400	1000	100	400	1000	100	400	1000
	MEAN	-0.50	-0.50	-0.50	-0.51	-0.50	-0.50	-0.51	-0.50	-0.50	-0.51	-0.50	-0.50
	MED	-0.50	-0.50	-0.49	-0.50	-0.50	-0.50	-0.51	-0.50	-0.50	-0.51	-0.51	-0.50
$\hat{\gamma}$	ESD	0.14	0.07	0.04	0.14	0.06	0.04	0.13	0.06	0.04	0.13	0.06	0.03
	ASD	0.15	0.07	0.04	0.14	0.06	0.04	0.13	0.06	0.04	0.13	0.06	0.04
	CP	95	94	95	94	95	95	95	95	94	95	94	94
	MEAN	0.52	0.49	0.49	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	MED	0.52	0.50	0.50	0.50	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.49
$\hat{\alpha}$	ESD	0.21	0.09	0.06	0.11	0.05	0.03	0.08	0.03	0.02	0.05	0.02	0.01
	ASD	0.21	0.10	0.06	0.11	0.05	0.03	0.08	0.03	0.02	0.05	0.02	0.01
	CP	95	95	95	94	94	95	95	95	94	94	96	94
	MEAN	1.19	0.56	0.52	0.62	0.51	0.49	0.53	0.49	0.49	0.49	0.49	0.49
	MED	0.49	0.49	0.49	0.48	0.49	0.48	0.46	0.48	0.48	0.47	0.49	0.49
$\hat{\beta}$	ESD	4.97	0.44	0.25	0.69	0.23	0.13	0.37	0.15	0.09	0.25	0.11	0.07
	ASD	6.43	0.48	0.25	0.76	0.22	0.13	0.37	0.15	0.09	0.25	0.11	0.07
	CP	90	93	94	93	93	95	93	93	93	93	94	94
	MEAN	-0.51	-0.49	-0.50	-0.51	-0.50	-0.50	-0.51	-0.50	-0.50	-0.51	-0.50	-0.50
	MED	-0.51	-0.49	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50	-0.51	-0.50	-0.49
$\hat{\gamma}$	ESD	0.15	0.07	0.04	0.14	0.07	0.04	0.14	0.06	0.04	0.14	0.07	0.04
	ASD	0.15	0.07	0.04	0.14	0.07	0.04	0.14	0.07	0.04	0.14	0.07	0.04
	CP	94	93	95	94	94	95	95	96	95	95	94	94
	MEAN	0.53	0.50	0.50	0.51	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	MED	0.53	0.50	0.50	0.51	0.50	0.49	0.49	0.49	0.49	0.49	0.50	0.49
$\hat{\alpha}$	ESD	0.27	0.10	0.06	0.14	0.06	0.04	0.10	0.04	0.03	0.08	0.04	0.02
	ASD	0.24	0.10	0.06	0.14	0.06	0.04	0.10	0.04	0.03	0.08	0.03	0.02
	CP	94	94	95	94	95	95	93	96	93	95	94	95
	MEAN	2.39	1.14	1.05	1.47	1.04	1.02	1.13	1.00	0.99	1.04	1.02	1.00
	MED	0.97	0.94	0.98	1.01	0.98	1.00	0.98	0.98	0.98	0.97	1.00	0.99
$\hat{\beta}$	ESD	7.05	0.97	0.45	3.31	0.38	0.23	1.10	0.28	0.18	0.53	0.23	0.14
	ASD	6.38	0.95	0.44	2.43	0.41	0.24	0.92	0.29	0.17	0.57	0.23	0.14
	CP	88	91	92	90	93	95	91	94	94	93	93	94

expected, the likelihood ratio test (LRT) has an inflated type-I error probability when the null hypothesis is true, since the usual regularity conditions are not satisfied. We observe the conservativeness of the Bonferroni correction (BC) under the same scenario, with the empirical size well below the nominal level. The statistical power of the Boferroni correction

Table 2.2: Simulation Results for Hypothesis Testing $H_0 : \alpha = 0$

n	100			400			1000		
β	$\hat{\gamma}$	$\hat{\alpha}$	*power*	$\hat{\gamma}$	$\hat{\alpha}$	*power*	$\hat{\gamma}$	$\hat{\alpha}$	*power*
0	-0.510	-0.012	5.2	-0.501	-0.001	4.0	-0.499	-0.002	4.9
0.25	-0.510	-0.014	5.5	-0.501	-0.001	3.6	-0.499	-0.003	5.0
0.5	-0.510	-0.016	5.1	-0.501	-0.001	4.3	-0.499	-0.004	4.8
0.75	-0.510	-0.017	5.3	-0.501	-0.001	4.5	-0.499	-0.005	4.7
1.0	-0.509	-0.018	4.7	-0.501	-0.001	4.3	-0.499	-0.006	5.2
BC			1.6			1.3			1.8
LRT			6.4			6.2			6.3
0	-0.506	0.061	9.7	-0.506	0.075	27.1	-0.499	0.075	57.7
0.25	-0.504	0.073	9.8	-0.504	0.089	30.3	-0.497	0.089	61.3
0.5	-0.503	0.082	10.7	-0.504	0.098	31.5	-0.498	0.098	62.3
0.75	-0.503	0.087	10.8	-0.505	0.104	30.8	-0.498	0.103	62.0
1.0	-0.504	0.091	10.3	-0.506	0.108	29.9	-0.500	0.107	60.6
BC			4.3			18.3			43.7
LRT			8.9			30.8			51.9
0	-0.510	0.154	29.7	-0.510	0.150	80.5	-0.504	0.151	99.2
0.25	-0.506	0.180	32.5	-0.506	0.176	84.9	-0.501	0.176	99.4
0.5	-0.506	0.196	33.7	-0.506	0.192	87.0	-0.500	0.192	99.5
0.75	-0.507	0.206	33.5	-0.507	0.202	86.1	-0.502	0.203	99.3
1.0	-0.509	0.213	33.7	-0.509	0.209	85.1	-0.504	0.210	98.9
BC			19.0			71.0			98.5
LRT			25.7			72.5			99.0

$\gamma = -0.5$, $\beta = 0.5$, and $\alpha = 0$ (first part), $\alpha = 0.1$ (second part), $\alpha = 0.2$ (third part)

tends to be lower than the likelihood ratio test when the null hypothesis is false, especially when the sample size is small. In contrast, the individual Wald tests with a given β have size close to 0.05 and have the highest powers in every scenario. This occurs because of the high correlation of the test statistics at different values of β, exceeding 0.90 for all pairs of β offering providing modest gains in power.

2.4 Analysis of Cystic Fibrosis Registry Data

Cystic fibrosis (CF) is an inherited disease of the secretory glands that causes thick and sticky mucus in lungs and blocks airways. The buildup of mucus facilitates bacterial growth in the lungs and repeated lung infections are common amongst CF patients. The most common pathogen observed in the lungs of CF patients is *Pseudomonas aeruginosa* (Pa) [34]. Recurrent Pa infections may be used to characterize the progression of chronic lung disease in young CF patients.

In the 2007 CF Registry data, there are $6,823$ subjects who were born after 1997 and had at least one follow-up before the end of year 2007. We considered gender, genotype, and a diagnostic group as the time-independent covariates. Among those patients, 50.3% were male. Regarding genotype, 68.8% were homozygous (ΔF508/ΔF508), 25.1% had a severe mutation in their genotype (ΔF508/non-ΔF508-I, II, III), and 6.1% had a mild mutation in their genotype (ΔF508/non-ΔF508-IV, V). The diagnostic group was defined by the method of diagnosis of CF. Among the CF registry patients, 23.9% were diagnosed by prenatal/neonatal screening (SCREEN), 18.6% were diagnosed by meconium ileus (MI), 5.7% were diagnosed by positive family history (FH), and 51.6% were diagnosed based on symptoms other than meconium ileus (SYMPTOM). We chose homozygous genotype and prenatal/neonatal screening groups as the reference categories for the genotype and diagnostic group variables, respectively. We excluded patients without complete information. The majority of patients who have missing data lack genotype information. The analysis below included $4,590$ individuals.

We applied and compared five different modeling approaches including AG, PWP-CP, PWP-GT, WLW and self-triggering Cox (STC) models. The AG model is a generalization of the proportional hazards model to the intensity function of the repeated events which assumes that the time-independent covariates have multiplicative effects. This model only

can introduce the influence of the prior events on future recurrences through the time-dependent covariates. The corresponding intensity function is given by

$$\lambda_i(t) = \lambda_0(t) \exp\{\gamma' Z_i\}$$

with the risk set indicator $Y_{is}(t) = I(T_{i,s-1} < t \leq T_{is})$ for $s = 1, \ldots, S$, $S = n_i + 1$ and the risk set at time t by $\sum_{is} Y_{is}(t)$. Prentice, Williams, and Peterson employed stratified AG models by considering two different time scales, a total time from the beginning of the study and a gap time from the most recent preceding occurrence of an event. The PWP models have restricted risk sets, i.e, the risk set for the $(s+1)$th event contains only subjects who have experienced s events and have stratum-specific hazards. The total time (PWP-CP) and gap time (PWP-GT) models are given, respectively, as follows:

$$\lambda_{is}(t) = \lambda_{0s}(t) \exp\{\gamma_s' Z_{is}\}$$

and $Y_{is}(t) = I(T_{i,s-1} < t \leq T_{is})$, and

$$\lambda_{is}(t) = \lambda_{0s}(t - t_{s-1}) \exp\{\gamma_s' Z_{is}\}$$

and $Y_{is}(t) = I(T_{is} - T_{i,s-1} > t)$, where λ_{0s} is an unspecified event-specific baseline hazard function for the sth event, $T_{i0} = 0$ and $T_{iS} = C_i$. The corresponding risk set at time t for each stratum s is $\sum_i Y_{is}(t)$. Thus, as s increases, the number of subjects in the risk set may decrease dramatically and it may be difficult to obtain stable parameter estimates for large values of s. The WLW model is based on the marginal Cox models. That is, each event is separately modeled on the total time scale with estimation within a given stratum ignoring information in other strata. To accommodate the dependence between the recurrent event times, a robust sandwich covariance matrix estimate is obtained. The hazard functions can be written

$$\lambda_{is}(t) = \lambda_{0s}(t) \exp\{\gamma_s' Z_{is}\}$$

with $Y_{is}(t) = I(T_{is} \geq t)$ and the risk set at time t in the sth stratum is $\sum_i Y_{is}(t)$. This method also has instability issues when the risk set becomes small for larger s. Finally, as

we mentioned in Section 2.2.1, the proposed STC model can be regarded as a generalized Cox-type model represented by the hazard function (2.2.3) with additional parameters α, β, and optionally m.

We divided the $4,590$ subjects into two groups: a training group and a test group. We randomly selected $3,590$ subjects as a training set for model estimation and building, and then used the rest of the subjects as a test set for the evaluation of prediction. The average number of events per subject in the training set is 1.8, with one individual having 39 repeated infections. Thus, the maximum number of possible strata based on the previous number of events for PWP and WLW methods is 40. Very few subjects would appear in some later strata if one uses a highly stratified approach. Several modifications have been proposed to deal with this issue [7, 35]. Firstly, one may ignore the risk set size issue and keep all strata in the analysis despite the fact that the within-strata hazard estimates are unstable. Secondly, one may delete the data after a certain number of events. Thirdly, one may aggregate the strata with a small number of subjects, especially those with a high frequency counts. Since among the 40 possible strata, 23 strata have less than 10 subjects, we adopted the third approach when applying PWP and WLW. We considered the two different strata schemes, denoted by s1 and s2, respectively. Up to the third event in s1 and the fifth event in s2, each stratum (event) has its own stratum-specific regression coefficients. However, a single model is considered for the later events.

For the conventional PWP and WLW methods, one can have either an overall estimate or event-specific estimates for the covariate effects. For WLW, the overall estimate is obtained as the weighted average of the event specific estimates minimizing the corresponding variance, while for PWP, the overall estimate is obtained by fitting a single model to all events with the same covariates. We chose six different lag numbers ($m = 1, 2, 3, 4, 5, \infty$) under the STC model. The estimation results under different models are shown in Table 2.3. To choose the number of lags (m) in the self-triggering models, $-2\log\{L(\hat{\theta})\}$ values were calculated, where

Table 2.3: Parameter Estimates under Various Models.

Model	Gender (0:male)	Severe	Mild	MI	FH	SYMPTOM	α	β
STC								
m=1	0.121*	-0.020	-0.395*	-0.048	-0.119*	0.013	1.731*	0.603*
	(0.024)	(0.028)	(0.073)	(0.041)	(0.060)	(0.029)	(0.033)	(0.031)
m=2	0.112*	-0.025	-0.386*	-0.047	-0.116	0.011	1.106*	0.549*
	(0.024)	(0.028)	(0.073)	(0.041)	(0.060)	(0.029)	(0.021)	(0.028)
m=3	0.110*	-0.023	-0.390*	-0.052	-0.118*	0.003	0.867*	0.514*
	(0.024)	(0.028)	(0.073)	(0.041)	(0.060)	(0.029)	(0.017)	(0.028)
m=4	0.107*	-0.025	-0.389*	-0.058	-0.126*	-0.006	0.743*	0.478*
	(0.024)	(0.028)	(0.073)	(0.041)	(0.060)	(0.029)	(0.016)	(0.027)
m=5	0.106*	-0.027	-0.396*	-0.063	-0.130*	-0.015	0.660*	0.452*
	(0.024)	(0.028)	(0.073)	(0.041)	(0.060)	(0.029)	(0.016)	(0.027)
m=∞	0.094*	0.001	-0.448*	-0.097	-0.178*	-0.082*	0.560*	0.801*
	(0.025)	(0.029)	(0.074)	(0.042)	(0.061)	(0.029)	(0.020)	(0.041)
AG	0.176*	-0.038	-0.595*	-0.114	-0.245*	-0.042		
	(0.031)	(0.036)	(0.088)	(0.052)	(0.075)	(0.037)		
PWP-CP	0.121*	-0.023	-0.374*	-0.026	-0.081	0.070*		
(s1)	(0.028)	(0.032)	(0.083)	(0.049)	(0.069)	(0.034)		
	0.110*	-0.031	-0.356*	-0.038	-0.102	0.051		
(s2)	(0.028)	(0.032)	(0.082)	(0.048)	(0.069)	(0.033)		
PWP-GT	0.106*	-0.026	-0.370*	-0.046	-0.113	0.002		
(s1)	(0.025)	(0.028)	(0.076)	(0.043)	(0.060)	(0.029)		
	0.102*	-0.031	-0.358*	-0.046	-0.122*	-0.010		
(s2)	(0.024)	(0.028)	(0.076)	(0.042)	(0.059)	(0.029)		
WLW	0.178*	-0.054	-0.660*	-0.144	-0.339*	-0.130*		
(s1)	(0.048)	(0.055)	(0.137)	(0.078)	(0.113)	(0.054)		
	0.198*	-0.051	-0.698*	-0.164*	-0.355*	-0.125*		
(s2)	(0.052)	(0.061)	(0.143)	(0.083)	(0.123)	(0.060)		

†Estimated standard errors in parentheses.
†† Superscript * indicates that the corresponding p-value is less than 0.05.

a model with a smaller $-2\log\{L(\hat{\theta})\}$ value is preferred. The corresponding $-2\log\{L(\hat{\theta})\}$ values for each m are 100266, 100027, 99974, 99931, 99937, and 100388 for $m = 1, \ldots, 6$, respectively. The smallest value occurs when $m = 4$, suggesting that the four most recent

events be used in the STC model.

To test the existence of the self-triggering effect, the Bonferroni correction method was used. When fitting the STC model with $m = 4$, referred to as STC(4), we estimated other parameters with β fixed at five different values, 0, 0.25, 0.478, 0.75, and 1.0, which is centered at the β estimate in the selected STC model. Five Wald-type test statistics and their corresponding p-values for $\alpha = 0$ were calculated. All five p-values are less than 0.01, which suggests that the self-triggering effect is statistically significant. The coefficient of gender $\hat{\gamma}_1 = 0.110$ indicates that the estimated relative risk for Pa infection is $\exp(0.110) = 1.116$. That is, the Pa infection is more likely to occur in females than in males. Similarly, one can conclude that the infection is less likely to occur in the mild mutation genotype group compared to the homozygous group and in the patients who are diagnosed via family history compared to the patients diagnosed by prenatal or neonatal screening. The same conclusions about covariate effects can be drawn under the AG model. The other models evidence somewhat different results even within the same stratification schemes. Additionally, we note that the coefficient estimates under the WLW model are greater than those from other competing models. Total time models, such as AG, PWP-CP, WLW, and STC tend to produce larger estimated covariate effects. This may occur because total times within a patient may be highly correlated, resulting in a *carry-over* effect. Such effects have been previously documented with the WLW model [5, 36–38].

