Multi-Stage Statistical Models for Cancer in Observational Studies and SMARTs

by

Bao Qui Tran

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2017

Doctoral Committee:

Professor Alexander Tsodikov, Co-Chair
Assistant Professor Kelley M. Kidwell, Co-Chair
Assistant Professor Rafael Meza
Professor Jeremy M. G. Taylor

Bao Qui Tran

quitran@umich.edu

ORCID iD: 0000-0002-0760-964X

ACKNOWLEDGEMENTS

I would like to thank my co-advisors, professors Alex Tsodikov and Kelley Kidwell. Alex has been the kindest and most patient advisor that I could have asked for. I am thankful to have the chance to work on joint modeling for cancer under his guidance. He introduced me to research ideas and helped me tackle the mathematical challenges I encountered. Kelley has taught me nearly everything I know about dynamic treatment regimens. She also generously edited all of my drafts and set weekly goals that ensured my on-time graduation. I would also like to thank my committee members, Professor Jeremy Taylor of the Department of Biostatistics and Professor Rafael Meza of the Department of Epidemiology, for their incredibly insightful comments that led to improvements in my research.

I would like to express my love to my parents Phong Tran and Mai-Anh Nguyen, especially my mother who does not speak English fluently but navigated half way across the world from Vietnam to visit me every year. Theirs and my older brother's support in the past 12 years allowed me all the carefree time to pursue my academic goals. I am also very grateful for my uncle Khoa and aunt Hieu, who raised me through my years in college. Special thanks to my aunt Thuan Nguyen for being my mentor; my aunts Chau-Anh Nguyen, Anh-Hoa Nguyen, and my uncle Long Ngo, Gene Yamamoto, Kim-Son Nguyen, who, in addition to being role models, have constasntly provided me with advices and encouragement. I have been enormously lucky to have you as my family.

My years in Ann Arbor would not have been be as happy if not for many of my

friends. Guangming Lang, in particular, not only provided me with invaluable career discussions but also helped me become more mature through our conversations about other aspects of life outside of statistical research.

This research was supported by National Cancer Institute grants 5U01CA157224 (CISNET), 1U01CA199338 (CISNET), and 5P50CA186786 (SPORE).

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

APPENDIX

ABSTRACT

Many diseases, especially cancer, are not static, but rather can be summarized by a series of events or stages (e.g. diagnosis, remission, recurrence, metastasis, death). Most available methods to analyze multi-stage data ignore intermediate events and focus on the terminal event or consider (time to) multiple events as independent. Competing-risk or semi-competing-risk models are often deficient in describing the complex relationship between disease progression events that are driven by a shared progression stochastic process. In the first chapter, we propose a semi-parametric joint model of diagnosis, latent metastasis, and cancer death and use nonparametric maximum likelihood to estimate covariate effects on the risks of intermediate events and death and the dependence between them. We illustrate the model using SEER prostate cancer data.

In the second chapter, we focus on the adverse effect of younger diagnosis age on cancer survival. We use a joint model with a shared gamma frailty term to interpret the effect as a consequence of correlation between diagnosis time and the post-diagnosis survival time. In the traditional analysis, diagnosis time is treated as the time origin for a model of overall survival that fails to utilize the full information leading up to diagnosis. Often the available covariates do not fully explain the correlation between time-to-diagnosis and time-to-death calling for use of joint modeling and frailties to extend the model. We show that the variance of the frailty term and covariate effects can be estimated by a nonparametric maximum likelihood method. Laplace transformation is used to derive likelihood contributions. The

model is applied to Michigan SEER breast cancer data.

In the third chapter, we compare dynamic treatment regimens from clinical trials with multiple rounds of treatment randomization (sequential multiple assignment randomized trials, SMARTs). Previously proposed methods to analyze data with survival outcomes from a SMART use inverse probability weighting and provide non-parametric estimation of survival rates, but no other information. We apply a joint modeling approach here to provide unbiased survival estimates and as a mechanism to include auxiliary covariates, treatment effects and their interaction within regimens. We address the multiple comparisons problem using multiple-comparisons-with-the-best (MCB).

Keywords: Disease natural history; Frailty model; Semiparametric regression; SMART designs; Survival analysis.

CHAPTER I

Introduction

The effects of clinical and demographic factors along with therapeutic interventions on the overall survival time of patients with any disease is not a new problem in survival analysis. Common methods of analysis include Cox proportional hazards and frailty models. Situations where a certain event prevents others from being observed (e.g. death due to other causes occurs before death from disease of interest) have been studied extensively within the competing risks framework. However, there is no agreed-upon procedure to model survival data where there are multiple time-to-events of interest that make up the disease process. This dissertation presents joint models to analyze multiple time-to-events in the observational (chapters II and III) and experimental (chapter IV) settings.

One setting where multiple time-to-events of are of interest is cancer progression. A cancer patient progresses from time of diagnosis, through various stages of cancer (e.g. metastasis, remission, relapse) to death. Because these events are symptoms of the same underlying disease, the times to each event are dependent on each other. As a result, the entire process cannot be studied using conventional survival analysis with a single time-to-event outcome. Our proposed methods for this setting are motivated by SEER data discussed in chapter II and III of this dissertation. Another

setting where analyses of multiple time-to-events are necessary is in the evaluation of dynamic treatment regimens, where subsequent treatments in a regimen are assigned based on an individual's time-to-response to the initial therapy. This setting and our proposed methods are discussed in chapter IV.

Chapter II is motivated by the SEER prostate cancer data which includes U.S. 315,722 men who were diagnosed with prostate cancer between 1988-2003. Age at diagnosis, metastasis status, and overall survival time were recorded. In the traditional analysis, diagnosis time is used as the time origin for a model with a single time-to-event (i.e. survival time after diagnosis) and ignores disease information leading up to diagnosis. Moreover, traditional analysis does not utilize time-to-metastasis which is an important clinical outcome that may depend on treatment at diagnosis (and thus cannot be used as a covariate). A framework that model these three stages simultaneously is necessary. Thus, we propose a joint model using non-parametric maximum likelihood estimator (NPMLE) to model the effects of treatment and other covariates on time-to-diagnosis, latent time-to-metastasis, and time-to-death while taking into account the dependence structure between them.

The difficulties of capturing latent dependence between breast cancer time-to-diagnosis and subsequent survival time is the main theme of chapter III. In the context of cancer studies, the age at which a patient is diagnosed with cancer can be considered as the time-to-diagnosis. From SEER registry of breast cancer patients in Michigan between 1973-2011, women diagnosed at younger ages (30-40 years old) have lower survival rates subsequently compared to those diagnosed at older ages (40-60, >60 years old). As demonstrated, current models do not adequately capture this non-constant adverse effect of diagnosis age on breast cancer survival. A joint model based on the one proposed in chapter II was not flexible enough to fit this

empirical observation from SEER data. A joint model using diagnosis age as one of the covariates for time-to-death, besides requiring a restrictive assumption about this age's effect over time (e.g. constant or piecewise), still could not predict the full magnitute of age's adverse effect. We proposed relaxing the assumption made in chapter II regarding parametrically related baseline hazards for different cancer stages, and add a shared gamma frailty term to interpret the effect as a consequence of correlation between diagnosis time and the post-diagnosis survival time. We show the procedure to estimate both baseline hazards for diagnosis and death non-parametrically. Using Laplace transformation to derive likelihood contributions averaged over the latent frailty term, we estimated the variance of the frailty term and other covariate effects by NPMLE. The proposed model's ability to capture latent dependence between the two stages is demonstrated via plots of post-diagnosis survival conditioning on different time-to-diagnosis.

Recent methods to model dynamic treatment regimens with survival outcomes from clinical trials with multiple rounds of treatment randomization (sequential multiple assignment randomized trials, SMARTs) are described in chapter IV. Previously proposed methods to analyze survival data from a SMART include non-parametric estimators and parametric models based on Cox proportional hazards models. All use inverse probability weighting to arrive at the estimation of survival rates, but the models do not provide any causal inference. Using the setting of a SMART with two treatment stages, we present a joint model of time between the first and second round of treatment randomization and the overall survival time to provide a mechanism to quantify treatment auxiliary covariate effects and their interaction within regimens. Furthermore, we demonstrate the joint model's robustness over previous methods for SMART in predicting dynamic treatment regimens's

survival rates under the presence of interaction between auxiliary covariates and treatments. Additionally, we address the issue of multiple comparisons for the inference concerning the dynamic treatment regimens within a SMART. Specifically, we implement multiple-comparisons-with-the-best to address the issue of multiple comparisons between embedded dynamic treatment regimens in a SMART with survival outcomes.

The concluding chapter summarizes the strengths and limitations of the proposed joint models for observational SEER cancer data and SMART settings. The method's applicability and further areas of extension for mutistaged survival analysis are presented.

CHAPTER II

A Joint Model of Cancer Incidence, Metastasis, and Mortality

2.1 Introduction

In a natural cancer progression, patients may be diagnosed with local disease and achieve remission for some time after treatment, potentially progress, develop metastasis and succumb to the cancer. Others may be diagnosed at a later stage and show metastasis at initial diagnosis. We focus on cancer diagnosis, metastasis, and death as a progressive sequence of cancer events (Figure 2.1) and how to model overall survival time accounting for dependence between time in each stage. Our motivating example is data on prostate cancer from the Surveillance, Epidemiology, and End Results (SEER) registry. The case file in SEER follows prostate cancer patients from diagnosis to death or end of last follow up. SEER provides metastasis status at time of diagnosis (i.e. whether cancer had metastasized prior to patients being diagnosed). The exact time of onset of metastasis remains unobserved. Because the SEER case file only includes subjects diagnosed with cancer, we make use of the second SEER dataset with population counts, matched by race and age, from which the case file originated. Although it is more straightforward to only focus on the case file, we want to make use of the available population data to (1) account for people who were not diagnosed with cancer and (2) gain information regarding the

latent metastasis process which occurs before diagnosis. This requires a joint model of cancer incidence, metastases and survival.

The model focuses on inferences and predictions of the cancer progression mechanism leading to cancer-specific death, treating other causes of death as nuisance. Implications of this approach for clinical decision making and model extensions into a competing risks framework are outlined in the Discussion.

As noted, the onset of metastasis is unobserved (latent) before the disease reaches a measureable size; therefore if metastasis is present at the time of cancer diagnosis, the onset must have occurred at some prior time point. Overall, metastases are the cause of over 90% of cancer deaths (Mehlen and Puisieux, 2006). For cancers that originate in a non-vital organ, such as prostate, virtually all deaths are caused by cancer spread to other organs. Indeed, 99% of men with prostate cancer survive after 5 year (95% after 15 years). However, only 28% of those diagnosed with prostate cancer that has metastasized survive to 5 years (American Cancer Society, 2016). Thus our study tailors to settings where death-due-to-cancer can only occur as a result of metastasis (Figure 2.1). We assume that if a patient was first diagnosed without metastasis, metastasis occur at some subsequent point prior to death. For the minor fraction of cancers where local growth can be lethal, such as brain cancer, dependent competing causes of cancer death would require substantial additional modeling that remains beyond the scope of this paper.

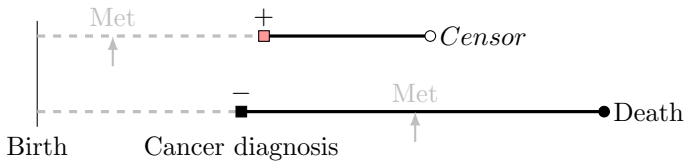


Figure 2.1: Diagram of cancer progression: Cancer diagnosis after metastasis ▫; diagnosis before metastasis ▪. Exact time of metastasis onset is unknown.

As a result of the multi-stage nature of cancer, there are many issues that make conventional survival analysis with a single time-to-event outcome inappropriate. First, the occurrence of certain stages may prevent others from being observed (e.g. metastasis may occur prior to cancer diagnosis, preventing observence of time-to-diagnosis-without-metastasis). This censoring due to other events should be accounted for when analyzing each time-to-event outcome. Second, part of the natural disease progression may be latent and thus cannot be observed (e.g. metastasis onset). Third, the dependence between time-to-observable-events (e.g. diagnosis and death) and the latent disease leading up to them (e.g. metastasis) make it difficult, and potentially misleading, to separately model the outcomes. The treatment effect has multiple causal paths, one of them being the effect on the risk of death through the latent metastatic process.

In cancers where death can occur without metastasis (from local disease in a vital organ), death and metastasis are semi-competing risks (Fine et al., 2001) and this area has received considerable attention. Because of the non-identifiability issue from competing risks, a full nonparametric solution is not available (Peng and Fine, 2007). Many have combined the semi-parametric approach with a copula model (Nelsen, 1999) to account for dependent competing risks. Clayton copula (Clayton, 1978) and bivariate survival models (Oakes, 1989) are popular choices for this approach. Recently, Chen (2012) proposed using a semi-parametric transformational model for the marginal distribution and a copula model for the joint distribution of time-to-each-event. In our setting, since death due to prostate cancer can only occur following metastasis but not vice versa, these events represent recurrent events that are partially unobserved, rather than semi-competing risks.

Multi-stage models have been a popular tool (Andersen et al., 1991; Xu et al.,

2010; Kalbfleisch and Prentice, 2011; Hougaard and Hougaard, 2000) where the simplest non-trivial example is the illness-death model (Andersen and Keiding, 2002). This method models the transition intensities between any two stages of the disease process. However, the limitation of these transition intensities is that they only consider two stages at a time rather than a joint progression process. This is usually referred to as the "Markovian" assumption (Andersen et al., 1991). There are recent developments to handle situations where the Markovian assumption is invalid including a nonparametric estimation of the transition intensities (Meira-Machado et al., 2006), the use of Gamma frailty (Xu et al., 2010), and a branch of shared-frailty effect (Nielsen et al., 1992; Hougaard and Hougaard, 2000; Liu et al., 2004; Govindarajulu et al., 2011).

To address the issue of latent stage, Hu and Tsodikov (2014b) proposed a semi-parametric regression model that is based on Markov modulated processes with non-parametric time transformation models in the setting of metastasis - recurrence paradigm where they model time-to-recurrence as a marked-end-point. We extend the model in Hu and Tsodikov (2014a) to a cancer process with more than two stages that can be latent or observed, terminal or non-terminal. Using non-parametric maximum likelihood (NPMLE) (Tsodikov, 2003; Zeng and Lin, 2006; Chen, 2009), we provide an estimating procedure for the effect of treatment and other covariates on time-to-metastasis, time-to-cancer-diagnosis and time-to-death, while taking into account the dependence structure between them.

In our data analysis example, there are deaths due to other causes. This presents the issue of competing risks. There are two common approaches to address this issue: Switch to cumulative incidence functions (Fine and Gray, 1999), or use copulas to explain the dependence between potential competing risks. The former is not ideal

because inference then depends on censoring, which is unpleasant to keep track of since this would change with different populations and cause distributions. On the other hand, because of the non-identifiability issue, defining a model based on copulas would depend on strong assumptions; thus we decided to perform a sensitivity analysis to assess the effect of informative censoring (dependent risks) in our model.

In Section 2.2, we provide notation and model description (2.2.1), data structure (2.2.2), counting processes (2.2.3), joint likelihood (2.2.4), score functions and estimating procedure (2.2.5), and asymptotic properties (2.2.6). In Section 2.3, we use Monte Carlo simulations to verify the model's performance and its robustness against dependent censoring. In Section 2.4, we present an application to prostate cancer using SEER data to demonstrate the model and its ability to handle issues in multi-stage time-to-event data.

2.2 Methods

2.2.1 Notation and Model

We are interested in modeling data of cancer progression including diagnosis, metastasis, and death. Let T_U be the unobservable time-to-metastasis-onset; T_1^* and T_2^* be the potentially (due to censoring) observable times of diagnosis and death, respectively. For clarity, we denote $\boldsymbol{Z_1}$ as the sets of covariates that affect all stages, while $\boldsymbol{Z_2}$ refers to variables that start after diagnosis (e.g. for SEER analysis, $\boldsymbol{Z_1}$ includes race categories and $\boldsymbol{Z_2}$ includes metastatic status at diagnosis and treatments). In the context of the joint model, part of $\boldsymbol{Z_2}$ (metastatic status) represents a component of the joint response, and part (treatment) is treated as a covariate. While formally a fixed covariate in the post-diagnosis submodel, treatment absorbs future internal processes and implied decisions not measured in SEER data, and is potentially subject to unmeasured confounding because the treatment decision at

disgnosis may be based on unmeasured clinical characteristics. We recognize the potential pitfall of including internal covariates (Kalbfleisch and Prentice, 2011) and advocate caution interpreting treatment effects. Let $\boldsymbol{Z} = (\boldsymbol{Z_1} \cup \boldsymbol{Z_2})$ be the covariate vector.

Hu and Tsodikov (2014b) provide mechanistic justification for use of a common baseline hazard across hazards to different stages as a surograte of the same underlying disease pattern driving the time to diagnosis, metastasis and death. The assumption can be understood as a single nonparametrically specified time scale for all disease-related events.

We relax this assumption of a common hazard by creating additional parameters δ_1 and δ_2 for the nonparametric baseline cummulative hazard function H_t (instantaneous hazard h_t). Denote the cummulative hazards for metastasis, diagnosis, and death at time t_x respectively as H_x, $H_x^{\delta_1}$, and $H_x^{\delta_2}$. Thus, the analogous instantaneous hazards are h_x, $\delta_1 H_x^{\delta_1 - 1} h_x$, and $\delta_2 H_x^{\delta_2 - 1} h_x$. This parameterization can flexibly account for the various stretching and shrinking in baseline hazards for different cancer events, while saving us the computational burden of estimating multiple baseline hazards nonparametrically. Using $\mathbb{1}(\cdot)$ as an indicator function and a series of Cox proportional hazards models, the hazards for cancer events are parameterized as follows:

- Time-to-metastasis-onset $\lambda_U(t_u|Z, T_1) = h_u(\eta \mathbb{1}(t_u < T_1) + \tilde{\eta}\mathbb{1}(t_u \geq T_1))$,
- Time-to-diagnosis $\lambda_1(t_1 \mid Z, T_U) = \delta_1 H_1^{\delta_1 - 1} h_1 \theta \mu^{\mathbb{1}(t_1 \geq T_u)}$,
- Time-to-death $\lambda_2(t_2 \mid Z, T_U, T_1) = \delta_2 H_2^{\delta_2 - 1} h_2 \gamma \mathbb{1}(t_2 \geq \max\{T_u, T_1\})$,

where $\eta = e^{\beta_{\eta_1}^T \boldsymbol{Z_1}}$, $\tilde{\eta} = e^{\beta_{\tilde{\eta}}^T \boldsymbol{Z}}$, $\theta = e^{\beta_\theta^T \boldsymbol{Z_1}}$, $\mu = e^{\beta_\mu^T \boldsymbol{Z_1}}$, and $\gamma = e^{\beta_\gamma^T \boldsymbol{Z}}$.

Let $\boldsymbol{\beta} = [\beta_\eta, \beta_\theta, \beta_\mu, \beta_\gamma, \delta_1, \delta_2]$ be the coefficients vector. Let $S_x(\mathrm{t}|\cdot) = e^{-\int_0^t \lambda_x(s|\cdot)ds}$ and $f_x(\mathrm{t}|\cdot) = \lambda_x(t \mid \cdot) S_x(t \mid \cdot)$ be the corresponding conditional survival and density

functions, respectively, for hazard functions $\lambda_x(t \mid \cdot)$.

2.2.2 Observed Data Structure

Let V^* be the censoring time, independent of (T_1^*, T_2^*) given Z; and ξ be the maximum follow-up time of the study. For each subject $i = 1, 2, ..., n$ in the study, we observe $\{T_{1i},\ T_{2i},\ \Delta_{1i}^-, \Delta_{1i}^+, \Delta_{2i}^-, \Delta_{2i}^+, Z_i\}$, where $T_{1i} = \min(T_{1i}^*, V_i^*, \xi)$ is the time-to-first-observable-event (diagnosis or censoring); $T_{2i} = \min(T_{2i}^*, V_i^*, \xi)$ is the time-to-second-observable-event (death or censoring); $\Delta_{1i}^- = \mathbb{1}(T_{1i} = T_{1i}^*, T_{Ui} > T_{1i})$ is an indicator of diagnosis without metastasis; $\Delta_{1i}^+ = \mathbb{1}(T_{1i} = T_{1i}^*, T_{Ui} \leq T_{1i})$ is an indicator of diagnosis with metastasis; $\Delta_{2i}^- = \mathbb{1}(T_{2i} = T_{2i}^*)\Delta_{1i}^-$ is an indicator of observing death after diagnosis without metastasis (i.e.. metastasis occurs some time between diagnosis and death); and $\Delta_{2i}^+ = \mathbb{1}(T_{2i} = T_{2i}^*)\Delta_{1i}^+$ is an indicator of observing death after diagnosis with metastasis.

2.2.3 Combine population risk-set and cancer events

SEER case data only include subjects who would eventually be diagnosed with cancer. This is sufficient if we are only interested in estimating the risk of death assuming everyone was diagnosed with cancer. However, we are also interested in estimating the risk of being diagnosed with cancer in the general population. Thus, we need to combine a separate dataset that tracks the size of the at-risk population from which our cancer cases originate. This essentially provides life-table data for cancer diagnosis, followed by subject-level data for cancer death.

Figure 2.2 outlines how we use population counts as the risk set for diagnosis events. Let $Y_{1i}(t) = \mathbb{1}(t \leq T_{1i}^\star)$ and $N_{1i}(t) = \mathbb{1}(T_{1i}^* < t)$ denote subject i's risk process and counting process for cancer diagnosis, respectively. Let $Y_{2i}(t) = \mathbb{1}(T_{1i} < t < T_{2i})$ and $N_{2i}(t) = \mathbb{1}(T_{2i}^* < t)N_{1i}(t)$ denote the risk process and counting process for death

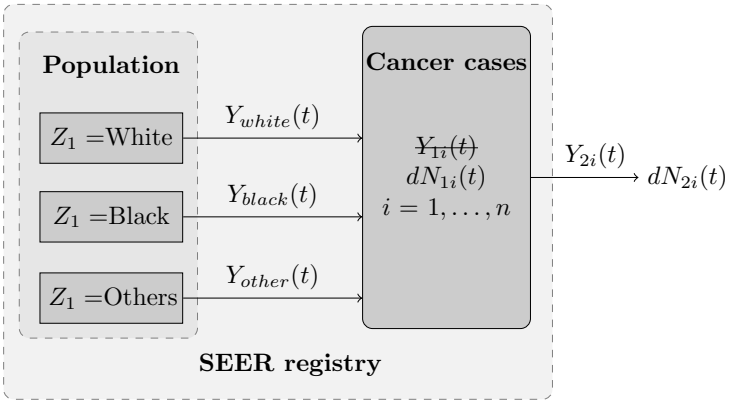


Figure 2.2: Using population counts as risk set for diagnosis events. Z_1 = race; $Y_{white}(t)$, $Y_{black}(t)$, and $Y_{other}(t)$ are population counts by race at age t; $dN_{1i}(t)$ denotes the counting process of breast cancer diagnosis for subject i from the population; $Y_{2i}(t)$ denotes subject i being at risk of death due to breast cancer; $dN_{2i}(t)$ is the counting process for this death event.

due to cancer. Since our prostate cancer diagnoses are from U.S. men between 1988-2003, we use counts of the male population during the same time period and in the same geographic location, stratified by $\boldsymbol{Z_1}$, to define the risk-set for cancer diagnosis. Let $Y_{z_1}(t)$ denote the population counts for subset that takes value $\boldsymbol{z_1}$ for covariate set $\boldsymbol{Z_1}$. Instead of summing up n cancer subjects, we sum up all categories of $\boldsymbol{Z_1}$ to get the whole at-risk population. Similarly, denote $N_{z_1}(t)=\sum_{i:Z_{1i}=z_1} \mathbb{1}(T^*_{1i} < t)$ as the counting process for the number of diagnoses from subgroup z_1.

2.2.4 Likelihood Construction

We integrate out time-to-metastasis-onset, T_U, by the appropriate time interval since the exact time is not observable. Denote $f_x f_y f_z(t_x, t_y, t_z)$ the joint probability of the sequence of events x, y, and z occuring at times t_x, t_y, t_z, respectively. The contribution of each subject to the likelihood falls into one of the following five scenarios (see Supplementary Materials A.1 for derivation):

(1) Subject is diagnosed without metastasis at t_1, then metastasizes, and dies at t_2:

$$\begin{aligned}\mathcal{L}_1^*(t_1,t_2) &= \int_{t_1}^{t_2} f_U f_1 f_2(t_u,t_1,t_2)dt_u \\ &= \delta_1\delta_2 H_1^{\delta_1-1}H_2^{\delta_2-1}h_1h_2\tilde{\eta}\theta\gamma e^{-H_1(\eta-\tilde{\eta})-H_1^{\delta_1}\theta-H_2^{\delta_2}\gamma}\int_{t_1}^{t_2} h_u e^{-H_u\tilde{\eta}+H_u^{\delta_2}\gamma}du.\end{aligned}$$

This joint probability can be written as $\mathcal{L}_1^* = \mathcal{L}_1\cdot\widetilde{\mathcal{L}}_1$, where $\mathcal{L}_1$ is the probability of observing up to time-of-diagnosis:

$$\mathcal{L}_1(t_1) = S_U f_1(t_1,t_1) = \delta_1 H_1^{\delta_1-1}h_1\theta e^{-H_1\eta-H_1^{\delta_1}\theta}.$$

(2) Subject is diagnosed without metastasis at t_1, may or may not have metastasis and is censored at t_2:

$$\begin{aligned}\mathcal{L}_2^*(t_1,t_2) &= \int_{t_1}^{t_2} f_U f_1 S_2(t_u,t_1,t_2)dt_u + \int_{t_2}^{\infty} f_U f_1 S_2(t_u,t_1,t_2)dt_u \\ &= \delta_1 H_1^{\delta_1-1}h_1\theta e^{-H_1(\eta-\tilde{\eta})-H_1^{\delta_1}\theta}\left[\int_{t_1}^{t_2}\tilde{\eta}h_u e^{-H_u\tilde{\eta}+H_u^{\delta_2}\gamma-H_2^{\delta_2}\gamma}du + e^{-H_2\tilde{\eta}}\right],\end{aligned}$$

where $\overline{\mu} = 1-\mu$. Alternatively, $\mathcal{L}_2^*$ can be rewritten as $\mathcal{L}_1\cdot\widetilde{\mathcal{L}}_2$.

(3) Subject is diagnosed with metastasis at t_1 and dies at t_2:

$$\begin{aligned}\mathcal{L}_3^*(t_1,t_2) &= \int_0^{t_1} f_U f_1 f_2(t_u,t_1,t_2)dt_u \\ &= \delta_1\delta_2 H_1^{\delta_1-1}H_2^{\delta_2-1}h_1h_2\eta\theta\mu\gamma e^{-H_1^{\delta_1}\theta\mu+H_1^{\delta_2}\gamma-H_2^{\delta_2}\gamma}\int_0^{t_1} h_u e^{-H_u\eta-H_u^{\delta_1}\theta\overline{\mu}}du.\end{aligned}$$

Alternatively, this can be specified as $\mathcal{L}_3^* = \mathcal{L}_3\cdot\widetilde{\mathcal{L}}_3$, where
$\mathcal{L}_3(t_1) = \int_0^{t_1} f_U f_1(t_u,t_1)dt_u = \delta_1 H_1^{\delta_1-1}h_1\eta\theta\mu e^{-H_1^{\delta_1}\theta\mu}\int_0^{t_1} h_u e^{-H_u\eta-H_u^{\delta_1}\theta\overline{\mu}}$.

(4) Subject is diagnosed with metastasis at t_1 and is censored at t_2:

$$\begin{aligned}\mathcal{L}_4^*(t_1,t_2) &= \int_0^{t_1} f_U f_1 S_2(t_u,t_1,t_2)dt_u \\ &= \delta_1 H_1^{\delta_1-1}h_1\eta\theta\mu e^{-H_1^{\delta_1}\theta\mu+H_1^{\delta_2}\gamma-H_2^{\delta_2}\gamma}\int_0^{t_1} h_u e^{-H_u\eta-H_u^{\delta_1}\theta\overline{\mu}}du.\end{aligned}$$

Alternatively, this can be specified as $\mathcal{L}_4^* = \mathcal{L}_3\cdot\widetilde{\mathcal{L}}_4$.

(5) Subject is censored at t_1 before any event is observed:

$$\begin{aligned}\mathcal{L}_5(t_1,t_2) &= \int_0^{t_1} f_U S_1 S_2(t_u,t_1,t_1)dt_u + S_U S_1 S_2(t_1,t_1,t_1)\\ &= \eta e^{-H_1^{\delta_1}\theta\mu}\int_0^{t_1} h_u e^{-H_u\eta - H_u^{\delta_1}\theta\overline{\mu}}du + e^{-H_1\eta - H_1^{\delta_1}\theta}.\end{aligned}$$

Thus, the contribution of subject i to the log-likelihood is:

$$\begin{aligned}\ell_i(\beta,H) &= \Delta_{1i}^-\Delta_{2i}^-\log(\mathcal{L}_{1i}^*) + \Delta_{1i}^-(1-\Delta_{2i}^-)\log(\mathcal{L}_{2i}^*) + \Delta_{1i}^+\Delta_{2i}^+\log(\mathcal{L}_{3i}^*)+\\ &\quad \Delta_{1i}^+(1-\Delta_{2i}^+)\log(\mathcal{L}_{4i}^*) + (1-\Delta_{1i}^- - \Delta_{1i}^+)\log(\mathcal{L}_{5i})\\ &= \left\{\Delta_{1i}^-\log(\mathcal{L}_{1i}) + \Delta_{1i}^+\log(\mathcal{L}_{3i}) + (1-\Delta_{1i}^- - \Delta_{1i}^+)\log(\mathcal{L}_{5i})\right\} +\\ &\quad \left\{\Delta_{1i}^-\left[\Delta_{2i}^-\log(\widetilde{\mathcal{L}}_{1i}) + (1-\Delta_{2i}^-)\log(\widetilde{\mathcal{L}}_{2i})\right]\right\} +\\ &\quad \left\{\Delta_{1i}^+\left[\Delta_{2i}^+\log(\widetilde{\mathcal{L}}_{3i}) + (1-\Delta_{2i}^+)\log(\widetilde{\mathcal{L}}_{4i})\right]\right\}. \qquad (2.1)\\ &= \int_0^{\xi}\left[\log\Theta_{1i}(H_t,\beta) + \log dH_t\right]dN_{1i}(t) - Y_{1i}(t)\Theta_{1i}(H_t,\beta)dH_t +\\ &\quad \left[\log\Theta_{2i}[H_t,\beta,t_1] + \log dH_t\right]dN_{2i}(t) - Y_{2i}(t)\Theta_{2i}[H_t,\beta,t_1]dH_t. \qquad (2.2)\end{aligned}$$

If we denote $\ell_{1i}, \ell_{2i}^-, \ell_{2i}^+$ as the quantities in each line of equation (2.1), it is easy to see that ℓ_1 is the contribution from diagnosis at t_1, while ℓ_2^- and ℓ_2^+ are the contribution from the subsequent time segments between diagnosis and death.

This is re-written using counting processes in equation (2.2), where $\Theta_{1i}(H_t,\beta)dH_t$ is the hazard of subject i being diagnosed at time t and $\Theta_{2i}(H_t,\beta,t_1)dH_t$ is the hazard of subject i dying of cancer, given diagnosis at time t_1 with (Met+) or without (Met-) metastasis. These quantities can be derived easily as follows:

$$\Theta_1(H_t,\beta) = \frac{\delta_1 H_t^{\delta_1-1}\theta + \delta_1 H_t^{\delta_1-1}\eta\theta\mu\int_0^t h_u e^{\eta(H_t-H_u)+\theta\overline{\mu}(H_t^{\delta_1}-H_u^{\delta_1})}du}{1+\eta\int_0^t h_u e^{\eta(H_t-H_u)+\theta\overline{\mu}(H_t^{\delta_1}-H_u^{\delta_1})}du} \qquad (2.3)$$

$$\Theta_2(H_t,\beta,t_1) = \begin{cases}\dfrac{\delta_2 H_t^{\delta_2-1}\tilde{\eta}\gamma\int_{t_1}^t h_u e^{\tilde{\eta}(H_t-H_u)-\gamma(H_t^{\delta_2}-H_u^{\delta_2})}du}{1+\tilde{\eta}\int_{t_1}^t h_u e^{\tilde{\eta}(H_t-H_u)-\gamma(H_t^{\delta_2}-H_u^{\delta_2})}du}, & \text{after Met-}\\ \delta_2 H_t^{\delta_2-1}\gamma, & \text{after Met+}\end{cases} \qquad (2.4)$$

(see Supplementary Materials A.2 for details of derivation).

For ease of notation in the rest of this paper, we shorten $\Theta_1(H_t, \beta)$ and $\Theta_2(H_t, \beta, t_1)$ to $\Theta_1(t)$ and $\Theta_2(t)$, respectively.

As discussed in section 2.2.3, we use population counts to account for the whole risk-set that includes people who never had prostate cancer. Let Z_1 denotes the set of covariates available at the population level before diagnosis. We stratified all subjects by Z_1 and set $\Theta_{z_1}(t) = \Theta_{1i}(t)$ where $Z_{1i} = z_1$. Once diagnosis occurs, SEER follows individual subjects; thus we have notation i for subject-level data. The joint log-likelihood in equation (2.2) is rewritten as follows:

$$\ell(\beta, H) = \int_0^\xi \sum_{z_1 \in Z_1} \left[\log\Theta_{z_1}(t) + \log dH_t\right] dN_{z_1}(t) - Y_{z_1}(t)\Theta_{z_1}(t)dH_t + \sum_{i=1}^{n} \left[\log\Theta_{2i}(t) + \log dH_t\right] dN_{2i}(t) - Y_{2i}(t)\Theta_{2i}(t)dH_t. \tag{2.5}$$

2.2.5 Nonparametric maximum likelihood estimator (NPMLE)

Score Functions

The score function with respect to β is

$$U_\beta = \frac{\partial \ell}{\partial \beta} = \sum_{z_1 \in Z_1} \int_0^\xi \left[\frac{\Theta_{z_1,\beta}(t)}{\Theta_{z_1}(t)} dN_{z_1}(t) - Y_{z_1}(t)\Theta_{z_1,\beta}(t)dH_t\right] + \sum_{i=1}^{n} \int_0^\xi \left[\frac{\Theta_{2i,\beta}(t)}{\Theta_{2i}(t)} dN_{2i}(t) - Y_{2i}(t)\Theta_{2i,\beta}(t)dH_t\right] \tag{2.6}$$

The score function with respect to H_t is

$$U_{H_t} = \frac{\partial \ell}{\partial dH_t} = \sum_{z_1 \in Z_1} \int_0^t \left[\frac{dN_{z_1}(x)}{dH_x} - Y_{z_1}(x)\Theta_{z_1}(x) + \int_{x+}^\xi \psi_{z_1}(u) dM_{z_1}(u)\right] + \sum_{i=1}^{n} \int_0^t \left[\frac{dN_{2i}(x)}{dH_x} - Y_{2i}(x)\Theta_{2i}(x) + \int_{x+}^\xi \psi_{2i}(u) dM_{2i}(u)\right], \tag{2.7}$$

where partial derivatives $\Theta_{\star,H}(t) = \frac{\partial \Theta_\star(t)}{\partial dH_t}$, $\Theta_{\star,\beta}(t) = \frac{\partial \Theta_\star(t)}{\partial \beta}$, and $\psi_\star(u) = \dfrac{\Theta_{\star,H}(u;\beta,H)}{\Theta_\star(u;\beta,H)}$.

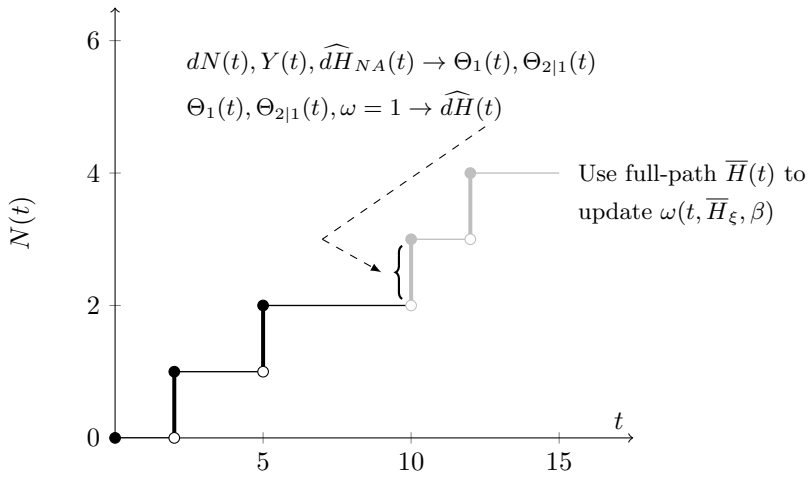


Figure 2.3: Calculate $\widehat{dH}(s)$ for all event time s, repeat until convergence.

Estimation Procedure

We set equations (2.7) to zero and solve for the NPMLE of the jump in the baseline hazard $d\hat{H}_x$ and the covariate effects $\hat{\beta}$. In particular,

$$d\hat{H}_x = \frac{\sum_{z_1 \in Z_1} dN_{z_1}(x) + \sum_{i=1}^{n} dN_{2i}(x)}{\sum_{z_1 \in Z_1} Y_{z_1}(x)\omega_{z_1}(\overline{H}_\xi, x, \beta)\Theta_{z_1}(x) + \sum_{i=1}^{n} Y_{2i}(x)\omega_{2i}(\overline{H}_\xi, x, \beta, t_{1i})\Theta_{2i}(x)}, \quad (2.8)$$

where $\omega_\star(\overline{H}_\xi, x, \beta) = 1 - \dfrac{\int_{x+}^{\xi} \psi_\star(t) dM_\star(t)}{\Theta_\star(x)}$, using the following martingales

$$dM_{z_1}(t) = dN_{z_1}(t) - Y_{z_1}(t)\Theta_{z_1}(t)dH_t, \text{ for } z_1 \in Z_1$$

$$dM_{2i}(t) = dN_{2i}(t) - Y_{2i}(t)\Theta_{2i}(t)dH_t, \text{ for } i = 1, 2, ..., n.$$

Denote $\boldsymbol{t} = \{t_{(1)}, t_{(2)}, ..., \xi\}$ as the set of unique time, with ξ being the last time point, where event occurs in our data. Maximization of the log-likelihood with respect to $H(\boldsymbol{t})$ and β simultaneously is done using a profile likelihood approach and an iterative reweighting algorithm (Chen, 2009) as follows:

1. Start with candidate $\boldsymbol{\beta}^\star$, weight $\omega^{(0)}(t) = 1$ and $\widehat{dH}^{(0)}(t)$ = Nelson-Aalen estimator.

2. As illustrated in Figure 2.3, repeat these steps until convergence of $\widehat{dH(\boldsymbol{t})}$:

   - Use $\widehat{dH}^{(0)}(x)$ to calculate $\Theta_{z_1}^{(0)}(x), \forall x \in \boldsymbol{t}$.
   - Use $\Theta_{z_1}^{(0)}(x)$ and $\omega^{(0)}$ and eq. (2.8) to calculate $\widehat{dH}^{(1)}(x)$. With the weight fixed, the right hand side of eq.(2.8) only depends on prior jumps $H(t)$, $\forall t < x$.
   - Sequentially do (b) along all $x \in \boldsymbol{t}$ to get full path $\widehat{H^{(1)}}$.
   - Use full path $\widehat{H^{(1)}}$ to update weights $\omega^{(1)}$.

3. Calculate log-likelihood $\ell_{\text{pr}}(\beta^\star, \widehat{H})$ using eq.(2.5).

4. Repeat step (1-4) to search $\hat{\beta} = \arg\max_{\beta} \ell(\beta)$.

Standard errors are estimated using the Hessian of the profile log-likelihood. See Supplementary Materials A.4 for justification.

2.2.6 Asymptotic properties

Martingale properties

From equations (2.6) and (2.7), the score functions for H_t and β can be rewritten as:

$$U_\beta = \sum_{z_1 \in Z_1} \int_0^\xi \frac{\Theta_{z_1,\beta}(t)}{\Theta_{z_1}(t)} dM_{z_1}(t) + \sum_{i=1}^n \int_0^\xi \frac{\Theta_{2i,\beta}(t)}{\Theta_{2i}(t)} dM_{2i}(t), \tag{2.9}$$

$$\begin{aligned} U_{H_t} = &\sum_{z_1 \in Z_1} \int_0^t \left\{ \frac{dM_{z_1}(x)}{dH_x} + \int_{x+}^\xi \psi_{z_1}(u) dM_{z_1}(u) \right\} \\ &+ \sum_{i=1}^n \int_0^t \left\{ \frac{dM_{2i}(x)}{dH_x} + \int_{x+}^\xi \psi_{2i}(u) dM_{2i}(u) \right\}, \end{aligned} \tag{2.10}$$

which are both martingales at the true model (see Supplementary Materials A.3 for proof).

Consistency and weak convergence

Consistency and weak convergence result of the NPMLE $\hat{\Omega} = (\hat{\beta}, \hat{H}_t)$ is adapted from Hu and Tsodikov (2014a). Assuming regularity conditions hold,

1. With probability one, $\hat{\beta}$ converges to β^0, $\hat{H}_t$ converges to H_t^0 uniformly in the interval $[0, \xi]$, where H_t^0, β^0 are the true values of H_t, β.
2. $n^{1/2}\{\hat{\beta} - \beta^0, \hat{H}_t - H_t^0\}$ converges weakly to a zero-mean Gaussian process. In addition, $n\mathcal{E}^T(\mathcal{I}_n)^{-1}\mathcal{E}$ converges in probability to the asymptotic variance-covariance function of the linear functional $n^{1/2}\left\{\int_0^\xi a^T(\hat{\beta} - \beta^0) + b(t)^T(\hat{H}_t - H_t^0)\right\}$, where a is real vector, $b(t)$ is a function with bounded total variation in $[0; \xi]$, $\mathcal{E}^T = (a^T, B^T)$ where B is a vector of the values of b at the jumps of $\hat{H}$, $\mathcal{I}_n$ is the negative Hessian matrix of the observed log-likelihood function with respect to $\hat{\Omega}$.

2.3 Simulation Studies

2.3.1 Different baseline hazards and sample sizes

We performed Monte Carlo simulations using R software to assess the proposed methodology. For each scenario, 1000 repetitions with sample sizes 1000 and 2000 were performed. The parameters were set as follow: Each scenario included covariates $Z_1 \sim \text{Bern}(0.5)$, $Z_2 \sim \text{N}(2, 0.5)$, and an increasing baseline hazard following a Weibull$(2, 2)$ distribution. Baseline hazard transformation is set at $\delta_1 = 0.8$ and $\delta_2 = 1.5$. Time-to-metastasis-onset, time-to-diagnosis, and time-to-death were generated under Cox proportional hazards using parameters $\beta_\eta = [1.5, 0.9]$, $\beta_\theta = [-0.6, 0.5]$, $\beta_\mu = [1, 1.2]$, $\beta_\gamma = [1.6, -0.7, 0.4]$, and are rounded to 2 decimal places.

Independent censoring time followed a uniform distribution $U(1.5, 3)$ for the increasing hazard, which yielded approximately 20% censoring before the first event

Table 2.1: Simulation result: Increasing baseline hazard

		Metastasis		Diagnosis				Death			H_t transform	
N	Statistics	$\beta_{\eta 1}$	$\beta_{\eta 2}$	$\beta_{\theta 0}$	$\beta_{\theta 1}$	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\gamma 0}$	$\beta_{\gamma 1}$	$\beta_{\gamma 2}$	δ_1	δ_2
	True	1.5	0.9	-0.6	0.5	1	1.2	1.6	-0.7	0.4	0.8	1.5
1000	Bias	0.03	0.00	.01	0.02	0.09	-0.02	0.04	-0.01	-0.01	0.01	-0.01
	ESE	0.24	0.12	0.11	0.20	0.61	0.51	0.42	0.10	0.08	0.19	0.21
	ASE	0.23	0.11	0.11	0.20	0.55	0.45	0.40	0.10	0.08	0.16	0.19
	CP	92	91.8	93.8	92.4	93.2	94.3	91.8	95.7	95.3	88.4	88.7
2000	Bias	0.03	0.00	.01	0.01	0.06	-0.04	0.03	-0.01	-0.00	0.00	-0.02
	ESE	0.16	0.08	0.07	0.14	0.41	0.33	0.29	0.07	0.05	0.11	0.12
	ASE	0.16	0.08	0.08	0.14	0.39	0.31	0.29	0.07	0.06	0.11	0.13
	CP	94.5	94.3	96.7	94.3	94.9	93.2	92.7	94.9	95.6	93.8	93.6

ESE: empirical standard errors based on 1,000 estimates
ASE: average of estimated standard errors
CP: 95% Coverage Probability

(diagnosis) and 30% censoring before the second event (death). The convergence criterion was set at 10^{-5} for change in baseline hazard estimation and improvement in the full likelihood.

The simulation results are summarized in Table 2.1. The iterative reweighting algorithm works well for both scenarios of increasing and decreasing baseline hazards (Weibull$(0.5, 2)$, results not shown). The log-hazard-ratio estimates for covariate effects on time-to-latent-metastasis, time-to-diagnosis, and time-to-death are almost unbiased at sample size n=1000 and get more accurate with larger sample size n=2000. The empirical standard errors (ESE) are close to the mean of estimated theoretical errors (ASE), which validate the performance of the variance estimators. Coefficient estimators improve with increasing sample size as the estimated variances decrease and 95% confidence intervals get narrower. The 95% coverage probabilities for all coefficient estimators remain as expected for both sample sizes.

Estimates for baseline hazard transformation δ_1 and δ_2 are also accurate in terms of small bias and consistent standard error estimation. The 95% coverages for these parameters slightly underperformed at sample size n=1000, and improved to the expected probabilities at n=2000.

2.3.2 Informative censoring

We performed sensitivity analyses to assess the performance of the estimating procedures and prediction function in the presence of time-dependent covariates and dependent censoring. The simulated variables are similar to the SEER example data as follows: $\boldsymbol{Z_1} \sim Unif(1,3)$ are 3 race indicators, $\boldsymbol{Z_2} \sim Unif(1,4)$ are 4 treatment indicators . Baseline hazard follows Weibull$(2,10)$ distribution and $\delta_1 = \delta_2 = 1$.

For the independent censoring scenario, censoring time $V^\star \sim Unif(0,30)$ results in 15 % censored before diagnosis and 30 % censored before death. For the dependent censoring scenarios, we use the following copula for bivariate survival data: $S(t,v) = G\left\{G^{-1}[S_2(t)] + G^{-1}[S_V(v)]\right\}$ where $G(s) = \left(\frac{\phi}{\phi+s}\right)^\rho$, S_2 and S_V are survival functions for death-due-to-prostate-cancer and death-due-to-other-causes (i.e. censored events) respectively. Thus, smaller ρ corresponds to higher dependence. This simulation was run 1000 times at sample size 500.

Table 2.2: Simulation result: Dependent censoring. LHR = log hazard ratio, CP = 95 % coverage probability

		LHR for overall survival time (γ)					
	True	**0.9**	**0.5**	**0.6**	**-0.6**	**-2**	**-1.4**
Independent	CP(%)	91.9	92.7	93.7	93.3	90.5	92.9
$\rho = 5$	CP(%)	91.1	95.3	93.7	92.1	90.7	92.9
$\rho = 2$	CP(%)	91.3	90.9	92.8	91.9	90.2	93.4
$\rho = 1$	CP(%)	90.5	92.3	90.0	92.1	89.6	90.2
$\rho = 0.5^*$	CP(%)	86.3	86.5	87.5	92.7	86.5	90.9

* smaller ρ corresponds to higher dependence

Table 2.2 shows the estimated log-hazard-ratios (LHRs) for overall survival time (LHRs for time-to-metastasis and time-to-diagnosis are mostly unaffected, see Sup-

plementary Materials A.5.2). In simulations where the dependence is moderate, LHRs and standard errors are accurately estimated with finite sample size. This performance, however, is effected by highly correlated censoring, In our most extreme case ($\rho = 0.5$), we see some deviation from the normal 95% coverage for 5 out of 16 estimated coefficients. Even then, their worst coverage fluctuates down to around 86%. A full-fledged dependent competing risks analyses is needed under stronger dependence.

2.4 Application to SEER prostate cancer data

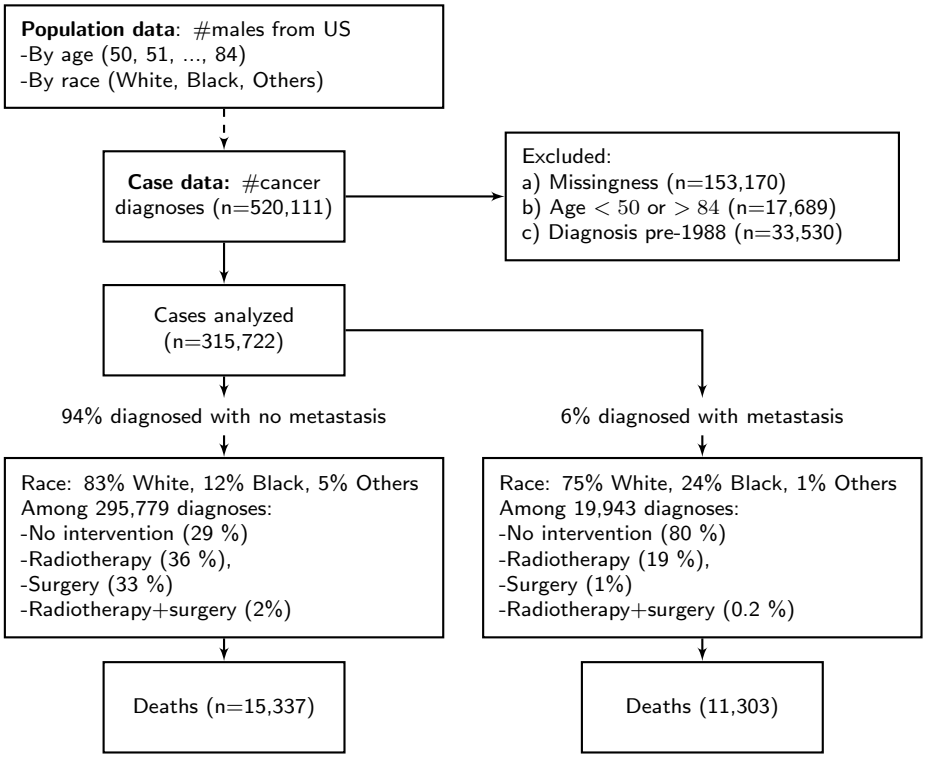
2.4.1 Data

We applied our proposed model to the motivating setting of prostate cancer using a case file and population-at-risk counts available from the SEER registry. The proposed model in its current specification (Figure 2.1) is tailored to this disease since virtually all deaths from prostate cancer occur after the cancer has metastasized. SEER, controlled by the National Cancer Institute, registers yearly incidence and population data on various cancers in the U.S. starting from time of diagnosis. Figure 2.4 outlines some summary statistics for the following 2 datasets:

SEER population data: A large part of the population at risk was missing since the incidence dataset only followed subjects who were diagnosed with cancer. Therefore, we used population counts from which the cancer cases originated (i.e. U.S. male population age 50-85 between 1988-2003) stratified by race, gender, age, and year as the risk set of cancer diagnosis ($Y_{z_1}(t)$ in eq.2.5). Although it is possible to focus only on the case file starting from diagnosis, important information regarding the latent metastasis process which occurs before diagnosis would be lost.

SEER incidence data: SEER case data provides race, gender, age, metastasis status at diagnosis, and subsequent survival time of men who were diagnosed with

Figure 2.4: SEER population and prostate cancer case data (USA, 1988-2003).



prostate cancer between 1973-2003. Out of all recorded prostate cancer diagnoses in this time period (n=520,111), we concentrated on men between the ages of 50-84. Subjects with unknown race, age, or treatment were removed under the assumption of missing at random (n=315,722 remained).

Besides clinical variables, another factor that could potentially confound the effect of treatment on survival is the use of prostate-specific antigen (PSA) screening which began around 1988. The benefits of PSA screening are still debatable since it can result in overdiagnosis (i.e. detection of disease that would not become symptomatic during a patient's lifetime in the absence of screening). Overdiagnosis can positively affect the overall survival from diagnosis under screening compared to no screening prior to 1988 (largely an artifact). To avoid confounding, we focused on cases with diagnosis after 1988. Amongst these men, 94% were diagnosed without metastasis

and 6% were diagnosed with metastasis. From 315,722 cancer diagnoses, 26,640 were recorded to die from prostate cancer by 12/31/2003. Deaths from other causes and those alive at the end of study were censored.

Metastasis status has a clear effect on treatment distribution. Radiotherapy and surgery (prostatectomy) were the most popular treatment options for patients diagnosed without metastasis. 29% received no treatment, 36% received radiotherapy, 33% had surgery, and 2% had a combination treatment of radiotherapy and surgery. On the other hand, most patients who were diagnosed with metastasis opted out of all treatments: 80% received no treatment, 19% received radiotherapy, and the rest received surgery.

Note that unlike race, treatments are time-dependent (zero prior to diagnosis and thus only affect subsequent time intervals between diagnosis and death). These specifications can easily be accommodated by taking race as a covariate Z_1 in the hazard of initial diagnosis $\lambda_1(t|Z_1)$. For patients who were diagnosed with local cancer (no metastasis), treatment are time-dependent covariate Z_2 effecting subsequent hazard of metastasis $\lambda_U(t|Z_1, Z_2)$. Because treatment are assigned dynamically based on response to prior therapy, special care has to be made to interpret the treatment effects on overall survival time. Beside the main treatment indicators, we add interaction terms with metastasis-status-at-diagnosis to account for the potential varying treatment effect due to the patient's underlying disease process.

2.4.2 Results

The estimated log-hazard ratios (LHR) for time-to-diagnosis, metastasis, and death are summarized in Tables 2.3, 2.4, and 2.5. These LHR's denote instantaneous relative risks with respect to the baseline group of white men who did not receive treatment (i.e. elected to wait after cancer diagnosis). A positive coefficient

signifies higher risk and thus a shorter time to event. The risk of metastasis (Table 2.3) changes at point of diagnosis (due to introduced interventions). Similarly, the risk of being diagnosed of cancer (Table 2.4) increases in all races once metastasis occured.

We found evidence of differences in the baseline hazards of three cancer stages beyond proportionality. Once we estimated the baseline hazard for metastasis events, the baseline hazards for diagnosis and death were estimated by taking exponent 1.66 and 1.02, respectively. This also demonstrate our model's improvement for relaxing the common baseline hazard assumption made in Hu and Tsodikov (2014a,b).

We found evidence of difference in baseline hazards of metastasis and diagnosis beyond proportionality. Once we estimated the baseline hazard for metastasis events, the baseline hazards for diagnosis was estimated by taking exponent 1.66 ($P<0.001$). There is no evidence of change beyond proportionality for baseline hazard of death ($P=0.27$). This also demonstrate our model's improvement for relaxing the common baseline hazard assumption made in Hu and Tsodikov (2014a,b).

Race and treatment were significantly associated with time-to-metastasis. White men were found to have lower risk of metastasis overall compared to same-age black men (LHR=0.08, $P<0.01$) and same-age men from other races (LHR=0.23, $P<0.001$). For patients diagnosed with prostate cancer but with no metastasis, treatment affected subsequent time-to-metastasis. Compared to no treatment (i.e. waiting), surgery (radical prostatectomy) by itself did not have any statistically significant effect ($P=0.16$). Radiotherapy alone and in combination with surgery increased metastasis (LHR=1.97, $P<0.001$). Similar effects of surgical removal of the primary tumor increasing risk of metastasis have been reported in cancer literature (O'Reilly et al., 1994; Smolle et al., 1997; Biki et al., 2008; Neeman and Ben-Eliyahu,

Table 2.3: SEER prostate cancer analysis: Time-to-metastasis-onset

LHR before diagnosis (β_η)

Covariate	Estimate	SE	P-value
Z_{Black}	0.08	0.03	<.01
Z_{Other}	0.23	0.05	<.001

LHR after diagnosis ($\beta_{\tilde{\eta}}$)

Covariate	Estimate	SE	P-value
$Intercept$	4.03	0.04	<.001
$Z_{Surgery}$	0.44	0.31	0.16
$Z_{Radiotherapy}$	0.81	0.10	<.001
Z_{Combo}	1.97	0.32	<.001

Table 2.4: SEER prostate cancer analysis: Time-to-diagnosis

Baseline hazard transformation

Parameter	Estimate	SE	P-value
δ_1	1.66	0.001	<.001

LHR before metastasis (β_θ)

Covariate	Estimate	SE	P-value
$Intercept$	4.53	0.02	<.001
Z_{Black}	0.42	0.01	<.001
Z_{Other}	0.60	0.01	<.001

Change to LHR after metastasis (β_μ)

Covariate	Estimate	SE	P-value
$Intercept$	2.95	0.02	<.001
Z_{Black}	0.81	0.05	<.001
Z_{Other}	0.41	0.07	<.001

Table 2.5: SEER prostate cancer analysis: Time-to-death

Baseline hazard transformation

Parameter	Estimate	SE	P-value
δ_2	1.02	0.02	0.27

LHR after diagnosis and metastasis (β_γ)

Covariate	Estimate	SE	P-value
$Intercept$	6.78	0.10	<.001
Z_{Black}	0.10	0.02	<.001
Z_{Other}	-0.45	0.04	<.001
$Z_{Surgery}$	-1.72	0.15	<.001
$Z_{Radiotherapy}$	0.09	0.02	<.001
Z_{Combo}	-1.82	0.41	<.001
$\mathbb{1}_{Local} * Z_{Waiting}$	1.52	0.06	<.001
$\mathbb{1}_{Local} * Z_{Surgery}$	0.28	0.29	0.33
$\mathbb{1}_{Local} * Z_{Radiotherapy}$	-0.46	0.11	<.001
$\mathbb{1}_{Local} * Z_{Combo}$	0.11	0.49	0.81

LHR: Log hazard ratio versus reference group (white men who elected "waiting" at diagnosis). SE: Standard error

2013) and could be overlooked in other models that do not account for the latent metastasis process. This may also be due to a confounder in the physicians' decision to perform surgeries only on more severe prostate cancer diagnoses or support the theory that removing the primary tumor along with its foreign antigens weakens the signal for the immune system's attack, making other body parts more susceptible to metastasis.