The model prediction using the test set with 1,000 subjects was implemented using the fitted models from the training set, where possible. The estimated probability of a new event occurring during 2004 is calculated based on data at the end of 2003. Among $1,000$ subjects in the test data, we do not consider 456 subjects who are censored before 2004. Also, the WLW method is based on the marginal models which do not provide the estimated predictions. The left panel of Figure 2.1 displays ROC curves and AUC values of each model, while the right panel is a zoom-in on the portion of the ROC curves where the false positive

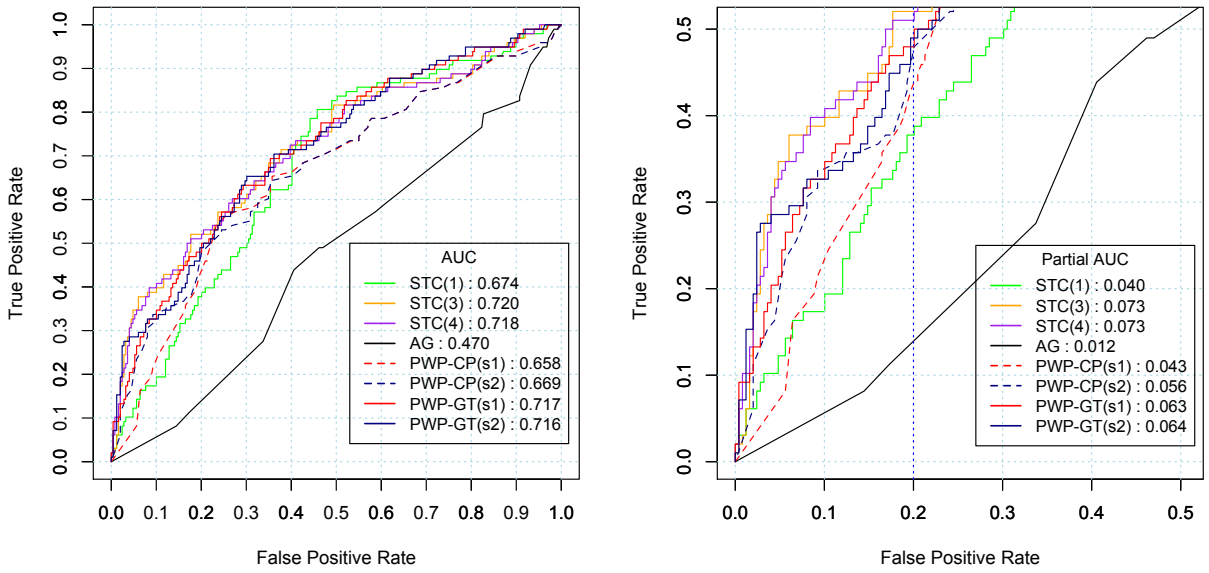

Figure 2.1: ROC curves and partial ROC curves

Table 2.4: Two Paired ROC Curves Comparison Tests' p-values by AUC and pAUC

Model	M1	M2	M3	M4	M5	M6	M7	M8	M9
M1: STC(1)		<.001	<.001	<.001	.002	.803	.113	<.001	<.001
M2: STC(2)	.001		.091	.187	<.001	.013	.190	.408	.484
M3: STC(3)	.004	.301		.936	<.001	.002	.040	.044	.067
M4: STC(4)	.013	.646	.504		<.001	<.001	.024	.035	.045
M5: AG	<.001	<.001	<.001	<.001		.001	<.001	<.001	<.001
M6: PWP-CP(s1)	.649	.050	.015	.011	<.001		.050	.028	.024
M7: PWP-CP(s2)	.880	.102	.037	.030	<.001	.175		.390	.255
M8: PWP-GT(s1)	.009	.797	.788	.948	<.001	.013	.037		.782
M9: PWP-GT(s2)	.012	.867	.694	.857	<.001	.014	.035	.832	

rate is less than or equal to 0.2.

We used the test method proposed by [39] for the comparison of the AUC and partial AUC using bootstrapped variance estimation [40]. The corresponding null hypothesis is $H_0 : A_1 = A_2$ and the alternative hypothesis is $H_0 : A_1 \neq A_2$, where A_1 and A_2 are the two

(partial) AUCs. All possible tests for the pairwise model comparisons are performed. Table 2.4 shows the corresponding p-values. Specifically, the lower triangular values represent p-values for the AUC comparisons, while the upper triangular values represent p-values for the partial AUC comparisons. The prediction performance of the STC(4) model is not significantly different from the PWP gap time model in terms of AUC, using either 4 (s1) or 6 (s2) strata. However, the underlined p-values in Table 2.4 indicate that the partial AUC of the STC(4) model is significantly different from all other competing models, including PWP gap time model. The STC(3) model has AUC and pAUC that are quite comparable to STC(4) and are significantly different from those for the non-STC models in Table 2.4.

2.5 Discussion

In this Chapter we examined the Cox model with self-triggering effects for recurrent event data, and compared the model to the currently existing methods. To test the existence of the self-triggering effect, the parameter beta describing the decay rate of the triggering effect was fixed with other unknown parameters estimated and used to construct a test statistic. This can address the non-identifiability of the STC model under the null of no triggering effect. The Bonferroni correction procedure was proposed to adjust for multiple testing at different values of beta. Interestingly, the adjustment appears to be rather conservative, with the simulations indicating that tests at fixed beta are highly correlated and a test at a single beta may provide greater power than the multiple testing approach. This requires further investigation. As an alternative, supremum score tests have been advocated in other testing scenarios with non-identifiability under the null and might be utilized in the current setting [30–32]. This is a topic for future research.

In analysis of the cystic fibrosis data, we demonstrated that the extended Cox model with a self-triggering scheme may yield significant gains in prediction of future events compared to available models. One particular characteristic of this dataset is that it includes fairly

large number of infections per subject, which creates difficulties in defining strata for the PWP and WLW methods. The STC model accommodates such data more naturally, with a growing number of recurrent events easily accommodated by the intensity function (2.2.3).

CHAPTER 3: COX MODEL FOR RECURRENT EVENT DATA WITH INSTRUMENTAL VARIABLE

3.1 Introduction

Recurrent failure events are frequently considered in a longitudinal study when subjects could possibly experience more than one event during the observation period. To explore the relationship between the effects of covariates and the correlated failure times, the Anderson-Gill (AG) model is commonly applied. However, when some patients do not adhere to their assigned treatments in a randomized trial, the standard intention-to-treat (ITT) analysis, which focuses on the causal effect of assignment of treatment rather than the causal effect of receipt of treatment, may not properly estimate the effect of treatment on the outcome. Another naive method is analyzing with received treatment. It is likely to be confounded by determinants of compliance [41]. The use of instrumental variable methods helps us to consistently estimate the average causal effect of an exposure on some outcome of interest even in the presence of latent confounding. Abadie [42] suggested new IV estimators for general response models with covariates. In this Chapter, we will apply the weighting scheme of Abadie [42] into the Cox and AG models for analyzing survival data with non-informative right censoring. We demonstrate our approach via comprehensive simulation studies and a colorectal-cancer screening data analysis.

Instrumental variable (IV) methodology is one approach to deal with the issue of unmeasured confounders. It has been actively explored and discussed over the last few decades in the econometrics, epidemiology, statistics, and biomedical sciences [10, 41, 43–46].

An important application of IV methods is to estimate the effect of receiving treatment

in randomized trials with non-adherence, i.e., some individuals do not comply with the assignment treatment. When there is noncompliance, the standard intention-to-treat (ITT) effect is different from the effect of receiving the treatment versus the control. Since the ITT measures the effect of assignment of treatment rather than the effect of actual receipt of treatment [42, 45, 47–49].

A special type of noncompliance, all-or-nothing (all-or-none) compliance, is that all subjects are immediately categorized whether they comply or not with their assigned treatment after randomization. That is, there is no partial compliance in this situation [50]. In contrast, an application in partial compliance requires strong assumptions for identifiability. Thus, the all-or-none compliance status is commonly considered [14, 18, 47, 48, 50, 51].

To deal with survival outcome under noncompliance, several methods have been developed and implemented. Baker [51] suggested a likelihood-based approach for discrete time survival data to estimate the difference between complier hazards in treatment and control group by considering death as competing risk. Loeys and Goetghebeur [14] proposed the marginal structural Cox proportional hazards model for a compliers proportional hazards effect of treatment and derived estimating equation for it under independent censoring. However, they did not consider covariates, that is a generalization of the Mantel-Haenszel estimator. Cuzick *et al.* [19] applied a partial likelihood method by accommodating covariates by assuming independence of covariates and compliance class. Also, they explored a full likelihood when covariates are not independent of compliance, but it was too complicated to estimate. Gong [50] developed several parametric potential outcome survival models with considering ignorable and non-ignorable censoring schemes. Based on Baker [51] and Nie *et al.* [52] estimated the effect of treatment on survival at specific times by adopting a nonparametric approach in the presence of noncompliance and administrative censoring. Nie *et al.* [52] included always-takers, unlikely that Baker [51] only considered compliers and never-takers. Additionally, they gained efficiency over the standard IV method by using the

mixture structure in the data. Lin *et al.* [53] considered a semi-parametric linear transformation model and proposed a two-stage estimation procedure to estimate the parameters. MacKenzie *et al.* [54] extended Cox's proportional hazards model by adding additional additive term into intensity function by assuming approximate orthogonality of an instrument with latent confounders. Thus, they derived an estimator from the score equation of the partial likelihood similar to Cox model. Yu *et al.* [18] used semi-parametric transformation models for the distribution of survival time, conditionally on covariates and latent compliance type. Maximum likelihood is used to estimate the parameters of the transformation models and applied expectation-maximization (EM) algorithm to overcome the computational difficulties from the mixture structure and the infinite dimensional parameter in the models. They considered the complier average causal effect, the complier effect on survival beyond time t, and the complier quantile effect. Like Cuzick *et al.* [19], they also allow for always-takers with a positive probability but they allow the association between covariates and response to vary with compliance class. A few papers have considered trials with repeated outcome measures in the presence of noncompliance. Yau *et al.* [55] extended Imbens and Rubin [56] by allowing the baseline covariates and missing in outcomes, Small *et al.* [57] proposed a random effects model approach for longitudinal binary outcomes, and OMalley [58] presented the concept of lagged predictors and outcomes to incorporate IV in longitudinal analysis.

The remaining sections are organized as follows. In Section 3.2 we introduce notation, model, estimand and IV assumptions mostly based on [42] and describes the method of estimation and asymptotic properties of the estimators. To demonstrate the feasibility of our proposed method, simulation studies are presented in Section 3.3. In Section 3.4, we apply the method into the colorectal- cancer data from Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial to evaluate the effect of flexible sigmoidoscopy in comparison with usual care on colorectal-cancer mortality.

3.2 Inference

3.2.1 Notations, Assumptions and Models

We consider two-armed randomized trials with all-or-nothing compliance and repeated failure time outcomes. Let X be a p-dimensional covariate vector. Suppose that Z is a binary instrument variable. If a subject is assigned to the control group then $Z = 0$, and if a subject is assigned to the treatment group, then $Z = 1$. D_z is a potential treatment status given $Z = z$. That is, under assignment $Z = z$, $D_z = 1$ represents a particular subject would take the treatment, but would not be given the treatment otherwise. The observed treatment status binary variable can be expressed as $D = ZD_1 + (1 - Z)D_0$. Similarly, let T_{zd} denote the potential event time vector when $Z = z$ and $D = d$.

Clearly, if $D_0 = 0$, we are only able to observe T_{00} not T_{01} without considering censoring. Let $L = l$ be the indicator of the potential compliance stratum. There are four different types given by [56] such as compliers, always-takers, never-takers, and defiers. By convention, the values $L = 1, 2, 3, 4$ refer to compliers, always-takers, never-takers, and defiers respectively. Since we cannot fully observe potential outcome vector, compliance type is also unobservable. These four compliance types can be written with using D_z as follows,

$$L = \begin{cases} 1 & \text{if } D_0 = 0 \text{ and } D_1 = 1 \\ 2 & \text{if } D_0 = 1 \text{ and } D_1 = 1 \\ 3 & \text{if } D_0 = 0 \text{ and } D_1 = 0 \\ 4 & \text{if } D_0 = 1 \text{ and } D_1 = 0. \end{cases}$$

We now make the following assumptions akin to [42]:

Assumption 3.2.1. *random assignment conditional on covariates*

(D_0, T_{00}, T_{01}, D_1, T_{10}, T_{11}) $\perp\!\!\!\perp$ $Z|X$, where $\perp\!\!\!\perp$ denotes independence.

Assumption 3.2.2. *conditional monotonicity*

$P(D_1 \geq D_0|X) = 1$.

Assumption 3.2.3. *exclusion restriction*

$P(T_{0d} > t|X) = P(T_{1d} > t|X)$ *for all* t *and* $d \in \{0, 1\}$.

Assumption 3.2.1 is a weaker assumption than (D_0, T_{00}, T_{01}, D_1, T_{10}, T_{11}, $X \perp\!\!\!\perp Z$) in a complete randomization setting. That is, it means that there is no other common factors of the IV and the outcome. Assumption 3.2.2 implies that there are no defiers. This assumption is based on consistent preference [46] and it is a key assumption under heterogeneous treatment effects [10]. Assumption 3.2.3 says that the probability of outcome does not depend on treatment in always-takers and never-takers strata since they receive the same treatment regardless of assignment; the experiment provides no information about the treatment effect in these strata. Thus, we can write $T_d \equiv T_{zd}$ for any z. That is, it guarantees that Z is not able to make an effect on the outcome directly only through the treatment status (D).

Assumption 3.2.4. *conditional first stage*

$0 < P(Z = 1|X) < 1$ *and* $P(D_1 = 1|X) > P(D_0 = 1|X)$.

Assumption 3.2.5. *independent censoring*

$(Z, D, T) \perp\!\!\!\perp C$, *where* C *is the censoring time.*

In addition to these assumptions, Abadie [42] presents a Lemma 2.1 an identification Theorem 3.1 summarized as following: Let $g(.)$ be any measurable real function of (Y, D, X) such that $E|g(Y, D, X)| < \infty$, where Y is an outcome of interest. Define

$$\kappa = 1 - \frac{D(1-Z)}{P(Z=0|X)} - \frac{(1-D)Z}{P(Z=1|X)},$$

and under the previous assumptions,

$$E[g(Y, D, X)|D_1 > D_0] = \frac{1}{P(D_1 > D_0)} E[\kappa g(Y, D, X)].$$

The κ can be considered as a weight that allows us to estimate expectations for compliers in terms of expectations for the entire population. However, κ takes a negative value when D is different from Z. Thus, we cannot directly use log partial likelihood as an objective function in the estimation process. This will be discussed more detail in Section 3.2.

Let $N^*(t)$ be the number of events that occur during the interval $[0, t]$ and $H(\cdot)$ be a $(p+1)$ dimensional covariate process including X and D. However, the observation period is mostly limited by some predetermined time τ so $N^*(\cdot)$ cannot be fully observed. Let C denote censoring time as previously defined. The censoring scheme is also assumed to be independent such that $E[dN^*(t)|H(t), C \geq t] = E[dN^*(t)|H(t)]$ for all $t \geq 0$. Anderson and Gill [3] suggested the following intensity model such as

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1' X + \beta_2 D\}, \tag{3.2.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, β_1 is a p-dimensional coefficient vector for X and β_2 is a scalar coefficient for D. These regression coefficients measure the effects of the corresponding covariates to the intensity on the log scale. Here, we only consider time invariant covariates, i.e., $H(t) = H$. Since Z and D for compliers, Z is ignorable under the Assumptions 3.2.1 and 3.2.3. Thus, we have

$$\lambda(T|D_1 > D_0, X, D = 1) = \lambda(T_1|D_1 > D_0, X, Z = 1) = \lambda(T_1|D_1 > D_0, X),$$

$$\lambda(T|D_1 > D_0, X, D = 0) = \lambda(T_0|D_1 > D_0, X, Z = 0) = \lambda(T_0|D_1 > D_0, X)$$

and

$$\frac{\lambda(T|D_1 > D_0, X, D = 1)}{\lambda(T|D_1 > D_0, X, D = 0)} = \frac{\lambda(T_1|D_1 > D_0, X)}{\lambda(T_0|D_1 > D_0, X)} = \exp(\beta_2).$$

We consider this hazard function as a local hazard function defined as $\lambda(T|D_1 > D_0, X, D)$. The β_2 is the parameter of interest and we refer to $\exp(\beta_2)$ as the causal proportional hazards effect corresponding to the conditions described by two levels of D within compliers.