The risk of being diagnosed with prostate cancer differs by race. In the U.S. population of men older than 50 years, white men had slightly lower risk compared to same age black men (LHR=0.42, P<0.001) and men from other races(LHR=0.60, P<0.001). The occurence of metastasis before diagnosis also quickened time-to-diagnosis: This risk increased among white men (LHR=2.95, P<0.001), more among black men (LHR = 2.95 + 0.81, P<0.001), and most among men of other races (LHR = 2.95 + 0.41, P<0.001).

The risk of death started after diagosis and metastasis occured. For patients who were diagnosed after metastases, surgery was found to significantly lower risk and prolong overall surival time (LHR=-1.72, P<.001). A combination of both surgery and radiotherapy was also beneficial (LHR=-1.82, P<0.001). Radiotherapy by itself slightly lower the survival rate but the effect was small (LHR=0.09, P=<0.001). For patients diagnosed prior to metastasis, we include interaction terms to account for the effect of subsequent treatment assigned dynamically after response to initial treatment. Among these patients, those assigned "waiting" and radiotherapy at diagnosis were found to have differences in risk-of-death compared to those diagnosed with metastasis (LHR=1.52, P<0.001; LHR=-0.46, P<0.001).

Due to the potential confounder of physician's subjectivity and other unaccounted covariates, the interpretation of treatment difference may suffer if interpreted as

treatment effect in the context of a randomized clinical trial. However, if we are willing to assume that treatment is given based on observation of the covariates included in the model, is a function of these covariates only, and ignore the extent to which treatment depends on unobserved characteristics that are not controlled in the model, then the model provides rough estimates of the treatment effect. Since this is observational registry data, the treatment effect is interpreted with caution.

If we adopt the view that there are unaccounted covariates that influence treatment choice, then we can use treatment as a surrogate for these covariates. This means we do not interpret the treatment coefficients as purely the treatment effect as in a clinical trial, but as an indication that perhaps more aggressive treatment is given to patients that are worse off. Then by adjusting for treatment, the other covariates would also be adjusted for in the model.

Beside race and the use of PSA screening, there are other potential confounders that could bias treatment effects which were not captured in our model. One such variable is the tumor grade (i.e. level of differentiation of tumor cells). There is a biological concensus that tumors containing well-differentiated cells are less aggressive due to the close resemblance of these cells to normal cells. Another factor that could not easily be captured is the doctor's decision in treatment choice which may reflect disease severity (not captured by stage), experience, and patient decision. For the purpose of treatment comparisons, statistical models assume similar representation of patients in each treatment group; however, this may not be the case in reality. Physicians usually recommend more aggressive treatments when patients are diagnosed with more advanced disease status and are young and healthy enough to endure it. Treatment regimens may also change when a patient's condition declines, improves, or experiences side effects. Currently SEER data does not include more

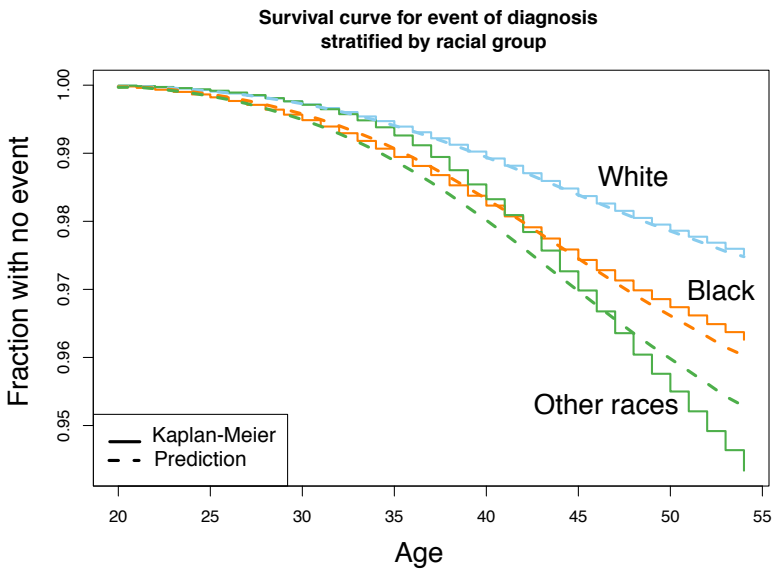


Figure 2.5: Fitted plot: Survival estimates for time-to-diagnosis. Proposed model (dotted lines, eq.2.11) closely matches Kaplan-Meier (KM) estimates (solid lines).

clinical information to adjust for this subjectivity or changing treatment regimen.

2.4.3 Predicted Survival Rates

From our fitted model, we can derive closed-form formulas to predict the time to each disease stage. This carries clinical significance because these survival rates could be used to inform treatment decisions at each time point for future patients diagnosed with prostate cancer. In this section, we derive the formulae for survival probabilities at time t for diagnosis and death in the SEER example. These predicted survival rates were compared against the empirical survival rates to validate our proposed model. The survival probability for latent metastasis is not shown since we do not have empirical time-to-metastasis-onset in SEER data to compare against.

The predicted probability of no cancer diagnosis ($\widetilde{\mathcal{L}}_5$ from section 2.2.4) is given by:

$$S_1(t_1|Z) = \eta e^{-H_1^{\delta_1}\theta\mu} \int_0^{t_1} h_u e^{-H_u\eta - H_u^{\delta_1}\theta\overline{\mu}} du + e^{-H_1\eta - H_1^{\delta_1}\theta}. \tag{2.11}$$

Figure 2.5 plots the Kaplan-Meier (KM) curve and predicted survival rate for prostate-cancer diagnosis in each racial group (eq. 2.11). Note that all the survival probabilities are very close to 1 because we used the full US male population ages 50-84 as the risk-set, which is much larger than the size of the incidence dataset. This figure validates the estimating procedure since the predicted and empirical survival rates match closely for white and black men. There is a slight deviation between the predicted survial rates and empirical KM curve amongst men from other races, which could be due to the relatively much smaller sample size in this group compared to white or black men.

The conditional survival rate given time-of-diagnosis can be derived and used to compare how different treatments affect time-to-death among those that were diagnosed with cancer at the same time. This conditional survival rate is derived from $\mathcal{L}_2^\star/\mathcal{L}_1$ (section 2.2.4), which gives:

$$S_2(t_2|t_1, T_u > t_1, Z) = \tilde{\eta} \int_{t_1}^{t_2} h_u e^{-\tilde{\eta}(H_u - H_1) - \gamma(H_2^{\delta_2} - H_u^{\delta_2})} du + e^{-\tilde{\eta}(H_2 - H_1)}. \tag{2.12}$$

Figure 2.6 compares the survival rates estimated from our proposed model (eq. 2.12) for death due to prostate cancer, conditioning on diagnosis time at 70 years old, with the empirical Kaplan-Meier curve. We plot the subset of white men who received no treatment (waiting), surgery, or radiotherapy. There were not enough events recorded for combination treatment once we stratified by diagnosis time (only 2 % received this type of treatment overall, 15 deaths recorded from 214 patients who were diagnosed at age 70). The validity of the prediction function was further confirmed via simulations that resemble the SEER setting but with equally distributed sample sizes between stratified groups (results not shown).

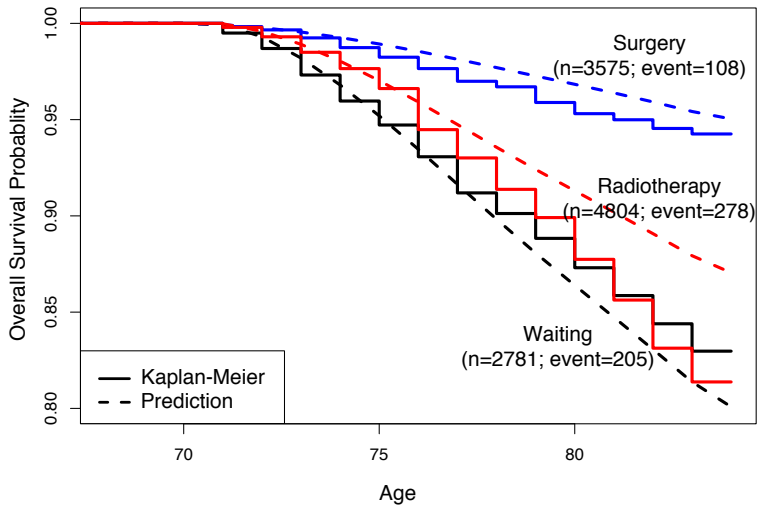


Figure 2.6: Fitted plot: Survival estimates (dotted lines, eq. 2.12) and empirical Kaplan-Meier curves (solid lines) for death from prostate cancer among men who were diagnosed at 70.

2.5 Discussion

We presented an NPMLE estimator and proposed a model that simultaneously estimated time to different events, where events were both latent and observable. We relaxed the assumption of one common baseline hazards at different stages. The two power parameters allow for flexible changes in baseline hazards of metastasis, diagnosis, and death. Moreover, because we estimated only one baseline hazard non-parametrically, our model was still more efficient compared to cause-specific models. In the likelihood, the latent disease event is integrated within the appropriate time interval anchored by observed events. The non-parametric baseline hazard and covariate effects are iteratively estimated using a weighted Breslow-type estimator derived from the score function. We verified the performance of this estimating algorithm through Monte Carlo simulations at two sample sizes and types of baseline hazards. Application of the proposed model to SEER prostate cancer cases among

U.S. men ages 50-84 revealed different ways through which treatments ultimately affected overall survival.

Our model differs from existing methods in that it can flexibly describe the relationship between outcomes in multiple stages beyond the fact that one outcome can prevent another from being observed. In SEER analysis, we specified that cancer death occurs after metastasis. This dependence can be even more nuanced by specifying, for example, that the quickness by which metastasis occurs affects how quickly death subsequently occurs. In cancer studies, where it is intuitively and biologically plausible that the "competing" stages are dependent, our model offers a convenient way to summarize these effects compared to existing methods such as cumulative incidence functions in a competing risk model or transitional intensity in an illness-death model. Moreoever, the alterable components in the proposed model's stage-specific hazards and incorporation of time-dependent covariates allow for easy adaptability to reflect covariates' effects on different parts of the disease process.

Traditional methods also fail to explicitly model the time-to-a-latent-event. In certain diseases, these latent outcomes may carry significant quality-of-life importance. Therefore, knowing how covariates affect these unseen events is sometimes as important as studying the observable outcomes. Our current model is equipped to infer this and incorporate scenarios where the event is partially missing in a subset of the population.

The model focuses on a mechanistic description of cancer progression leading to cancer-specific death. Its clinical value lies in the ability to study treatment and other effects on the biological process, and the ability to predict the latent process given clinical information available on the subject. This prediction can be dynamically updated as more information becomes available in the course of subject's life

pre-diagnosis, diagnosis and subsequent follow-up. At the same time, there are limitations in the clinical utility of this model. In particular, studying the potential progression of cancer for a subject who died from other causes carries limited clinical relevance. Other causes of death were treated as a nuisance in this paper. The model's extension to a competing risks framework is an interesting future development that would allow better characterization of clinically relevant effects and potential exploration of possible association between cancer of interest, and its treatments, and the risks of comorbidities and death due to other causes.

Currently, the earliest event of interest in our model is metastasis (or diagnosis if it happens prior). We recognize that there is a disease-free period before a tumor is initiated. However, modeling this process means making assumption about covariates that are never observed. Our model addresses the challenge of latent variable time-to-metastasis because this is observed indirectly via metastasis status at cancer diagnosis time. If we introduce a point of tumor initiation, where risk of diagnosis and metastasis are zero, we would need even stronger assumptions than the current model.

While we denote these stages as unique, the demarcation of these stages is not always clear. Such is the case with metastasis in our SEER example. Treatments taken from the time-of-diagnosis, while not affecting time-to-metastasis in patients whose cancer had already metastasized prior to diagnosis, should have affect time-to-metastasis in the rest of patients. This ability to select different covariates for each stage, and at the same time allow latter stages to be dependent upon earlier stages, may reveal the effect of treatment at different stages and help guide clinical decision making.

The optional pathways in the joint likelihood further makes our proposed model

adapatable to describe different disease processes. In some cancers, death occurs after metastasis. In others, death may occur at any point after diagnosis. To model the latter, an additional term to the likelihood can reflect the possibility of a patient dying without intermediate stages, while the estimation procedure remains the same. This can also be generalized to model different segments of the diseases, such as the time between recurrence and death.

In this paper treatment is a choice that the physician makes at diagnosis, based on the observed information at that point, under the assumption of no unmeasured confounding. It is common to consider treatment as a fixed (external) covariate for the post-diagnosis model in this context. In reality though treatment represents a label for what essentially is a dynamic treatment regimen that will follow diagnosis. One potential problem is that this future information depends on internal processes in the subject that are unmeasured in registry data, and are not included in the current model. Another problem is that the assumption of no unmeasured confounding is likely unrealistic, and treatment decisions at diagnosis occur in part based on some unmeasured clinical information. All this calls for caution interpreting treatment effects in this paper as they absorb these effects.

CHAPTER III

Shared Frailty in Joint Model of Cancer Incidence and Mortality

3.1 Introduction

It is common for diseases to progress through multiple events between clinical onset and death. In the context of cancer studies, the age at which a patient is diagnosed with cancer can also be described as the time-to-diagnosis. After diagnosis, there may be other events of interest such as remission, recurrence, and death. Usually these events are analyzed independently using the time-to-event as an outcome. However, the relationship between the events driven by the same disease calls for a joint analysis.

In this paper, we focus on the time-to-diagnosis and the time-to-death. Most existing methods use time-between-diagnosis-and-death as the outcome of interest to draw inference about treatment effect. This circumvents the need to look at the two different outcomes, but assumes that there is no difference in post-diagnosis survival conditioning on diagnosis time. This assumption is not valid for many diseases. Intuitively, one may expect patients who are diagnosed at younger ages to survive longer compared to those who are diagnosed at older ages. Yet, positive correlation between diagnosis time and post diagnosis survival time have been observed in some diseases. For example, consider the motivating setting of SEER breast cancer data.

It has been shown that younger breast cancer patients have worse survival compared to women diagnosed at an older age Tsodikov (2002). The speculation was that more aggressive cancers grow faster causing them to surface earlier and have shorter survival. As we will show (section 3.5), even models that include diagnosis age as a covariate for risk of death is inadequate to capture this correlation. Thus, we advocate for joint modeling time-to-diagnosis (and potentially other intermediate events) and time-to-death using frailties to explain the correlation induced by the shared disease process.

Existing models often do not adequately capture the relationship between various stages in the natural progression of some diseases. Semi-competing risks (Fine et al., 2001) can model events in which one occurrence prevents the other from being observed, but cannot model events that occur sequentially. The illness-death model (Andersen and Keiding, 2002) was introduced as the simplest variation of a multi-stage model (Andersen et al., 1991; Xu et al., 2010; Kalbfleisch and Prentice, 2011; Hougaard and Hougaard, 2000). This models the probabilities of moving between any two specific stages, but it is not able to describe how the time spent in one stage might affect time spent between other stages (e.g. breast cancer patients diagnosed at younger ages tend to have shorter post-diagnosis survival time). Recent works in joint modeling proposed a semiparametric joint likelihood approach with a common baseline hazard for time to two different events (Hu and Tsodikov, 2014a; Rice and Tsodikov, 2016). This model, however, is also unable to predict the positive correlation between diagnosis time and post-diagnosis survival as illustrated in breast cancer patients described in Tsodikov (2002).

Our model incorporates two nonparametrically specified baseline hazards for diagnosis and death. We found in analyzing SEER data that there is a level of dependence

between disease stages that could not be accounted for by using common covariates and indicators. Therefore, we have included a frailty term as a latent factor that correlates time-to-diagnosis and post-diagnosis survival time. The shared-frailty approach is discussed in Hougaard and Hougaard (2000), Klein (1992), Andersen et al. (1997), and Zeng et al. (2009). The expectation of the joint likelihood with respect to this latent frailty term is evaluated via Laplace transforms (Hougaard, 1986). From this averaged joint likelihood, we are able to derive a nonparametric maximum likelihood (NPMLE) estimating procedure to estimate unique baseline hazards for diagnosis and death, treatment effects, and the variance of the frailty term. We derive a closed form expression for post-diagnosis survival prediction and capture survival dependence on the diagnosis time.

Figure 3.1 shows a typical cancer's progression as recorded in the National Cancer Institutes Surveillance Epidemiology and End Results (SEER) registry. To analyze such data, we use population counts that include healthy individuals to account for those at risk of a breast cancer diagnosis from birth. Once diagnosed with cancer, we use individual level data to access risk of death due to breast cancer. Patients lost to follow-up are considered censored at that time.

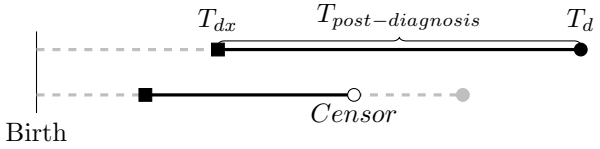


Figure 3.1: Diagram of cancer progression: Patients are diagnosed with cancer at time T_{dx} and die at time T_d. $T_{post-diagnosis}$ indicates the survival time after diagnosis.

In section 3.2, we provide notation and the model description (3.2.1), data structure (3.2.2), counting processes (3.2.3), joint likelihood (3.2.4), score functions and estimating procedure (3.2.5), and asymptotic properties (3.2.6). In Section 3.3, we use Monte Carlo simulations to verify the model's perfomance and its robustness

against dependent censoring. In section 3.4, we present our model fit for breast cancer using SEER data to demonstrate its interpretation, closed form prediction functions, and its ability to handle latent correlation in multi-stage time-to-event data. In section 3.5, we show that the model where diagnosis age is used as covariate for risk of death post diagnosis still could not fit the data as well as the proposed joint model.

3.2 Method

3.2.1 Model

Let T_1^* and T_2^* be the potentially observable times of diagnosis and death, respectively; $\boldsymbol{Z}$ be the set of covariates, which consists of $\boldsymbol{Z_1}$ - covariates that are available at time 0 (e.g. race, gender, etc.) and $\boldsymbol{Z_2}$ - additional covariates after diagnosis (e.g. treatments); and A be the common frailty term among all disease stages. Denote $h_{dx}(t)$ and $h_d(t)$ as the baseline hazards of diagnosis and death, respectively, to be estimated non-parametrically; and $\mathbb{1}(\cdot)$ as an indicator function. We define the following conditional hazards:

- Time-to-diagnosis: $\lambda_{dx}(t \mid A, \boldsymbol{Z_1}) = Ah_{dx}(t)\theta(\boldsymbol{Z_1})$,
- Time-to-death: $\lambda_d(t \mid T_{dx}, A, \boldsymbol{Z}) = Ah_d(t)\gamma(\boldsymbol{Z})\mathbb{1}(t \geq T_{dx})$,

where $\theta(\boldsymbol{Z_1}) = e^{\beta_\theta^T \boldsymbol{Z_1}}$, $\gamma(\boldsymbol{Z}) = e^{\beta_\gamma^T \boldsymbol{Z}}$, and $A \sim \text{Gamma}(\frac{1}{\phi(\boldsymbol{Z_1})}, \phi(\boldsymbol{Z_1}))$, where $\phi(\boldsymbol{Z_1}) = e^{\beta_\phi^T \boldsymbol{Z_1}}$. The advantage of this type of frailty is that we can eventually average (i.e. find the expectation) with respect to this latent term using a Laplace transform.

Let $\boldsymbol{\beta} = [\beta_\theta, \beta_\gamma, \beta_\phi]$ be the coefficient vector that we would like to estimate. Denote $S_\star(t)$ and $f_\star(t)$ as the corresponding survival and density functions, respectively, for hazard functions $\lambda_\star(t)$.

3.2.2 Observed Data Structure

Let $V^\star$ be the censoring time, independent of $(T_1^\star, T_2^\star)$ given $\boldsymbol{Z}$; and ξ be the maximum follow-up time of the study. For each subject $i = 1, 2, ..., n$ in the study, we observe $\{T_{1i}, T_{2i}, \Delta_{1i}, \Delta_{2i}, \boldsymbol{Z_i}\}$, where $T_{1i} = \min(T_{1i}^*, V_i^\star, \xi)$ is the time to the first event (i.e. diagnosis or censoring); $T_{2i} = \min(T_{2i}^\star, V_i^\star, \xi)$ is the time to the second event (i.e. death or censoring); $\Delta_{1i} = \mathbb{1}(T_{1i} = T_{1i}^\star)$ is an indicator of observing diagnosis; $\Delta_{2i} = \mathbb{1}(T_{2i} = T_{2i}^\star)\Delta_{1i}$ is an indicator of observing death after diagnosis.

3.2.3 Combine population risk-set and cancer events

Let $Y_{1i}(t) = \mathbb{1}(t \leq T_{1i}^\star)$ and $N_{1i}(t) = \mathbb{1}(T_{1i}^* < t)$ denote subject i's risk process and counting process, respectively, for a cancer diagnosis. Let $Y_{2i}(t) = \mathbb{1}(T_{1i} < t < T_{2i})$ and $N_{2i}(t) = \mathbb{1}(T_{2i}^* < t)N_{1i}(t)$ denote the risk process and counting process for death due to cancer.

SEER case data only include subjects who are diagnosed with cancer. This is sufficient if we are only interested in estimating the risk of death in such subjects. However, we are also interested in estimating the risk of being diagnosed with cancer in the general population. Thus, we need to invoke a separate dataset that tracks the size of the at-risk population from which our cancer cases originate. This separate dataset essentially provides cross-sectional life-table data for cancer diagnosis, and the SEER data includes subject-level follow-up for cancer death.

Figure 3.2 outlines how we use population counts as the risk set for diagnosis events. Since our breast cancer diagnoses are from females in Michigan between 1973-2011, we use counts of the female population during the same time period and in the same geographic location, stratified by $\boldsymbol{Z_1}$, to define the risk-set for breast cancer diagnosis. Let $Y_{z_1}(t)$ denote the population counts for the subset that takes

value $\boldsymbol{z_1}$ for covariate set $\boldsymbol{Z_1}$. Instead of summing up n cancer subjects, we sum up all categories of $\boldsymbol{Z_1}$ to get the whole at-risk population count. Similarly, denote $N_{z_1}(t)=\sum_{i:Z_{1i}=z_1} \mathbb{1}(T^*_{1i} < t)$ as the counting process for the number of diagnoses from subgroup z_1.

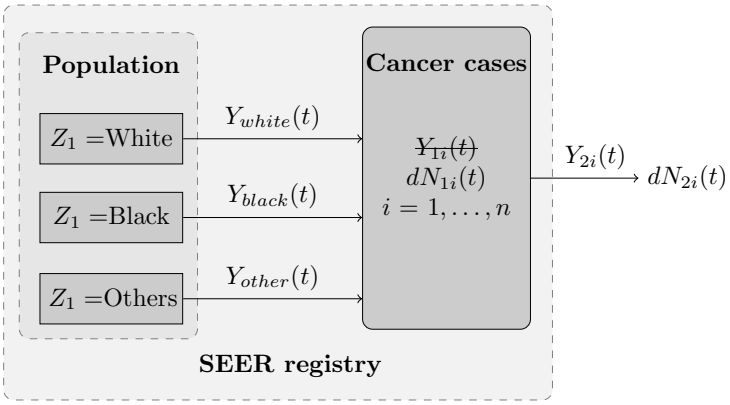


Figure 3.2: Using population counts as risk set for diagnosis events. Z_1 = race; $Y_{white}(t)$, $Y_{black}(t)$, and $Y_{other}(t)$ are population counts by race at age t; $dN_{1i}(t)$ denotes breast cancer diagnosis for subject i from the population; $Y_{2i}(t)$ denotes subject i being at risk of death due to breast cancer; $dN_{2i}(t)$ is the recorded death.

3.2.4 Likelihood Construction

For ease of notation, denote $f_{dx}f_d(t_1, t_2)$ as the joint probability of the sequence of diagnosis at t_1 followed by death at t_2. Dropping subscript i, we can derive the probability of observing each of the following sequences of events (see Appendix B.1 for derivation):

(1) Subject is diagnosed at t_1:

$$L_1(t_1) = \mathbb{E}_A\left[f_{dx}(t_1 \mid A)\right] = -h_{dx}(t_1)\theta\mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta}$$

(2) Subject is censored at t_1 before any event is observed:

$$L_2(t_1) = \mathbb{E}_A\left[S_{dx}S_d(t_1, t_1 \mid A)\right] = \mathfrak{L}^{(0)}_{H_{dx}(t_1)\theta}.$$

(3) Subject is diagnosed at t_1, then dies at t_2:

$$L_3(t_1, t_2) = \mathbb{E}_A \left[f_{dx} f_d(t_1, t_2 \mid A) \right] = h_{dx}(t_1) h_d(t_2) \theta\gamma \mathfrak{L}^{(2)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}.$$

(4) Subject is diagnosed at t_1 and is censored at t_2:

$$L_4(t_1, t_2) = \mathbb{E}_A \left[f_{dx} S_d(t_1, t_2 \mid A) \right] = -h_{dx}(t_1) \theta \mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}.$$

where $\mathfrak{L}^{(k)}_s$ is the k^{th} derivative of the Laplace transform of A with respect to s. For $A \sim \text{Gamma}(\frac{1}{\phi}, \phi)$, the first three Laplace transformations are as follow: $\mathfrak{L}^{(0)}_s = (1+\phi s)^{-\frac{1}{\phi}}$; $\mathfrak{L}^{(1)}_s = -(1+\phi s)^{-\frac{1}{\phi}-1}$; $\mathfrak{L}^{(2)}_s = (1+\phi)(1+\phi s)^{-\frac{1}{\phi}-2}$. With some algebra, the joint log-likelihood can be expressed in counting process form as follows:

$$\begin{aligned}
\ell(H_{dx}, H_d, \boldsymbol{Z}) = \sum_{i=1}^{n} \int_0^{\xi} & \log \left\{ \Theta_{1i} \left[H_{dx}(t), \boldsymbol{Z_1} \right] dH_{dx}(t) \right\} dN_{1i}(t) \\
& - Y_{1i}(t) \Theta_{1i} \left[H_{dx}(t), \boldsymbol{Z_1} \right] dH_{dx}(t) \\
& + \log \left\{ \Theta_{2i} \left[H_{dx}(t_{1i}), H_d(t), \boldsymbol{Z} \right] dH_d(t) \right\} dN_{2i}(t) \\
& - Y_{2i}(t) \Theta_{2i} \left[H_{dx}(t_{1i}), H_d(t), \boldsymbol{Z} \right] dH_d(t)
\end{aligned} \tag{3.1}$$

It is easy to recognize that eq.(3.1) follows the usual counting process notation for a time-to-event likelihood, but terms are repeated to accomodate both time-to-diagnosis and time-to-death. For ease of notation in the rest of this paper, we shorten $\Theta_1 \left[H_{dx}(t), \boldsymbol{Z_1} \right]$ and $\Theta_2 \left[H_{dx}(t_1), H_d(t), \boldsymbol{Z} \right]$ to $\Theta_1(t)$ and $\Theta_2(t; t_1)$, respectively. $\Theta_1(t) dH_{dx}(t)$ is the hazard of diagnosis at time t and $\Theta_2(t; t_1) dH_d(t)$ is the hazard of death due to cancer at t after diagnosis at t_1. These quantities can be derived (see Appendix B.2 for details) as follows:

$$\Theta_1(t) = \frac{L_1(t)}{L_2(t) dH_{dx}(t)} = \frac{-\theta \mathfrak{L}^{(1)}_{H_{dx}(t)\theta}}{\mathfrak{L}^{(0)}_{H_{dx}(t)\theta}} \tag{3.2}$$

$$\Theta_2(t; t_1) = \frac{L_3(t_1, t)}{L_4(t_1, t) dH_d(t)} = \frac{-\gamma \mathfrak{L}^{(2)}_{H_{dx}(t_1)\theta + H_d(t)\gamma - H_d(t_1)\gamma}}{\mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta + H_d(t)\gamma - H_d(t_1)\gamma}}. \tag{3.3}$$

As discussed in section 3.2.3, we use a separate population count dataset to obtain the number of people at risk for the event of diagnosis. Because we use $\boldsymbol{Z_1}$ =race as a covariate for the risk of cancer diagnosis, we stratify the population count by $\boldsymbol{Z_1}$ (this stratification is available in SEER using race as a covariate) and substitute the at-risk count $Y_{1i}(t)$ by $Y_{z_1}(t)$ and the event count $N_{1i}(t)$ by $N_{z_1}(t)$, where $Z_{1i} = z_1$. The joint log-likelihood in equation 3.1 is rewritten as follows:

$$\begin{aligned}\ell(H_{dx}, H_d, \boldsymbol{Z}) = &\sum_{\boldsymbol{z_1}} \int_0^{\xi} \log\left[\Theta_{\boldsymbol{z_1}}(t) dH_{dx}(t)\right] dN_{\boldsymbol{z_1}}(t) - Y_{\boldsymbol{z_1}}(t)\Theta_{1i}(t) dH_{dx}(t) + \\ &\sum_{i=1}^{n} \int_0^{\xi} \log\left[\Theta_{2i}(t; t_1) dH_d(t)\right] dN_{2i}(t) - Y_{2i}(t)\Theta_{2i}(t; t_1) dH_d(t).\end{aligned} \tag{3.4}$$

3.2.5 Nonparametric maximum likelihood estimator (NPMLE)

Score Functions

We use a profile likelihood approach to perform NPMLE with respect to $H_{dx}(t)$, $H_d(t)$, and $\boldsymbol{\beta}$. After some algebra, the score function with respect to $H_{dx}(x)$ is:

$$U_{H_{dx}(x)} = \sum_{z_1} \int_0^{\xi} \frac{dN_{z_1}(x)}{dH_{dx}(x)} - Y_{z_1}(x)\Theta_{z_1}(x) + \int_{x+}^{\xi} \psi_{z_1}(u) dM_{z_1}(u) \tag{3.5}$$

where $\psi_{z_1}(u) = \frac{\Theta_{z_1, H_{dx}}(u)}{\Theta_{z_1}(u)}$ and $\Theta_{z_1, H_{dx}}(t)$ denotes the partial derivative $\frac{\partial \Theta_{z_1}(t)}{\partial dH_{dx}(t)}$. Functional derivatives with respect to dH follow the intuition of differentiating over jumps of a step-function in the discrete case, and are rigorously defined in Hu and Tsodikov (2014a,b) to be valid for continuous functions as well.