3.2.2 Estimation and Asymptotic Properties

Abadie [42] presented two methods to estimate parameters for LARF such as Least Squares (LS) and Maximum Likelihood (ML). ML can be easily applied to the partial likelihood to deal with time-to-event data, so it will be adopted in this paper. Define $N(t) = N^*(t \wedge C)$ and $Y(t) = I\{C \geq t\}$, where $a \wedge b = \min(a, b)$, and $I(\cdot)$ is an indicator function. Suppose that the observed data $\{N_i(\cdot), Y_i(\cdot), H_i\}$ for $i = 1, \ldots, n$ are independent and identically distributed. Let's denote $\omega(x) = P(Z = 1|X = x)$ as a nuisance function for κ. If we know the function $\omega(x)$, then we can calculate κ_i with $\kappa_i = 1 - d_i(1 - z_i)/\{1 - \omega(x_i)\} - (1 - d_i)z_i/\omega(x_i)$. The following equation is a partial likelihood after weighting by κ based on [3]

$$PL(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left\{ \frac{\kappa_i \exp(\theta' H_i)}{\sum_{k \in R(t_{ij})} \kappa_k \exp(\theta' H_k)} \right\}^{dN_i(t_{ij})}, \tag{3.2.2}$$

where $H_i = (X_i', D_i)'$, $\theta = (\beta_1', \beta_2)'$, $j = 1, \ldots, n_i$, and $n_i = N_i(C_i)$. As we mentioned in Section 3.2, we cannot directly use this weighted partial likelihood as an objective function for ML. Since κ_i can be negative values, taking the logarithm of this function is not valid. Under Eq.(3.2.2), the corresponding weighted score function for θ is given by

$$U_\kappa(t; \theta) = \sum_{i=1}^{n} \kappa_i \int_0^t \{H_i - E_\kappa(s; \theta)\} \, dN_i(s), \tag{3.2.3}$$

where $E_\kappa(s; \theta) = \frac{\sum_{j=1}^{n} H_j Y_j(s) \kappa_j \exp(\theta' H_j)}{\sum_{j=1}^{n} Y_j(s) \kappa_j \exp(\theta' H_j)}$.

Based on [42], the following assumption guarantees the usual identification condition.

Assumption 3.2.6. *$U(\tau; \theta | D_1 > D_0) = 0$ has a unique solution at θ_0 over $\theta \in \Theta$, where τ denotes the end of the study time, $U(\tau; \theta) = \sum_{i=1}^{n} \int_0^\tau \{H_i - E(s; \theta)\} dN_i(s)$, and $E(s; \theta) = \frac{\sum_{j=1}^{n} H_j Y_j(s) \exp(\theta' H_j)}{\sum_{j=1}^{n} Y_j(s) \exp(\theta' H_j)}$.*

Then, the parameter θ_0 can be estimated by $\hat{\theta}$ the solution to the equation $U_\kappa(\tau; \theta) = 0$

with using Theorem 3.1 in [42]. This implies that $n^{-1}U_\kappa(\theta)$ converges almost surely to a function $U(\theta|D_1 > D_0)$, that is $\|U_\kappa(\theta) - U(\theta|D_1 > D_0)\| \longrightarrow_{a.s.} 0$.

As we mentioned, if we know the function ω_0, then $\kappa_i(d_i, z_i, \omega_0(x_i, \gamma_0))$ is observed and θ_0 can be directly estimated with a single step. However, $\omega(\cdot)$ is commonly unknown in practice. Therefore, we need one more step to estimate γ_0 which could be estimated by specifying a parametric model $\omega(X, \gamma)$. Suppose that we consider $E[Z|X] = \Phi(\gamma' X)$, where $\Phi(\cdot)$ is a standard normal cumulative distribution function. This probit linear model is often used when the dependent variable is binary. Then, γ_0 can be estimated by solving

$$\frac{\partial}{\partial \gamma} \sum_{i=1}^{n} \{z_i \ln \Phi(\gamma' x_i) + (1 - z_i) \ln \Phi(-\gamma' x_i)\} = 0. \tag{3.2.4}$$

Let $\hat{\gamma}$ denote the solution of the Eq.(3.2.4) and κ_i can be estimated by $\hat{\kappa}_i(\hat{\gamma}) = \kappa_i(d_i, z_i, \omega(x_i, \hat{\gamma}))$. The weighted score function re-expressed by adding γ is following

$$U_\kappa(\theta, \gamma) = \sum_{i=1}^{n} \kappa_i(\gamma) \int_0^\tau [H_i - E_\kappa(s; \theta, \gamma)] \, dN_i(s), \tag{3.2.5}$$

where $E_\kappa(s; \theta, \gamma) = \frac{\sum_{j=1}^{n} H_j Y_j(s) \kappa_j(\gamma) \exp(\theta' H_j)}{\sum_{j=1}^{n} Y_j(s) \kappa_j(\gamma) \exp(\theta' H_j)}$. To obtain $\hat{\theta}(\hat{\gamma})$, we solve $U_\kappa(\theta, \hat{\gamma}) = 0$. The following regularity conditions are imposed to construct Theorem 3.2.1 and 3.2.2, for $i = 1, \ldots, n$:

c1) $P(Y_i(\tau) = 1) > 0$.

c2) $N_i(\tau)$ are bounded by a constant.

c3) H_i are bounded and time invariant.

c4) Let $S_\kappa^{(r)}(t; \theta, \gamma) = n^{-1} \sum_{i=1}^{n} H_i^{\otimes r} Y_i(t) \kappa_i(\gamma) \exp(\theta' H_i)$ for $r = 0, 1, 2$, where for any vector a, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa'$. Also, let $E_\kappa(t; \theta, \gamma) = S_\kappa^{(1)}(t; \theta, \gamma) / S_\kappa^{(0)}(t; \theta, \gamma)$, and $e_\kappa(t; \theta, \gamma)$ be the corresponding limit.

c5) $\Sigma_\theta = E\left[\int_0^\tau \kappa(\gamma_0)\{H - e_\kappa(\theta_0, \gamma_0)\}^{\otimes 2} Y(t) \exp(\theta_0' H) d\Lambda_0(t)\right]$ is positive definite.

Theorem 3.2.1. *(Consistency of $\hat{\theta}$)*

Suppose that all Assumptions 3.2.1-3.2.6 and the regularity conditions previously listed and that (i) $\omega(\cdot)$ belongs to some parametric functions such that $\omega(X,\gamma)$ with $\omega_0 = \omega(X,\gamma_0)$ for some $\gamma_0 \in \mathbb{R}^q$; there exists $\eta > 0$ such that for $\|\gamma - \gamma_0\| < \eta$, $\omega(X,\gamma)$ is bounded away from zero and one and is continuous at each γ on the support of X; (ii) $\hat{\gamma} \longrightarrow_p \gamma_0$. Then $\hat{\theta}(\hat{\gamma}) \longrightarrow_p \theta_0$.

To establish the asymptotic normality of $\hat{\theta}$, we need to consider the corresponding distribution of $U_\kappa(\theta_0,\gamma_0) = \sum_{i=1}^n \kappa_i(\gamma_0)\int_0^\tau [H_i - E_\kappa(s;\theta_0,\gamma_0)]\, dM_i(s)$, where $dM_i(s) = dN_i(s) - Y_i(s)\exp(\theta_0' H_i) d\Lambda_0(s)$.

Theorem 3.2.2. *(Asymptotic normality of $\hat{\theta}$)*

Under the assumptions in Theorem 3.2.1, and (i) for $\|\gamma - \gamma_0\| < \eta$, $\omega(X,\gamma)$ is continuously differentiable at each γ, $\partial\omega(X,\gamma)/\partial\gamma$ is bounded; (ii) $\hat{\gamma}$ is asymptotically linear with influence function $\psi(B)$, i.e., $\sqrt{n}(\hat{\gamma} - \gamma_0) = \sqrt{n}^{-1}\sum_{i=1}^n \psi(b_i) + o_p(1)$, where $E[\psi(b_i)] = 0$ and $E[\psi(b_i)'\psi(b_i)] < \infty$. Then, $\sqrt{n}(\hat{\theta}(\hat{\gamma}) - \theta_0) \longrightarrow_d N(0,\Sigma)$, where

$$\Sigma = \Sigma_\theta^{-1} E\left[\{U_\kappa(\theta_0,\gamma_0) + \Sigma_\gamma\psi\}\{U_\kappa(\theta_0,\gamma_0) + \Sigma_\gamma\psi\}'\right]\Sigma_\theta^{-1},$$

$$\Sigma_\theta = E[\frac{\partial}{\partial\theta}U_\kappa(\theta_0,\gamma_0)], \Sigma_\gamma = E[U(\theta_0,\gamma_0)\{\partial\kappa(\gamma_0)/\partial\gamma\}],$$

and

$$U(\theta_0,\gamma_0) = \sum_{i=1}^n \int_0^\tau \{H_i - E_\kappa(s;\theta_0,\gamma_0)\} dM_i(s).$$

The Σ can be consistently estimated by

$$\hat{\Sigma} = \hat{\Sigma}_\theta^{-1}\left[\frac{1}{n}\sum_{i=1}^n \{U_{i\kappa}(\hat{\theta},\hat{\gamma}) + \hat{\Sigma}_\gamma\hat{\psi}_i\}\{U_{i\kappa}(\hat{\theta},\hat{\gamma}) + \hat{\Sigma}_\gamma\hat{\psi}_i\}'\right]\hat{\Sigma}_\theta^{-1}, \text{ where}$$

$$\hat{\Sigma}_\theta = -n^{-1}\partial U_\kappa(\hat{\theta},\hat{\gamma})/\partial\theta = n^{-1}\sum_{i=1}^n \int_0^\tau \kappa_i(\hat{\gamma})\{H_i - E_\kappa(\hat{\theta},\hat{\gamma})\}^{\otimes 2} Y_i(s)\exp(\hat{\theta}' H_i) d\hat{\Lambda}_0(s),$$

$$U_{i\kappa}(\hat{\theta},\hat{\gamma}) = \int_0^\tau \kappa_i(\hat{\gamma})\{H_i - E_\kappa(s;\hat{\theta},\hat{\gamma})\} d\hat{M}_i(s), \text{ where } d\hat{M}_i(s) = dN_i(s) - Y_i(s)\exp(\hat{\theta}' H_i) d\hat{\Lambda}_0(s),$$

$$\hat{\Sigma}_\gamma = n^{-1} \sum_{i=1}^{n} \int_0^\tau \{H_i - E_\kappa(s; \hat{\theta}, \hat{\gamma})\} d\hat{M}_i(s) \{\partial \kappa_i(\hat{\gamma}) / \partial \gamma\}'.$$

The estimate for $U_{i\kappa}$ is defined as the score residual for each subject and each variable [59]. The estimator of the cumulative baseline intensity function $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is similar to the Breslow estimator in [60]. However, some κ_i's are negative so it is given by

$$\hat{\Lambda}_0(t; \hat{\theta}, \hat{\gamma}) = \int_0^t \max \left\{ \frac{\sum_{i=1}^n \kappa_i(\hat{\gamma}) dN_i(s)}{\sum_{i=1}^n \kappa_i(\hat{\gamma}) Y_i(s) \exp(\hat{\theta}' H_i)}, 0 \right\}. \tag{3.2.6}$$

3.3 Simulation Studies

Extensive simulation experiments are explored in this section to demonstrate the feasibility of our proposed method by comparing it with the existing method. We adopted the ways to generate u_i, d_i and t_{ij} based on [53] and [18]. For each subject $i = 1, \ldots, n$, we do the following steps:

1) Draw x_i from some choice of $f(x)$.

2) Draw z_i from some choice of $P(z = 1|x; \gamma)$. We chose $P(z = 1|x) = \Phi(\gamma' x)$, where Φ is the cumulative distribution function (CDF) of the standard normal distribution.

3) Draw l from the multinomial distribution with probability $p = (p_1, p_2, p_3)$, in which $l = 1$ (compliers), 2 (always-takers), or 3 (never-takers), and the subgroup probabilities are calculated by using the multinomial logistic regression model given by

$$p_l = P(L = l|x; \alpha) = \frac{\exp(\alpha_{0l} + \alpha_l' x)}{1 + \sum_{l=1}^{2} \exp(\alpha_{0l} + \alpha_l' x)}, \text{ for } l = 1, 2$$

, and $p_3 = 1 - p_1 - p_2$.

4) Determine d_i by $f(z, l) = zI\{l \leq 2\} + (1 - z)I\{l = 2\}$.

5) Generate t_{ij} by applying different coefficients depending on potential class types.

   a) Draw r_i from the uniform distribution from 0 to 1.

b) Solve the following equation to get a gap time g_{ij},

$$\sum_{m=1}^{3} g_{ij} \exp(\theta_l' h_i) I\{l_i = l\} + \log(r_i) = 0, \text{ where } h_i = (x_i, d_i)'.$$

c) Calculate $t_{ij} = t_{i(j-1)} + g_{ij}$.

d) Recursively do step a)-c) while t_{ij} is less than some predetermined value τ.

6) Generate c_i from the uniform distribution from 0 to τ.

As a brief illustration, we performed a simulation study where 1000 samples were generated according to the aforementioned algorithm, each with $n = 500, 1000, 2000$ observations and $\tau = 5$. We considered two covariates except D. One is a binary variable drawn from the binomial distribution with probability 0.5 and the other one is a continuous covariate drawn from the truncated normal distribution whose range is from -2 to 2. By changing the parameter α's in step 3, we are able to adjust the proportion of compliers, p_1. The proportion of compliers can measure the strength of the IV which refers to how strongly the IV is associated with the treatment after controlling for the measured confounders X [10, 45]. We conducted simulations under two different p_1's ranges such as $[.49, .68]$ and $[.61, .89]$. Additionally, each stratum has a different θ_m, $m = 1, 2, 3$ in step 5 b), but we will use the same value for $m = 2$ and 3 for simplicity. Since heterogeneity of treatment effects is not restricted by the identification conditions in this IV model [42]. Data was generated according to the assumptions that we have in Section 3.2.

Table 3.1 and 3.2 show the simulation results when $\theta_1 = (\beta_{11c}, \beta_{12c}, \beta_{2c})' = (0.5, 0.5, 1.0)'$ and $\theta_2 = \theta_3 = (\beta_{11o}, \beta_{12o}, \beta_{2o})' = (0.3, 0.3, 0.5)'$, with complier probability ranges $[.49, .68]$ and $[.61, .89]$, respectively. The proposed IV model is compared with the AG models, 1) with D; 2) with Z, i.e., ITT, in every six settings. The IV parameters are estimated by using initial values from the AG model with D. We report mean (MEAN); median (MED);

Table 3.1: Simulation Results with $\beta_{11c} = \beta_{12c} = 0.5$, $\beta_{11o} = \beta_{12o} = 0.3$, $\beta_{2c} = 1.0$, $\beta_{2o} = 0.5$ and $p_1 = [.49, .68]$.

		Proposed IV Model			AG Model			AG Model (ITT)		
	N	500	1000	2000	500	1000	2000	500	1000	2000
$X(\beta_{11c})$	MEAN	0.511	0.501	0.500	0.540	0.538	0.536	0.472	0.471	0.470
	MED	0.504	0.498	0.497	0.536	0.538	0.536	0.470	0.470	0.471
	ESD	0.199	0.106	0.069	0.068	0.049	0.036	0.070	0.051	0.036
	TSD	0.191	0.109	0.074	0.069	0.049	0.035	0.059	0.042	0.029
	CP	96.9	96.4	96.6	86.0	79.9	73.4	88.2	84.6	77.3
$X(\beta_{12c})$	MEAN	0.518	0.503	0.503	0.480	0.481	0.481	0.433	0.433	0.434
	MED	0.502	0.495	0.500	0.481	0.483	0.481	0.433	0.432	0.433
	ESD	0.137	0.077	0.042	0.044	0.030	0.021	0.045	0.031	0.021
	TSD	0.119	0.070	0.045	0.042	0.030	0.021	0.035	0.024	0.017
	CP	94.9	95.2	96.0	82.1	79.1	73.2	50.0	25.2	7.8
$D(\beta_{2c})$	MEAN	1.010	1.004	0.999	0.708	0.709	0.706	0.688	0.688	0.686
	MED	1.005	1.008	0.998	0.708	0.709	0.707	0.687	0.690	0.686
	ESD	0.115	0.075	0.053	0.068	0.046	0.032	0.074	0.049	0.036
	TSD	0.123	0.082	0.053	0.066	0.046	0.033	0.060	0.042	0.030
	CP	95.3	95.7	94.8	0.8	0	0	0.5	0	0
Outlier/No Convergence		22	9	3	0	0	0	0	0	0

empirical standard deviation (ESD), defined by the sample standard deviation of the replicated estimates; average of the replicated theoretical standard deviation estimates (TSD); and empirical coverage probability (CP) at a 0.95 nominal level. Additionally, we present the number of outliers or no-convergence cases out of $1,000$ replications. The outliers are defined by using the median absolute deviation (MAD), $MAD(x) = b * MED(|x - MED(x)|)$, where $b = 1.4826$. If $|x_i - MED(x)|/MAD(x)$ is greater than 20, then we define x_i as an outlier. Table 3.1 and 3.2 contain all estimates based on the proposed IV method. The estimates have biases close to 0 and the biases decrease as the sample size increases. The estimated variances are similar to the corresponding empirical variances, so the empirical coverage probabilities are close to the nominal level 95. In contrast, the biases by the naive AG model are relatively large and do not noticeably decrease by increasing sample size.

Table 3.2: Simulation Results with $\beta_{11c} = \beta_{12c} = 0.5$, $\beta_{11o} = \beta_{12o} = 0.3$, $\beta_{2c} = 1.0$, $\beta_{2o} = 0.5$ and $p_1 = [.61, .89]$.