Similarly, the score function with respect to $H_d(x)$ is

$$U_{H_d(x)} = \sum_{i=1}^{n} \int_0^{\xi} \frac{dN_{2i}(x)}{dH_d(x)} - Y_{2i}(x)\Theta_{2i}(x; t_{1i}) + \int_{x+}^{\xi} \psi_{2i}(u) dM_{2i}(u). \tag{3.6}$$

The score function with respect to $\boldsymbol{\beta}$ is

$$U_\beta = \sum_{z_1 \in Z_1} \int_0^\xi \left[\frac{\Theta_{z_1,\beta}(t)}{\Theta_{z_1}(t)} dN_{z_1}(t) - Y_{z_1}(t)\Theta_{z_1,\beta}(t) dH_{dx}(t) \right] + \sum_{i=1}^{n} \int_0^\xi \left[\frac{\Theta_{2i,\beta}(t;t_{1i})}{\Theta_{2i}(t)} dN_{2i}(t) - Y_{2i}(t)\Theta_{2i,\beta}(t;t_{1i}) dH_d(t) \right] \tag{3.7}$$

where martingales (at the true model) $dM(t)$ are defined as follows:

$$dM_{z_1}(u) = dN_{z_1}(u) - Y_{z_1}(u)\Theta_{z_1}(t)dH_{dx}(t), \forall z_1 \text{ values of } \boldsymbol{Z_1}$$

$$dM_{2i}(u) = dN_{2i}(u) - Y_{2i}(u)\Theta_{2i}(t;t_{1i})dH_d(t).$$

Estimation Procedure

Given $\boldsymbol{\beta}$, we derive Breslow-type estimators for $H_{dx}(t)$ and $H_d(t)$ by solving the score functions given in equations 3.5 and 3.6. The jump in baseline hazards $\widehat{dH_{dx}}$ and $\widehat{dH_d}$ at time x are

$$\widehat{dH_{dx}}(x) = \frac{\sum_{z_1} dN_{z_1}(x)}{\sum_{z_1 \in Z_1} Y_{z_1}(x)\omega_{z_1}(\overline{H}_{dx}, x)\Theta_{z_1}(x)}, \text{ and} \tag{3.8}$$

$$\widehat{dH_d}(x) = \frac{\sum_{i=1}^{n} dN_{2i}(x)}{\sum_{i=1}^{n} Y_{2i}(x)\omega_{2i}(\overline{H}_d, \overline{H}_{dx}, x)\Theta_{2i}(x;t_{1i})}, \tag{3.9}$$

where $\omega_\star(\overline{H}, x) = 1 - \frac{\int_{x+}^{\xi} \psi_\star(t) dM_\star(t)}{\Theta_\star(x)}$. Here we use $\overline{H}$ to denote the full trajectory of the function H as opposed to a single value of the function.

Denote $\boldsymbol{t} = \{t_{(1)}, t_{(2)}, ..., \xi\}$ as the set of unique times, with ξ the last time point, where events occur in the data. Simultaneous maximization of the log-likelihood with respect to H_{dx}, H_d, and $\boldsymbol{\beta}$ is done using a profile likelihood approach and an iterative reweighting algorithm (Chen, 2009) as follows:

1. For a candidate $\boldsymbol{\beta}^\star$, start with weight $\omega^{(0)}(t) = 1$ and $\widehat{dH}_{dx}^{(0)}(t)$ = Nelson-Aalen estimator.

2. Repeat until convergence of $\widehat{dH_{dx}(\boldsymbol{t})}$:

   - Use $\widehat{dH}_{dx}^{(0)}(x)$ to calculate $\Theta_{z_1}^{(0)}(x), \forall x \in \boldsymbol{t}$.
   - Use $\Theta_{z_1}^{(0)}(x)$, $\omega^{(0)} = 1$, and eq.3.8 to calculate $\widehat{dH}_{dx}^{(1)}(x)$. With the weight fixed, the right hand side of eq.3.8 only depends on prior jumps $H_{dx}(t)$, $\forall t < x$.
   - Sequentially calculate the jump size in baseline hazard along all time $x \in \boldsymbol{t}$ to get the full path $\widehat{H_{dx}^{(1)}}$.
   - Use full path $\widehat{H_{dx}^{(1)}}$ to update weights $\omega^{(1)}$.

3. Use converged $\widehat{H_{dx}}$ and a similar process as in the previous step to find $\widehat{H_d}$.

4. Calculate the profile log-likelihood $\ell_{\text{pr}}(\beta^\star, \widehat{H_{dx}}, \widehat{H_d})$ using eq.3.4.

5. Repeat steps (1-4) to search for $\hat{\beta} = \arg\max_{\beta} \ell(\beta)$.

Standard errors are estimated using the Hessian of the profile log-likelihood (see Appendix B.4 for outline of proof).

3.2.6 Asymptotic properties

Martingale properties The score functions for $H_{dx}(t)$, $H_d(t)$, and β can be rewritten as:

$$U_{H_{dx}(t)} = \sum_{z_1} \int_0^t \left\{ \frac{dM_{z_1}(x)}{dH_x} + \int_{x+}^{\xi} \psi_{z_1}(u) dM_{z_1}(u) \right\}, \tag{3.10}$$

$$U_{H_d(t)} = \sum_{i=1}^{n} \int_0^t \left\{ \frac{dM_{2i}(x)}{dH_x} + \int_{x+}^{\xi} \psi_{2i}(u) dM_{2i}(u) \right\}, \text{ and} \tag{3.11}$$

$$U_\beta = \sum_{z_1 \in Z_1} \int_0^\xi \frac{\Theta_{z_1,\beta}(t)}{\Theta_{z_1}(t)} dM_{z_1}(t) + \sum_{i=1}^{n} \int_0^\xi \frac{\Theta_{2i,\beta}(t)}{\Theta_{2i}(t)} dM_{2i}(t), \tag{3.12}$$

which are martingales under the true model (see Appendix B.3).

Consistency and weak convergence Consistency and weak convergence result of the NPMLE $\hat{\Omega} = (\hat{\beta}, \widehat{H_{dx}}(t), \widehat{H_d}(t))$ is adapted from Rice and Tsodikov (2016). Assuming regularity conditions hold,

(i) With probability one, $\hat{\beta}$ converges to β^0, $\hat{H}_{dx}$ converges to H^0_{dx}, and $\hat{H}_d$ converges to H^0_d uniformly in the interval $[0, \xi]$, where H^0_{dx}, H^0_d, β^0 are the true values of H_{dx}, H_d, β.

(ii) $n^{1/2}\{\hat{\beta} - \beta^0, \hat{H}_{dx}(t) - H^0_{dx}(t), \hat{H}_d(t) - H^0_d(t)\}$ converges weakly to a zero-mean Gaussian process. In addition, $n\mathcal{E}^T(\mathcal{I}_n)^{-1}\mathcal{E}$ converges in probability to the asymptotic variance-covariance function of the linear functional $n^{1/2}\left\{\int_0^\xi a^T(\hat{\beta} - \beta^0) + b(t)^T(\hat{H}_{dx}(t) - H^0_{dx}(t)) + c(t)^T(\hat{H}_d(t) - H^0_d(t))\right\}$, where a is real vector, $b(t)$ and $c(t)$ are functions with bounded total variation in $[0; \xi]$, $\mathcal{E}^T = (a^T, B^T)$ where B is a vector of the values of b at the jumps of $\hat{H}$, and $\mathcal{I}_n$ is the negative Hessian matrix of the observed log-likelihood function with respect to $\hat{\Omega}$.

3.3 Simulation Studies

We performed Monte Carlo simulations to assess the proposed methodology. For each scenario, 1000 sample replicates with sample sizes 500 and 1000 were performed. The parameters were set as follow: Frailty term $A \sim Gamma$ with mean 1 and variance $e^{\beta_\phi^T \boldsymbol{Z_1}}$ where $\beta_\phi = [-1, 0.4, 0.7]$; covariates $Z_1 \sim \text{Binom}(3, 0.5)$, $Z_2 \sim \text{N}(2, 0.5)$, and $Z_3 \sim \text{Bern}(0.5)$; increasing baseline hazard $H_{dx} \sim \text{Weibull}(5, 10)$ and decreasing $H_d \sim \text{Weibull}(0.5, 4)$. Time-to-diagnosis and time-to-death were generated under Cox proportional hazards using $\beta_\theta = [1.6, 0.5]$, and $\beta_\gamma = [2, -0.7, 1.4]$. Independent

censoring time followed a uniform distribution $U(8, 15)$, which yielded approximately 25% censoring before diagnosis and 50% censoring before death.

Table 3.1: Simulation results

N	Statistics	Diagnosis β_θ		Death β_γ			Frailty β_ϕ		
	True	**1.6**	**0.5**	**2**	**-0.7**	**1.4**	**-1**	**0.4**	**0.7**
500	Bias	0.02	0.01	0.05	-0.02	0.03	-0.05	-0.002	-0.01
	SD($\hat{\beta}$)	0.26	0.40	0.36	0.39	0.35	0.58	0.36	0.39
	Avg($\widehat{SE}$)	0.19	0.24	0.23	0.25	0.20	1.52	1.14	0.41
	95% CP	94.5	93.9	92.0	94.1	94.3	93.3	94.8	94.2
1000	Bias	0.01	-0.004	0.03	-0.02	0.02	0.01	-0.003	0.001
	SD($\hat{\beta}$)	0.14	0.18	0.17	0.18	0.13	0.34	0.07	0.11
	Avg($\widehat{SE}$)	0.13	0.17	0.16	0.17	0.14	0.32	0.08	0.11
	95% CP	95.3	93.5	93.5	93.8	94.8	93.4	95.3	94.9

SD($\hat{\beta}$): Empirical standard deviation of estimated $\hat{\beta}$ coefficients

Avg($\widehat{SE}$): Average of 1000 estimated standard errors (using Hessian matrix)

95% CP: 95% coverage probability.

The simulation results are summarized in Table 3.1. The estimated log-hazard-ratio (LHR) for time-to-diagnosis and time-to-death are very close to the true β and the bias decreases as sample size increases. The empirical standard deviation of 1000 Monte Carlo replicates are similar to the average of estimated standard error, validating our estimation of standard error. The differences decrease further with large sample size n=1000.

The estimation for parameter of frailty term A (β_ϕ) also performed well. The biases were small at sample size n=500, and get smaller with larger sample size. There is a discrepancy in the standard error estimation (difference between $SD(\hat{\beta})$ and $Avg(\widehat{SE})$ at smaller sample sizes, but it vanished for sample sizes as large as 1000. The performance of our proposed model is robust when switching between increasing and decreasing baseline hazards for $h_{dx}(t)$, $h_d(t)$, and different values of β_ϕ (results not shown).

3.4 Application to SEER breast cancer data

3.4.1 Data

We applied our proposed model to the motivating setting of SEER breast cancer data. We use Michigan data because the Michigan population is sufficiently diverse, allowing us to include race as a covariate in the model.

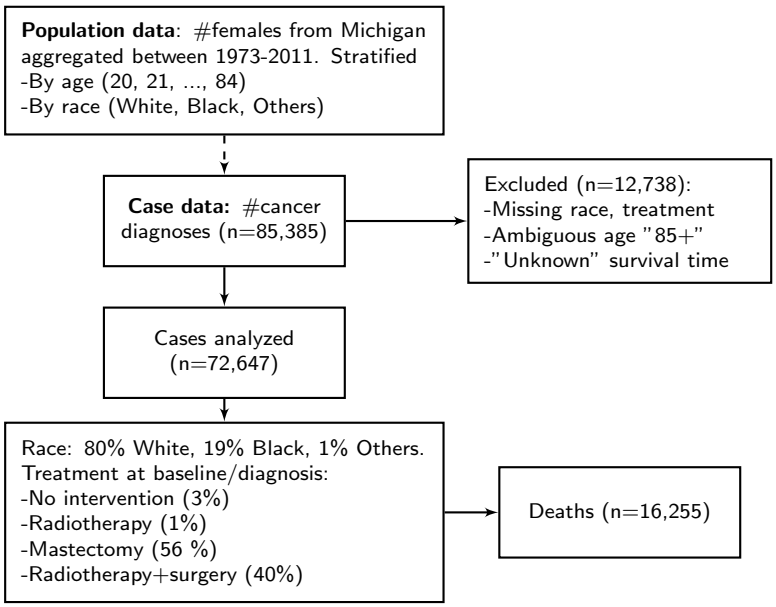
SEER cancer case: The SEER-18 database provided information on race, age at diagnosis, and the survival time of women who were diagnosed with breast cancer between 1973-2011. Out of all recorded cases in this time period, we concentrated on women between the ages of 20-84. Subjects with unknown race, age, or treatment were removed under the assumption of missing at random. Deaths from other causes and those alive at the end of data collection period were censored. The breakdown of data collected and used for analysis by race and treatment is summarized in Figure 3.3.

SEER population data: We used a separate SEER dataset to obtain counts of the female population by race in Michigan for each age between 20 and 84. These numbers were summed across the span of years recorded in the data (1973-2011) to define the at-risk set for each age. Modification to the log-likelihood to incorporate this new risk-set is shown in eq.(3.4).

3.4.2 Results

Figure 3.4 shows our estimated baseline cummulative hazards for diagnosis and death from SEER breast cancer data. The baseline hazard of diagnosis is on a smaller magnitude (left y-axis) compared to that of death because the population at risk of cancer diagnosis (i.e. Michigan female population) is much bigger than that of death (i.e. Michigan breast cancer patients). More importantly, we note that the shapes

Figure 3.3: SEER population and breast cancer case data (Michigan, 1973-2011) using in our analysis. # denotes number, n denotes number of observations.



of these two baseline hazards are substantially different and would not satisfied the common baseline hazard assumption from Hu and Tsodikov (2014a,b). Additionally, most parametric transformations could not adequately relate one baseline hazard to the other, thus strenthening the use of our model.

Table 3.2 shows the estimated covariate effects on time-to-diagnosis and time-to-death, respectively. The coefficients in table (a) and (b) denote log-hazard ratios with respect to the reference group (white women). A positive coefficient signifies higher risk and thus a faster time-to-event.

White women were found to have slightly lower risk of breast cancer diagnosis compared to black women ($\hat{\beta}_{\theta Black} = 0.56, P < 0.001$). There was not significant evidence of a difference compared to women of other races ($\hat{\beta}_{\theta Other} = 0.11, P = 0.21$).

Once diagnosed with breast cancer, race was also found to be an important factor affecting survival time. Black women had a slightly lower risk of death due to cancer (i.e. survive longer) compared to same-age white women ($\hat{\beta}_{\gamma Black} = -0.64, P <$

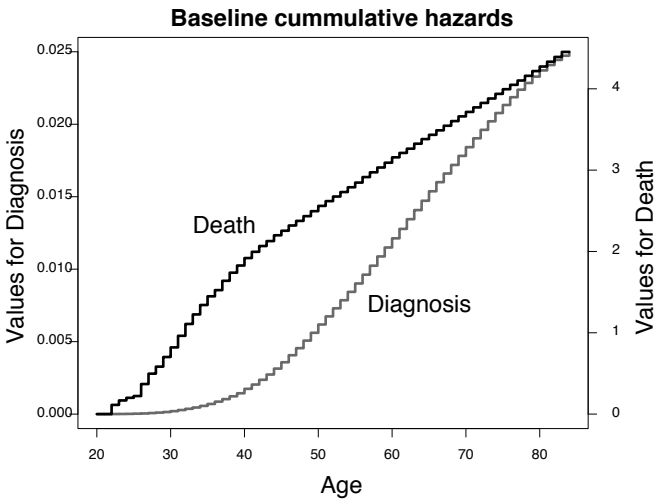


Figure 3.4: The shapes of baseline cummulative hazards for diagnosis events and death events in the SEER breast cancer data are too distinct to get from one to the other using transformation.

0.001). Women from other races had the lowest risk relative to black and white patients ($\hat{\beta}_{\gamma_{Other}} = -3.12, P < 0.001$).

Table 3.2: SEER breast cancer analysis

(a) LHR for time-to-diagnosis (β_θ)

Covariate	Estimate	SE	P-value
Z_{Black}	0.56	0.02	<.001
Z_{Other}	0.11	0.09	0.21

(b) LHR for time-to-death (β_γ)

Covariate	Estimate	SE	P-value
Z_{Black}	-0.64	0.07	<.001
Z_{Other}	-3.12	0.15	<.001
$Z_{Surgery}$	-2.61	0.05	<.001
$Z_{Radiotherapy}$	-0.04	0.09	0.67
$Z_{Combination}$	-2.73	0.05	<.001

(c) Gamma frailty parameters (β_ϕ)

Covariate	Estimate	SE	P-value
Z_{White}	1.61	0.10	<.001
Z_{Black}	3.10	0.05	<.001
Z_{Other}	5.16	0.08	<.001

LHR: Log-hazard ratio versus reference group (white women who received no treatment).
SE: Estimated standard errors.

Treatment options that involved surgery (mastectomy) were found to improve sur-

vival time compared to no immediate treatment. Specifically, surgery in combination with radiotherapy was most effective ($\hat{\beta}_{\gamma_{Combo}} = -2.73, P < 0.001$) and marginally better than surgery by itself ($\hat{\beta}_{\gamma_{Surgery}} = -2.61, P < 0.001$). There was not enough evidence for the effectiveness of radiotherapy alone ($\hat{\beta}_{\gamma_{Radiotherapy}} = -0.04, P = 0.67$).

The latent correlation between time-to-diagnosis and time-to-death previously unaccounted for was detected in our proposed model. The common frailty term in the hazards for diagnosis and death followed a Gamma distribution with mean 1 and variance $e^{1.61}$ in white patients, $e^{3.10}$ in black patients, and $e^{5.16}$ in patients from other races. The effect of this frailty term is more apparent when we explore the correlation between time-to-diagnosis and time-to-death (Figure 3.6).

3.4.3 Predicted Survival Rates and Diagnosis Plots

One of the advantages of our proposed model is that there are closed form expressions for the survival functions of the time-to-each-event (or between events) of interest.

Figure 3.5 shows the empirical Kaplan-Meier curves for the diagnosis of breast cancer among different racial groups in Michigan. The predicted rates from our proposed model, as given by equation 3.13, are very similar to the empirical rates for all racial groups.

$$S_{dx}(t|Z) = L_3(t) = \mathfrak{L}^{(0)}_{H_{dx}(t)\theta} \tag{3.13}$$

We also plotted the post-diagnosis survival time conditioned on age at diagnosis (eq.3.14) in Figure 3.6.

$$\hat{S}(T_d - T_{dx} = t \mid t_{dx}) = \frac{L_4}{L_1} = \frac{\mathfrak{L}^{(1)}_{\theta H_{dx}(t_{dx}) + \gamma H_d(t_{dx}+t) - \gamma H_d(t_{dx})}}{\mathfrak{L}^{(1)}_{\theta H_{dx}(t_{dx})}} \tag{3.14}$$

Our model is able to predict the adverse effect of age at diagnosis on survival time without explictly using diagnosis-age as a covariate in an overall survival model. For

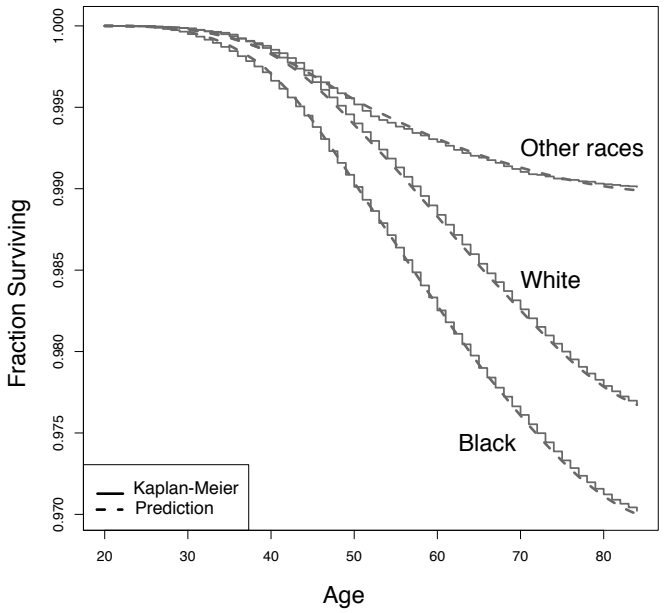


Figure 3.5: Percentage of women without a breast cancer diagnosis in the Michigan population for each racial group from 1973-2011. Proposed model predictions $\hat{S}_1(t_1|Z)$ (dotted lines) from eq. (3.13) are compared to empirical Kaplan-Meier estimates (solid lines).

example, patients diagnosed at 35 years old have worse survival compared to those diagnosed at 50 (or 70) years old after the same time period from diagnosis. This adverse age effect diminishes in older groups. Those diagnosed at 50 have slightly worse prognosis compared to 70 year-old patients. Figure 3.6 includes only the subset of white women who received surgery post diagnosis, but results for other treatment options and racial groups were similar (not shown).

3.5 SEER Analysis: Reset clock to 0 at diagnosis time

We also tried the following 2 models where time-to-death is reset to 0 at diagnosis time (T_{dx}). Denote s as the time-since-diagnosis.

M1: $\lambda_d(s \mid T_{dx}, A, \boldsymbol{Z}) = Ah_d(s)e^{\beta_\gamma^T \boldsymbol{Z}}$

M2: $\lambda_d(s \mid T_{dx}, A, \boldsymbol{Z}) = Ah_d(s)e^{\beta_\gamma^T \boldsymbol{Z} + \beta_2 \mathbb{1}(20 \leq T_{dx} \leq 40) + \beta_1 \mathbb{1}(40 < T_{dx} \leq 60) + \beta_3 \mathbb{1}(T_{dx} > 60)}$

Table 3.3 shows the log-hazard ratios from 2 models where time-to-death was reset

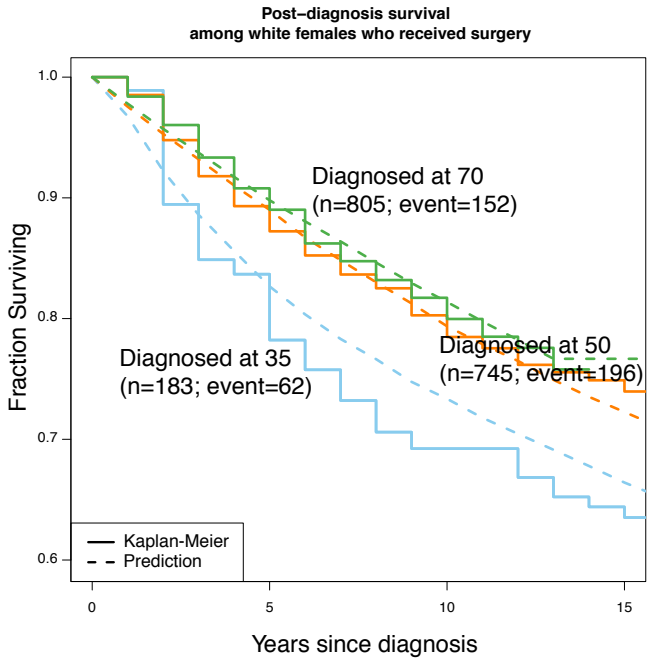


Figure 3.6: Women diagnosed at a younger age have worse survival post diagnosis. This plot shows post-diagnosis survival among white women who received surgery. We focus on those that were diagnosed at younger age (35 years old) versus older age (50 or 70 years old). The prediction from our proposed model $\hat{S}_{T_d - T_{dx}}(t \mid t_{dx})$ (dash lines) from eq. (3.13) are plotted against the empirical Kaplan-Meier curve (solid lines).

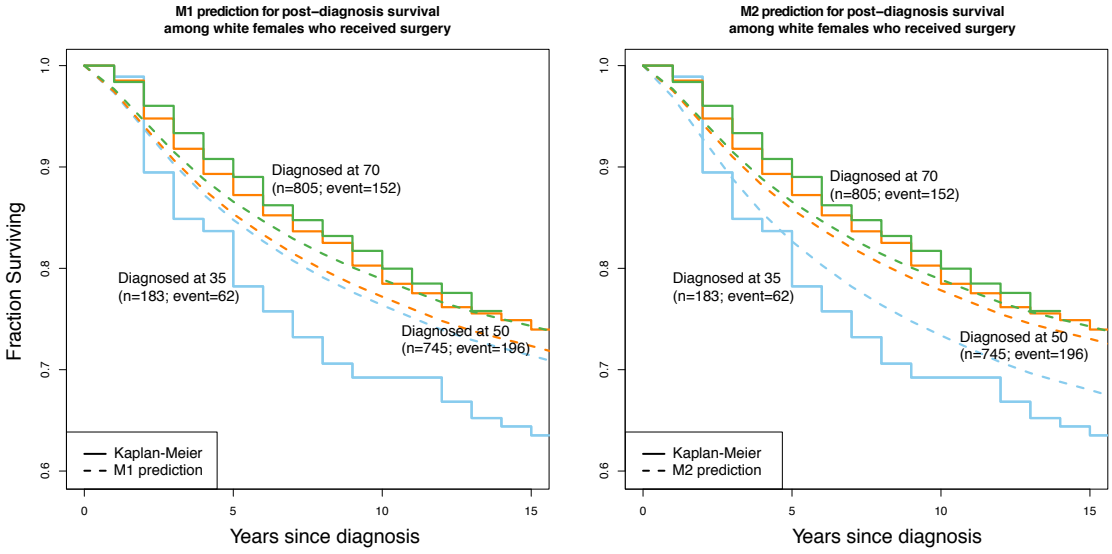


Figure 3.7: Predicted curves for post-diagnosis survival $\hat{S}_{T_d - T_{dx}}(t \mid t_{dx})$ among white women who received surgery using model M1 (left plot, dash lines) and model M2 (right plot, dash lines) are plotted against the empirical Kaplan-Meier curve (solid lines).

(a) Log hazard ratios for time-to-diagnosis (β_θ) and SE

Covariate	Proposed	M1	M2
Z_{Black}	0.56 (0.02)	0.61 (0.02)	0.60 (0.02)
Z_{Other}	0.11 (0.09)	0.15 (0.09)	0.13 (0.09)

(b) Log hazard ratios for time-to-death (β_γ) and SE

Covariate	Proposed	M1	M2
Z_{Black}	-0.64 (0.07)	-0.35 (0.11)	-0.67 (0.17)
Z_{Other}	-3.12 (0.15)	-2.64 (0.17)	-3.04 (0.23)
$Z_{Surgery}$	-2.61 (0.05)	-2.52 (0.05)	-2.49 (0.06)
$Z_{Radiotherapy}$	-0.04 (0.09)	-0.02 (0.09)	-0.02 (0.09)
$Z_{Combination}$	-2.73 (0.05)	-2.68 (0.06)	-2.66 (0.06)
$\mathbb{1}(40 < T_{dx} \leq 60)$	-	-	-0.20 (0.03)
$\mathbb{1}(T_{dx} > 60)$	-	-	-0.21 (0.04)

(c) Gamma frailty parameters (β_ϕ) and SE

Covariate	Proposed	M1	M2
Z_{White}	1.61 (0.10)	2.14 (0.16)	1.67 (0.25)
Z_{Black}	3.10 (0.05)	3.33 (0.05)	3.22 (0.06)
Z_{Other}	5.16 (0.08)	5.23 (0.07)	5.18 (0.08)

Table 3.3: Comparing proposed model to M1 and M2 using SEER breast cancer analysis. (LHR: Log-hazard ratio versus reference group (white women who received no treatment). SE: Estimated standard errors.)

to 0 at diagnosis. This is possible because we estimate the baseline hazard for diagnosis and death separately. The estimated log-hazard-ratios are consistent with our proposed model where we did not reset the clock. However, this new model requires explicit inclusion of diagnosis age as covariates for risk-of-death. This necessitates deciding whether to treat diagnosis age as continuous covariates or adding splines. Even with the later approach, the model's prediction for overall survival time, conditioned on diagnosis time, does not fit as well as our proposed model where we did not have to include age or diagnosis time as a covariate.