		Proposed IV Model			AG Model			AG Model (ITT)		
	N	500	1000	2000	500	1000	2000	500	1000	2000
$X(\beta_{11c})$	MEAN	0.502	0.502	0.500	0.556	0.556	0.556	0.523	0.522	0.524
	MED	0.496	0.497	0.500	0.557	0.556	0.555	0.526	0.523	0.524
	ESD	0.116	0.074	0.052	0.065	0.044	0.033	0.072	0.050	0.035
	TSD	0.110	0.074	0.051	0.062	0.044	0.031	0.059	0.042	0.030
	CP	95.6	95.9	94.7	82.7	71.6	52.2	88.5	86.8	82.7
$X(\beta_{12c})$	MEAN	0.504	0.500	0.502	0.490	0.488	0.489	0.475	0.473	0.473
	MED	0.496	0.497	0.501	0.490	0.488	0.489	0.474	0.473	0.472
	ESD	0.072	0.044	0.031	0.036	0.025	0.018	0.040	0.028	0.021
	TSD	0.062	0.042	0.029	0.035	0.025	0.018	0.035	0.024	0.017
	CP	96.1	95.8	94.5	91.6	91.0	87.0	83.4	77.3	62.7
$D(\beta_{2c})$	MEAN	1.001	0.999	0.999	0.969	0.968	0.968	0.875	0.873	0.874
	MED	0.999	0.999	1.000	0.966	0.968	0.968	0.875	0.874	0.874
	ESD	0.079	0.054	0.037	0.061	0.042	0.029	0.068	0.046	0.032
	TSD	0.079	0.055	0.038	0.060	0.042	0.030	0.059	0.042	0.029
	CP	94.9	95.4	95.9	91.1	88.0	79.0	43.9	16.8	1.5
Outlier/No Convergence		2	2	0	0	0	0	0	0	0

Additionally, even though the variance estimation is close to the empirical variance, the empirical coverage probability has a rather poor value due to bias. We expected intuitively some values between 0.3 and 0.5 for the estimates of β_{11} and β_{12} in AG model. However, the β_{12} is included in that range but not for the β_{11}. The bias of β_{11} is even larger when the compliance rate is higher. Through this simulation, we can also confirm that IV estimators have a larger asymptotic variance than the conventional ones, since the IV introduces an additional source of uncertainty. In terms of proportion of compliance, the less the rate of noncompliance, the less the ITT effect and the average treatment effect among compliers tend to differ, which is the same results in [10, 44, 45]. The number of outliers or divergence cases decreases as the rate of compliance, or sample size increase. We also applied this proposed method to the Cox proportional hazard model and present corresponding simulation

results in the Appendix 5.

3.4 Colorectal-Cancer Mortality with Screening Flexible Sigmoidoscopy

Colorectal-cancer is cancer that initiates in the colon or rectum which are parts of the large intestine. It is known as the second most common cause of cancer death in the United States after lung cancer [61]. Screening is often suggested as one of the best way to protect colorectal-cancer. Because it can early detect precancerous growths, called polyps, that can transform into cancer. There are several recommended screening options, such as fecal occult blood test (FOBT), fecal immunochemical test (FIT), colonoscopy, virtual colonoscopy or flexible sigmoidoscopy.

The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial is a multi-center, two-armed randomized trial, sponsored by the National Cancer Institute, of screening tests for prostate, lung, colorectal and ovarian cancers. Ten centers across the U.S. recruited participants between November 1993 and July 2001. Data were collected until December 31, 2009. One objective of the trial is evaluating the effectiveness of the screening with flexible sigmoidoscopy on mortality from colorectal-cancer by comparing with usual-care. Prorok *et al.* reported further details about this trial [62].

The original data consist of $154,897$ individuals aged 55 to 74 years. They were randomly assigned to either the usual-care (control, $N = 77,453$) group or the screening with flexible sigmoidoscopy (intervention, $N = 77,444$) group. For the intervention group, subjects were offered the screening at baseline and 3 or 5 years later. Among them, 187 participants who left study, dead, diagnosed cancer, or removed organ before the first intervention and 4 participants who have 0 day from trial entry (randomization) to the last follow-up were ignored. Thus, we only considered $154,706$ individuals in this analysis. The screening assignment (Z) can be used as an instrument and the treatment variable (D) becomes

an indicator of screen received at baseline. Table 3.3 presents descriptive statistics for the baseline characteristics of the participants by the instrumental variable (Z) and the treatment variable (D). In particular, the main factor of interest is the flexible sigmoidoscopy screening. We also consider other risk factors such as: sex, age (year), family history of any cancer, family history of colorectal cancer, colorectal polyps, colon comorbidities, and diabetes.

Table 3.3: Characteristics of the Study Participants

Characteristics	Control ($Z = 0$) $N = 77449$	Intervention ($Z = 1$) $N = 77257$	Not Screened ($D = 0$) $N = 90056$	Screened ($D = 1$) $N = 64650$
	Number of Participants (%)			
Sex				
Male	38340 (49.5)	38229 (49.5)	43529 (48.3)	33040 (51.1)
Female	39109 (50.5)	39028 (50.5)	46527 (51.7)	31610 (48.9)
Age§				
	62.60 (5.37)	62.59 (5.39)	62.65 (5.39)	62.52 (5.33)
Age Level				
55-59 yr	25838 (33.4)	25789 (33.4)	29902 (33.2)	21725 (33.6)
60-64 yr	23767 (30.7)	23736 (30.7)	27451 (30.5)	20052 (31.0)
65-69 yr	17473 (22.6)	17402 (22.5)	20352 (22.6)	14523 (22.5)
70-74 yr	10371 (13.4)	10330 (13.4)	12351 (13.7)	8350 (12.9)
Family History of Any Cancer				
No	32742 (42.3)	33327 (43.1)	37798 (42.0)	28271 (43.7)
Yes	41305 (53.3)	41971 (54.3)	47137 (52.3)	36139 (55.9)
Unknown	3402 (4.4)	1959 (2.5)	5121 (5.7)	240 (0.4)
Family History of Colorectal Cancer				
No	64504 (83.3)	65203 (84.4)	73997 (82.2)	55710 (86.2)
Yes †	7320 (9.5)	7627 (9.9)	8331 (9.3)	6616 (10.2)
Possibly ‡	1925 (2.5)	2108 (2.7)	2262 (2.5)	1771 (2.7)
Unknown	3700 (4.8)	2319 (3.0)	5466 (6.1)	553 (0.9)
Colorectal Polyps				
No	68690 (88.7)	69910 (90.5)	78705 (87.4)	59895 (92.6)
Yes	4947 (6.4)	5185 (6.7)	5739 (6.4)	4393 (6.8)
Unknown	3812 (4.9)	2162 (2.8)	5612 (6.2)	362 (0.6)
Colon Comorbidities				
No	72351 (93.4)	73786 (95.5)	82905 (92.1)	63232 (97.8)
Yes	1052 (1.4)	1090 (1.4)	1247 (1.4)	895 (1.4)
Unknown	4046 (5.2)	2381 (3.1)	5904 (6.6)	523 (0.8)
Diabetes				
No	68028 (87.8)	69371 (89.8)	77773 (86.4)	59626 (92.2)
Yes	5699 (7.4)	5810 (7.5)	6776 (7.5)	4733 (7.3)
Unknown	3722 (4.8)	2076 (2.7)	5507 (6.1)	291 (0.5)

§ denotes a continuous variable. Mean and standard deviation are reported.
† indicates colorectal-cancer family history in immediate family member.
‡ indicates colorectal-cancer family history in relatives or unclear cancer type.

This trial assumed that at least 85% compliance with screening in the intervention group and no more than 15% contamination among participants in the usual-care group [62, 63].

However, note that individuals who are assigned to the usual-care group ($Z = 0$) do not have records of treatment by assuming $P(D_0 = 0|X) = 1$. It is a special case referred as a perfect exclusion of the control group from the treatment [42]. The Assumption 3.2.2 holds trivially and it implies that only two possible types of compliance strata, compliers and never-takers, exist. In this case, we have

$$\lambda(T|D_1 > D_0, X, D = 1) = \lambda(T_1|D_1 = 1, X, Z = 1) = \lambda(T_1|D = 1, X)$$

and

$$\begin{aligned}\lambda(T|D_1 > D_0, X, D = 0) &= \lambda(T_0|D_1 = 1, X, Z = 0) \\ &= \lambda(T_0|D_1 = 1, X, Z = 1) = \lambda(T_0|D = 1, X).\end{aligned}$$

Thus, the proposed estimator describes the effect of the treatment for the treated given X.

The observed time from trial entry (randomization) to death for participants known to be dead, or to trial exit for participants not known to be dead is given in days. By dividing 365.25, the observed times (days) are transformed into years. The censoring indicator, denoted by Δ, is equal to 1 if the individual died only due to colorectal-cancer and 0 otherwise.

Table 3.4: Analysis Results without Covariates

Data	N	$\hat{p}_1$	Proposed IV Model	Cox Model	Cox Model (ITT)
(Subgroup)				Estimates (S.E)	
Total	154706	0.84	-0.427 (0.100)***	-0.442 (0.088)***	-0.343 (0.083)***
Family History of Any Cancer	83276	0.86	-0.294 (0.125)*	-0.237 (0.114)*	-0.258 (0.111)*
Family History of Colorectal Cancer	14947	0.87	-0.105 (0.260)	-0.010 (0.241)	-0.097 (0.239)
Colorectal Polyps	10132	0.85	0.335 (0.369)	0.315 (0.305)	0.288 (0.309)
Colon Comorbidities	2142	0.82	-2.127 (1.060)*	-1.772 (1.061)†	-2.125 (1.062)*
Diabetes	11509	0.81	-0.603 (0.412)	-1.036 (0.311)***	-0.335 (0.253)

p-value: *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05, † ≤ 0.1

Table 3.4 shows the estimation results with different subgroup data sets by risk factors of colorectal-cancer. Also, three models such as the proposed IV model, the Cox model with D and the Cox model with Z (ITT) were applied for each data set. For each data set, the

proportion of compliers p_1 is estimated by using the following formula in [41]: $p_1 = p_{11} - p_{01}$, where p_{11} is the proportion of participants who would receive treatment ($D = 1$) if assigned treatment ($Z = 1$) and p_{01} is the proportion of participants who would receive treatment ($D = 1$) if assigned treatment ($Z = 0$). This estimated complier proportion could explain each others' distances between three different estimates in Table 3.4. When $\hat{p}_1$ has a relatively high value, there is no big difference within the values of hazard ratio by exponentiating the parameter estimates. For example, the subgroup of individuals who have a colorectal-cancer family history has the highest $\hat{p}_1$ value with 0.87 and the corresponding three exponentiated values are 0.900, 0.990 and 0.908, respectively. In contrast, the subgroup including subjects who have a diabetes has the lowest $\hat{p}_1$ value with 0.81. The IV estimate has the middle value between the estimates of Cox model with D and ITT. In the Cox model, the hazard ratio is $\exp(-1.036) = 0.355$. A hazard ratio value smaller than 1 says that an increase in one unit for that particular variable, will decrease the rate of experiencing an event (end point) throughout the observation period. That is, a screened individual who has not yet experienced death by colorectal-cancer or not yet censored by a certain time has the decreased rate of being dead caused by colorectal-cancer at the next point in time by $100\% - 35.5\% = 64.5\%$ compared to an individual who did not get the screening. Also, there is a statistically significant association between the screening and mortality caused by colorectal-cancer with p-value 0.001 which is less than 0.05. In the ITT analysis, the corresponding hazard ratio is $\exp(-0.335) = 0.715$. Similarly, colorectal-cancer mortality rate is decreased in the treatment assigned group with $100\% - 71.5\% = 28.5\%$ compared to the usual-care assigned group. It is not statistically significant with p-value 0.185 which is greater than 0.05. The hazard ratio of ITT is nearly two times greater than one of the Cox model, and the statistical significances are different. The IV estimate -0.603 lies between the two estimates of the naive models. The hazard ratio of $\exp(-0.603) = 0.547$ corresponds to a decreased rate of colorectal-cancer mortality by $100\% - 54.7\% = 45.3\%$ as D is altered from the usual-care to the screening within the population subset that has received

screening. Also, the results of subset that contains individuals who have colon comorbidities show difference in parameter estimates. Thus, we will further examine these two subgroups with considering covariates.

Table 3.5: Analysis Results with Covariates

	Colon Comorbidities		
Covariate	Proposed IV Model	Cox Model	Cox Model (ITT)
D	-2.148 (1.062)*	-1.696 (1.061)	-2.143 (1.061)*
Gender (Female)	2.013 (1.153)†	1.855 (1.063)†	1.910 (1.064)†
Age	0.224 (0.095)*	0.218 (0.076)**	0.220 (0.074)*
	Diabetes		
D	-0.641 (0.409)	-1.076 (0.311)***	-0.361 (0.253)
Gender (Female)	-0.846 (0.597)	-0.522 (0.270)†	-0.475 (0.270)†
Age	0.068 (0.043)	0.097 (0.024)***	0.096 (0.024)***

Estimates (S.E)
p-value: *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05, † ≤ 0.1

Table 3.5 shows the parameter estimation results with considering two covariates such as gender and age. Note that the both naive Cox models with four categories of age variable do not converge in subgroup of colon comorbidities, so the continuous age variable is used. After accounting for gender and age, there is no statistically significant association between D and colorectal cancer mortality, but there is a statistically significant association between Z and colorectal cancer mortality in colon comorbidities subgroup. Conditioning on the covariates, the average treatment hazard ratio for the treated is $\exp(-2.148) = 0.117$. Consistent with the ITT analysis results, it is statistically significant. All IV estimates in the subset of colon comorbidities are similar to the ITT estimates. However, the IV estimates in the subset of diabetes are different from the others.

3.5 Discussion

Non-compliance is a common issue in randomized clinical trials or observational studies. The intention-to-treat analysis ignores non-compliers and the analysis which only focuses on treatment (D) cannot provide reliable inferences regarding the true effect of the treatment. To overcome the problem, IV methods can be applied. However, there are limitations in using IV in recurrent event data analysis. Thus, we suggested an extended Abadie's IV method into AG model by using weighting scheme. We could use the similar arguments with AG model to develop asymptotic properties of the proposed estimators. Note that we applied this method into PLCO data, which does not include recurrent event data.

However, there are limitations of this method, that is, we could not use full observed data. For example, the PLCO data have several time variant covariates including the exposure, and the IV also could be changing by time in longitudinal data. Since we only consider a binary IV and a binary treatment at baseline in this Chapter, incorporating time variant covariates and time varying IV into this model will be the future research.

Acknowledgements

I would like to express my special thanks of gratitude to Jerome Mabie, Tom Riley, Ryan Nobel and Josh Rathmell, Information Management Services (IMS) Inc, for supporting and managing the PLCO data. I would also like to thank Dr. Stuart G. Baker, National Cancer Institute of USA, for kindly introducing the IMS team for this research.

CHAPTER 4: PARAMETRIC LIKELIHOOD INFERENCE FOR INTERVAL CENSORED AND LEFT TRUNCATED COMPETING RISKS DATA

4.1 Introduction

Hudgens *et al.* [64] developed parametric modeling of the cumulative incidence function for interval censored competing risks data. We extend their parametric models to additionally account for left truncation by changing time scale. That is, instead of using the elapsed time from the study entry to failure time, we applied the time adjusted for age. A full likelihood estimator is still valid under a mixed case interval censoring model and an independent inspection process model. However, a naive likelihood method is shown to be invalid in both settings, while it is not valid only under the independent inspection process model without considering truncation. This extended parametric method is demonstrated via comprehensive simulation studies and is applied to data from the Study of Osteoporotic Fractures to obtain bone mineral density (BMD) testing interval by age as a time scale.

In this chapter, we develop parametric estimation methods for competing risks survival data subject to interval censoring and left truncation. These methods are motivated by clinical practice guidelines for bone mineral density (BMD) screening to identify and treat osteoporosis (very low bone density) for fracture prevention in postmenopausal women. Clinical practice guidelines agree that women aged 65 and older should receive BMD testing to detect and treat osteoporosis [65–67]. However, no standard BMD screening interval has been recommended. To guide decisions about the interval of BMD tests for women aged 65 and older, Gourlay *et al.* [68] conducted a competing risk analysis of BMD screening intervals using data from the Study of Osteoporotic Fractures (SOF) observational study of

BMD and fractures in women aged 65 and older at study entry. For four different risk strata (Normal BMD, Mild osteopenia, Moderate osteopenia, Advanced osteopenia) reflecting an increasing fracture risk correlating with baseline BMD testing scores, they provided cumulative incidence curves of osteoporosis by applying Hudgens' parametric method [64] with the standard approach using the elapsed time from the baseline to the occurrence of the event as the time scale. Interval censoring arose, because BMD was measured intermittently. Thus, the time to osteoporosis was not directly observed, but was only known up to some intervals. Incident hip and clinical vertebral fractures and pharmacologic treatment were considered to be competing risks, because individuals who have hip or clinical vertebral fractures are always treated regardless of their BMD level, and treated individuals no longer need BMD risk stratification.

Unlike the work done in the previous study, we seek to estimate the cumulative incidence curves of osteoporosis with age as a time scale instead of using time in the study. This creates a left truncation issue because subjects entered into the study at different ages and the necessary BMD measurements were not initiated until 2 years after baseline enrollment. For example, women entering the study at age 65 who had the event (osteoporosis) before age 67 (when their BMD measurements began) were not considered as incident osteoporosis cases. That is, left truncation occurs when the subjects have been at risk before beginning BMD measurements in this study (or before study entry in other cases). Lamarca *et al.* [69] indicated that the usage of age as time is more appropriate for survival analysis of the elderly population where the goal is to describe the risk factors that modify the hazard of the failure after a specific age, 65 years in SOF data. Cain *et al.* [70] also supported that bias can be increased and that standard errors can be underestimated by ignoring delayed entries. Thus, bias can be reduced when analyses account for left truncation, although the results are unstable when there are higher levels of truncation.