3.6 Discussion

The effects of demographic factors and therapeutic interventions on the overall survival time of patients in any disease is not a new problem in survival analysis. Common methods of analysis include Cox proportional hazards and frailty models.

Situations where a certain event (e.g. death due to a particular cause) prevents others from being observed (e.g. death from other causes) have been studied extensively within the competing risks framework. However, there is no agreed-upon procedure to model cancer data where there are multiple stages (i.e. diagnosis, metastasis, remission, relapse, and death) of interest that make up the entire disease natural history. Intuitively, the stages of interest have the same underlying biological cause and thus their time-to-events should be analyzed simultaneously.

The current common practice fails to take into account the relationship between stages. Due to its convenient availability, Cox models are often used to analyze time-to-stages as independent outcomes. Less naive approaches using competing-risks or semi-competing-risks model only stages that directly compete with each other and ignore their relationships with other stages that come earlier or later. The non-identifiability issue in competing-risks prevents full estimation of the non-parametric baseline hazard.

Instead of using the same baseline hazard for all stages to construct the joint likelihood, we extended the joint modeling approach to estimates a baseline hazard for the first observed event (i.e. diagnosis) and a different baseline hazard for the next event (i.e. death) conditioning on the former. In practical terms, this assumption would increase computing time due to having to estimate two baseline hazards non-parametrically. But the advantage is that it relaxes the restriction and captures the full dependence between stages. Earlier versions of our model that estimate one baseline hazard non-parametrically and use a parametric transformation (e.g. raise the estimated cummulative baseline hazard to a power) to arrive at the second baseline hazard were also not flexible enough to predict the positive correlation between diagnosis time and post-diagnosis survival.

Our proposed model can easily be generalized to include more cancer stages such as metastasis, remission, or recurrence. We can estimate unique baseline hazards for each of these stages as long as the onset times are observed. In the case where a stage is not observed (e.g. positive metastasis status at diagnosis time indicates metastasis onset at unknown time prior to diagnosis) preventing the non-parametric estimation of its baseline hazard, we can arrive at its estimate using a parametric transfomation of the baseline hazard from another observed stage. We found this to be a good approach to model diagnosis, interval-censored metastasis, and death in prostate cancer. This, of course, requires making the assumption that the unobserved stage and observed stage have a similar enough underlying disease process that can be modeled via a transformation.

Future research in this area includes modeling recurring stages and partial missingness. More development to speed up the estimation procedure would also help make this joint modeling approach more attractive to researchers.

CHAPTER IV

Joint Modeling and Multiple Comparisons with the Best of Data from a SMART with Survival Outcomes

4.1 Introduction

Many diseases, chronic and acute, require ongoing treatment. Multiple first- and second-line treatment options are available for many diseases, but most evidence for optimal treatment focuses on one specific time point. The best treatment strategy over the course of the disease, however, may include sequences of treatments that depend on previous treatments. A dynamic treatment regimen (DTR) (Murphy et al., 2001) is a guideline providing first-line and second-line treatment, where second-line treatment is specified based on response to first-line treatment. For diseases and disorders where DTRs are standard practice, evidence for optimal clinical decision rules considering more than one stage of treatment is necessary. A sequential multiple assigment randomized trial (SMART) (Murphy, 2005; Lavori et al., 2000; Lavori and Dawson, 2004) is one such type of trial design that can assess entire treatment regimens and build effective DTRs. In such a multi-stage trial, second-line treatment may be randomized based on a patient's outcome to a first-line treatment. DTRs are embedded within a SMART for estimation and comparsion. The goal of a SMART is to identify decision rules or DTRs that result in the best overall outcome.

Figure 4.1 shows three common SMART designs. In design I, there are 8 em-

bedded DTRs: $\{A_1B_1C_1\}$, $\{A_1B_1C_2\}$, $\{A_1B_2C_1\}$, $\{A_1B_2C_2\}$, $\{A_2B_1C_1\}$, $\{A_2B_1C_2\}$, $\{A_2B_2C_1\}$, and $\{A_2B_2C_2\}$, where each set consists of first stage treatment (A_1 or A_2), second stage treatment for responders (B_1 or B_2), and second stage treatment for non-responders (C_1 or C_2). Note that every person is consistent with two DTRs (e.g. A_1B_1 group is part of both regimens $\{A_1B_1C_1\}$ and $\{A_1B_1C_2\}$; A_1C_1 group is consistent with $\{A_1B_1C_1\}$ and $\{A_1B_2C_1\}$, etc.) This feature of a SMART allows for efficient use of each subject's information to estimate the regimen outcome (Ko and Wahed, 2012). Design II shows a variation of a SMART design where second-line treatments are only assigned to responders resulting in 4 DTRs: $\{A_1B_1C_1\}$, $\{A_1B_2C_1\}$, $\{A_2B_1C_1\}$, and $\{A_2B_2C_1\}$. Design III has second-line treatments only for responders from A_1, resulting in 3 DTRs $\{A_1B_1C_1\}$, $\{A_1B_2C_1\}$ and $\{A_2B_1C_1\}$. Note that "response" and "non-response" may switch in different context, but this change is trivial. Further, many other possible SMART designs exist with different number of stages and treatments.

Many published methods for the analysis of data from a SMART consider continuous outcomes. This is due to the ease of model parameterization and robust theoretical properties (Murphy, 2005; Oetting et al., 2011; Orellana et al., 2010). Since many DTRs can be embedded within a SMART, multiple comparisons can be an issue. The issue of multiple DTR comparisons has been addressed for continuous outcomes using Ertefaie et al. (2016), but it is an open probelm for survival outcomes.

Survival outcomes have been assessed in SMART literature. Many estimators have been built upon the inversed probability weighting (IPW) framework: Lunceford et al. (2002) used IPW to average responders and non-responders in each regimen and estimate survival probabilties non-parametrically; Guo and Tsiatis (2005)

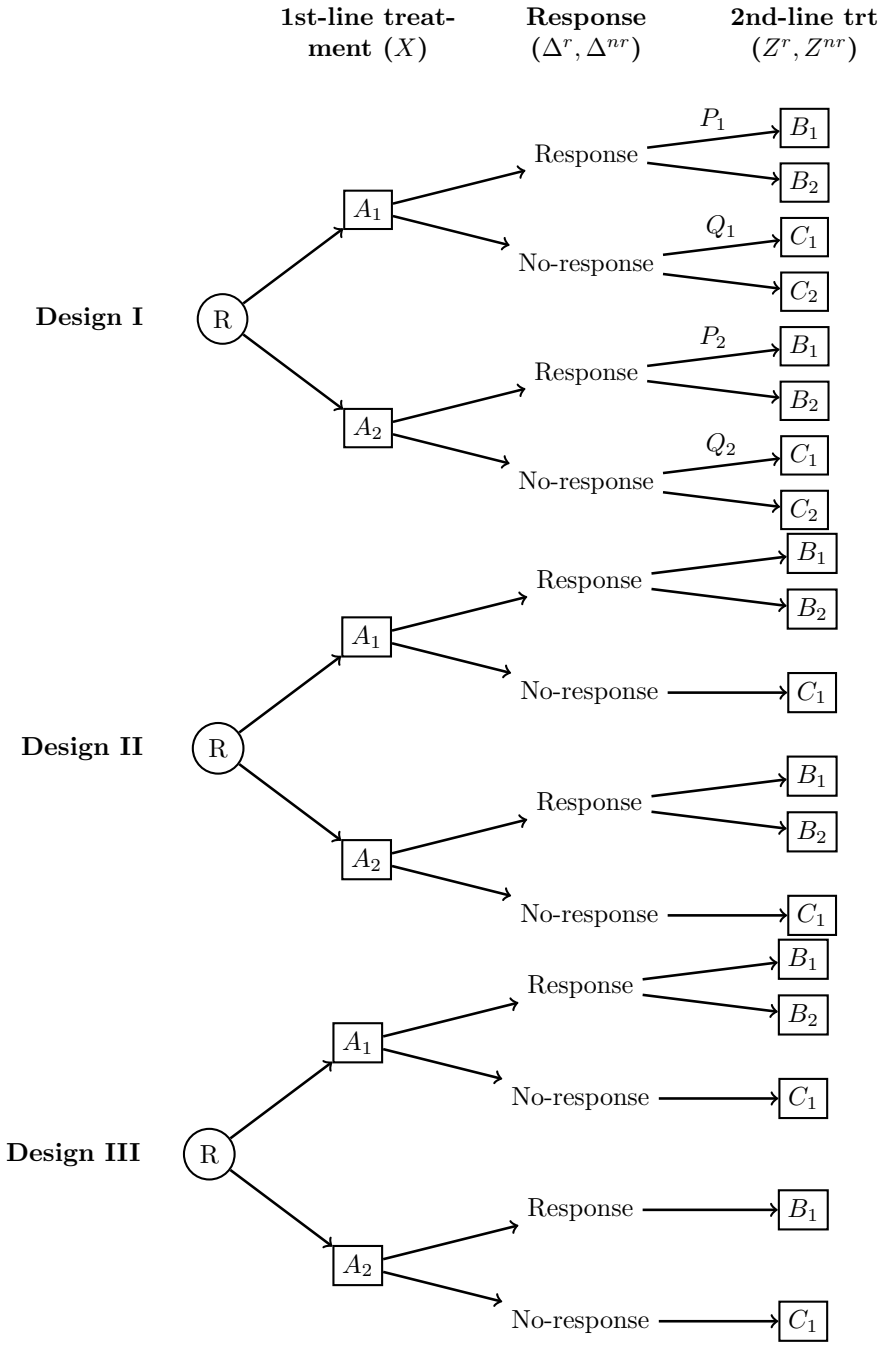


Figure 4.1: Three common two-stage SMART designs. R is first randomization to first-line tretament A_1 ($X = 0$) or A_2 ($X = 1$). Δ^r and Δ^{nr} are indicators of responders and non-responders to first-line treatment, respectively. B_1 ($Z^r = 0$) and B_2 ($Z^r = 1$) are second-line treatment options for responders. C_1 ($Z^{nr} = 0$) and C_2 ($Z^{nr} = 1$) are second-line treatment options for non-responders. $P_j = P(Z^r = 1 \mid A_j)$ and $Q_j = P(Z^{nr} = 1 \mid A_j)$ are second-line randomization probabilities after treatment A_j, $j = 1, 2$. In design I, both responders and non-responders are re-randomized to second treatment. In design II, only responders are re-randomized. In design III, only responders from A_1 are re-randomized.

combined time-dependent IPW and Nelson-Aalen estimators to estimate regimens' survival rates; Feng and Wahed (2008) proposed a modified supremum weighted log-rank test and Miyahara and Wahed (2010) introduced a weighted Kaplan-Meier estimator. Wahed (2010) used a mixture of exponential, normal, and Weibull distributions to model survival times from different regimens. Some of these works were followed up with large sample properties and sample-size formulae including methods by Feng and Wahed (2009) and Li and Murphy (2011). Even though these methods estimate the averaged survival outcomes of DTRs (e.g. mean survival time or survival probabilities), none of them offer a way to parameterize treatment effects, interactions, and incorporate auxiliary variables beside treatment assignment.

Methods that do incorporate auxiliary variables extend the standard Cox model. Lokhnygina and Helterbrand (2007) incorporated IPW into a Cox model to compare separate-path DTRs (e.g. independent DTRs with different first-line treatment such as $A_1B_1C_1$ and $A_2B_1C_1$ in design II), but could not compare shared-path DTRs (Kidwell and Wahed (2013), e.g. DTRs with overlapping patients such as $A_1B_1C_1$ and $A_1B_2C_1$ in design II). Thall et al. (2007) proposed a Bayesian joint model that regressed the overall survival time on intermediate response time, but assume a rather restrictive Weibull distribution for both time points. Recently, Tang and Wahed (2015) assigned different Cox baseline hazards for each DTR and modeled auxiliary variables as covariates effects on the hazard ratios. This approach used different baseline hazards to represent the overall difference between DTRs but did not offer a way to measure each treatment's effect and their interaction within DTRs. Moreover, Tang and Wahed (2015) assumed that the auxiliary covariates' effects were constant over different treatment stages within each DTR and across different DTRs.

In this manuscript, we propose a joint modeling approach that consider the at

time-to-response and time-to-death together. The joint model offers a way to measure first- and second-line treatments within a regimen and their interactions while naturally adjusting for SMART randomization probabilities. Unless otherwise stated, the notation and simulation results are based on design II. As we will show, adaptations for designs I and III are straightforward. In section 4.2.1, we introduce the notation for a SMART setting with a survival outcome. In section 4.2.2, we introduce our joint model, the resulting joint log-likehood (section 4.2.3), and estimation procedure (section 4.2.4). We illustrate the estimation performance using Monte Carlo simulations (section 4.3.1). We briefly discuss previously proposed non-parametric survival estimators and compare them against our model's predictive performance (section 4.3.2). Finally, we implement multiple comparisons with the best (MCB) to compare survival rates at some time t between DTRs (section 4.4) and provide some discussion (section 4.5).

4.2 Joint model and NPMLE

4.2.1 Notation for SMARTs data structure

Assume a SMART design with two intervention options for first- and second-line treatment for responders, and 1 option for second-line non-responders (Design II, Figure 4.1). We use the counterfactural framework to summarize the set of covariates and potential outcomes from each subject i:

$$\{V_i, X_i, Z_i^r, Z_i^{nr}, T_i^r, T_i^d, T_i^c, \Delta_i^r, \Delta_i^{nr}, \Delta_i^d\}$$

where V is the set of baseline covariates, X is an indicator of first-line treatment ($X = 0$ for A_1, $X = 1$ for A_2), Z^r is an indicator of second-line treatment for responders ($Z^r = 0$ for B_1, $Z^r = 1$ for B_2), Z^{nr} is an indicator of second-line treatment for non-responders ($Z_{nr} = 0$ for C_1). We define non-responders as those

who have not responded by a cut-off time t^{nr}. Denote $P_j = P(Z^r = 1 \mid A_j)$ and $Q_j = P(Z^{nr} = 1 \mid A_j) = 0$ as the second-line randomization probabilities among responders and non-responders, respectively. These notations can be generalized for design I and III.

T^r, T^d, and T^c are potentially observed time-to-response, time-to-death, and censoring time, respectively. If a patient does not respond by t^{nr} or dies first, T^r is not observed. Thus, we can relate the first observed event time as $T_1 = \min(T^r, t^{nr}, T^d, T^c)$. Indicators $\Delta^r = \mathbb{1}(T^r = T_1)$ and $\Delta^{nr} = \mathbb{1}(t^{nr} = T_1)$ denote whether the patient is a responder or non-responder, respectively. Note that $\Delta^r + \Delta^{nr} = 0$ indicates that a patient was censored or died before response assessment.

If $\Delta^r + \Delta^{nr} = 1$, a patient is re-randomized to receive a second-line treatment. Subsequent survival time or censoring time is observed at $T_2 = min(T^d, T^c)$. Indicator $\Delta^d = \mathbb{1}(T^d = T_2)$ denotes that a death is observed.

4.2.2 Model

We introduce conditional hazards to model time-to-response and time-to-death. Denote $h_r(t)$ and $h_d(t)$ as the non-parametric baseline hazards ($H_r(t)$ and $H_r(t)$ are the cummulative hazards) of response and death, respectively, and $\mathbb{1}(\cdot)$ as an indicator function. Then, we define the following:

- time-to-response:

$$\lambda_r(t) = h_r(t) \exp\{\beta_1 V + \beta_2 X\}, \tag{4.1}$$

- time-to-death:

$$\lambda_d(t) = h_d(t) \exp\{\beta_3 V + \beta_4 X + \Delta^r \mathbb{1}(t > t^r)(\beta_5 + \beta_6 X + \beta_7 Z^r + \beta_8 X Z^r + \beta_9 X Z^r V) + \Delta^{nr} \mathbb{1}(t > t^{nr})(\beta_9 + \beta_{10} X)\}. \tag{4.2}$$

Note that in design II, all regimens have $Z^{nr} = 1$, thus its effect is absorbed into the baseline hazard. For ease of notation, we denote $\theta = e^{\beta_1 V + \beta_2 X}$, $\gamma = e^{\beta_3 V + \beta_4 X}$, $\eta = \gamma e^{\beta_5 + \beta_6 X + \beta_7 Z^r + \beta_8 X Z^r + \beta_9 X Z^r V}$, and $\mu = \gamma e^{\beta_{10} + \beta_{11} X}$. Denote $S(t)$ and $f(t)$ as the corresponding survival and density functions, respectively, for hazard functions $\lambda(t)$.

The current form of eq. (4.2) can easily be adapted for designs I and III as follows.

Design I $\lambda_d(t) = h_d(t) \exp\{\beta_3 V + \beta_4 X + \Delta^r \mathbb{1}(t > t^r)(\beta_5 + \beta_6 X + \beta_7 Z^r + \beta_8 X Z^r + \beta_9 X Z^r V) + \Delta^{nr} \mathbb{1}(t > t^{nr})(\beta_{10} + \beta_{11} X + \beta_{12} Z^{nr} + \beta_{13} X Z^{nr} + \beta_{14} X Z^{nr} V)\}$.

Design III $\lambda_d(t) = h_d(t) \exp\{\beta_3 V + \beta_4 X + \mathbb{1}(A_1)\Delta^r \mathbb{1}(t > t^r)(\beta_5 + \beta_6 Z^r + \beta_8 X Z^r)\}$.

4.2.3 Counting Processes and Likelihood Construction

Denote $f_r f_d(t_1, t_2)$ as the joint probability of the sequence of response at t_1, followed by death at t_2. With some algebra, the joint log-likelihood can be expressed in counting process form as follows:

$$\begin{aligned}
\ell = \sum_{i=1}^{n} \int_0^{\xi} & \log\{\Theta_{1i}^r(t) dH_r(t)\} \, dN_{1i}^r(t) - \log\left\{\Theta_{1i}^d(t) dH_d(t)\right\} dN_{1i}^d(t) \\
& - Y_{1i}(t) \left[\Theta_{1i}^r(t) dH_r(t) + \Theta_{1i}^d(t) dH_d(t)\right] \\
& + \log\{\Theta_{2i}^r(t \mid t_1) dH_d(t)\} \, dN_{2i}^r(t) - Y_{2i}^r(t) \Theta_{2i}^r(t \mid t_1) dH_d(t) \\
& + \log\{\Theta_{2i}^{nr}(t \mid t_1) dH_d(t)\} \, dN_{2i}^{nr}(t) - Y_{2i}^{nr}(t) \Theta_{2i}^{nr}(t \mid t_1) dH_d(t), \qquad (4.3)
\end{aligned}$$

where ξ is the maximum follow-up time, $\Theta_1^r(t)$ and $\Theta_1^d(t)$ are the hazards of response (counting process $dN_1^r(t)$) or death (counting process $dN_1^d(t)$) first at time t, respectively. $\Theta_2^r(t \mid t_1)$ and $\Theta_2^{nr}(t)(t \mid t_1)$ are the hazards of death after response (counting process $dN_2^r(t)$) or non-response (counting process $dN_2^{nr}(t)$), respectively. $Y_1(t)$ is the at-risk process for the first event. $Y_2^r(t)$ is the at-risk process for death after the first event is response. $Y_2^{nr}(t)$ is the at-risk process for death after the first event is non-response.

It is easy to see that the first three terms in eq. (4.3) of the joint log-likelihood express the probability of observing the first event (response, non-response, death, or censoring). The next four terms express the probability of observing the second event (death or censoring) given the first event was response (subscript r) or non-response (subscript nr).

The following quantities can be derived (see Appendix C.1 for details):

$$\Theta_1^r(t) = \text{P(response first at t)} = \frac{f_r S_d(t,t)}{S_r S_d(t,t) dH_r(t)} = \theta \frac{e^{-H_r(t)\theta - H_d(t)\gamma}}{e^{-H_r(t)\theta - H_d(t)\gamma}},$$

$$\Theta_1^r(t) = \text{P(die first at t)} = \frac{S_r f_d(t,t)}{S_r S_d(t,t) dH_d(t)} = \gamma \frac{e^{-H_r(t)\theta - H_d(t)\gamma}}{e^{-H_r(t)\theta - H_d(t)\gamma}},$$

$$\Theta_2^r(t) = \text{P(die at t} \mid \text{response at } t_1) = \frac{f_r f_d(t_1,t)}{f_r S_d(t_1,t) dH_d(t)} = \eta \frac{e^{-H_r(t_1)\theta - H_d(t_1)(\gamma-\eta) - H_d(t)\eta}}{e^{-H_r(t_1)\theta - H_d(t_1)(\gamma-\eta) - H_d(t)\eta}},$$

$$\Theta_2^{nr}(t) = \text{P(die at t} \mid \text{no-response}) = \frac{S_r f_d(t^{nr},t)}{S_r S_d(t^{nr},t) dH_d(t)} = \mu \frac{e^{-H_r(t^{nr})\theta - H_d(t^{nr})(\gamma-\mu) - H_d(t)\mu}}{e^{-H_r(t^{nr})\theta - H_d(t^{nr})(\gamma-\mu) - H_d(t)\mu}}.$$

4.2.4 Estimation Procedure

We derive Breslow-type estimators for $H_r(t)$ and $H_d(t)$ by solving their score functions. See Supplemental Material Section C.2 for the score functions and a proof of the martingale properties.

The jump in baseline hazards $\widehat{dH_r}$ and $\widehat{dH_d}$ at time x are

$$\frac{\partial \ell}{\partial dH_r(x)} = 0 \Leftrightarrow \widehat{dH_r}(x) = \frac{\sum_{i=1}^{n} dN_{1i}^r(x)}{\sum_{i=1}^{n} Y_{1i}(x)\Theta_{1i}^r(x)\omega_{1i}^r}, \text{ and} \tag{4.4}$$

$$\frac{\partial \ell}{\partial dH_d(x)} = 0 \Leftrightarrow \widehat{dH_d}(x) = \frac{\sum_{i=1}^{n} dN_{1i}^d(x) + dN_{2i}^r(x) + dN_{2i}^{nr}(x)}{\sum_{i=1}^{n} Y_{1i}(x)\Theta_{1i}^d(x)\omega_{1i}^d + Y_{2i}^r(x)\Theta_{2i}^r(x)\omega_{2i}^r + Y_{2i}^{nr}(x)\Theta_{2i}^{nr}(x)\omega_{2i}^{nr}}, \tag{4.5}$$

where the exact forms of ω_1^r, ω_1^d, ω_2^r, and ω_2^{nr} are

$$\omega_1^r(x) = 1 - \frac{\int_{x+}^{\xi} \frac{\partial \log \theta_1^r(t)}{\partial dH_r(t)} dM_1^r(t) + \frac{\partial \log \theta_1^d(t)}{\partial dH_r(t)} dM_1^d(t)}{\theta_1^r(x)},$$

$$\omega_1^d(x) = 1 - \frac{\int_{x+}^{\xi} \frac{\partial \log \theta_1^r(t)}{\partial dH_d(t)} dM_1^r(t) + \frac{\partial \log \theta_1^d(t)}{\partial dH_d(t)} dM_1^d(t)}{\theta_1^d(x)},$$

$$\omega_2^r(x) = 1 - \frac{\int_{x+}^{\xi} \frac{\partial \log \theta_2^r(t)}{\partial dH_d(t)} dM_2^r(t)}{\theta_2^r(x)}, \text{ and } \omega_2^{nr}(x) = 1 - \frac{\int_{x+}^{\xi} \frac{\partial \log \theta_2^{nr}(t)}{\partial dH_d(t)} dM_2^{nr}(t)}{\theta_2^{nr}(x)},$$

using the following martingales

$$dM_1^r(t) = dN_1^r(t) - Y_1(t)\Theta_1^r(t)dH_r(t) \text{ and } dM_1^d(t) = dN_1^d(t) - Y_1(t)\Theta_1^d(t)dH_d(t),$$

$$dM_2^r(t) = dN_2^r(t) - Y_2^r(t)\Theta_2^r(t)dH_d(t) \text{ and } dM_2^{nr}(t) = dN_2^{nr}(t) - Y_2^{nr}(t)\Theta_2^{nr}(t)dH_d(t).$$

Denote $\boldsymbol{t} = \{t_{(1)}, t_{(2)}, ..., \xi\}$ as the set of unique times where events occur in the data. Simultaneous maximization of the log-likelihood with respect to $H_r(\boldsymbol{t})$, $H_d(\boldsymbol{t})$, and $\boldsymbol{\beta}$ is done using a profile likelihood approach and an iterative reweighting algorithm (Chen, 2009) as follows:

1. For a candidate $\boldsymbol{\beta}^\star$, start with weight $\omega^{(0)}(t) = 1$ and Nelson-Aalen estimators for $\widehat{dH}_r^{(0)}(\boldsymbol{t})$ and $\widehat{dH}_d^{(0)}(\boldsymbol{t})$.
2. Repeat until convergence of $\widehat{dH_r(\boldsymbol{t})}$ and $\widehat{dH_d(\boldsymbol{t})}$:
   - Use $\widehat{dH}_r^{(0)}(x)$ and $\widehat{dH}_d^{(0)}(x)$ to calculate $\Theta_1^r(x), \Theta_1^d(x), \Theta_2^r(x), \Theta_2^{nr}(x) \forall x \in \boldsymbol{t}$.
   - Use $\Theta_1^r(x)$, weight $\omega_1^{r(0)} = 1$, and eq. (4.4) to calculate $\widehat{dH}_r^{(1)}(x)$ at each time $x \in \boldsymbol{t}$. With the weight fixed, the right hand side of eq. (4.4) only depends on prior jumps $H_r(t)$, $\forall t < x$. Sequentially calculate the jump size in baseline hazard along all time $x \in \boldsymbol{t}$ to get the full path $\widehat{H_r^{(1)}}$.
   - Similarly, use $\Theta_1^d(x), \Theta_2^r(x), \Theta_2^{nr}(x)$, weight $\omega^{(0)} = 1$ and eq. (4.5) to calculate $\widehat{H_d^{(1)}}$.
   - Use full path $\widehat{H_r^{(1)}}$ and $\widehat{H_d^{(1)}}$ to update weights $\omega_1^{r(1)}, \omega_1^{d(1)}, \omega_2^{r(1)}, \omega_2^{nr(1)}$.

3. Calculate the profile log-likelihood $\ell_{\text{pr}}(\beta^\star, \widehat{H_r}, \widehat{H_d})$ using eq. (4.3).

4. Repeat steps (1-3) to search for $\hat{\beta} = \arg\max_{\beta} \ell(\beta)$.

Standard errors are estimated using the Hessian of the profile log-likelihood (see Appendix B.4 for proof of consistency).

4.3 Simulation studies

4.3.1 Performance of estimation procedure

There are many situations where a subgroup of patients react differently to treatment regimens, thus the ability to adjust for auxiliary covariates is crucial to identify the best regimen. Motivated by this, we designed simulations following SMART design II where patients with different values of baseline covariates $V \sim \text{Bern}(0.75)$ had different optimal regimens.

We performed Monte Carlo simulations to assess the proposed methodology in estimating covariate effects from a SMART with survival outcome. The parameters were set as follow: n=1000, first treatment randomization probability was $P(X = 1) = 0.5$. Second treatment randomization probabilities were $P_1 = 0.5$, and $P_2 = 0.5$. Baseline hazard $H_r \sim \text{Weibull}(0.75, 4)$ and $H_d \sim \text{Weibull}(3, 4)$. Time-to-response and time-to-death were generated under Cox proportional hazards

- $\lambda_r(t) = h_r(t) \exp\{\beta_1 V + \beta_2 X\}$
- $\lambda_d(t) = h_d(t) \exp\{\beta_3 V + \beta_4 X + \Delta^r \mathbb{1}(t > t^r)(\beta_5 + \beta_6 X + \beta_7 Z^r + \beta_8 X Z^r + \beta_9 X Z^r V)\}$

where $\boldsymbol{\beta} = [-1.5, 0.6, -0.5, -1.3, -1.2, -0.7, -1, 1.8, -2.4]$.

We chose $\boldsymbol{\beta}$ to demonstrate the situation where: (i) a particular first-line and second-line treatment were superior on their own, but their interaction was not as beneficial overall compared to another combination from less effective single-line treatments and (ii) subgroups of patients reacted to regimens differently. The for-

mer justifies the evaluation of DTRs instead of single-line treatment comparisons. Including the interaction between V and treatment further motivates our model over existing methods.

In particular, A_2 (ie. $X = 1$) was better than A_1: $\beta_2 = 0.6$ indicated faster time-to-response and $\beta_4 = -1.3$ indicated A_2 resulted in a lower risk of death overall for non-responders and in the period pre-response for responders. After response and second treatment randomization, $\beta_7 = -1$ indicated B_2 (ie. $Z^r = 1$) was better than B_1. However, $\beta_8 = 1.8$ increased risk and made the combination A_2B_2 undesirable. Thus, $A_2B_1C_1$ was the best regimen for reference subgroup where $V = 0$. On the other hand, $\beta_9 = -2.4$ made $A_2B_2C_1$ the best regimen for subgroup where $V = 1$.

The cut-off time for response was set at $t^{nr} = 5$. Censoring time was generated from Unif$(0, 20)$, which resulted in 10% censored, 25% death, 50% response and 15% non-response by time t^{nr}. Of the 65% that received second-line treatment, 25% were censored before death.