It is a common interest to estimate cumulative incidence function (CIF), which is the

probability of a specific event occurrence by time t, in the presence of other competing events. Hudgens *et al.* [71] computationally derived a nonparametric maximum likelihood estimator (NPMLE) of the CIFs for competing risks data subject to interval censoring and truncation by generalizing Turnbull's estimator [72]. The NPMLE and the naive estimator of the CIF for current status data had been studied by Jewell *et al.* [73] and Groeneboom *et al.* [74, 75]. Li and Fine [76] applied kernel smoothing to estimate the NPMLE of the CIF and the cause-specific hazard function (CSHF) in current status data with competing risks. Since the nonparametric estimation require intense computation [64, 74], parametric models can be preferred in this case.

Based on Jeong and Fine [77] which suggested parametric estimation method of the CIF for right censored competing risks data, Hudgens *et al.* [64] extended the parametric models for allowing interval censoring with considering both full maximum likelihood estimators (MLEs) and naive estimators. In this Chapter 4, we conduct competing risks analysis additionally accounting for left truncation by extending the parametric estimation method from Hudgens *et al.* [64]. The remaining sections are organized as follows. We begin by introducing notations and the proposed parametric modeling, and its estimation and asymptotic properties are given in Section 4.2. In Section 4.3, we present simulation studies conducted to examine the feasibility of the extended method, and results of the SOF data analysis are reported in Section 4.4.

4.2 Inference

4.2.1 Notations and Parametric Modelling

Let $K \in \{1, \ldots, n_K\}$ denote the cause of failure. An individual can only experience one of n_K distinct mutually exclusive competing causes. Let T denote a continuous random variable representing the time of failure, which is only known up to some interval. The CIF

for type k failure is defined as $F_k(t) = Pr[T \leq t, K = k]$, i.e., the cumulative probability of type k failure in the presence of other competing events. It is also expressed in terms of CSHF, such that $F_k(t) = \int_0^t S(u)\lambda_k^*(u)$, where $S(t) = \exp\{-\int_0^t \sum_{k=1}^{n_K} \lambda_k^*(u)du\}$ is the overall survival function and $\lambda_k(t) = \lim_{dt\to 0}\{Pr(t \leq T < t + dt, K = k|t \leq T)/dt\}$ is known to be as CSHF of the type k.

To overcome complication due to interval censoring, the direct parameterization of the CIF without covariates by Jeong and Fine [77] is applied. Also, it gives natural interpretation concerning the probability of an event of interest. A parametric model $F_k(t; \Theta_k)$ is separately specified for each CIF with distinct Θ_k for each k. We also adopt the Gompertz model in Jeong and Fine [77], which is defined by

$$F_k(t; \Theta_k) = 1 - \exp[\beta_k\{1 - \exp(\alpha_k t)\}/\alpha_k] \tag{4.2.1}$$

with $\Theta_k = (\alpha_k, \beta_k)$ where $\alpha_k < 0$ and $\beta_k > 0$ so that the function (4.2.1) is an improper distribution function. That is, the CIF satisfies $\lim_{t\to\infty} F_k(t) < 1$ when $n_K > 1$ and the probability of experiencing each cause has positive value. The corresponding hazard function is given by

$$\lambda_k(t) = \beta_k \exp(\alpha_k t). \tag{4.2.2}$$

This model satisfies

$$0 < F_k(t; \Theta_k) < 1 \text{ for all } t > 0 \text{ and } k = 1, \ldots, n_K, \text{ and} \tag{4.2.3}$$

$$F_k(t; \Theta_k) \text{ is monotone increasing function of } t \text{ for } k = 1, \ldots, n_K. \tag{4.2.4}$$

Similar to Hudgens *et al.* [64], we consider two observation processes such as the mixed case and independent inspection process (IIP) models by revisiting notations. Additionally, we consider that the T is left truncated at V_0 with a time origin that is the same for all failure types. If the time origin is smaller than the minimum value of V_0, then the estimated CIF is used for extrapolation i.e. predicting the response to an input which lies outside of the range

of the values of the observed intervals used to fit the model. Conversely, when one is bigger, we will assume that there are no event between the corresponding V_0's and the time origin by ignoring the periods. The former will be considered and then we need to consider the conditional distribution of T given $T > V_0$. Let $V = (V_0, V_1, \ldots, V_M)$ be the vector of ordered observation times where M is the random number of observation times for an individual, $V_0 \geq 0$ and $V_{M+1} = \infty$ such that $V_{l-1} < V_l$ for $l = 1, \ldots, M+1$. Define $\Delta_{kl} = 1(V_{l-1} < T \leq V_l, K = k)$ for $k = 1, \ldots, n_K$ and $l = 1, \ldots, M$. That is, Δ_{kl} equals 1 if a subject has an event of type k during $(V_{l-1}, V_l]$ and 0 otherwise. Let $\Delta_{M+1} = 1 - \sum_{k=1}^{n_K} \sum_{l=1}^{M} \Delta_{kl}$. When $\Delta_{M+1} = 1$, the event type is unknown and right censored. Instead of observing (T, K) directly, we observe copies of $Y = (M, V, \Delta)$ where $\Delta = (\Delta_{11}, \ldots, \Delta_{1M}, \Delta_{21}, \ldots, \Delta_{n_K M}, \Delta_{M+1})$. Under the mixed case interval censoring model, $(M, V) \perp\!\!\!\perp (T, K)$ is assumed. It implies that the observation process is determined independently of the failure time and the cause of failure.

For $l = 1, 2, \ldots$, define the history of observation times and failure information by $H_l = (V_1, \ldots, V_{l-1}, \Delta_{11}, \ldots, \Delta_{n_K 1}, \ldots, \Delta_{1,l-1}, \ldots, \Delta_{n_K, l-1} | V_0)$, where $H_1 = V_0$. Under the IIP model, it is assumed $V_l \perp\!\!\!\perp (T, K) | H_l$ implying that the next observation times is independent with the failure time and cause given the history of observation times and failure information. That is, the IIP stops if a failure is detected, such that $\Delta_{jl} = 0$ for all $l < M$ and $j \in \{1, \ldots, n_K\}$. Thus, the IIP model is more appropriate when future observation times depend on the history of the observed data [78].

4.2.2 Estimation and Asymptotic Properties

Let $Y_1, \ldots, Y_n$ be a random sample of n independent and identically distributed copies of Y. Therefore, the corresponding log likelihood functions for $Y_1, \ldots, Y_n$ under the mixed case interval censoring model and the IIP model are following

$$\log L(\Theta) = \sum_{i=1}^{n} \log \ell(Y_i; \Theta), \tag{4.2.5}$$

where Θ is the vector consisting of elements of $\Theta_1 \cup \ldots \cup \Theta_{n_K}$ and

$$\ell(Y_i;\Theta) = \prod_{k=1}^{n_K}\prod_{l=1}^{M}\left\{\frac{F_k(v_l;\Theta_k) - F_k(v_{l-1};\Theta_k)}{1-\sum_{k=1}^{n_K}F_k(v_0;\Theta_k)}\right\}^{\Delta_{kl}}\left\{\frac{1-\sum_{k=1}^{n_K}F_k(v_m;\Theta_k)}{1-\sum_{k=1}^{n_K}F_k(v_0;\Theta_k)}\right\}^{\Delta_{M+1}}. \quad (4.2.6)$$

The full likelihood estimator of Θ, $\hat{\Theta}$, is defined by the value that maximizes (4.2.5) with the assumptions 4.2.3 and 4.2.4. They cannot be separately estimated by the failure type k. Under certain regularity conditions, the maximum likelihood estimates of Θ have consistency and asymptotic normality with mean Θ and covariance matrix Σ_Θ. By taking the negative second derivatives of the log likelihood function (4.2.5) with respect to Θ, the observed information matrices can be computed and inverted to estimate the Σ_Θ. The estimated $F_k(t;\hat{\Theta}_k)$ is also approximately Normal distribution with mean $F_k(t;\Theta_k)$ and covariance matrix $\Sigma(t)$ by using the multivariate delta method. The $\Sigma(t)$ can be estimated by

$$\widehat{var}\{F_k(t;\hat{\Theta}_k)\} = \hat{\Sigma}(t) = \left(\frac{\partial F_k(t;\Theta_k)}{\partial\Theta_k}\right)\hat{\Sigma}_{\hat{\Theta}}\left(\frac{\partial F_k(t;\Theta_k)}{\partial\Theta_k}\right)'\Bigg|_{\Theta=\hat{\Theta}} \quad (4.2.7)$$

, for $k = 1,\ldots,n_K$. The $\hat{\Sigma}_{\hat{\Theta}}$ is the inverse of the observed Fisher information, evaluated at the parameter estimates $\hat{\Theta}$. A pointwise $(1-\alpha)\%$ confidence interval (CI) for $F_k(t;\hat{\Theta}_k)$ at time t is given by

$$F_k(t;\hat{\Theta}_k) \pm z_{1-2/\alpha}\sqrt{\hat{\Sigma}(t)}, \quad (4.2.8)$$

where z_q is the q quantile of the standard normal distribution. Note that the estimation is only valid on the support of the observation time. To test differences between the probability of a particular failure type by time t between two subgroups, i.e., the null hypothesis is $H_0 : F_k^1(t;\Theta_k^1) = F_k^2(t;\Theta_k^2)$, where $F_k^g(t;\Theta_k^g)$ denote the CIF for a failure of type k and Θ_k^g are corresponding parameter in subgroup $g = 1, 2$. The Wald type test statistics is following:

$$F_k^1(t;\hat{\Theta}_k^1) - F_k^2(t;\hat{\Theta}_k^2)/\sqrt{\widehat{var}\{F_k^1(t;\hat{\Theta}_k^1)\} + \widehat{var}\{F_k^2(t;\hat{\Theta}_k^2)\}} \quad (4.2.9)$$

which follows a standard Normal distribution under the null hypothesis.

Note that in addition to this full likelihood method, Hudgens *et al.* [64] considered a naive likelihood for estimating the CIF by using a reduced data. Since the naive estimator

enables separate estimation of models for each cause, unlike the MLEs where all models are fitted simultaneously. Thus, it is computationally easier to estimate than the MLEs. However, we need to estimate all Θ_k's for $k = 1, \ldots, n_K$ to calculate $Pr[T > v_0]$. Therefore, the naive likelihood method cannot be applied.

4.3 Simulation Studies

We adopted simulation settings from Hudgens *et al.* [64]. There are two causes of failure, i.e., $n_K = 2$. The event type k and the failure time T given k are generated by using $Pr[K = k; \Theta_k] = 1 - \exp(\beta_k/\alpha_k)$ and $F_k(t; \Theta_k) = Pr[T \leq t | K = k; \Theta_k] Pr[K = k; \Theta_k]$, where $\Theta_k = (\alpha_k, \beta_k)$ for $k = 1, 2$. The left truncation time V_0 is randomly drawn from a uniform distribution between 0 and u. Two different values 5 and 10 for u have different truncation rates 0.16 and 0.29, respectively. The truncation rate for each iteration is obtained by taking an average of cases when the event times are less than the truncation time point, V_0, and then take average of the truncation rates over the iterations. The following observation times $V_1 < \ldots < V_7$ are randomly generated with $V_1 = V_0 + \text{Unif}\,(3, 5), V_2 = V_0 + \text{Unif}\,(7, 9), \ldots, V_6 = V_0 + \text{Unif}\,(23, 25)$, and $V_7 = V_0 + 28$. We start with an vector, $(-0.01, 0.01, -0.01, 0.01)$, for the initial values of $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. We consider sample size $n = 500, 1000, 2000$ and each scenario is replicated $1,000$ times.

We report bias (BIAS), empirical standard deviation (ESD), defined by the sample standard deviation of the replicated estimates, average of the replicated standard deviation estimates (TSD), and empirical coverage probability (CP) at a 0.95 nominal level in Table 4.1. The columns of LT(5) and LT(10) indicate the results of accounting for two different left truncation proportions. To check appealing features by considering left truncation, the results without adjusting left truncation are also presented in last three columns denoted by FL. All estimators are approximately unbiased. The estimates with considering left truncation have biases closer to 0 than the ones without considering left truncation, and the

Table 4.1: Simulation Results with $(\alpha_1, \beta_1, \alpha_2, \beta_2) = (-0.058, 0.0093, -0.035, 0.067)$

		LT(5)			LT(10)			FL		
	n	500	1000	2000	500	1000	2000	500	1000	2000
α_1	BIAS	0.008	-0.060	-0.020	0.004	-0.068	-0.020	-0.032	-0.096	-0.046
	ESD	1.839	1.318	0.940	1.840	1.283	0.954	1.504	1.038	0.750
	TSD	1.852	1.304	0.92	1.890	1.331	0.939	1.607	1.132	0.798
	CP	0.949	0.955	0.951	0.959	0.968	0.941	0.962	0.977	0.965
β_1	BIAS	0.019	0.018	0.006	0.027	0.024	0.008	0.012	0.017	0.007
	ESD	0.251	0.180	0.128	0.291	0.203	0.152	0.197	0.137	0.099
	TSD	0.253	0.179	0.126	0.297	0.21	0.147	0.198	0.141	0.099
	CP	0.939	0.949	0.938	0.941	0.955	0.935	0.946	0.970	0.941
α_2	BIAS	-0.002	-0.007	-0.010	-0.008	0.018	0.005	0.007	0.004	-0.003
	ESD	0.780	0.559	0.414	0.801	0.582	0.419	0.691	0.469	0.332
	TSD	0.800	0.566	0.400	0.825	0.582	0.412	0.692	0.489	0.346
	CP	0.959	0.949	0.949	0.956	0.951	0.946	0.963	0.955	0.964
β_2	BIAS	0.014	0.015	0.021	0.027	-0.001	0.008	-0.001	0.008	0.009
	ESD	0.739	0.528	0.388	0.873	0.626	0.445	0.607	0.420	0.300
	TSD	0.744	0.526	0.372	0.878	0.618	0.438	0.577	0.410	0.289
	CP	0.951	0.950	0.941	0.940	0.943	0.946	0.946	0.950	0.947

BIAS ($\times 10^2$), ESD ($\times 10^2$), TSD ($\times 10^2$)

bias is decreasing as the fraction of truncation is getting lower. In addition, standard errors are underestimated when left truncation is ignored. These findings are consistent with Cain *et al.* [70] as we mentioned in Section 4.1. The variance estimates using the observed information are similar to the corresponding empirical variances of the estimators, so the empirical coverage probabilities are nearly reached up to the nominal level 95.

4.4 Application: The Study of Osteoporotic Fractures

For clinical applications, bone mineral density (BMD) measurements are converted to T-scores, i.e.

$$\frac{\text{BMD of participant} \quad \text{mean BMD of young reference population}}{\text{SD of BMD of a young female reference population}}.$$

A BMD test can identify osteoporosis, the most important marker of fracture risk which is defined as lowest BMD T-score ≤ -2.50 at one or more of three anatomical sites in the lower spine and hip [79]. T-scores > -2.50 represent lesser degrees of fracture risk, i.e., T-score -1.00 and higher at all hip and spine sites is considered normal or healthy, and lowest T-score between -1.01 and -2.49 indicates osteopenia. The BMD testing interval was defined as the estimated time during which osteoporosis developed in 10% of women to make the transition to osteoporosis from normal BMD or osteopenia at baseline before having a hip or clinical vertebral fracture.

The BMD testing to screen for osteoporosis is recommended for women 65 years of age or older. To determine the BMD testing interval via age scale, we studied 4957 women, aged 67 years or older and recruited between 1986 and 1988, who did not have osteoporosis at baseline and who were followed prospectively for up to 15 years. The follow-up period included study examinations at year 2 (1989-1990), year 6 (1992-1994), year 8 (1995-1996), year 10 (1997-1999), and year 16 (2002-2004), which represented the time period during which BMD and fracture were followed concurrently. Thus, interval censoring occurs since a random variable of interest, i.e., time to osteoporosis, is known only to lie within an interval between BMD examinations at intermittent study visits instead of being observed exactly. Also, our goal is to estimate the cumulative incidence curves of osteoporosis with age as time scale, which accompanies left truncation. Because individuals who have already experienced osteoporosis or having a hip or clinical vertebral fracture were not included in the study and they entered into the study at different ages. Therefore, we need to account for the probability that the osteoporosis has not been occurred before the entry age years. Thus, age minus 65 was taken as the time scale. Similar to the Gourlay *et al.* [68], the analysis included the 513 women who made the transition from normal BMD to osteopenia and had at least one subsequent examination with BMD recorded. In other words, total $4957 + 513 = 5470$ records are used by assuming independence. We consider two failure types ($n_K = 2$) such as osteoporosis and others including incident hip or clinical vertebral fractures and the first reported use of a

Food and Drug Administration (FDA) approved agent for the treatment of osteoporosis.

Table 4.2: Characteristics of the Study Participants

		Failure Types		
Characteristics	Total $N = 5470$ (100)	No Event $N = 3704$ (67.7)	Osteoporosis $N = 1224$ (22.4)	Others $N = 542$ (9.9)
Age†				
	72.67 (4.38)	72.64 (4.39)	73.12 (4.41)	71.89 (4.07)
Age Level				
67-74 yr	3838 (70.2)	2610 (68.0)	821 (21.4)	407 (10.6)
≥75 yr	1632 (29.8)	1094 (67.0)	403 (24.7)	135 (8.3)
T-score				
≥-1, normal	1255 (22.9)	1171 (93.3)	10 (0.8)	74 (5.9)
(-1.50, -1), mild osteopenia	1386 (25.3)	1195 (86.2)	64 (4.6)	127 (9.2)
(-2, -1.50], moderate osteopenia	1478 (27.0)	972 (65.8)	309 (20.9)	197 (13.3)
≤-2, advanced osteopenia	1351 (24.7)	366 (27.1)	841 (62.3)	144 (10.7)
BMI (4)				
<18.5, underweight	31 (0.6)	14 (45.2)	15 (48.4)	2 (6.5)
[18.5-25), normal	1940 (35.6)	1149 (59.2)	572 (29.5)	219 (11.3)
[25-30), overweight	2164 (39.7)	1493 (69.0)	451 (20.8)	220 (10.2)
≥30, obese	1310 (24.1)	1029 (78.5)	182 (13.9)	99 (7.6)
BMI (2)				
<30	4160 (76.1)	2675 (64.3)	1042 (25.0)	443 (10.6)
≥30	1310 (23.9)	1029 (78.5)	182 (13.9)	99 (7.6)
Family History of Hip Fracture				
No	4845 (88.6)	3301 (68.1)	1065 (22.0)	479 (9.9)
Yes	625 (11.4)	403 (64.5)	159 (25.4)	63 (10.1)

Number of Participants (%)

† denotes a continuous variable. Mean and standard deviation are reported.

Body mass index (BMI)= $\frac{mass(kg)}{height^2(m^2)} = \frac{mass(lb)}{height^2(in^2)} \times 703$

Table 4.2 presents descriptive statistics of several clinical risk factors' characteristics at the baseline by the failure types. Note that age cannot be used as a covariate. The age range at baseline is from 67 to 91. The corresponding largest V_M is 99, so the range of observation time is [67, 99] in age scale. We applied the Gompertz model with two parameters. Since the Gompertz model is suitable when there is a subset of the population can never experience the event of interest.

Figure 4.1 shows the estimates of the CIFs (solid lines) and the 95% CIs (dotted lines) for the two parameters Gompertz models accounting for BMI (2) and family history of hip fracture at baseline, separately. The grey horizontal dashed line represents the 10% threshold for the transition to osteoporosis. This line intersects each cumulative incidence curve and the corresponding values of x axis would be the estimated testing intervals in age scale between 65 and 99, respectively. The times in age scale for 10% of women without osteoporosis to make the transition to osteoporosis decreased with lower BMI or with family history of hip fracture. The estimates are 68.57 years for women whose BMI less than 30 and 71.26 years for those with obese. By contrast, the estimates for women with family history and for those without family history are very similar with 68.47 and 69.02, respectively. The grey vertical dashed line represents the minimum observed age in this study, 67 years.

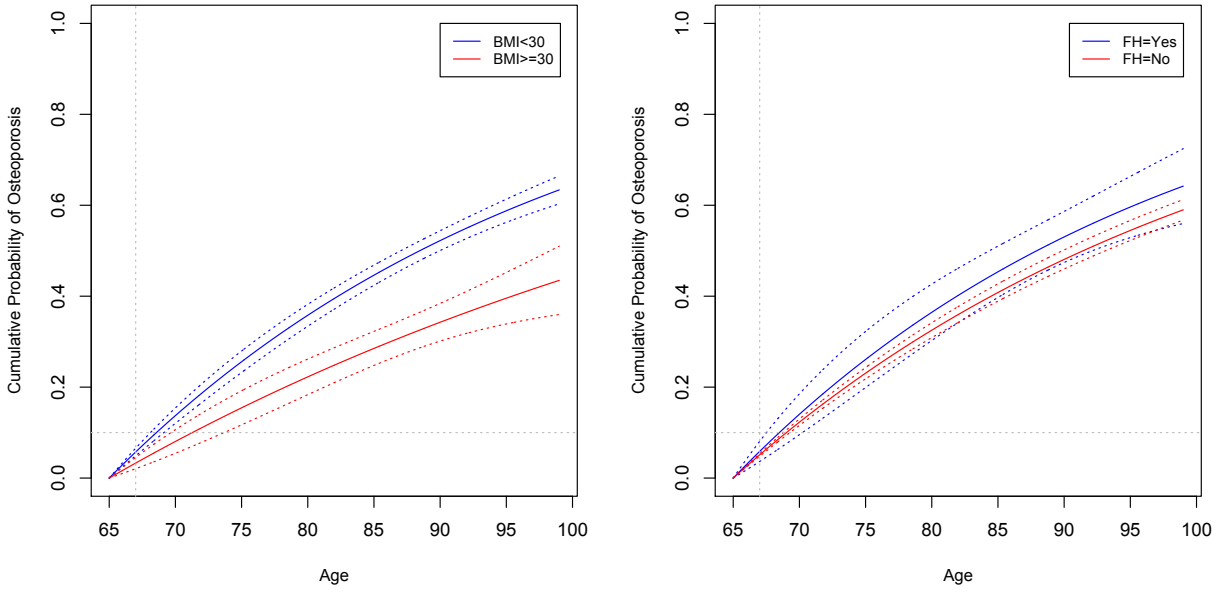

Figure 4.1: Estimated CIFs and CIs for Osteoporosis by BMI (2) and Family History

Then, the Figure 4.2 shows the CIF estimates by considering combination with these two BMI and family history and by adjusting with four categories of BMI instead of two. The estimated ages for 10% of subjects having the transition to osteoporosis are 68.23, 68.61,

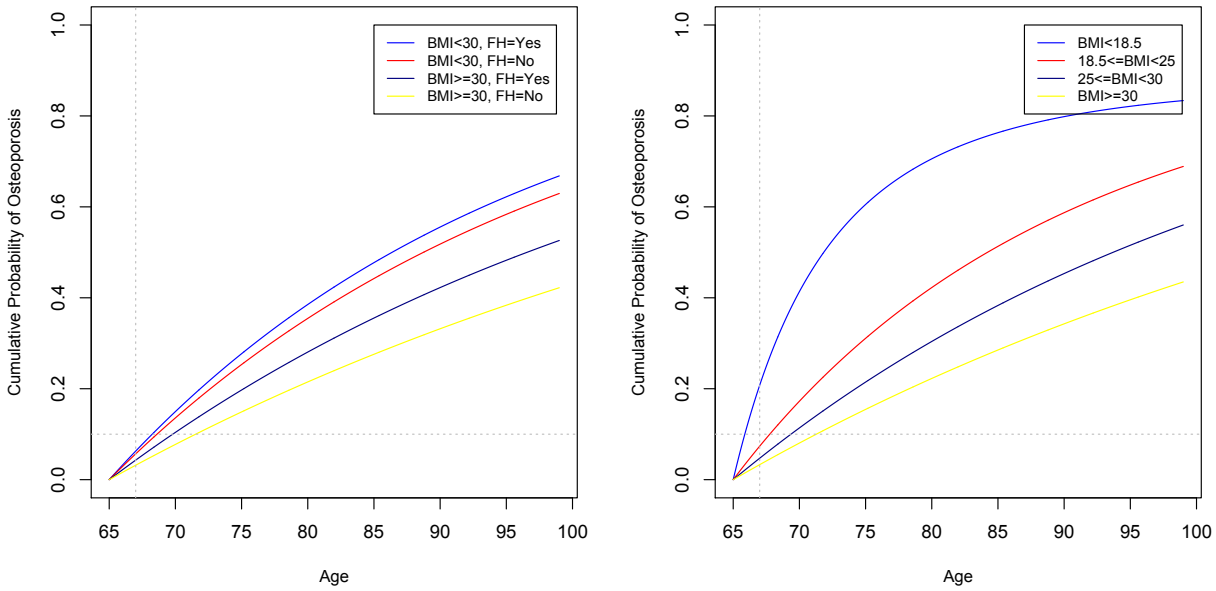

Figure 4.2: Estimated CIFs for Osteoporosis by Combining with BMI (2) and Family History and by BMI (4)

Table 4.3: Estimates of the Osteoporosis Development in 10% of Participants

Subgroup	N	Age	Estimates	95% CI
BMI < 30 & Family History=Yes	476	68.23	0.10	(0.063, 0.136)
BMI < 30 & Family History=No	3684	68.61	0.10	(0.094, 0.106)
BMI ≥ 30 & Family History=Yes	149	69.80	0.10	(0.061, 0.139)
BMI ≥ 30 & Family History=No	1161	71.53	0.10	(0.069, 0.131)
BMI < 18.5	31	65.88	0.10	(0.000, 0.217)
$18.5 \leq$ BMI < 25	1940	67.76	0.10	(0.082, 0.118)
$25 \leq$ BMI < 30	2164	69.36	0.10	(0.091, 0.109)
BMI ≥ 30	1310	71.26	0.10	(0.070, 0.130)

69.80 and 71.53 in the left and 65.88, 67.76, 69.36 and 71.26 in the right with the same order of the legend. The corresponding CIF estimates and CIs are reported in Table 4.3. Note that the number of women who have less than 18.5 BMI is only 31, so it has a wide 95% CI. However, we still can check that the BMI gives more noticeable separation between subgroups than the family history. We might conclude that the BMI is a key variable to explain occurrence of osteoporosis.

Existing analyses of osteoporosis risk assessment tools suggest that our findings are robust and clinically relevant. Ravn *et al.* [80] concluded that thinness, defined by low percentage of body fat, low BMI, or low body weight, is an risk factor for low bone mass and fast bone loss in postmenopausal women. Also, DeLaet *et al.*'s meta-analysis of BMI as a predictor of fracture concluded that low BMI confers a risk of substantial importance for all fractures that is largely independent of age and sex, but dependent on BMD [81]. Systematic reviews of osteoporosis risk assessment tools have demonstrated that simple risk tools perform as well as complicated risk tools to identify postmenopausal women aged 50 and older with osteoporosis [82, 83]. For example, the Osteoporosis Self-Assessment Tool based on age and body weight alone performed as well or better (as demonstrated by the area under the ROC curve) than risk tools including as many as 6 clinical risk factors [84–86]. This is consistent with our finding of improved osteoporosis risk stratification (more distinct separation of cumulative incidence curves for osteoporosis) by using BMI rather than BMI combined with family history of fracture for the baseline predictor.

4.5 Conclusion

The goal of this Chapter is to estimate the cumulative incidence functions of osteoporosis with age as time scale instead of gap time from the beginning of the study to the event. Thus, left truncation has been considered with interval censoring in competing risk setting by extending Hudgens *et al.*'s parametric estimation method [64]. The numerical studies sug-

gested that the proposed estimation performs well when the parametric models are correctly specified, and also bias can be reduced with a relatively lower truncation rate. Additionally, because left truncation of the outcome measure at study entry is common in the data sets used for analyses of many risk assessment tools (not just for osteoporosis), we conclude that our left truncation methods have potential for expanded use in clinical studies of screening tests, especially those based on continuous measures/scores.

However, as we mentioned in Section 4.2, we could not apply the naive estimation method, which can separately estimate for each event type by treating other events as independent censoring events, since we still need to estimate all other parameters for different causes to calculate the truncation probability in denominator. Thus, it could be a future study to utilize the naive likelihood method.

CHAPTER 5: SUMMARY AND FUTURE STUDIES

In this dissertation we have studied about recurrent event data within different circumstances. We focused on recurrence of the same type of events in Chapter 2 and Chapter 3, and we slightly changed it to the competing risks setting allowing different types of events so that the occurrence of event caused by one type may prohibit to observe the other type of events in Chapter 4.

In Chapter 2, we proposed the self-triggering Cox (STC) model for recurrent event data, and compared with the existing methods. Also, to address the non-identifiability of the STC model under the null of no triggering effect, the Bonferroni correction procedure was suggested to adjust for multiple testing at given some values for the non-identifiable parameter. The simulation studies indicate that the adjustment appears to be rather conservative and a test at a single values may provide greater power than the multiple testing approach. In addition, we demonstrated that the extended Cox model may give significant gains in prediction of future events compared to available models.

In Chapter 3, we considered non-compliance issue within recurrent time-to event data by using instrumental variable. Because the intention-to-treat analysis ignores non-compliers and the analysis which only focuses on treatment cannot provide reliable inferences regarding the true effect of the treatment. We suggested to extend Abadie's IV method into Anderson-Gill (AG) model and additionally into Cox model. We showed that the similar arguments with AG model can be used to develop asymptotic properties of the proposed estimators by using weighting scheme. Through PLCO data analysis, we demonstrated the proposed method with Cox proportional hazards model instead of AG model.

In Chapter 4, to estimate the cumulative incidence functions by accommodating left truncation and interval censoring within competing risks data setting, we extended Hudgens *et al.*'s parametric estimation method. The numerical studies suggested that the bias reduced compared to the one from the estimation without considering left truncation. In addition, we applied the extended parametric modeling for analyzing data from the study of osteoporotic fractures.

There are several future researches that would be invoked by the presented works in this dissertation. Firstly, we can generalize the STC model with using different intensity functions, or we could incorporate the number of lag as an additional parameter in the current model. Also, to deal with non-identifiability under the null, more rigorous testing method would be required. Secondly, the suggested IV method has not been verified with time varying covariates and time varying IV, it could be another candidate for the future study. Additionally, we still need to demonstrate the method with AG model by applying it into the real data analysis. Lastly, we could not apply the naive estimators by adding left truncation. Thus, we can propose pairwise likelihood method to utilize the naive likelihood method which makes estimation procedure simpler by allowing separate estimation for each failure type.

APPENDIX I: ADDITIONAL TABLES FOR CHAPTER 2

Table 5.1, 5.2 and 5.3 show the Anderson-Gill (AG) model estimation results when $\theta_1 = (\beta_{1c}, \beta_{2c})' = (0.5, 1.0)'$ and $\theta_2 = \theta_3 = (\beta_{1o}, \beta_{2o})' = (0.3, 0.5)'$ with different p_1 intervals, respectively, and Table 5.4 shows the simulation results with the same setting except β_{2c} to test $H_0 : \beta_{2c} = 0$.

We additionally report Table 5.5, which shows the Cox model estimation results when $\theta_1 = (\beta_{1c}, \beta_{2c})' = (0.5, 1.0)'$ and $\theta_2 = \theta_3 = (\beta_{1o}, \beta_{2o})' = (0.3, 0.5)'$ with different p_1 intervals, respectively.