Table 4.1: Simulation for Design II: Estimated log hazard ratios and SE

		Response		Death first		Death after 2nd randomization				
N	Statistics	β_θ		β_γ		β_η				
	True	**-1.5**	**0.6**	**-0.5**	**-1.3**	**-1.2**	**-0.7**	**-1**	**1.8**	**-2.4**
200	Bias	0.00	0.01	-0.02	-0.01	0.00	-0.07	-0.03	0.10	-1.03
	SD($\hat{\beta}$)	0.24	0.24	0.24	0.23	0.36	0.53	0.54	0.80	2.75
	Avg($\widehat{SE}$)	0.24	0.24	0.24	0.22	0.36	0.52	0.51	0.77	6.05
	95% CP	94.9	94.8	95.8	95.0	94.8	94.9	93.9	95.0	97.4
500	Bias	0.02	0.01	0.00	0.01	0.02	-0.03	0.00	0.02	-0.19
	SD($\hat{\beta}$)	0.15	0.16	0.14	0.14	0.22	0.32	0.28	0.44	0.96
	Avg($\widehat{SE}$)	0.15	0.15	0.15	0.14	0.22	0.31	0.28	0.43	0.81
	95% CP	94.9	94.4	95.3	95.9	94.6	94.9	95.7	95.1	96.1

SD($\hat{\beta}$): Empirical standard deviation of estimated $\hat{\beta}$ coefficients
Avg($\widehat{SE}$): Average of 1000 estimated standard errors (using Hessian matrix)
95% CP: 95% coverage probability.

The simulation results are summarized in Table 4.1. The estimated log-hazard-ratios (LHRs) for time-to-response and time-to-death were very close to the true β. The averaged estimated standard errors were consistent with empirical standard

errors of $\hat{\beta}$. Both were smaller as sample size increased. The only discrepency was with the interaction term between V and regimen A_2B_2 ($\beta_\eta = -2.4$) in the smaller sample size (n=200). This is understandable because of our setup with equal treatment randomization probabilities and the distribution of V, only 18.75% of the sample experienced this interaction effect ($\sim$ 37 individuals). At sample size n=500, the interaction was adequately estimated.

4.3.2 Predicted survival rates: Comparison of the joint model with existing methods

As noted in Section 4.1, a DTR includes both responders and non-responders that are consistent with the regimens. In design II, regimen $\{A_1B_1C_1\}$ includes the A_1 responders who get B_1 and the A_1 non-responders who get C_1. However, traditional Kaplan-Meier estimation based on the combined sample of these two subgroups is biased because it overweights the non-responders. To correct for this bias, the responders who receives B_1 have to represent themselves and be up-weighted to account for those that are re-randomized to B_2. The same bias has to be corrected in design I if randomization probabilities P or Q are not balanced (1/2 for each treatment arm).

Lunceford et al. (2002) corrected this bias by using the randomization probabilities to assign weight for each patient's contribution to regimen $A_jB_kC_l$ as $W_{jkl} = \mathbb{1}(A_j)\left\{\frac{\Delta^r\mathbb{1}(B_k)}{P(B_k|A_j)} + \frac{\Delta^{nr}\mathbb{1}(C_l)}{P(C_l|A_j)}\right\}$. Patients who died before response were classified as non-responders. Using these inverse probability weights, they proposed a survival estimator for each regimen as

$$\hat{S}^{IPWE}_{jkl}(t) = 1 - n^{-1}\sum_{i=1}^{n}\frac{\Delta_i^d W_{jkli}}{\hat{K}(t)}\mathbb{1}(T_{2i} \le t) + \frac{\alpha}{n}\sum_{i=1}^{n}\frac{\Delta_i^d}{\hat{K}(t)}(W_{jkli} - 1), \qquad (4.6)$$

where $\hat{K}(t)$ is the Kaplan-Meier estimate from censoring time to adjust for censored patients and α is chosen so that the estimator has minimum variance.

Guo and Tsiatis (2005) defined the response indicator by $\Delta_i^r(t) = \Delta_i^r \mathbb{1}(T_i^r \leq t)$ making the weighting scheme time dependent. This means that before the time of response, responders to A_1 who would eventually be re-randomized to B_2 were also consistent with regimen $A_1B_1C_1$. Using this weight, they proposed the weighted risk set estimator (WRSE) for the survival of regimen $A_jB_kC_l$ as follows:

$$\hat{S}_{jkl}^{WRSE}(t) = \exp\left\{-\int_0^t \frac{\sum_{i=1}^n W_{jkli}(u)dN_i(u)}{\sum_{i=1}^n W_{jkli}(u)Y_i(u)}\right\}. \tag{4.7}$$

It is easy to see that if we ignore weights W_{jkl} and $\hat{K}(t)$, IPWE is the proportion of subjects that have not yet failed before t, and WRSE is a Nelson-Aalen survival estimator. The disadvantage of both estimates is that they can not be adjusted for auxiliary covariates such as V, only that one can perform subset analyses for different values of V.

Tang and Wahed (2015) proposed a model based on Cox with IPW that can incorporate baseline covariates. Using their proposed cumulative hazards, we calculated the survival prediction for each DTR and compared it against our model in the following simulation.

The advantage of our proposed joint model is that not only do we have a model to parameterize how auxiliary covariates and treatments in each stage affect time-to-response and time-to-death and their interaction, but also that we can derive closed-form expressions for the marginal survival function (eq. 4.8) of each DTR, analogous to eq.(4.6) and (4.7), without having to reconfigure the weighting scheme

for different SMART designs.

$$\hat{S}(t) = \begin{cases} \int_0^t f_r S_d(r,t)dr + S_r S_d(t,t), & t < t^{nr} \\ \int_0^{t^{nr}} f_r S_d(r,t)dr + S_r S_d(t^{nr},t), & t \geq t^{nr} \end{cases}$$

$$= \begin{cases} \int_0^t h_r(r)\theta e^{-\theta H_r(r)-(\gamma-\eta)H_d(r)-\eta H_d(t)}dr + e^{-\theta H_r(t)-\gamma H_d(t)}, & t < t^{nr} \\ \int_0^{t^{nr}} h_r(r)\theta e^{-\theta H_r(r)-(\gamma-\eta)H_d(r)-\eta H_d(t)}dr + e^{-\theta H_r(t^{nr})-H_d(t^{nr})(\gamma-\mu)-H_d(t)\mu}, & t \geq t^{nr} \end{cases} \tag{4.8}$$

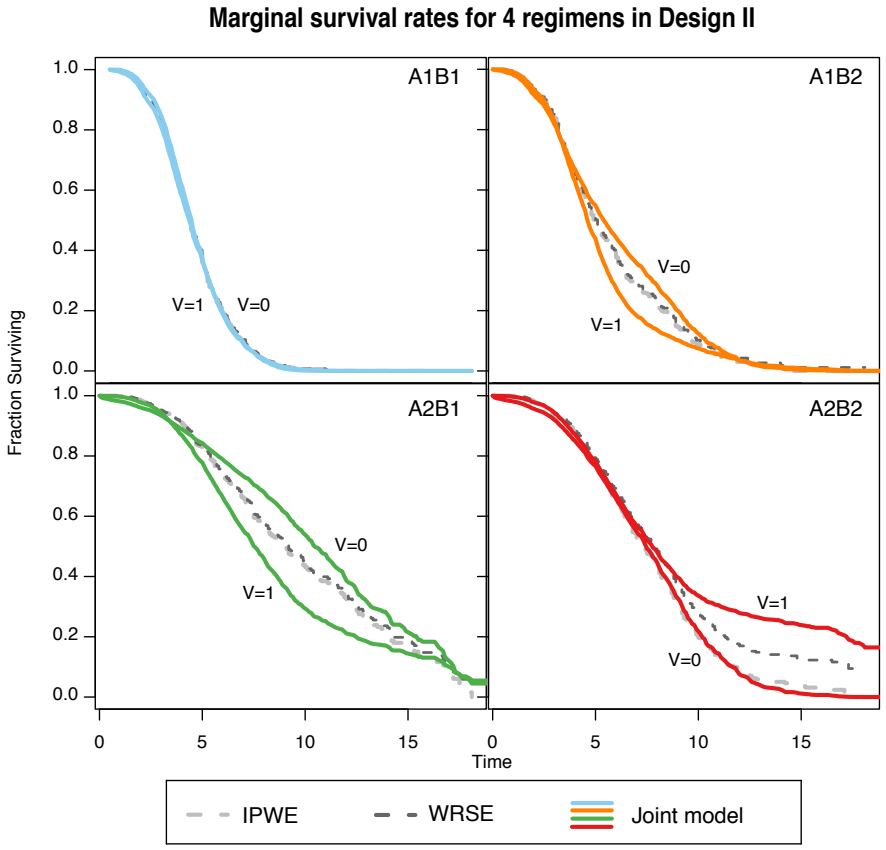


Figure 4.2: Estimated survival probabilities for 4 DTRs in Design II: Joint modeling (solid lines, eq. 4.8) compared with IPWE (gray dashed lines, eq. 4.6) and WRSE (black dashed lines, eq. 4.7).

Figure 4.2 plots the survival estimates for the four DTRs in Design II using our proposed model and non-parametric estimators from Lunceford et al. (2002) and Guo and Tsiatis (2005). The solid lines denote our model survival prediction conditioning

on regimen and baseline covariate V (eq. 4.8). The gray and black dashed lines denote the IPWE (eq. 4.6) and WRSE (eq. 4.7), respectively. The latter two estimators do not have a way to adjust for baseline V, thus only provide an overall prediction for each regimen.

It is easy to see that the ability to adjust for auxiliary covariates such as V at baseline hazards (and potentially other relevant covariates at second randomization) is important. In the top left panel of Figure 4.2, the effect of reigmen $A_1B_1C_1$ is identical for both subgroups of V, thus our survival prediction conditioning on V overlaps that of IPWE and WRSE. However, this is not the case for regimens $A_1B_2C_1$, $A_2B_1C_1$, and $A_2B_2C_1$. We can see that different strata of V react differently to these regimens. Group $V = 0$ had the highest survival rates from $A_2B_1C_1$, while group $V = 1$ benefited from $A_2B_2C_1$. This information would be lost if we ignored V and used IPWE and WRSE, where the best conclusion would be that $A_2B_1C_1$ and $A_2B_2C_1$ were equally optimal for the whole population on average. If we average our model's survival prediction over V, we get consistent prediction with WRSE. IPWE, on the other hand, does not perform well for $A_2B_2C_1$ regimen (Figure 4.2 bottom right, gray dashed line). This is because the weights in IPWE are time-independent, thus ignoring the information from people who receive $A_2B_1C_1$ (e.g. assigning weight 0) in the survival estimate for $A_2B_2C_1$. WRSE, on the other hand, assigns non-zero weight for $A_2B_1C_1$ during the period before second randomization (i.e. when only A_2 is administered). Our joint model naturally achieves this advantage of WRSE. Regardless, only our proposed model is able to handle situations with multiple and/or continuous auxiliary covariates where subset analysis is not feasible.

We should note that in special cases where V is categorical with a small number of subgroups, IPWE and WRSE can adjust for V by subset analysis. In such cases, our

model's prediction is consistent with the survival predictions from subset analysis using IPWE and WRSE (see Appendix C.3 for simulation results).

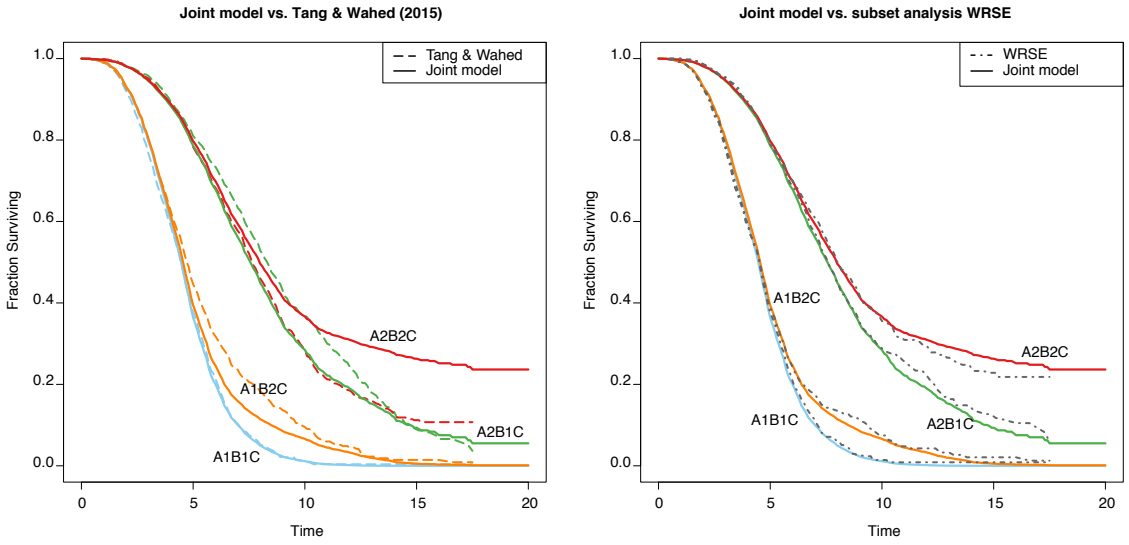


Figure 4.3: Methods by Tang and Wahed (2015) cannot adjust for interactions between V and regimens (left plot). The proposed joint model's predictions are consistent with subset analysis using WRSE (right plot).

Unlike our joint model, the Cox model proposed by Tang and Wahed (2015) only performs well under certain conditions. Even though this model can adjust for baseline covariates V, it requires the assumption that the effect of V is constant across different DTRs (i.e. there is no interaction between covariates V and regimens). In simulations where there this is the case, the survival prediction from Tang and Wahed (2015) is consistent with our joint model (Appendix C.3). However, when different V subgroups have different optimal DTRs, only our model's predictions and subset analyses of WRSE (when categorical) remain consistent. The model in Tang and Wahed (2015) fails to account for the interaction between covariates and treatment which leads to the inaccurate survival predictions shown in the left plot of Figure 4.3.

4.4 Multiple comparisons with the best (MCB)

4.4.1 Methods

Many DTRs are embedded within a SMART, and often the comparisons between DTRs are of interest. For example, there are four embedded DTRs in design II leading to six pairwise comparisons. Design I has eight DTRs leading to potentially 28 pairwise comparisons. As the number of DTRs increases, standard comparison methods begin to lose statistical power. The ultimate goal of a SMART may be to identify the best DTR of all embedded DTRs or to find a set of best DTRs if their differences are minimal. Multiple comparisons with the best (MCB) (Hsu, 1996) was introduced to address this issue in SMART with continuous outcomes (Ertefaie et al., 2016), but this method has not been extended to address survival outcomes in the SMART setting.

For ease of notation, we use $\hat{S}_m$ to denote $\hat{S}(t)$ for DTR m, where $m = 1, \ldots, M$ and t is the time at which we want to compare all DTR survival probabilities. Let $\mathfrak{B}$ be the set of DTRs that results in best $\hat{S}(t)$ and others that can not be differentiated from the former at a certain threshold c. If we plot the survival rate from each DTR at a particular time of interest (Figure 4.4), regimen l with highest survival rate should be selected for inclusion in set $\mathfrak{B}$. Regimen m that is very close to the best is also of interest (e.g. it may be less expensive, easier to administer, or have fewer side effects). Therefore, we want to implement MCB to identify regimens within a distance c from the best one.

The steps to find c_m, the threshold for each regimen m, are as follows: Since regimens with different first-line treatments are independent, their covariances are 0. Thus, the vector of survival probabilities at time t for all regimens $\hat{\boldsymbol{S}} = (\hat{S}_1(t), \ldots, \hat{S}_M(t))$

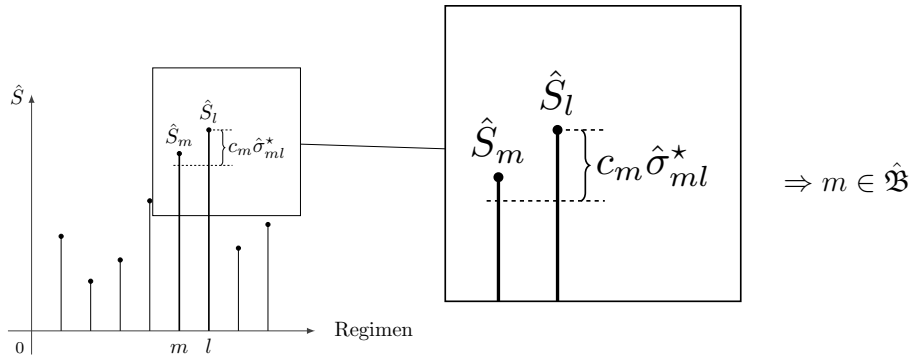


Figure 4.4: Multiple comparisons with the best

has the following covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{4.9}$$

where Σ_1 is the covariance matrix for $\hat{S}(t)$ from DTRs that start with first-line treatment A_1 and Σ_2 is from DTRs that start with A_2. In design I, Σ_1 and Σ_2 are 4×4 matrices. In design II, they are both 2×2. In design III, Σ_1 is 2×2 and Σ_2 is a scalar. Denote $\sigma_{m,l}$ the element of Σ_1 and Σ_2 (i.e. the covariance between $\hat{S}_m(t)$ and $\hat{S}_l(t)$).

The set of best DTRs is $\hat{\mathfrak{B}} = \{m \mid (\hat{S}_m - \hat{S}_l)/\sigma^\star_{ml} \geq c_m, \forall l \neq m\}$, where

$$\hat{\sigma}^\star_{ml} = \sqrt{\text{var}(\hat{S}_m - \hat{S}_l)} = \sqrt{(\hat{\sigma}^2_m + \hat{\sigma}^2_l - 2\hat{\sigma}_{m,l})}. \tag{4.10}$$

A natural choice for c_m is a value that satisfies α Type 1 error rate (i.e. if the differences between DTRs are 0, each has $(1-\alpha)$ chance of being included in $\hat{\mathfrak{B}}$). Then c_m satisfies

$$\begin{aligned} 1-\alpha =& P(Z_m \geq Z_l - c_m\sigma^\star_{ml}, \forall l \neq m) \\ =& P(c_m \geq (Z_l - Z_m)/\sigma^\star_{ml}, \forall l \neq m) \\ =& \int P\left[c_m \geq \max_{l \neq m}\left(\frac{Z_l - z}{\sigma^\star_{ml}}\right)\right] d\phi(z) \end{aligned}$$

where $Z_1, \ldots, Z_M$ are MVN$(0, \Sigma)$, and $\phi(z)$ is the CDF of Z_m. The procedure to find c_m is adapted from the procedure used for SMART data with continuous outcomes (Ertefaie et al., 2016):

1. Estimate the boostrapped covariance matrix $\hat{\Sigma}$ between DTR survival rates at any time t of interest.
2. Simulate $B = 1000$ samples of $Z_{1b}, \ldots, Z_{Mb} \sim \text{MVN}(0, \hat{\Sigma})$.
3. For each $b = 1, \ldots, B$, calculate $\max_{l \neq m} [(Z_{lb} - Z_{mb})/\sigma^{\star}_{ml}]$, where $\sigma^{\star}_{ml}$ can be calculated from $\hat{\Sigma}$ and eq.(4.10).
4. c_m is the $(1-\alpha)^{th}$ quantile of the $(1 \times B)$ vector found in previous step.

4.4.2 Simulation study

We applied MCB to simulated data from SMART design II with four possible DTRs. Implementation of MCB for designs I and III is identical.

The parameters are as follows: sample size was 500; randomization probabilities was set to 0.5 for both first-line X and second-line treatments Z^r; censoring time $C \sim \text{Unif}(0, 20)$. The hazard of response was $\lambda_r(t) = h_r(t) \exp\{\delta X\}$, where $h_r(t) \sim$ Weibull$(0.75, 4)$. With $h_d(t) \sim$ Weibull$(3, 4)$, the hazard of death was

$$\lambda_d(t) = h_d(t) \exp\left\{-\delta X + \Delta^r \mathbb{1}(t > t^r)\left(-\frac{\delta}{4}X - \frac{\delta}{2}Z^r\right)\right\}.$$

When $\delta = 0$, all four DTRs were equally effective. As δ increased, the individuals following DTR $A_2B_2C_1$ $(X = 1, Z^r = 1)$ responded faster and had lower risk of death (both before and after response) compared to those following other DTRs. For each δ between 0 and 2, we simulated 500 datasets and applied MCB to identify the set of best DTRs. Using survival rates at time $t = 7$ and $t = 15$ as the metric of comparison, the average number of DTRs in $\hat{\mathfrak{B}}$ over 500 datasets was recorded.

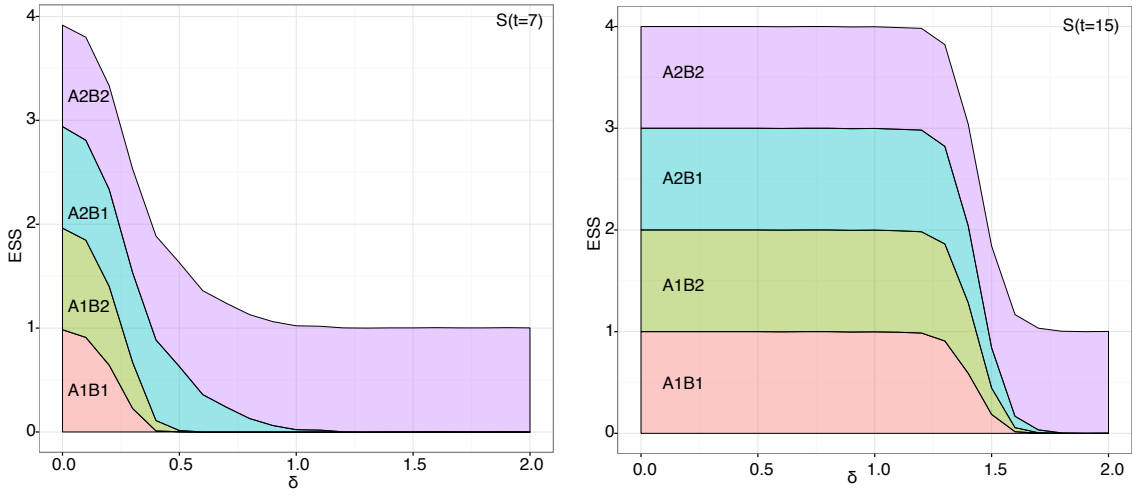


Figure 4.5: MCB selection for simulated SMART design II: the y-axis is the expected size of best DTRs set (ESS) and x-axis is various values of δ.

Figure 4.5 shows the probabilities that each of the four DTRs was chosen to be in the set of "best" DTRs for different δ values. The sum of these probabilities is the expected size of set $\hat{\mathfrak{B}}$ (ESS). As expected, ESS converged to 1 as treatment effect δ gets larger. At $\delta = 0$, all regimens were equally beneficial; thus all four DTRs were chosen to be included in $\hat{\mathfrak{B}}$. As δ increased, $A_2B_2C_1$ became the optimal regimen with the lowest risk of death, making $\hat{\mathfrak{B}}$ a set of size 1. If interest is in the comparison of DTRs at time 7, the left plot shows MCB results using survival rates at this time. ESS quickly converged to 1 with $A_2B_2C_1$ being the chosen regimen. The right plot compares DTRs at a later time, $t = 15$, where for $\delta < 1.3$, most patients died before time 15 and all four DTRs were comparable. At larger δ, $A_2B_2C_1$ had better survival rates compared to the other three DTRs.

4.5 Discussion

We proposed a joint modeling framework to evaluate dynamic treatment regimens embedded in a SMART. Previously proposed methods for data from a SMART with survival outcomes employ inverse probability weighting and either treat survival

time as a continuous outcome or compare survival rates between DTRs using non-parametric estimators. Our proposed model is based on Cox models with flexible non-parametric baseline hazards for time-to-response and time on second-line treatment. We demonstrated that our model provides equally unbiased surival estimates as the non-parametric methods, while offering many other advantages.

Compared to the non-parametric estimators, our model offers the following features: (i) adjusts for auxiliary covariates that includes not only baseline information but also potentially new covariates at the end of first-line treatment, (ii) has a mechanism to parameterize/measure each treatment's effect within the regimen, their interactions, and interactions with auxiliary covariates (without assumptions of these effects within or between DTRs), (iii) easy adaptation between different SMART designs by setting certain β's to 0, and (iv) the ability to estimate conditional survival functions conditioning on not only the particular DTR of interest, but also on what has been observed so far (e.g. $\hat{S}(t \mid \text{response at } s)$ or $\hat{S}(t \mid \text{no response by } s)$).

The limitation of our model is that the non-parametric estimation of baseline hazards requires larger sample sizes (depending on the total number of unique time-points from all subjects in the data) and longer computational time compared to other models where a distribution is assumed. If only a small sample size is available, we can fit our model assuming some parametric distributions for the baseline hazards (e.g. Weibull). This will decrease the number of parameters needed to be estimated, yet still offers all of the advantages discussed.

The issue of multiple comparisons has not been fully addressed for survival data in the SMART setting. In order to identify the optimal DTR(s), SMARTs with many embedded DTRs may require a large number of comparisons resulting in a loss of statistical power. We adapt the previously proposed approach using multiple

comparisons with the best for continous outcomes to compare DTRs with survival outcomes. We compare each DTR's predicted survival rate at time t with the best of others, then identify a set of DTRs that contains the true best DTR with a certain type I error rate. We outlined the procedure to find rejection rule that would satisfy such type I error and demonstrated the method with simulations.

CHAPTER V

Conclusions and Future Work

This dissertation presented joint models that simultaneously analyze multiple time-to-events, where events are related stages in two settings: cancer progression stages (latent metastasis, observed diagnosis and death) and sequential rounds of treatment in a SMART.

In chapter II, we presented a joint model of time-to-diagnosis, time-to-metastasis, and time-to-death motivated by prostate cancer progression. Assuming the same underlying biological cause drives the non-terminal and terminal events, previously proposed joint models used the same baseline hazard for a series of nested proportional hazard models to construct the joint likelihood of the entire disease timeline (Hu and Tsodikov, 2014a). This assumption assists the incorporation of latent events (e.g. closed-form integration of interval censored metastasis status), but it is too restrictive as the number of disease stages increases (e.g. three stages in the prostate cancer analysis, section 2.4). To address this problem without compromising computational efficiency and a mechanism to relate the risk of one event to another, the joint model in chapter II estimated the baseline hazards for one stage non-parametrically and for other stages using parametric transformations of the former. Extension to handle latent disease stages where a change in status is only known at a certain time

points (e.g. metastasis status at diagnosis time) is also demonstrated.

The removable pathways in constructing the joint likelihood further makes the proposed model adapatable to describe different disease processes. In some cancers, death occurs after metastasis. In others, death may occur at any point after diagnosis. To model the latter, an additional term to the likelihood can reflect the possibility of a patient dying without intermediate stages, while the estimation procedure remains the same. This can also be generalized to model different segments of the diseases, such as the time between recurrence and death.

Chapter III was motivated by the need to model the observed adversed effect of early diagnosis age on breast cancer survival. We extended the joint model in chapter II to allow for multiple non-parametric baseline hazards and an additional common latent frailty term to capture the unaccounted dependence between cancer stages. The model can be adapted to allow for different sets of covariates, both time-dependent and time-independent, at each stage. Due to the complexity of the joint modeling and non-parametric component, the NPMLE estimation can be time-consuming, but incorporating C++ via R package "Rcpp" speeds up the process significantly.

Future research in this area includes modeling recurring stages, partial missingness, incorporating screening history, and competing risks due to death from other causes. More consideration regarding the intepretation of the treatment effect in analysis of observational data is also needed. As treatment assignment is not known during pre-diagnosis stages in the disease progression, it can be argued to be dependent on the patient's underlying disease process and thus cannot be used as external covariates. Development to speed up the estimation procedure would also help make this joint modeling approach more attractive to researchers.

In chapter IV, the joint modeling approach was applied to survival data from a SMART. Compared to previously proposed methods, this approach can adjust for auxiliary covariates beside treatments (e.g. baseline covariate and new covariates at second randomization). It also provides a mechanism to parameterize treatment effects within the regimen and interactions between treatment and covariates. Simulations show that the proposed joint model allows for flexible modeling of various interactions between treatments and auxiliary covariates. This model benefits researchers because accurate conditional survival predictions can be used to inform treatment decisions for ongoing patients. As individualized treatment and SMART designs gain popularity, joint models offer an approach that can be tailored to many settings with different types of outcomes. Its application to study treatment regimens can be extended to SMART with more than two stages. Additional research is required to adapt these models for smaller samples.

APPENDICES

APPENDIX A

Supplementary Materials for Chapter II

A.1 Derivation of likelihood terms

For ease of notation, we suppress the subscript i for each individual subject. The likelihood for different scenarios of diagnosis, metastasis, and death are:

- Subject is diagnosed without metastasis at t_1, metastasizes, and dies at t_2:

$$\begin{aligned}
&\mathcal{L}_1^*(t_1,t_2) = P(T_1^*=t_1, T_2^*=t_2, t_1<T_u<t_2) = \int_{t_1}^{t_2} f_U f_1 f_2(t_u,t_1,t_2)dt_u \\
&= \int_{t_1}^{t_2} \left[h_u\tilde{\eta}e^{-\eta H_1-\tilde{\eta}(H_u-H_1)}\right]\left[\delta_1 H_1^{\delta 1-1}h_1\theta e^{-\theta H_1^{\delta_1}}\right]\left[\delta_2 H_2^{\delta 2-1}h_2\gamma e^{-\gamma(H_2^{\delta_2}-H_u^{\delta_2})}\right]du \\
&= \delta_1\delta_2 H_1^{\delta_1-1}H_2^{\delta_2-1}h_1h_2\tilde{\eta}\theta\gamma e^{-H1(\eta-\tilde{\eta})-H_1^{\delta_1}\theta-H_2^{\delta_2}\gamma}\int_{t_1}^{t_2} h_u e^{-H_u\tilde{\eta}+H_u^{\delta_2}\gamma}du.
\end{aligned}$$

- Subject is diagnosed without metastasis at t_1 and is censored at t_2

$$\begin{aligned}
\mathcal{L}_2^*(t_1,t_2) &= P(T_1^*=t_1, T_2^*>t_2, T_u>t_2) + P(T_1^*=t_1, T_2^*>t_2, t_1<T_u<t_2) \\
&= S_U f_1 S_2(t_2,t_1,t_2)dt_u + \int_{t_1}^{t_2} f_U f_1 S_2(t_u,t_1,t_2)dt_u \\
&= \left[e^{-\eta H_1-\tilde{\eta}(H_2-H_1)}\right]\left[\delta_1 H_1^{\delta_1-1}h_1\theta e^{-\theta H_1^{\delta_1}}\right][1] \\
&\quad + \int_{t_1}^{t_2}\left[h_u\tilde{\eta}e^{-\eta H_1-\tilde{\eta}(H_u-H_1)}\right]\left[\delta_1 H_1^{\delta 1-1}h_1\theta e^{-\theta H_1^{\delta_1}}\right]\left[e^{-\gamma(H_2^{\delta_2}-H_u^{\delta_2})}\right]du \\
&= \delta_1 H_1^{\delta_1-1}h_1\theta e^{-H_1(\eta-\tilde{\eta})-H_1^{\delta_1}\theta}\left[e^{-H_2\tilde{\eta}} + \tilde{\eta}\int_{t_1}^{t_2} h_u e^{-H_u\tilde{\eta}+H_u^{\delta_2}\gamma-H_2^{\delta_2}\gamma}du\right].
\end{aligned}$$

- Subject is diagnosed with metastasis at t_1 and dies at t_2

$$\mathcal{L}_3^*(t_1,t_2) = P(T_u < t_1, T_1^* = t_1, T_2^* = t_2) = \int_0^{t_1} f_U f_1 f_2(t_u, t_1, t_2) dt_u$$
$$= \int_0^{t_1} \left[h_u \eta e^{-\eta H_u}\right] \left[\delta_1 H_1^{\delta 1 - 1} h_1 \theta\mu e^{-\theta\mu H_1^{\delta_1} - \theta\overline{\mu} H_u^{\delta_1}}\right] \left[\delta_2 H_2^{\delta_2 - 1} h_2 \gamma e^{-\gamma(H_2^{\delta_2} - H_1^{\delta_2})}\right] du$$
$$= \delta_1\delta_2 H_1^{\delta_1 - 1} H_2^{\delta_2 - 1} h_1 h_2 \eta\theta\mu\gamma e^{-H_1^{\delta_1}\theta\mu + H_1^{\delta_2}\gamma - H_2^{\delta_2}\gamma} \int_0^{t_1} h_u e^{-H_u\eta - H_u^{\delta_1}\theta\overline{\mu}} du.$$

- Subject is diagnosed with metastasis at t_1 and is censored at t_2

$$\mathcal{L}_4^*(t_1,t_2) = P(T_u < t_1, T_1^* = t_1, T_2^* > t_2) = \int_0^{t_1} f_U f_1 S_2(t_u, t_1, t_2) dt_u$$
$$= \int_0^{t_1} \left[h_u \eta e^{-\eta H_u}\right] \left[\delta_1 H_1^{\delta 1 - 1} h_1 \theta\mu e^{-\theta\mu H_1^{\delta_1} - \theta\overline{\mu} H_u^{\delta_1}}\right] \left[e^{-\gamma(H_2^{\delta_2} - H_1^{\delta_2})}\right] du$$
$$= \delta_1 H_1^{\delta_1 - 1} h_1 \eta\theta\mu e^{-H_1^{\delta_1}\theta\mu + H_1^{\delta_2}\gamma - H_2^{\delta_2}\gamma} \int_0^{t_1} h_u e^{-H_u\eta - H_u^{\delta_1}\theta\overline{\mu}} du.$$

- Subject is censored at t_1 before any event is observed

$$\mathcal{L}_5(t_1) = P(T_u < t_1, T_1^* > t_1, T_2^* > t_1) + P(T_u > t_1, T_1^* > t_1, T_2^* > t_1)$$
$$= \int_0^{t_1} f_U S_1 S_2(t_u, t_1, t_1) dt_u + S_U S_1 S_2(t_1, t_1, t_1)$$
$$= \int_0^{t_1} \left[h_u \eta e^{-\eta H_u}\right] \cdot \left[e^{-\theta\mu H_1^{\delta_1} - \theta\overline{\mu} H_u^{\delta_1}}\right] du + \left[e^{-\eta H_1}\right] \cdot \left[e^{-\theta H_1^{\delta_1}}\right] du$$
$$= \eta e^{-H_1^{\delta_1}\theta\mu} \int_0^{t_1} h_u e^{-H_u\eta - H_u^{\delta_1}\theta\overline{\mu}} du + e^{-H_1\eta - H_1^{\delta_1}\theta}.$$

A.2 Derivation of hazard terms

For ease of notation, we suppress the subscript i for each individual subject.

1. The probability of diagnosis at time t:

   - If diagnosed before metastasis

$$\Theta_1(H_t,\beta) dH_t = Pr[dN_1^-(t) = 1 \mid Y_1(t) = 1] = \mathcal{L}_1/\mathcal{L}_5$$
$$\Rightarrow \Theta_1(H_t,\beta) = \frac{\mathcal{L}_1}{\mathcal{L}_5 h_1} = \frac{\delta_1 H_t^{\delta_1 - 1}\theta}{1 + \eta\int_0^t h_u e^{\eta(H_t - H_u) + \theta\overline{\mu}(H_t^{\delta_1} - H_u^{\delta_1})} du}.$$

- If diagnosed after metastasis

$$\Theta_1(H_t,\beta)dH_t = Pr[dN_1^+(t)=1 \mid Y_1(t)=1] = \mathcal{L}_3/\mathcal{L}_5$$
$$\Rightarrow \Theta_1(H_t,\beta) = \frac{\mathcal{L}_3}{\mathcal{L}_5 h_1} = \frac{\delta_1 H_t^{\delta_1-1}\eta\theta\mu\int_0^t h_u e^{\eta(H_t-H_u)+\theta\overline{\mu}(H_t^{\delta_1}-H_u^{\delta_1})}du}{1+\eta\int_0^t h_u e^{\eta(H_t-H_u)+\theta\overline{\mu}(H_t^{\delta_1}-H_u^{\delta_1})}du}.$$

2. The probability of death at time t:

- If diagnosed without metastasis at time t_1 prior to t

$$\Theta_2(H_t,\beta,t_1)dH_t = Pr[dN_2^-(t)=1 \mid Y_2^-(t)=1] = \widetilde{\mathcal{L}_1}/\widetilde{\mathcal{L}_2} = \mathcal{L}_1^\star/\mathcal{L}_2^\star$$
$$\Rightarrow \Theta_2(H_t,\beta,t_1) = \frac{\mathcal{L}_1^\star}{\mathcal{L}_2^\star h_2} = \frac{\delta_2 H_t^{\delta_2-1}\tilde{\eta}\gamma\int_{t_1}^t h_u e^{\tilde{\eta}(H_t-H_u)-\gamma(H_t^{\delta_2}-H_u^{\delta_2})}du}{1+\tilde{\eta}\int_{t_1}^t h_u e^{\tilde{\eta}(H_t-H_u)-\gamma(H_t^{\delta_2}-H_u^{\delta_2})}du}.$$

- If diagnosed with metastasis at time t_1 prior to t

$$\Theta_2(H_t,\beta,t_1)dH_t = Pr(dN_2^+(t)=1 \mid Y_2^+(t)=1) = \widetilde{\mathcal{L}_3}/\widetilde{\mathcal{L}_4} = \mathcal{L}_3^\star/\mathcal{L}_4^\star$$
$$\Rightarrow \Theta_2(H_t,\beta,t_1) = \frac{\mathcal{L}_3^\star}{\mathcal{L}_4^\star h_2} = \delta_2 H_t^{\delta_2-1}\gamma.$$

A.3 Proof of Martingale Properties:

From the score functions (2.9) and (2.10),

$$U_\beta = \sum_{z_1\in Z_1}\int_0^\xi \frac{\Theta_{z_1,\beta}(x;\beta,H)}{\Theta_{z_1}(x;\beta,H)}dM_{z_1}(x) + \sum_{i=1}^n\int_0^\xi \frac{\Theta_{2i,\beta}(x;\beta,H,t_{1i})}{\Theta_{2i}(x;\beta,H,t_{1i})}dM_{2i}(x), \tag{A.1}$$

$$U_{H_t} = \sum_{z_1\in Z_1}\int_0^t\left[\frac{dM_{z_1}(x)}{dH_x} + \int_{x^+}^\xi \psi_{z_1}(u;\beta,H)dM_{z_1}(u)\right] + \sum_{i=1}^n\int_0^t\left[\frac{dM_{2i}(x)}{dH_x} + \int_{x^+}^\xi \psi_{2i}(u;\beta,H,t_{1i})dM_{2i}(u)\right] \tag{A.2}$$

Exchange the integrals over x and u, (A.2) becomes

$$\begin{aligned} U_{H_t} &= \sum_{z_1\in Z_1}\int_0^\xi \left[1_{(u\le t)} + H(u\wedge t)\psi_{z_1}(u;\beta,H)\right]dM_{z_1}(u) + \\ &\quad \sum_{i=1}^n\int_0^\xi \left[1_{(u\le t)} + H(u\wedge t)\psi_{2i}(u;\beta,H,t_{1i})\right]dM_{2i}(u) \\ &= \sum_{z_1\in Z_1}\int_0^\xi \varepsilon_{z_1}(u,t;\beta,H)dM_{z_1}(u) + \sum_{i=1}^n\int_0^\xi \varepsilon_{2i}(u,t;\beta,H,t_{1i})dM_{2i}(u), \end{aligned}$$

where $\varepsilon_\star(u,t;\beta,H) = \mathbb{1}(u \le t) + \int_0^{u\wedge t} \psi_\star(u;\beta,H)dH_x$.

As shown in Supplementary Materials E.3 of Hu and Tsodikov (2014a), the linear transform $\int_0^\xi \varepsilon_\star(u,t;H,\boldsymbol{\beta})\, dM_\star(u)$ is a martingale under the true model because

$$E\left[\int_0^\xi \varepsilon'(u,t)dM(u)|\mathcal{F}_{t-}\right] = dt\int_0^\xi \varepsilon'(u,t)E\{dM(u)|\mathcal{F}_{t-}\} = dt\int_0^t \varepsilon'(u,t)dM(u),$$

which equal 0 if $\varepsilon'(u,t) = \frac{\partial\varepsilon(u,t)}{\partial t} = 0$ for $u \le t$ (true by definition).

Therefore, the score function U_β and U_{H_t} are both martingales at the true model.

A.4 Profile likelihood Hessian

The following proof to justify the use of Hessian matrix to estimate the standard errors of $\boldsymbol{\beta}$ is based on (Rice and Tsodikov, 2016, Appendix F).

Let $\hat{H}_{\boldsymbol{\beta}}$ be the solution to $\mathcal{U}_H = 0$ for fixed $\boldsymbol{\beta}$, we want to show that $\mathbb{1}\mathrm{pr}^{-1} = \left(\frac{\partial^2\ell_{\mathrm{pr}}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)^{-1}$ is consistent for the covariance matrix of $\hat{\boldsymbol{\beta}}$, that is, the $\boldsymbol{\beta}\boldsymbol{\beta}$ submatrix of $\mathcal{I}_\infty^{-1}$, where

$$\mathcal{I}_\infty \equiv \begin{bmatrix} \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathcal{I}_{\boldsymbol{\beta}H} \\ \mathcal{I}_{H\boldsymbol{\beta}} & \mathcal{I}_{HH} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & -\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH_s} \\ -\frac{\partial^2\ell_\infty}{\partial dH_t\partial\boldsymbol{\beta}'} & -\frac{\partial^2\ell_\infty}{\partial dH_t\partial dH_s} \end{bmatrix}$$

is the asymptotic covariance matrix of the score for the full model. This may be expressed as $\mathrm{var}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\right] = Q^{-1}$, where

$$Q = \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \mathcal{I}_{\boldsymbol{\beta}H}\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\boldsymbol{\beta}}. \tag{A.3}$$

Denote by $J_{H\boldsymbol{\beta}}$ the Jacobian $\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}$. The full derivative (including over $\boldsymbol{\beta}$ in $\hat{H}_{\boldsymbol{\beta}}$) is

$$\frac{d\hat{\ell}}{d\boldsymbol{\beta}} = \int_s \frac{\partial\hat{\ell}}{\partial dH_s}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}} + \frac{\partial\hat{\ell}}{\partial\boldsymbol{\beta}}$$

so that the profile score is

$$\mathcal{U}_{\mathrm{pr}} = \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}} + \mathcal{U}_{\boldsymbol{\beta}}|_{H=\hat{H}_{\boldsymbol{\beta}}} = \mathcal{U}_{\boldsymbol{\beta}}|_{H=\hat{H}_{\boldsymbol{\beta}}} \tag{A.4}$$

since $\mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}} = 0$.

The profile Hessian is

$$\begin{aligned}
\frac{d^2\hat{\ell}}{d\boldsymbol{\beta}d\boldsymbol{\beta}'} &= \int_y\int_s \frac{\partial}{\partial dH(y)}\left[\frac{\partial\hat{\ell}}{\partial dH_s}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\right]\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} \\
&\quad + \int_y \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} + \int_s \frac{\partial}{\partial\boldsymbol{\beta}'}\left[\frac{\partial\hat{\ell}}{\partial dH_s}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\right] + \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \\
&= \int_y\int_s \left[\frac{\partial^2\hat{\ell}}{\partial dH_s\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} + \frac{\partial\hat{\ell}}{\partial dH_s}\frac{\partial^2 d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'}\right] \\
&\quad + \int_s \left[\frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial dH_s}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}'} + \frac{\partial^2\hat{\ell}}{\partial dH_s\partial\boldsymbol{\beta}'}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}} + \frac{\partial\hat{\ell}}{\partial dH_s}\frac{\partial^2 d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] + \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \\
&= -J_{\boldsymbol{\beta}H}\hat{I}_{HH}J_{H\boldsymbol{\beta}} + \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} - J_{\boldsymbol{\beta}H}\hat{I}_{H\boldsymbol{\beta}} + \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}.
\end{aligned}$$

Since $\mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}} = 0$, we have

$$\mathbb{1}\text{pr} = J_{\boldsymbol{\beta}H}\hat{I}_{HH}J_{H\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} + J_{\boldsymbol{\beta}H}\hat{I}_{H\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} \tag{A.5}$$

To express $\mathbb{1}$pr without Jacobians, we note that

$$\begin{aligned}
0 &= \frac{d}{d\boldsymbol{\beta}}\left(\frac{\partial\ell}{\partial dH_s}\bigg|_{H=\hat{H}_{\boldsymbol{\beta}}}\right) \\
&= \frac{\partial^2\hat{\ell}}{\partial dH_s\partial\boldsymbol{\beta}} + \int \frac{\partial^2\hat{\ell}}{\partial dH_s\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}} \\
&= -\hat{I}_{H\boldsymbol{\beta}} - \hat{I}_{HH}J_{H\boldsymbol{\beta}},
\end{aligned}$$

implying that

$$J_{H\boldsymbol{\beta}} = -\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}. \tag{A.6}$$

Substitution of (A.6) into (A.5) yields

$$\begin{aligned}
\mathbb{1}\text{pr} &= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \left[-\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\right]\hat{I}_{HH}\left[-\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}\right] + \hat{I}_{\boldsymbol{\beta}H}\left[-\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}\right] + \left[-\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\right]\hat{I}_{H\boldsymbol{\beta}} \\
&= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} - 2\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} \\
&= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}.
\end{aligned} \tag{A.7}$$

That is, we see by comparing (A.7) with (A.3) that $\mathcal{I}_\infty^{-1}$ has $\boldsymbol{\beta}$ submatrix $Q^{-1} = \mathbb{1}\mathrm{pr}^{-1}$. Since

$$\sqrt{n}\left(\hat{\Omega} - \Omega^0\right) = \mathcal{I}_\infty^{-1}\sqrt{n}\mathcal{U}^0 + o_p(1),$$

we have

$$\mathrm{var}\left[\sqrt{n}\left(\hat{\Omega} - \Omega^0\right)\right] = \mathcal{I}_\infty^{-1}\mathrm{cov}\left(\sqrt{n}\mathcal{U}^0\right)\mathcal{I}_\infty^{-1} = \mathcal{I}_\infty^{-1}\left[\mathcal{I}_\infty + o_p(1)\right]\mathcal{I}_\infty^{-1} = \mathcal{I}_\infty^{-1}.$$

Therefore,

$$\mathrm{var}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\right] = Q^{-1} = \mathbb{1}\mathrm{pr}^{-1}.$$

A.5 Simulation not included in main paper

A.5.1 Smaller sizes needed when $\delta_1 = \delta_2 = 1$

When $\delta_1 = \delta_2 = 1$, we can solve the integral in Θ_1 and Θ_2 mathematically instead of using numerical summation, thus estimation accuracy can be achieved at smaller sample sizes. We performed 1000 repetitions with sample sizes 250 and 500. Each scenario included covariates $Z_1 \sim U(0, 5)$ and an increasing or decreasing baseline hazard following a Weibull distribution $h_t = 0.02t$ (increasing) or $h_t = 0.05(t/10)^{-1/2}$ (decreasing). Time-to-metastasis-onset, time-to-diagnosis, and time-to-death were generated under Cox proportional hazards using parameters $\beta_\eta = 0.9$, $\beta_\theta = [0.5, 0.8]$, $\beta_\mu = [0.5, 0.6]$, and $\beta_\gamma = [0.1, 0.7]$.

Independent censoring time followed a uniform distribution $U(0, 20)$ for the increasing hazard and $U(0, 10)$ for the decreasing hazard scenario, which yielded approximately 15% censoring before the first event (diagnosis) and 30% censoring before the second event (death). The convergence criterion was set at 10^{-5} for a change in the baseline hazard estimation and improvement in the full likelihood.

The simulation results are summarized in Table A.1. The iterative reweighting algorithm works well for both scenarios of increasing and decreasing baseline hazards

Table A.1: Simulation result: Independent censoring

N	Statistics	Metastasis	Diagnosis				Death	
		$\beta_{\eta 1}$	$\beta_{\theta 0}$	$\beta_{\theta 1}$	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\gamma 0}$	$\beta_{\gamma 1}$
	True	0.9	0.5	0.8	0.5	0.6	0.1	0.7
Increasing baseline hazard $h_t = 0.02t$								
200	Bias	0.013	0.023	0.008	0.061	0.006	-0.045	0.020
	ESE	0.094	0.312	0.088	0.740	0.380	0.433	0.096
	ASE	0.093	0.301	0.087	0.742	0.332	0.437	0.095
	CP(%)	94.2	94.8	95.7	95.5	94.4	95.7	95.1
500	Bias	0.006	0.003	0.006	0.027	-0.004	-0.021	0.007
	ESE	0.059	0.190	0.058	0.471	0.223	0.276	0.058
	ASE	0.058	0.188	0.054	0.449	0.200	0.271	0.058
	CP(%)	94.1	94.6	92.7	94.8	93.5	95.4	95.7
Decreasing baseline hazard $h_t = 0.05(t/10)^{-1/2}$								
200	Bias	0.003	-0.021	0.010	0.004	0.032	-0.064	0.018
	ESE	0.094	0.295	0.083	0.741	0.371	0.454	0.094
	ASE	0.094	0.291	0.085	0.790	0.346	0.448	0.092
	CP(%)	94.7	94.6	95.8	96.1	94.0	94.7	94.8
500	Bias	0.004	0.002	0.005	-0.010	0.004	-0.010	0.005
	ESE	0.058	0.188	0.055	0.497	0.219	0.296	0.057
	ASE	0.059	0.183	0.053	0.473	0.206	0.281	0.057
	CP(%)	95.6	94.8	94.3	94.7	94.6	93.3	94.3

ESE: empirical standard errors based on 1,000 estimates
ASE: average of estimated standard errors
CP: 95% Coverage Probability

(not shown). The estimates for covariate effects on time-to-latent-metastasis, time-to-diagnosis, and time-to-death are almost unbiased at small sample size (n=200) and further reduce bias for larger sample size (n=500). The empirical standard errors (ESE) are close to the mean of estimated theoretical errors (ASE), which validate the performance of the variance estimators. Coefficient estimators improve with increasing sample size as the estimated variances decrease and 95% confidence intervals tighten. The 95% coverage probabilities for all coefficient estimators remain as expected for both sample sizes.

A.5.2 Full simulation results when censoring time and survival time are dependent

This is the result of sensitivity analyses for dependent censoring discussed in section 2.3.2. There are 16 covariates in this simulation, 5 of which are directly related to risk of death and are shown in Table 2.2 of the main paper.

Table A.2: Simulation with dependence censoring and time-varying covariates (n=500)

Dependence		Metastasis						Diagnosis						Death					
(censor %)	Statistics	$\beta_{\eta 1}$	$\beta_{\eta 2}$	$\beta_{\eta 3}$	$\beta_{\eta 4}$	$\beta_{\eta 5}$	$\beta_{\eta 6}$	$\beta_{\theta 0}$	$\beta_{\theta 1}$	$\beta_{\theta 2}$	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\mu 2}$	$\beta_{\gamma 1}$	$\beta_{\gamma 2}$	$\beta_{\gamma 3}$	$\beta_{\gamma 4}$	$\beta_{\gamma 5}$	$\beta_{\gamma 6}$
	True	0.2	0.5	0.4	-0.3	-1	-0.7	1	0.6	0.1	0.8	0.3	0.4	0.9	0.5	0.6	-0.6	-2	-1.4
Independent	Bias	-0.01	0.00	0.00	0.00	-0.02	-0.03	0.01	0	0.01	0.04	-0.01	-0.06	-0.02	0.04	0.04	-0.02	-0.06	-0.05
(30%)	ESE	0.23	0.23	0.31	0.41	0.53	0.4	0.21	0.15	0.18	0.54	0.72	0.72	0.43	0.27	0.27	0.37	0.4	0.36
	ASE	0.22	0.23	0.29	0.4	0.45	0.4	0.21	0.15	0.18	0.48	0.65	0.69	0.4	0.24	0.25	0.37	0.35	0.34
	CP(%)	93.9	94.7	93.5	95.9	92.3	95.5	95.1	94.7	95.5	93.1	95.5	95.7	91.9	92.7	93.7	93.3	90.5	92.9
$\rho = 5$	Bias	-0.01	0.01	0.03	-0.03	-0.08	-0.07	-0.01	0.01	0.00	0.06	-0.01	-0.04	-0.08	0.02	0.04	-0.01	-0.05	-0.02
(30%)	ESE	0.22	0.22	0.32	0.41	0.49	0.43	0.20	0.15	0.19	0.48	0.74	0.69	0.38	0.22	0.23	0.36	0.34	0.33
	ASE	0.20	0.22	0.27	0.36	0.41	0.36	0.20	0.15	0.18	0.46	0.64	0.67	0.35	0.21	0.22	0.33	0.31	0.30
	CP(%)	92.70	95.10	95.30	96.30	93.70	93.90	93.90	94.90	94.10	94.30	97.40	95.50	91.10	95.30	93.70	92.10	90.70	92.90
$\rho = 2$	Bias	-0.03	-0.01	0.02	-0.02	0.01	-0.01	-0.01	0.00	0.00	0.02	0.01	0.03	-0.12	0.06	0.07	0.01	-0.13	-0.07
(30%)	ESE	0.22	0.22	0.27	0.37	0.50	0.39	0.20	0.15	0.18	0.49	0.70	0.70	0.36	0.22	0.22	0.35	0.33	0.31
	ASE	0.20	0.22	0.26	0.36	0.42	0.36	0.20	0.15	0.18	0.46	0.64	0.68	0.35	0.20	0.21	0.33	0.31	0.30
	CP(%)	92.10	95.10	96.00	96.00	92.60	96.40	95.30	94.90	95.30	93.40	95.10	97.00	91.30	90.90	92.80	91.90	90.20	93.40
$\rho = 1$	Bias	-0.01	-0.01	0.01	-0.01	-0.05	-0.09	-0.01	0.00	0.01	0.01	0.05	0.04	-0.16	0.08	0.09	0.00	-0.18	-0.08
(30%)	ESE	0.20	0.23	0.29	0.44	0.51	0.43	0.20	0.15	0.18	0.44	0.66	0.75	0.38	0.21	0.24	0.36	0.34	0.33
	ASE	0.20	0.22	0.27	0.37	0.44	0.36	0.20	0.15	0.18	0.45	0.63	0.68	0.35	0.20	0.21	0.33	0.32	0.30
	CP(%)	95.20	94.00	95.40	94.60	90.70	94.00	94.40	95.60	95.60	93.80	95.20	96.70	90.50	92.30	90.00	92.10	89.60	90.20
$\rho = 0.5$	Bias	0.01	0.02	-0.02	-0.03	-0.18	-0.14	0.02	0.00	0.00	0.05	0.05	0.00	-0.24	0.15	0.15	0.00	-0.23	-0.09
(30%)	ESE	0.22	0.24	0.31	0.49	0.57	0.45	0.21	0.15	0.18	0.53	0.76	0.76	0.38	0.23	0.24	0.35	0.34	0.32
	ASE	0.21	0.22	0.27	0.37	0.44	0.36	0.20	0.15	0.18	0.47	0.67	0.69	0.36	0.21	0.22	0.33	0.32	0.30
	CP(%)	92.7	91.7	94	94.2	86.1	90.1	92.3	95.2	94.2	94.2	95	95	86.3	86.5	87.5	92.7	86.5	90.9

ESE: empirical standard errors based on 1,000 estimates
ASE: average of estimated standard errors
CP: 95% Coverage Probability

APPENDIX B

Supplementary Materials for Chapter III

B.1 Derivation of likelihood probabilities

For ease of notation, we suppress the subscript i for each individual subject. The likelihood for different scenarios of diagnosis, metastasis, and death are:

- Subject is diagnosed at t_1:

$$\begin{aligned}\mathcal{L}_1 &= P(T_1^* = t_1) = \mathbb{E}_A[f_{dx}(t_1 \mid A)] \\ &= \mathbb{E}_A\left[Ah_{dx}(t_1)\theta e^{-A\theta H_{dx}(t_1)}\right] = -h_{dx}(t_1)\theta \mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta}.\end{aligned}$$

- Subject is diagnosed at t_1 and dies at t_2:

$$\begin{aligned}\mathcal{L}_1^* &= P(T_1^* = t_1, T_2^* = t_2) = \mathbb{E}_A[f_{dx} f_d(t_1, t_2 \mid A)] \\ &= \mathbb{E}_A\left\{\left[Ah_{dx}(t_1)\theta e^{-A\theta H_{dx}(t_1)}\right]\left[Ah_d(t_2)\gamma e^{-A\gamma[H_{dx}(t_2)-H_{dx}(t_1)]}\right]\right\} \\ &= \mathbb{E}_A\left\{A^2 h_{dx}(t_1)h_d(t_2)\theta\gamma e^{-A[H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma]}\right\} \\ &= h_{dx}(t_1)h_d(t_2)\theta\gamma \mathfrak{L}^{(2)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}.\end{aligned}$$

- Subject is diagnosed at t_1 and is censored at t_2

$$\begin{aligned}\mathcal{L}_2^* &= P(T_1^* = t_1, T_2^* > t_2) = \mathbb{E}_A[f_{dx} S_d(t_1, t_2 \mid A)] \\ &= \mathbb{E}_A\left\{\left[A h_{dx}(t_1)\theta e^{-A\theta H_{dx}(t_1)}\right]\left[e^{-A\gamma[H_{dx}(t_2)-H_{dx}(t_1)]}\right]\right\} \\ &= \mathbb{E}_A\left\{A h_{dx}(t_1)\theta e^{-A[H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma]}\right\} \\ &= -h_{dx}(t_1)\theta \mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}.\end{aligned}$$

- Subject is censored at t_1 before any event is observed

$$\begin{aligned}\mathcal{L}_3 &= P(T_1^* > t_1, T_2^* > t_1) = \mathbb{E}_A[S_{dx} S_d(t_1, t_1 \mid A)] \\ &= \mathbb{E}_A\left[e^{-A\theta H_{dx}(t_1)}\right] = \mathfrak{L}^{(0)}_{H_{dx}(t_1)\theta}.\end{aligned}$$

B.2 Derivation of hazard terms

1. The hazard of being diagnosed with cancer at time t:

$$\begin{aligned}\Theta_1[H_{dx}(t), \beta] dH_{dx}(t) &= P[dN_1(t) = 1 \mid Y_1(t) = 1] = \frac{\mathcal{L}_1}{\mathcal{L}_3} \\ \Rightarrow \Theta_1[H_{dx}(t), \beta] &= \frac{\mathcal{L}_1}{\mathcal{L}_3 h_{dx}(t)} = \frac{-\theta \mathfrak{L}^{(1)}_{H_{dx}(t)\theta}}{\mathfrak{L}^{(0)}_{H_{dx}(t)\theta}}.\end{aligned}$$

2. The hazard of death at time t, given diagnosis at t_1:

$$\begin{aligned}\Theta_2[H_d(t), \beta, t_1] dH_d(t) &= P[dN_2(t) = 1 \mid Y_2(t) = 1] = \frac{\mathcal{L}_1^\star / \mathcal{L}_1}{\mathcal{L}_2^\star / \mathcal{L}_1} = \frac{\mathcal{L}_1^\star}{\mathcal{L}_2^\star} \\ \Rightarrow \Theta_2[H_d(t), \beta, t_1] &= \frac{\mathcal{L}_1^\star}{\mathcal{L}_2^\star h_d(t)} = \frac{-\gamma \mathfrak{L}^{(2)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}}{\mathfrak{L}^{(1)}_{H_{dx}(t_1)\theta + H_d(t_2)\gamma - H_d(t_1)\gamma}}.\end{aligned}$$

B.3 Martingale Properties of Nonparametric Maximum Likelihood Estimators

From the score functions eqns. (3.5), (3.6), and (3.7),

$$U_\beta = \sum_{z_1 \in Z_1} \int_0^\xi \frac{\Theta_{z_1,\beta}(x;\beta,H)}{\Theta_{z_1}(x;\beta,H)} dM_{z_1}(x) + \sum_{i=1}^n \int_0^\xi \frac{\Theta_{2i,\beta}(x;\beta,H,t_{1i})}{\Theta_{2i}(x;\beta,H,t_{1i})} dM_{2i}(x), \tag{B.1}$$

$$U_{H_{dx}(t)} = \sum_{z_1 \in Z_1} \int_0^t \left[\frac{dM_{z_1}(x)}{dH_{dx}(x)} + \int_{x^+}^\xi \psi_{z_1}(u) dM_{z_1}(u) \right], \tag{B.2}$$

$$U_{H_d(t)} = \sum_{i=1}^n \int_0^t \left[\frac{dM_{2i}(x)}{dH_d(x)} + \int_{x^+}^\xi \psi_{2i}(u) dM_{2i}(u) \right] \tag{B.3}$$

Exchange the integrals over x and u, eqn. (B.2) becomes

$$\begin{aligned} U_{H_{dx}(t)} &= \sum_{z_1 \in Z_1} \int_0^\xi \left[\mathbb{1}(u \le t) + H_{dx}(u \wedge t)\psi_{z_1}(u) \right] dM_{z_1}(u) \\ &= \sum_{z_1 \in Z_1} \int_0^\xi \varepsilon_{z_1}(u,t;\beta,H_{dx}) dM_{z_1}(u), \end{aligned}$$

where $\varepsilon_{z_1}(u,t;\boldsymbol{\beta},H_{dx}) = I(u \le t) + \int_0^{u\wedge t} \psi_{z_1}(u;\boldsymbol{\beta},H_{dx}) dH_{dx}(x)$.