Table 5.1: Simulation Results (AG model) with $p_1 = [.41, .49]$

Prob.Complier (p_1)		[.41, .49]								
		Known weight			Unknown weight			AG model		
	N	500	1000	2000	500	1000	2000	500	1000	2000
$X(\beta_{1c})$	MEAN	0.517	0.503	0.500	0.517	0.500	0.500	0.459	0.457	0.460
	MED	0.502	0.498	0.496	0.496	0.499	0.498	0.458	0.456	0.461
	ESD	0.195	0.131	0.089	0.193	0.134	0.088	0.067	0.046	0.033
	TSD	0.223	0.136	0.095	0.223	0.144	0.091	0.062	0.044	0.031
	CP	97.7	97.3	96.1	98.3	97.5	95.9	86.7	82.2	72.3
$D(\beta_{2c})$	MEAN	1.020	1.014	1.004	1.007	1.009	1.002	0.629	0.627	0.627
	MED	1.020	1.014	1.003	1.008	1.009	1.000	0.625	0.628	0.627
	ESD	0.157	0.115	0.074	0.152	0.110	0.070	0.066	0.049	0.033
	TSD	0.200	0.115	0.077	0.170	0.112	0.074	0.062	0.044	0.031
	CP	97.3	95.9	95.9	96.5	95.0	95.3	0	0	0
Outlier/No convergence		29	10	2	25	12	5	0	0	0

Table 5.2: Simulation Results (AG model) with $p_1 = [.51, .67]$

Prob.Complier (p_1)		[.51, .67]								
		Known weight			Unknown weight			AG model		
	N	500	1000	2000	500	1000	2000	500	1000	2000
$X(\beta_{1c})$	MEAN	0.500	0.505	0.500	0.502	0.504	0.501	0.516	0.577	0.548
	MED	0.491	0.501	0.500	0.491	0.502	0.498	0.514	0.576	0.550
	ESD	0.147	0.099	0.065	0.156	0.100	0.064	0.066	0.046	0.044
	TSD	0.154	0.099	0.068	0.159	0.096	0.068	0.062	0.043	0.031
	CP	97.2	94.6	95.6	97.3	95.6	95.7	92.0	57.4	57.7
$D(\beta_{2c})$	MEAN	1.008	1.004	1.003	1.006	1.001	1.001	0.680	0.712	0.695
	MED	1.009	1.000	1.003	1.006	0.996	1.001	0.681	0.712	0.695
	ESD	0.111	0.077	0.053	0.106	0.075	0.051	0.064	0.044	0.035
	TSD	0.121	0.079	0.054	0.113	0.073	0.053	0.060	0.042	0.030
	CP	96.4	95.5	94.3	96.7	94.8	95.4	0.3	0	0
Outlier/No convergence		4	2	4	7	3	2	0	0	0

Table 5.3: Simulation Results (AG model) with $p_1 = [.63, .84]$

Prob.Complier (p_1)		[.63, .84]								
		Known weight			Unknown weight			AG model		
	N	500	1000	2000	500	1000	2000	500	1000	2000
$X(\beta_{1c})$	MEAN	0.503	0.500	0.502	0.504	0.500	0.502	0.567	0.567	0.569
	MED	0.499	0.498	0.502	0.499	0.497	0.503	0.566	0.565	0.568
	ESD	0.126	0.081	0.059	0.117	0.081	0.059	0.069	0.048	0.032
	TSD	0.124	0.080	0.056	0.120	0.081	0.056	0.063	0.044	0.031
	CP	96.3	95.8	94.2	95.7	95.5	94.1	79.6	66.1	40.6
$D(\beta_{2c})$	MEAN	1.002	1.000	0.999	1.001	0.998	0.999	0.884	0.881	0.882
	MED	1.001	1.001	1.000	1.000	1.001	1.001	0.885	0.882	0.883
	ESD	0.086	0.057	0.041	0.084	0.056	0.041	0.065	0.045	0.032
	TSD	0.086	0.059	0.041	0.086	0.059	0.041	0.060	0.042	0.030
	CP	95.7	96.2	95.2	95.7	96.3	94.6	50.9	21.9	3.3
Outlier/No convergence		3	0	0	3	1	0	0	0	0

Table 5.4: Simulation Results for Hypothesis Testing $H_0 : \beta_{2c} = 0$ with $(\beta_{1c}, \beta_{1o}, \beta_{2o}) = (0.5, 0.3, 0.5)$

		$\beta_{2c} = 0$			$\beta_{2c} = 0.1$			$\beta_{2c} = 0.2$		
	N	500	1000	2000	500	1000	2000	500	1000	2000
Prob.Complier (p_1)	Method	Type I error			Power			Power		
[.41, .49]	Known weight	2.29	3.83	4.54	7.56	10.91	21.45	17.79	35.47	60.80
	Unknown weight	2.97	4.04	4.41	8.64	12.46	20.90	20.95	37.51	63.95
	AG model	100	100	100	100	100	100	100	100	100
[.51, .67]	Known weight	3.03	3.90	4.20	12.02	18.33	33.80	35.05	58.33	87.99
	Unknown weight	2.63	4.30	4.70	12.69	18.88	32.83	36.71	59.62	86.86
	AG model	99.2	100	100	99.7	100	100	99.9	100	100
[.63, .84]	Known weight	5.73	3.32	5.20	15.68	30.58	47.30	50.10	81.80	97.90
	Unknown weight	6.14	3.31	5.32	16.98	30.95	47.55	51.92	81.30	98.50
	AG model	61.3	88.9	99.5	89.1	99.1	100	98.1	100	100

Table 5.5: Simulation Results (Cox model) with $(\beta_{1c}, \beta_{1o}, \beta_{2c}, \beta_{2o}) = (0.5, 0.3, 1.0, 0.5)$

Prob.Complier (p_1)		[.41, .49]				
	N	500	1000	2000	4000	8000
$X(\beta_{1c})$	MEAN	0.530	0.538	0.512	0.509	0.505
	MED	0.519	0.514	0.500	0.501	0.507
	ESD	0.440	0.348	0.241	0.131	0.088
	TSD	1.054	0.488	0.304	0.272	0.129
	CP	96.0	96.7	96.3	96.0	95.8
$D(\beta_{2c})$	MEAN	1.102	1.099	1.057	1.029	1.012
	MED	0.977	1.015	1.003	1.017	1.000
	ESD	0.595	0.462	0.335	0.197	0.145
	TSD	0.826	0.523	0.312	0.235	0.199
	CP	91.4	93.3	94.6	95.5	95.5
Outlier/No convergence		223	133	87	50	28
Prob.Complier (p_1)		[.51, .67]				
$X(\beta_{1c})$	MEAN	0.516	0.519	0.505	0.498	0.501
	MED	0.506	0.511	0.507	0.499	0.501
	ESD	0.289	0.217	0.123	0.086	0.057
	TSD	0.430	0.281	0.157	0.099	0.078
	CP	96.7	97.0	96.5	96.3	94.9
$D(\beta_{2c})$	MEAN	1.092	1.050	1.014	1.002	0.999
	MED	1.022	1.011	0.996	0.997	0.999
	ESD	0.438	0.292	0.156	0.100	0.070
	TSD	0.504	0.287	0.160	0.107	0.080
	CP	93.4	95.8	96.6	94.6	95.3
Outlier/No convergence		142	67	29	20	10
Prob.Complier (p_1)		[.63, .84]				
$X(\beta_{1c})$	MEAN	0.542	0.512	0.504	0.499	0.501
	MED	0.540	0.506	0.501	0.499	0.500
	ESD	0.262	0.172	0.105	0.072	0.049
	TSD	0.460	0.190	0.152	0.081	0.061
	CP	95.8	96.1	95.7	95.9	95.2
$D(\beta_{2c})$	MEAN	1.028	1.007	1.001	1.003	0.999
	MED	1.017	1.012	0.998	1.003	1.000
	ESD	0.227	0.134	0.092	0.067	0.045
	TSD	0.425	0.139	0.105	0.066	0.051
	CP	95.6	95.4	95.4	94.4	94.4
Outlier/No convergence		76	33	29	15	6

APPENDIX II: PROOFS FOR CHAPTER 4

Appendix A: Likelihood

a: The Mixed Case Interval Censoring Model

The probability of the observed data $Y = (M, V, \Delta)$ for $\Delta_{kl} = 1$ equals to

$$
\begin{aligned}
&Pr[M = m, V = v, \Delta_{kl} = 1 | T > v_0] \\
&= Pr[M = m, V = v, T \in (V_{l-1}, V_l], K = k | T > v_0], \\
&= \frac{Pr[T \in (V_{l-1}, V_l], T > v_0, K = k, M = m, V = v]}{Pr[T > v_0]} \\
&= \frac{Pr[T \in (v_{l-1}, v_l], K = k | M = m, V = v] Pr[M = m, V = v]}{Pr[T > v_0]} \text{ , since } (M, V) \perp\!\!\!\perp (T, K) \\
&= \frac{Pr[T \in (v_{l-1}, v_l], K = k] Pr[M = m, V = v]}{Pr[T > v_0]}.
\end{aligned}
$$

Since the $Pr[M = m, V = v]$ does not involve with parameters from Θ, the likelihood is only related to $Pr[T \in (v_{l-1}, v_l], T > v_0, K = k] = F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)$ and $Pr[T > v_0] = 1 - \sum_{k=1}^{n_K} F_k(v_0; \Theta_k)$.

Similarly, for right censored observations, i.e., $\Delta_{M+1} = 1$, we have

$$
\begin{aligned}
&Pr[M = m, V = v, \Delta_{M+1} = 1 | T > v_0] \\
&= Pr[M = m, V = v, T > V_M | T > v_0], \\
&= \frac{Pr[T > V_M, T > v_0, M = m, V = v]}{Pr[T > v_0]}, \\
&= \frac{Pr[T > v_M | M = m, V = v] Pr[M = m, V = v]}{Pr[T > v_0]} \text{ , since } (M, V) \perp\!\!\!\perp (T, K) \\
&= \frac{Pr[T > v_M] Pr[M = m, V = v]}{Pr[T > v_0]}.
\end{aligned}
$$

The corresponding likelihood equals to $\frac{Pr[T>v_M]}{Pr[T>v_0]}$. Therefore,

$$\log L(\Theta) = \sum_{i=1}^{n} \log \ell(Y_i; \Theta),$$

where

$$\ell(Y_i; \Theta) = \prod_{k=1}^{n_K} \prod_{l=1}^{M} \left\{ \frac{F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)}{1 - \sum_{k=1}^{n_K} F_k(v_0; \Theta_k)} \right\}^{\Delta_{kl}} \left\{ \frac{1 - \sum_{k=1}^{n_K} F_k(v_m; \Theta_k)}{1 - \sum_{k=1}^{n_K} F_k(v_0; \Theta_k)} \right\}^{\Delta_{M+1}}.$$

b: The IIP Model

The probability of the observed data (M, V, Δ) for $\Delta_{kM} = 1$ equals to the following

$$
\begin{aligned}
&\prod_{l=1}^{m-1} \frac{1}{Pr[T > v_0]} \times Pr[V_l = v_l | H_l = h_l] Pr[\Delta_{1l} = \ldots = \Delta_{n_K l} = 0, T > v_0 | H_l = h_l, V_l = v_l] \\
&\times Pr[V_M = v_m | H_M = h_m] Pr[\Delta_{kM} = 1, T > v_0 | H_M = h_m, V_M = v_m], \\
&\sim \prod_{l=1}^{m-1} \frac{1}{Pr[T > v_0]} Pr[\Delta_{1l} = \ldots = \Delta_{n_K l} = 0, T > v_0 | H_l = h_l, V_l = v_l] \\
&\times Pr[\Delta_{kM} = 1, T > v_0 | H_M = h_m, V_M = v_m].
\end{aligned}
$$

For $\Delta_{kl} = 1$, the corresponding probability is

$$
\begin{aligned}
&Pr[\Delta_{kl} = 1, T > v_0 | H_l = h_l, V_l = v_l] \\
&= Pr[v_{l-1} < T \leq v_l, K = k, T > v_0 | H_l = h_l, V_l = v_l] \text{ , since } V_l \perp\!\!\!\perp (T, K) | H_l \\
&= Pr[v_{l-1} < T \leq v_l, K = k, T > v_0 | H_l = h_l] \\
&= Pr[v_{l-1} < T \leq v_l, K = k, T > v_0 | H_{l-1} = h_{l-1}, V_{l-1} = v_{l-1}, \Delta_{1,l-1} = \ldots = \Delta_{n_K,l-1} = 0, T > \tau] \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k, T > v_0, \Delta_{1,l-1} = \ldots = \Delta_{n_K,l-1} = 0 | H_{l-1} = h_{l-1}, V_{l-1} = v_{l-1}]}{Pr[\Delta_{1,l-1} = \ldots = \Delta_{n_K,l-1} = 0, T > v_0 | H_{l-1} = h_{l-1}, V_{l-1} = v_{l-1}]} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k, T > v_0 | H_{l-1} = h_{l-1}, V_{l-1} = v_{l-1}]}{Pr[T > v_{l-1}, T > v_0 | H_{l-1} = h_{l-1}, V_{l-1} = v_{l-1}]} \text{ , since } \tau \leq v_{l-1} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k | H_{l-1} = h_{l-1}]}{Pr[T > v_{l-1} | H_{l-1} = h_{l-1}]} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k | H_{l-2} = h_{l-2}, V_{l-2} = v_{l-2}, \Delta_{1,l-2} = \ldots = \Delta_{n_K,l-2} = 0, T > v_0]}{Pr[T > v_{l-1} | H_{l-2} = h_{l-2}, V_{l-2} = v_{l-2}, \Delta_{1,l-2} = \ldots = \Delta_{n_K,l-2} = 0, T > v_0]} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k, \Delta_{1,l-2} = \ldots = \Delta_{n_K,l-2} = 0, T > v_0 | H_{l-2} = h_{l-2}, V_{l-2} = v_{l-2}]}{Pr[T > v_{l-1}, \Delta_{1,l-2} = \ldots = \Delta_{n_K,l-2} = 0, T > v_0 | H_{l-2} = h_{l-2}, V_{l-2} = v_{l-2}]} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k | H_{l-2} = h_{l-2}]}{Pr[T > v_{l-1} | H_{l-2} = h_{l-2}]} \\
&= \ldots = \frac{Pr[v_{l-1} < T \leq v_l, K = k]}{Pr[T > v_{l-1}]} \\
&= \frac{Pr[v_{l-1} < T \leq v_l, K = k]}{Pr[T > v_{l-1}]} \\
&= \frac{F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)}{1 - \sum_{k=1}^{n_K} F_k(v_{l-1}; \Theta_k)}
\end{aligned}
$$

and

$$Pr[\Delta_{1l} = \ldots = \Delta_{n_K l} = 0, T > v_0 | H_l = h_l, V_l = v_l] = 1 - \sum_{k=1}^{n_K} Pr[\Delta_{kl} = 1, T > v_0 | H_l = h_l, V_l = v_l]$$

$$= 1 - \sum_{k=1}^{n_K} \frac{F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)}{1 - \sum_{k=1}^{n_K} F_k(v_{l-1}; \Theta_k)}$$

$$= 1 - \sum_{k=1}^{n_K} \frac{F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)}{1 - F(v_{l-1}; \Theta_k)}$$

$$= 1 - \frac{F(v_l) - F(v_{l-1}; \Theta_k)}{1 - F(v_{l-1}; \Theta_k)} = \frac{1 - F(v_l; \Theta_k)}{1 - F(v_{l-1}; \Theta_k)}.$$

For $\Delta_{kM} = 1$,

$$\frac{1}{Pr[T > \tau]} \left\{ \prod_{l=1}^{m-1} \frac{1 - F(v_l; \Theta_k)}{1 - F(v_{l-1}; \Theta_k)} \right\} \frac{F_k(v_m; \Theta_k) - F_k(v_{m-1}; \Theta_k)}{1 - F(v_{m-1}; \Theta_k)}$$

$$= \frac{F_k(v_m; \Theta_k) - F_k(v_{m-1}; \Theta_k)}{Pr[T > v_0]} = \frac{F_k(v_m; \Theta_k) - F_k(v_{m-1}; \Theta_k)}{1 - F(v_0; \Theta_k)},$$

and for $\Delta_{1M} = \Delta_{2M} = \ldots = \Delta_{n_K M} = 0$,

$$\frac{1}{Pr[T > v_0]} \prod_{l=1}^{m} Pr[V_l = v_l | H_l = h_l] Pr[\Delta_{1M} = \ldots = \Delta_{n_K M} = 0, T > v_0 | H_l = h_l, V_l = v_l]$$

$$= \frac{1}{Pr[T > v_0]} \prod_{l=1}^{m} Pr[V_l = v_l | H_l = h_l] Pr[T > v_m]$$

$$\sim \frac{Pr[T > v_m]}{Pr[T > v_0]} = \frac{1 - F(v_m; \Theta_k)}{1 - F(v_0; \Theta_k)}.$$

Therefore, the log likelihood is expressed by

$$\log L(\Theta) = \sum_{i=1}^{n} \log l(Y_i; \Theta),$$

where

$$l(Y_i; \Theta) = \prod_{k=1}^{n_K} \prod_{l=1}^{M} \left\{ \frac{F_k(v_l; \Theta_k) - F_k(v_{l-1}; \Theta_k)}{1 - F(v_0; \Theta_k)} \right\}^{\Delta_{kl}} \left\{ \frac{1 - F(v_m; \Theta_k)}{1 - F(v_0; \Theta_k)} \right\}^{\Delta_{M+1}}.$$

Appendix B: Asymptotic Variance of $F_k(t;\hat{\Theta}_k)$ for Gompertz Model

The approximate variance of $F_k(t;\hat{\Theta}_k) = 1-\exp[\hat{\beta}_k\{1-\exp(\hat{\alpha}_k t)\}/\hat{\alpha}_k]$ for $k=1,\ldots,n_K$ has the following form

$$\widehat{var}\{F_k(t;\hat{\Theta}_k)\} = \left(\frac{\partial F_k(t;\Theta_k)}{\partial \Theta_k}\right)\hat{\Sigma}_{\hat{\Theta}}\left(\frac{\partial F_k(t;\Theta_k)}{\partial \Theta_k}\right)'\Bigg|_{\Theta=\hat{\Theta}},$$

where $\hat{\Sigma}_{\hat{\Theta}}$ is the inverse of the following observed Fisher information matrix

$$-\begin{bmatrix} \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1^2} & \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1 \partial \beta_1} & \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1 \partial \alpha_2} & \cdots & \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1 \partial \beta_{n_K}} \\ \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1 \partial \beta_1} & \frac{\partial^2 \log L(\Theta)}{\partial \beta_1^2} & \frac{\partial^2 \log L(\Theta)}{\partial \alpha_2 \partial \beta_1} & \cdots & \frac{\partial^2 \log L(\Theta)}{\partial \beta_1 \partial \beta_{n_K}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \log L(\Theta)}{\partial \alpha_1 \partial \beta_{n_K}} & \frac{\partial^2 \log L(\Theta)}{\partial \beta_1 \partial \beta_{n_K}} & \frac{\partial^2 \log L(\Theta)}{\partial \alpha_2 \partial \beta_{n_K}} & \cdots & \frac{\partial^2 \log L(\Theta)}{\partial \beta_{n_K}^2} \end{bmatrix}.$$