As shown in Supplementary Materials E.3 of Hu and Tsodikov (2014a), the linear transform $\int_0^\xi \varepsilon(u,t;H,\boldsymbol{\beta})\, dM(u)$ is a martingale under the true model because

$$E\left[\int_0^\xi \varepsilon'(u,t) dM(u) | \mathcal{F}_{t-} \right] = dt \int_0^\xi \varepsilon'(u,t) E\{dM(u)|\mathcal{F}_{t-}\} = dt \int_0^t \varepsilon'(u,t) dM(u),$$

which equal 0 if $\varepsilon'(u,t) = \frac{\partial \varepsilon(u,t)}{\partial t} = 0$ for $u \le t$ (true by definition).

Therefore, the score function $U_{H_{dx}(t)}$ is martingale at the true model. The proofs for U_β and $U_{H_d(t)}$ follow along the same line.

B.4 Profile likelihood Hessian

Let $\hat{H}_{\boldsymbol{\beta}} = \left[\hat{H}_{dx\boldsymbol{\beta}},\ \hat{H}_{d\boldsymbol{\beta}} \right]$ be the solutions to $\mathcal{U}_{H_r} = 0$ and $\mathcal{U}_{H_d} = 0$ for fixed $\boldsymbol{\beta}$, we want to show that $I_{\text{pr}}^{-1} = \left(\frac{\partial^2 \ell_{\text{pr}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right)^{-1}$ is consistent for the covariance matrix of $\hat{\boldsymbol{\beta}}$, that

is, the $\boldsymbol{\beta\beta}$ submatrix of $\mathcal{I}_\infty^{-1}$, where

$$\mathcal{I}_\infty \equiv \begin{bmatrix} \mathcal{I}_{\boldsymbol{\beta\beta}} & \mathcal{I}_{\boldsymbol{\beta}H_{dx}} & \mathcal{I}_{\boldsymbol{\beta}H_d} \\ \mathcal{I}_{H_{dx}\boldsymbol{\beta}} & \mathcal{I}_{H_{dx}H_{dx}} & \mathcal{I}_{H_{dx}H_d} \\ \mathcal{I}_{H_d\boldsymbol{\beta}} & \mathcal{I}_{H_dH_{dx}} & \mathcal{I}_{H_dH_d} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & -\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH_r(s)} & -\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH_d(s)} \\ -\frac{\partial^2\ell_\infty}{\partial dH_r(t)\partial\boldsymbol{\beta}'} & -\frac{\partial^2\ell_\infty}{\partial dH_r(t)\partial dH_r(s)} & -\frac{\partial^2\ell_\infty}{\partial dH_r(t)\partial dH_d(s)} \\ -\frac{\partial^2\ell_\infty}{\partial dH_d(t)\partial\boldsymbol{\beta}'} & -\frac{\partial^2\ell_\infty}{\partial dH_d(t)\partial dH_r(s)} & -\frac{\partial^2\ell_\infty}{\partial dH_d(t)\partial dH_d(s)} \end{bmatrix}$$

is the asymptotic covariance matrix of the score for the full model. This may be expressed as $\text{Var}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\right] = Q^{-1}$, where

$$Q = \mathcal{I}_{\boldsymbol{\beta\beta}} - \mathcal{I}_{\boldsymbol{\beta}H}\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\boldsymbol{\beta}}. \tag{B.4}$$

where $\mathcal{I}_{\boldsymbol{\beta}H} = [\mathcal{I}_{\boldsymbol{\beta}H_{dx}}\ \mathcal{I}_{\boldsymbol{\beta}H_d}]$ and $\mathcal{I}_{HH}$ is the bottom right 2×2 matrix of $\mathcal{I}_\infty$.

Denote by $J_{H\boldsymbol{\beta}}$ the Jacobian $\left[\frac{\partial d\hat{H}_{dx\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\ \frac{\partial d\hat{H}_{d\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\right]$. The full derivative is

$$\frac{d\hat{\ell}}{d\boldsymbol{\beta}} = \int_s \frac{\partial\hat{\ell}}{\partial dH(s)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}} + \frac{\partial\hat{\ell}}{\partial\boldsymbol{\beta}}$$

so that the profile score is

$$\mathcal{U}_{\text{pr}} = \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}} + \mathcal{U}_{\boldsymbol{\beta}}|_{H_{dx}=\hat{H}_{dx\boldsymbol{\beta}},H_d=\hat{H}_{d\boldsymbol{\beta}}} = \mathcal{U}_{\boldsymbol{\beta}}|_{H_{dx}=\hat{H}_{dx\boldsymbol{\beta}},H_d=\hat{H}_{d\boldsymbol{\beta}}} \tag{B.5}$$

since $\mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}} = \left[\mathcal{U}_{\hat{H}_{dx\boldsymbol{\beta}}}, \mathcal{U}_{\hat{H}_{d\boldsymbol{\beta}}}\right] = 0$.

The profile Hessian is

$$\begin{aligned}
\frac{d^2\hat{\ell}}{d\boldsymbol{\beta}d\boldsymbol{\beta}'} &= \int_y\int_s \frac{\partial}{\partial dH(y)}\left[\frac{\partial\hat{\ell}}{\partial dH(s)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\right]\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} \\
&\quad + \int_y \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} + \int_s \frac{\partial}{\partial\boldsymbol{\beta}'}\left[\frac{\partial\hat{\ell}}{\partial dH(s)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\right] + \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \\
&= \int_y\int_s \left[\frac{\partial^2\hat{\ell}}{\partial dH(s)\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'} + \frac{\partial\hat{\ell}}{\partial dH(s)}\frac{\partial^2 d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial\boldsymbol{\beta}'}\right] \\
&\quad + \int_s \left[\frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial dH(s)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}'} + \frac{\partial^2\hat{\ell}}{\partial dH(s)\partial\boldsymbol{\beta}'}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}} + \frac{\partial\hat{\ell}}{\partial dH(s)}\frac{\partial^2 d\hat{H}_{\boldsymbol{\beta}}(s)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] + \frac{\partial^2\hat{\ell}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \\
&= -J_{\boldsymbol{\beta}H}\hat{I}_{HH}J_{H\boldsymbol{\beta}} + \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} - J_{\boldsymbol{\beta}H}\hat{I}_{H\boldsymbol{\beta}} + \mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}}J_{H\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta\beta}}.
\end{aligned}$$

Since $\mathcal{U}_{\hat{H}_{\boldsymbol{\beta}}} = 0$, we have

$$I_{\text{pr}} = J_{\boldsymbol{\beta}H}\hat{I}_{HH}J_{H\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} + J_{\boldsymbol{\beta}H}\hat{I}_{H\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} \tag{B.6}$$

To express I_{pr} without Jacobians, we note that

$$\begin{aligned} 0 &= \frac{d}{d\boldsymbol{\beta}}\left(\left.\frac{\partial \ell}{\partial dH(s)}\right|_{H=\hat{H}_{\boldsymbol{\beta}}}\right) \\ &= \frac{\partial^2 \hat{\ell}}{\partial dH(s)\partial \boldsymbol{\beta}} + \int \frac{\partial^2 \hat{\ell}}{\partial dH(s)\partial dH(y)}\frac{\partial d\hat{H}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}} \\ &= -\hat{I}_{H\boldsymbol{\beta}} - \hat{I}_{HH}J_{H\boldsymbol{\beta}}, \end{aligned}$$

implying that

$$J_{H\boldsymbol{\beta}} = -\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}. \tag{B.7}$$

Substitution of (B.7) into (B.6) yields

$$\begin{aligned} I_{\text{pr}} &= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \left[-\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\right]\hat{I}_{HH}\left[-\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}\right] + \hat{I}_{\boldsymbol{\beta}H}\left[-\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}\right] + \left[-\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\right]\hat{I}_{H\boldsymbol{\beta}} \\ &= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} - 2\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} \\ &= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}}. \end{aligned} \tag{B.8}$$

That is, we see by comparing (B.8) with (B.4) that $\mathcal{I}_\infty^{-1}$ has $\boldsymbol{\beta}$ submatrix $Q^{-1} = I_{\text{pr}}^{-1}$.

Since

$$\sqrt{n}\left(\hat{\Omega} - \Omega^0\right) = \mathcal{I}_\infty^{-1}\sqrt{n}\mathcal{U}^0 + o_p(1),$$

we have

$$\begin{aligned} Var\left[\sqrt{n}\left(\hat{\Omega} - \Omega^0\right)\right] &= \mathcal{I}_\infty^{-1} Cov\left(\sqrt{n}\mathcal{U}^0\right)\mathcal{I}_\infty^{-1} \\ &= \mathcal{I}_\infty^{-1}\left[\mathcal{I}_\infty + o_p(1)\right]\mathcal{I}_\infty^{-1} \\ &= \mathcal{I}_\infty^{-1}. \end{aligned}$$

Therefore,

$$Var\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\right] = Q^{-1} = I_{\text{pr}}^{-1}.$$

APPENDIX C

Supplementary Materials for Chapter IV

C.1 Derivation of likelihood probabilities

For ease of notation, we suppress the subscript i for each individual subject. We assume Design II, but modifications for other designs are trivial (by including/excluding treatment covariates in γ and μ). The likelihood for different scenarios of response, non-response, and death are:

(1) Subject responds at t_1, is re-randomized in stage 2, and dies at t_2:

$$\begin{aligned} L_1(t_1, t_2) &= f_r f_d(t_1, t_2) = \left[h_r(t_1)\theta e^{-H_r(t_1)\theta}\right] \cdot \left[h_d(t_2)\eta e^{-H_d(t_1)(\gamma-\eta)-H_d(t_2)\eta}\right] \\ &= h_r(t_1)h_d(t_2)\theta\eta e^{-\theta H_r(t_1)-(\gamma-\eta)H_d(t_1)-\eta H_d(t_2)} \end{aligned}$$

(2) Subject responds at t_1, is re-randomized in stage 2, and is censored at t_2:

$$\begin{aligned} L_2(t_1, t_2) &= f_r S_d(t_1, t_2) = \left[h_r(t_1)\theta e^{-H_r(t_1)\theta}\right] \cdot \left[e^{-H_d(t_1)(\gamma-\eta)-H_d(t_2)\eta}\right] \\ &= h_r(t_1)\theta e^{-\theta H_r(t_1)-(\gamma-\eta)H_d(t_1)-\eta H_d(t_2)}. \end{aligned}$$

(3) Subject does not respond by t^{nr}, receives stage 2 treatment as a non-responder, then dies at t_2:

$$\begin{aligned} L_3(t^{nr}, t_2) &= S_r f_d(t^{nr}, t_2) = \left[e^{-H_r(t^{nr})\theta}\right] \cdot \left[h_d(t_2)\mu e^{-H_d(t^{nr})(\gamma-\mu)-H_d(t_2)\mu}\right] \\ &= h_d(t_2)\mu e^{-\theta H_r(t^{nr})-(\gamma-\mu)H_d(t^{nr})-\mu H_d(t_2)}. \end{aligned}$$

(4) Subject does not respond by t^{nr}, receives stage 2 treatment as a non-responder, and is censored at t_2:

$$L_4(t^{nr}, t_2) = S_r S_d(t^{nr}, t_2) = \left[e^{-H_r(t^{nr})\theta}\right] \cdot \left[e^{-H_d(t^{nr})(\gamma-\mu)-H_d(t_2)\mu}\right]$$
$$= e^{-\theta H_r(t^{nr})-(\gamma-\mu)H_d(t^{nr})-\mu H_d(t^{nr})}.$$

(5) Subject dies at t_1 before the second randomization:

$$L_5(t_1) = S_r f_d(t_1, t_1) = \left[e^{-H_r(t_1)\theta}\right] \cdot \left[h_d(t_1)\gamma e^{-H_d(t_1)\gamma}\right]$$
$$= h_d(t_1)\gamma e^{-\theta H_r(t_1)-\gamma H_d(t_1)}$$

(6) Subject is censored at t_1 before second randomization:

$$L_6(t_1) = S_r S_d(t_1, t_1) = e^{-\theta H_r(t_1)-\gamma H_d(t_1)}$$

C.2 Martingale Properties of Nonparametric Maximum Likelihood Estimators

The estimating equations (4.4) and (4.5) are derived from the following score functions:

$$\begin{aligned} U_{H_r(t)} =& \sum_{i=1}^{n} \int_0^t [dN_{1i}^r(x) - Y_1(x)\Theta_{1i}^r(x) dH_r(x) \\ &+ dH_r(x) \int_{x+}^{\xi} \psi_{1i}^r(u) dM_{1i}^r(u) + \psi_{1i}^d(u) dM_{1i}^d(u)\Big] \\ =& \sum_{i=1}^{n} \int_0^t dM_{1i}^r(x) + \int_0^t \left[\int_{x+}^{\xi} \psi_{1i}^r(u) dM_{1i}^r(u) + \psi_{1i}^d(u) dM_{1i}^d(u)\right] dH_r(x), \end{aligned} \quad \text{(C.1)}$$

where $\psi_{1i}^r(u) = \frac{\partial \log \Theta_{1i}^r(u)}{\partial dH_r(u)}$ and $\psi_{1i}^d(u) = \frac{\partial \log \Theta_{1i}^d(u)}{\partial dH_r(u)}$.

Exchanging the integrals over x and u, (C.1) becomes

$$\begin{aligned}
U_{H_r(t)} &= \sum_{i=1}^{n} \int_0^t dM_{1i}^r(u) + \int_0^t \int_0^u \psi_{1i}^r(u) dH_r(x) dM_{1i}^r(u) + \psi_{1i}^d(u) dH_r(x) dM_{1i}^d(u) \\
&\quad + \int_{t+}^{\xi} \int_0^t \psi_{1i}^r(u) dH_r(x) dM_{1i}^r(u) + \psi_{1i}^d(u) dH_r(x) dM_{1i}^d(u) \\
&= \sum_{i=1}^{n} \int_0^t \left[1 + \int_0^u \psi_{1i}^r(u) dH_r(x) + \psi_{1i}^d(u) dH_r(x) \frac{dM_{1i}^d(u)}{dM_{1i}^r(u)} \right] dM_{1i}^r(u) \\
&\quad + \int_{t+}^{\xi} \int_0^t \left[\psi_{1i}^r(u) dH_r(x) + \psi_{1i}^d(u) dH_r(x) \frac{dM_{1i}^d(u)}{dM_{1i}^r(u)} \right] dM_{1i}^r(u) \\
&= \sum_{i=1}^{n} \int_0^{\xi} \left[\mathbb{1}(u \le t) + \psi_{1i}^r(u) dH_r(u \wedge t) + \psi_{1i}^d(u) dH_r(u \wedge t) \frac{dM_{1i}^d(u)}{dM_{1i}^r(u)} \right] dM_{1i}^r(u) \\
&= \sum_{i=1}^{n} \int_0^{\xi} \varepsilon_{z_1}(u, t; \beta, H_r) dM_{1i}^r(u),
\end{aligned}$$

where $\varepsilon_{1i}^r(u, t; \boldsymbol{\beta}, H_r) = \mathbb{1}(u \le t) + \int_0^{u \wedge t} \left[\psi_{1i}^r(u) + \psi_{1i}^d(u) \frac{dM_{1i}^d(u)}{dM_{1i}^r(u)} \right] dH_r(x)$.

As shown in Hu and Tsodikov (2014a) (Supplementary Materials E.3), the linear transform $\int_0^{\xi} \varepsilon(u, t; H, \boldsymbol{\beta})\, dM(u)$ is a martingale under the true model because

$$E\left[\int_0^{\xi} \varepsilon'(u,t) dM(u) | \mathcal{F}_{t-} \right] = dt \int_0^{\xi} \varepsilon'(u,t) E\{dM(u) | \mathcal{F}_{t-}\} = dt \int_0^{t} \varepsilon'(u,t) dM(u),$$

which equal 0 if $\varepsilon'(u,t) = \frac{\partial \varepsilon(u,t)}{\partial t} = 0$ for $u \le t$ (true by definition).

Therefore, the score function $U_{H_r(t)}$ is martingale at the true model. The proofs for U_β and $U_{H_d(t)}$ follow along the same line.

C.3 Simulations: No interaction effect between V and regimens

The model in Tang and Wahed (2015) can adjust for baseline covariates. However, these covariates' effects are constant across regimens. We simulate SMART survival data under this assumption to compare this model's predictive performance and validate the survival predictions from our proposed model.

The simulation parameters are identical to those used in Section 4.3.1, but with interaction term $\beta_9 = 0$ (i.e. the effect of a baseline covariate V is the same across

different DTRs). Regimen $A_2B_1C_1$ is the best for both V subgroups.

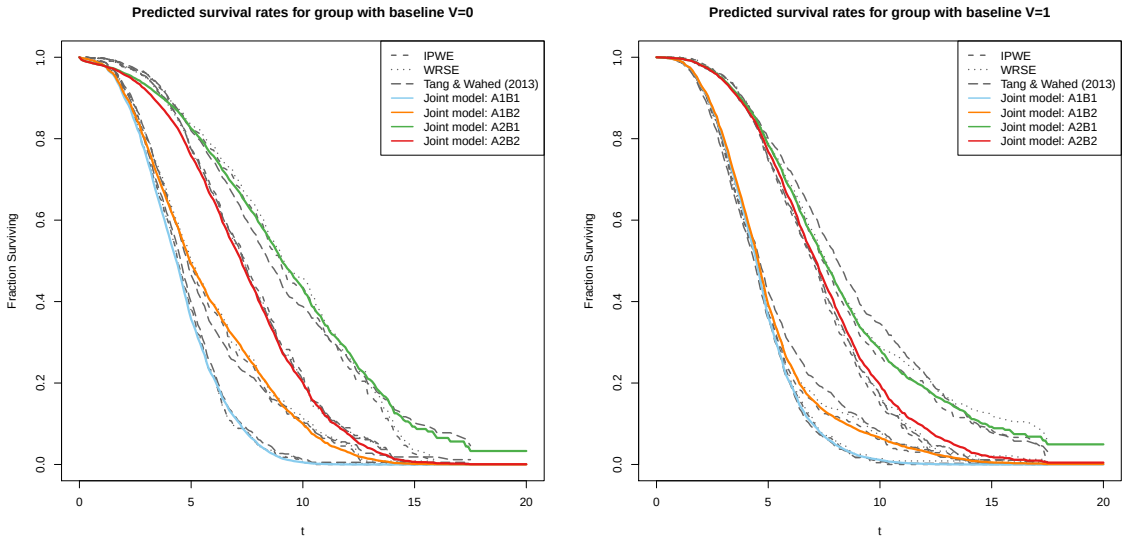


Figure C.1: Survival estimates in a scenario with no interaction between baseline covariate V and regimens.

Figure C.1 plots the survival estimates for each DTR. The left plot is for $V = 0$ and the right plot is for $V = 1$. Our model's estimates are solid lines. Tang and Wahed (2015) estimates are long dashed lines. Subset analysis (by V) using IPWE and WRSE from Lunceford et al. (2002) and Guo and Tsiatis (2005) are plotted in short dashed lines and dotted lines, respectively. We see that our model's prediction is consistent with those from previously proposed methods in scenarios where there is no interaction between baseline covariates and regimens.

BIBLIOGRAPHY

BIBLIOGRAPHY

American Cancer Society. Cancer facts & figures 2016. *Atlanta: American Cancer Society*, 2016.

P. K. Andersen and N. Keiding. Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11(2):91–115, 2002.

P. K. Andersen, L. S. Hansen, and N. Keiding. Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process. *Scandinavian Journal of Statistics*, pages 153–167, 1991.

P. K. Andersen, J. P. Klein, K. M. Knudsen, and R. T. y Palacios. Estimation of variance in cox's regression model with shared gamma frailties. *Biometrics*, pages 1475–1484, 1997.

B. Biki, E. Mascha, D. C. Moriarty, J. M. Fitzpatrick, D. I. Sessler, and D. J. Buggy. Anesthetic technique for radical prostatectomy surgery affects cancer recurrence: A retrospective analysis. *The Journal of the American Society of Anesthesiologists*, 109(2):180–187, 2008.

Y.-H. Chen. Weighted breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika*, 96(3):591–600, 2009.

Y.-H. Chen. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis*, 18(1):36–57, 2012.

D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.

A. Ertefaie, T. Wu, K. G. Lynch, and I. Nahum-Shani. Identifying a set that contains the best dynamic treatment regimes. *Biostatistics*, 17(1):135, 2016.

W. Feng and A. S. Wahed. Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies. *Biometrika*, 95(3):695–707, 2008.

W. Feng and A. S. Wahed. Sample size for two-stage studies with maintenance therapy. *Statistics in Medicine*, 28(15):2028–2041, 2009.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

J. P. Fine, H. Jiang, and R. Chappell. On semi-competing risks data. *Biometrika*, 88(4):907–919, 2001.

U. S. Govindarajulu, H. Lin, K. L. Lunetta, and R. D'Agostino. Frailty models: applications to biomedical and genetic studies. *Statistics in Medicine*, 30(22):2754–2764, 2011.

X. Guo and A. Tsiatis. A weighted risk set estimator for survival distributions in two-stage randomization designs with censored survival data. *The International Journal of Biostatistics*, 2005.

P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

P. Hougaard and P. Hougaard. *Analysis of multivariate survival data*, volume 564. Springer New York, 2000.

J. Hsu. *Multiple comparisons: theory and methods*. CRC Press, 1996.

C. Hu and A. Tsodikov. Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime Data Analysis*, 20(4):563–583, 2014a.

C. Hu and A. Tsodikov. Semiparametric regression analysis for time-to-event marked endpoints in cancer studies. *Biostatistics*, 15(3):513–525, 2014b.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

K. M. Kidwell and A. S. Wahed. Weighted log-rank statistic to compare shared-path adaptive treatment strategies. *Biostatistics*, 14(2):299–312, 2013.

J. P. Klein. Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, pages 795–806, 1992.

J. H. Ko and A. S. Wahed. Up-front versus sequential randomizations for inference on adaptive treatment strategies. *Statistics in Medicine*, 31(9):812–830, 2012.

P. W. Lavori and R. Dawson. Dynamic treatment regimes: practical design considerations. *Clinical Trials*, 1(1):9–20, 2004.

P. W. Lavori, R. Dawson, and A. J. Rush. Flexible treatment strategies in chronic disease: clinical and research implications. *Biological psychiatry*, 48(6):605–614, 2000.

Z. Li and S. A. Murphy. Sample size formulae for two-stage randomized trials with survival outcomes. *Biometrika*, 98(3):503–518, 2011.

L. Liu, R. A. Wolfe, and X. Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

Y. Lokhnygina and J. D. Helterbrand. Cox regression methods for two-stage randomization designs. *Biometrics*, 63(2):422–428, 2007.

J. K. Lunceford, M. Davidian, and A. A. Tsiatis. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 58(1):48–57, 2002.

P. Mehlen and A. Puisieux. Metastasis: a question of life or death. *Nature Reviews*, 6:449–458, 2006.

L. Meira-Machado, J. de Uña-Álvarez, and C. Cadarso-Suárez. Nonparametric estimation of transition probabilities in a non-markov illness–death model. *Lifetime Data Analysis*, 12(3):325–344, 2006.

S. Miyahara and A. S. Wahed. Weighted kaplan–meier estimators for two-stage treatment regimes. *Statistics in Medicine*, 29(25):2581–2591, 2010.

S. A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, 2005.

S. A. Murphy, M. J. van der Laan, and J. M. Robins. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

E. Neeman and S. Ben-Eliyahu. Surgery and stress promote cancer metastasis: new outlooks on perioperative mediating mechanisms and immune involvement. *Brain, Behavior, and Immunity*, 30:S32–S40, 2013.

R. B. Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 1999.

G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, pages 25–43, 1992.

D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493, 1989.

A. Oetting, J. Levy, R. Weiss, and S. Murphy. Statistical methodology for a smart design in the development of adaptive treatment strategies. *Causality and psychopathology: Finding the determinants of disorders and their cures*, pages 179–205, 2011.

M. S. O'Reilly, L. Holmgren, Y. Shing, C. Chen, R. A. Rosenthal, M. Moses, W. S. Lane, Y. Cao, E. H. Sage, and J. Folkman. Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a lewis lung carcinoma. *Cell*, 79(2):315–328, 1994.

L. Orellana, A. Rotnitzky, and J. M. Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *The International Journal of Biostatistics*, 6(2), 2010.

L. Peng and J. P. Fine. Regression modeling of semicompeting risks data. *Biometrics*, 63(1):96–108, 2007.

J. D. Rice and A. Tsodikov. Semiparametric time-to-event modeling in the presence of a latent

progression event. *Biometrics*, 2016.

J. Smolle, H. P. Soyer, F. Smolle-Jüttner, E. Rieger, and H. Kerl. Does surgical removal of primary melanoma trigger growth of occult metastases? an analytical epidemiological approach. *Dermatologic surgery*, 23(11):1043–1046, 1997.

X. Tang and A. S. Wahed. Cumulative hazard ratio estimation for treatment regimes in sequentially randomized clinical trials. *Statistics in Biosciences*, 7(1):1–18, 2015.

P. F. Thall, L. H. Wooten, C. J. Logothetis, R. E. Millikan, and N. M. Tannir. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in Medicine*, 26(26):4687–4702, 2007.

A. Tsodikov. Semi-parametric models of long-and short-term survival: an application to the analysis of breast cancer survival in utah by age and stage. *Statistics in Medicine*, 21(6):895–920, 2002.

A. Tsodikov. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):759–774, 2003.

A. S. Wahed. Inference for two-stage adaptive treatment strategies using mixture distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1):1–18, 2010.

J. Xu, J. D. Kalbfleisch, and B. Tai. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725, 2010.

D. Zeng and D. Lin. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640, 2006.

D. Zeng, Q. Chen, and J. G. Ibrahim. Gamma frailty transformation models for multivariate survival times. *Biometrika*, 96(2):277–291, 2009.