By letting $D_{kl}(\Theta_k) = \{F_k(V_l;\Theta_k) - F_k(V_{l-1};\Theta_k)\}$, we have

$$\begin{aligned}\frac{\partial \log l(Y_i;\Theta)}{\partial \alpha_k} &= \sum_{l=1}^{M}\Delta_{kl}\left[\frac{1}{D_{kl}(\Theta_k)}\frac{\partial D_{kl}(\Theta_k)}{\partial \alpha_k} + \frac{1}{1-F(\tau;\Theta_k)}\frac{\partial F_k(\tau;\Theta_k)}{\partial \alpha_k}\right] \\ &\quad + \Delta_{M+1}\left[\frac{-1}{1-F(v_M;\Theta_k)}\frac{\partial F_k(v_M;\Theta_k)}{\partial \alpha_k} + \frac{1}{1-F(\tau;\Theta_k)}\frac{\partial F_k(\tau;\Theta_k)}{\partial \alpha_k}\right],\end{aligned}$$

$$\begin{aligned}&\frac{\partial^2 \log l(Y_i;\Theta)}{\partial \alpha_k^2} \\ &= \sum_{l=1}^{M}\Delta_{kl}\left[\frac{1}{D_{kl}^2(\Theta_k)}\left\{\frac{\partial^2 D_{kl}(\Theta_k)}{\partial \alpha_k^2}D_{kl}(\Theta_k) - \left(\frac{\partial D_{kl}(\Theta_k)}{\partial \alpha_k}\right)^2\right\} + \right. \\ &\quad \left.\frac{1}{\{1-F(\tau;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(\tau;\Theta_k)}{\partial \alpha_k^2}\{1-F(\tau;\Theta_k)\} + \left(\frac{\partial F_k(\tau;\Theta_k)}{\partial \alpha_k}\right)^2\right\}\right] \\ &+\Delta_{M+1}\left[\frac{-1}{\{1-F(v_M;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(v_M;\Theta_k)}{\partial \alpha_k^2}\{1-F(v_M;\Theta_k)\} + \left(\frac{\partial F_k(v_M;\Theta_k)}{\partial \alpha_k}\right)^2\right\} + \right. \\ &\quad \left.\frac{1}{\{1-F(\tau;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(\tau;\Theta_k)}{\partial \alpha_k^2}\{1-F(\tau;\Theta_k)\} + \left(\frac{\partial F_k(\tau;\Theta_k)}{\partial \alpha_k}\right)^2\right\}\right]\end{aligned}$$

and

$$
\begin{aligned}
&\frac{\partial^2 \log l(Y_i;\Theta)}{\partial\alpha_k\partial\beta_k} \\
=&\sum_{l=1}^{M}\Delta_{kl}\left[\frac{1}{D_{kl}^2(\Theta_k)}\left\{\frac{\partial^2 D_{kl}(\Theta_k)}{\partial\alpha_k\partial\beta_k}D_{kl}(\Theta_k)-\frac{\partial D_{kl}(\Theta_k)}{\partial\alpha_k}\frac{\partial D_{kl}(\Theta_k)}{\partial\beta_k}\right\}+\right. \\
&\left.\frac{1}{\{1-F(\tau;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(\tau;\Theta_k)}{\partial\alpha_k\partial\beta_k}\{1-F(\tau;\Theta_k)\}+\frac{\partial F_k(\tau;\Theta_k)}{\partial\alpha_k}\frac{\partial F_k(\tau;\Theta_k)}{\partial\beta_k}\right\}\right] \\
+\Delta_{M+1}&\left[\frac{-1}{\{1-F(v_M;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(v_M;\Theta_k)}{\partial\alpha_k\partial\beta_k}\{1-F(v_M;\Theta_k)\}+\frac{\partial F_k(v_M;\Theta_k)}{\partial\alpha_k}\frac{\partial F_k(v_M;\Theta_k)}{\partial\beta_k}\right\}+\right. \\
&\left.\frac{1}{\{1-F(\tau;\Theta_k)\}^2}\left\{\frac{\partial^2 F_k(\tau;\Theta_k)}{\partial\alpha_k\partial\beta_k}\{1-F(\tau;\Theta_k)\}+\frac{\partial F_k(\tau;\Theta_k)}{\partial\alpha_k}\frac{\partial F_k(\tau;\Theta_k)}{\partial\beta_k}\right\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial F_k(t;\Theta_k)}{\partial\alpha_k}&=\{1-F_k(t;\Theta_k)\}\{1+\exp(\alpha_k t)(\alpha_k t-1)\}\beta_k/\alpha_k^2, \\
\frac{\partial F_k(t;\Theta_k)}{\partial\beta_k}&=\{1-F_k(t;\Theta_k)\}\{\exp(\alpha_k t)-1\}/\alpha_k, \\
\frac{\partial^2 F_k(t;\Theta_k)}{\partial\alpha_k^2}&=\frac{-\partial F_k(t;\Theta_k)}{\partial\alpha_k}\left[\frac{\beta_k}{\alpha_k^2}+\frac{\beta_k}{\alpha_k^2}\exp(\alpha_k t)(\alpha_k t-1)\right] \\
&\qquad+\{1-F_k(t;\Theta_k)\}\left[\frac{-2\beta_k}{\alpha_k^3}+\left\{\beta_k\alpha_k^{-1}t^2+2\beta_k\alpha_k^{-3}-2\beta_k\alpha_k^{-2}t\right\}\exp(\alpha_k t)\right], \\
\frac{\partial^2 F_k(t;\Theta_k)}{\partial\beta_k^2}&=-\{\exp(\alpha_k t)-1\}^2\{1-F_k(t;\Theta_k)\}/\alpha_k^2, \\
\frac{\partial^2 F_k(t;\Theta_k)}{\partial\alpha_k\partial\beta_k}&=-\{1-F_k(t;\Theta_k)\}\{(\alpha_k t-1)\exp(\alpha_k t)+1\}\{(\exp(\alpha_k t)-1)\beta_k-\alpha_k\}/\alpha_k^3.
\end{aligned}
$$

The other mixed partial second derivatives are omitted.

BIBLIOGRAPHY

1. Cox DR. Regression models and life-tables. *Royal Statistical Society. Series B (Methodological)* 1972; **34**(2):187–220.

2. Cox DR. Partial likelihood. *Biometrika* 1975; **62**(2):269–276.

3. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* Dec 1982; **10**(4):1100–1120.

4. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**(2):373–379.

5. Kelly PJ, Lim LLY. Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine* Jan 2000; **19**(1):13–33.

6. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**(408):1065–1073.

7. Therneau TM, Hamilton SA. rhDNase as an example of recurrent event analysis. *Statistics in Medicine* 1997; **16**(18):2029–2047.

8. Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine* 1997; **16**:833–839.

9. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Royal Statistical Society. Series B (Methodological)* 2000; **62**(4):711–730.

10. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and Drug Safety* 2010; **19**(6):537–554.

11. Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics-Theory and Methods* 1991; **20**(8):2609–2631.

12. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**(5):561–570.

13. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* 2001; **96**(454):440–448.

14. Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* 2003; **59**(1):100–105.

15. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**(3):779–788.

16. Cox DR, Oakes D. *Analysis of survival data*. Chapman and Hall, 1984.

17. Loeys T, Goetghebeur E, Vandebosch A. Causal proportional hazards models and time-constant exposure in randomized clinical trials. *Lifetime Data Analysis* 2005; **11**:435–449.

18. Yu W, Chen K, Sobel ME, Ying Z. Semiparametric transformation models for causal inference in time-to-event studies with all-or-nothing compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2015; **77**(2):397–415.

19. Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* Sep 2007; **69**(4):565–588.

20. Tsiatis A. A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences of the United States of America* 1975; **72**(1):20–22.

21. Prentice RL, Kalbfleisch JD, Jr AVP, Flournoy N, Farewelland VT, Breslow NE. The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics* 1978; **34**:541–554.

22. Gray RJ. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics* 1988; **16**(3):1141–1154.

23. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**(446):496–509.

24. Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics* 2001; **2**(1):85–97.

25. Lee EW, Wei LJ, Amato DA. Cox-type regression analysis for large numbers of small groups of correlated failure time observations, edited by J.P. Klein, P.K. Goel. *Survival analysis: State of the art* 1992; **211**:237–247.

26. Oakes D. Frailty models for multiple event times, edited by J.P. Klein, P.K. Goel. *Survival Analysis, State of the Art. Kluwer Academic Publisher; Netherlands* 1992; **211**:371–379.

27. Lin DY. On the Breslow estimator. *Lifetime Data Analysis* 2007; :471–480.

28. Castañeda J, Gerritse B. Appraisal of several methods to model time to multiple events per subject: Modeling time to hospitalizations and death. *Revista Colombiana de Estadística* 2010; **33**(1):43–61.

29. Chen F, Chen K. Modeling event clustering the m-Memory Cox-type self-exciting intensity model. *International Journal of Statistics and Probability* 2014; **3**(3):126–137.

30. Hansen BE. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 1996; **64**(2):413–430.

31. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 1977; **64**(2):247–254.

32. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 1987; **74**(1):33–43.

33. McClish DK. Analyzing a portion of the ROC curve. *Medical Decision Making* 1989; **9**(3):190–195.

34. Davies JC. Pseudomonas aeruginosa in cystic fibrosis: Pathogenesis and persistence. *Paediatric Respiratory Reviews* 2002; **3**:128–134.

35. Liu X. *Survival analysis: models and applications*. Wiley, 2012.

36. Lin DY. Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* 1994; **13**:2233–2247.

37. Cook RJ, Lawless JF. Discussion of paper by Wei and Glidden. *Statistics in Medicine* 1997; **16**:841–851.

38. Villegas R, Julià O, Ocaña J. Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring. *BMC Medical Research Methodology* 2013; **13**(1):95.

39. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; **44**(3):837–845.

40. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; **12**(77):doi:10.1186/1471–2105–12–77.

41. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**(4):722–729.

42. Abadie A. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* Apr 2003; **113**(2):231–263.

43. Newhouse JP, McClellan M. Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health* 1998; **19**(1):17–34.

44. Hernán MA, Robins JM. Instruments for causal inference. *Epidemiology* 2006; **17**(4):360–372.

45. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Statistics in Medicine* Jun 2014; **33**(13):2297–2340.

46. Baker SG, Kramer BS, Lindeman KS. Latent class instrumental variables: A clinical and biostatistical perspective. *Statistics in Medicine* 2015; **35**(1):147–160.

47. Frangakis C. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999; **86**(2):365–379.

48. Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000; **1**(1):69–88.

49. Tan Z. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association* 2006; **101**(476):1607–1618.

50. Gong Z. Parametric potential-outcome survival models for causal inference. PhD Thesis, Ph. D. dissertation, University of Canterbury 2008.

51. Baker SG. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association* 1998; **93**(443):929.

52. Nie H, Cheng J, Small DS. Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* 2011; **67**(4):1397–1405.

53. Lin H, Li Y, Jiang L, Li G. A semiparametric linear transformation model to estimate causal effects for survival data. *Canadian Journal of Statistics* 2014; **42**(1):18–35.

54. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Health Services and Outcomes Research Methodology* 2014; **14**(1-2):54–68.

55. Yau LHY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing Data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 2001; **96**(456):1232–1244.

56. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* Feb 1997; **25**(1):305–327.

57. Small DS, Ten Have TR, Joffe MM, Cheng J. Random effects logistic models for analyzing efficacy of a longitudinal randomized treatment with non-adherence. *Statistics in Medicine* 2006; **25**(12):1981–2007.

58. O'Malley AJ. Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets. *Health Services and Outcomes Research Methodology* 2012; **12**(4):254–272.

59. Therneau TM, Grambsch PM, Fleming TR. Martingale-based residuals for survival models. *Biometrika* 1990; **77**(1):147–160.

60. Breslow NE. Contribution to the discussion of paper by D.R. Cox. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1972; **34**:216–217.

61. American Cancer Society: Colorectal Cancer Facts & Figures 2016. *Atlanta, GA: American Cancer Society* 2016; :1–72.

62. Design of the prostate, lung, colorectal and ovarian (plco) cancer screening trial. *Controlled Clinical Trials* 2000; **21**(6):273S – 309S.

63. Schoen RE, Pinsky PF, Weissfeld JL, Yokochi LA, Church T, Laiyemo AO, Bresalier RS, Andriole GL, Buys SS, Crawford ED, *et al.*. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *The New England Journal of Medicine* 2012; **366**(25):2345–2357.

64. Hudgens MG, Li C, Fine JP. Parametric likelihood inference for interval censored competing risks data. *Biometrics* 2014; **70**(1):1–9.

65. Screening for osteoporosis: U.S. preventive services task force recommendation statement. *Ann Intern Med* 2011; **154**(5):356–364.

66. Cosman F, de Beur S, LeBoff M, Lewiecki E, Tanner B, Randall S, Lindsay R. Clinician's Guide to Prevention and Treatment of Osteoporosis. *Osteoporos Int.* 2014; **25**(10):2359–2381.

67. Baim S, Binkley N, Bilezikian J, Kendler D, Hans D, Lewiecki E, Silverman S. Official Positions of the International Society for Clinical Densitometry and Executive Summary of the 2007 ISCD Position Development Conference. *J Clin Densitom* 2008; **11**.

68. Gourlay ML, Fine JP, Preisser JS, May RC, Li C, Lui LY, David F Ransohoff JAC, Ensrud KE. Bone-Density Testing Interval and Transition to Osteoporosis in Older Women. *The New England Journal of Medicine* 2012; **366**(3):225–233.

69. Lamarca R, Alonso J, Gómez G, , Álvaro Munõz. Left-truncated Data With Age as Time Scale: An Alternative for Survival Analysis in the Elderly Population. *Journal of Gerontology: MEDICAL SCIENCES* 1988; **53A**(5):337–343.

70. Cain KC, Harlow SD, Little RJ, Nan B, Yosef M, Taffe JR, Elliott MR. Bias Due to Left Truncation and Left Censoring in Longitudinal Studies of Developmental and Disease Processes. *American Journal of Epidemiology* 2011; **173**(9):1078–1084.

71. Hudgens MG, Satten GA, Ira M Longini J. Nonparametric Maximum Likelihood Estimation for Competing Risks Survival Data Subject to Interval Censoring and Truncation. *Biometrics* Mar 2001; **57**:74–80.

72. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)* 1976; **38**(3):290–295.

73. Jewell NP, Laan MVD, Henneman T. Nonparametric estimation from current status data with competing risks. *Biometrika* 2003; **90**(1):183–197.

74. Groeneboom P, Maathuis MH, Wellner JA. Current Status Data with Competing Risks: Consistency and Rates of Convergence of the MLE. *The Annals of Statistic* 2008; **36**(3):1031–1063.

75. Groeneboom P, Maathuis MH, Wellner JA. Current Status Data with Competing Risks: Limiting Distribution of the MLE. *The Annals of Statistic* 2008; **36**(3):1064–1089.

76. Li C, Fine JP. Smoothed nonparametric estimation for current status competing risks data. *Biometrika* 2013; **100**(1):173–187.

77. Jeong JH, Fine JP. Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society, Series C.* 2006; **55**:187–200.

78. Lawless JF. *Statistical Models and Methods for Lifetime Data, 2nd.* John Wiley & Sons, 2003.

79. Assessment of osteoporotic fracture risk and its role in screening for postmenopausal osteoporosis. *WHO Technical report series* 1994; (843).

80. Ravn P, Cizza G, Bjarnason N, Thompson D, Daley M, Wasnich R, Mcclung M, Hosking D, Yates A, Christiansen C. Low Body Mass Index Is an Important Risk Factor for Low Bone Mass and Increased Bone Loss in Early Postmenopausal Women. *Journal of Bone and Mineral Research* 1999; **14**(9):1622–1627.

81. De Laet C, Kanis J, Oden A, Johanson H, Johnell O, Delmas P, Eisman J, Kroger H, Fujiwara S, Garnero P, *et al.*. Body mass index as a predictor of fracture risk: a meta-analysis. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 2005; **16**(11):1330–1338.

82. Rud B, Hilden J, Hyldstrup L, Hrobjartsson A. The Osteoporosis Self-Assessment Tool versus alternative tests for selecting postmenopausal women for bone mineral density assessment: a comparative systematic review of accuracy. *Osteoporos Int* 2009; **20**(4):599–607.

83. Nayak S, Edwards D, Saleh A, Greenspan S. Systematic review and meta-analysis of the performance of clinical risk assessment instruments for screening for osteoporosis or low bone density. *Osteoporos Int* 2015; **26**(5):1543–1554.

84. Gourlay M, Powers J, Lui L, Ensrud K. Clinical performance of osteoporosis risk assessment tools in women aged 67 years and older. *Osteoporos Int* 2008; **19**(8):1175–1183.

85. Gourlay M, Miller W, Richy F, Garrett J, Hanson L, Reginster JY. Performance of osteoporosis risk assessment tools in postmenopausal women aged 45 to 64 years. *Osteoporos Int* 2005; **16**:921–927.

86. Crandall C, Larson J, Gourlay M, Donaldson M, Lacroix A, Cauley J, Wactawski-Wende J, Gass M, Robbins J, Watts N, *et al.*. Osteoporosis Screening in Postmenopausal Women 50-64 Years-Old: Comparison of U.S. Preventive Services Task Force Strategy and Two Traditional Strategies in the Women's Health Initiative [Epub ahead of print]. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* Jan 2016; .