UNIVERSITY OF CALIFORNIA SAN DIEGO

Causal Inference for Competing Risks and Semi-competing Risks Data

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biostatistics

by

Yiran Zhang

Committee in charge:

Professor Ronghui Xu, Chair
Professor Ery Arias-Castro
Professor Parambir Dulai
Professor Steven D. Edland
Professor Wesley Thompson

2022

The Dissertation of Yiran Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To Mom, Dad, my two Grandfathers and two Grandmothers.

EPIGRAPH

The future depends on what you do today

Mahatma Gandhi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Ronghui (Lily) Xu. It's my great honor to work with a Professor with such explicit and comprehensive understanding for statistical methodology, yet so focused and perceptive. You showed me the world to combine the beautiful mathematical statistics and impactful biomedical research. As an advisor, you taught me step by step when facing a scientific research project, from first principle. You also taught me to explore multiple areas in novel statistics and to be curious by reading papers. You managed to make me to explore beyond my comfort zone and think critically, while still being encouraging and considerate. I have no doubt that I learn and grow from your instruction, education and intelligence in the past few years, and you inspired me to become a good women researcher. I feel very fortunate to have the opportunity to be one of your student.

I would like to thank Wesley Thompson for being the best scientific research advisor and collaborator. You taught me how to overcome the difficulties when it comes to applying the novel statistical model to the complicated data structure, and also how to be cautious yet flexible as a scientific researcher. You showed me how much you dedicate to your work, and how passionate you are for biomedical research. I want to show my appreciation for your patience and hard-working for our collaborate works, and hopefully our work can show the importance for developing flexible statistical modeling.

I would also like to thank the other members of my PhD committee: Dulai Parambir for being a great principle investigator, generously helping me understand the background of medical research and encouraging me when I performed the statistical analysis, it's an amazing experience collaborate with you; Steve Edland for being a great advisor when it comes to Alzheimer's research, and providing insightful suggestion when I use the HAAS data; and Ery Arias-Castro for being a fantastic educator for my fundamental statistical theory learning and always try to make the explanation easier when I have questions.

I also want to show my thanks to many others from the University of California, San Diego that helped me over the past few years. I want to thank Florin Vaida for the extraordinary

mentorship on EM related theories. You gave me great advise on how to proceed with my problem and help me think broadly for extending the existing theory. I would like to thank Karen Messer as my faculty advisor and give me some valuable suggestion for both my research and career path. Thanks Loki Natarajan for being an amazing department advisor, you helped me to understand the complicated administrative process of getting a PhD, and encourage me every time when I achieved something during this process. I also want to thank CTRI for giving an opportunity working as a statistician, and provide me opportunities to work with some great principle investigator and researchers.

Finally, I would like to thank all the people that I love and love me in my life. My mom, you always stand by my side even if we are thousands of miles apart. You always cheer me up and listen to me when I need you. You are my biggest fan and always encourage me to pursue what I love. And you are a such great role model to me as an independent and successful women. My dad, who also unconditionally support every decision I made and be proud of me every moment. I'm truly grateful for everything you give in my life, especially your love and care. My beloved grandfathers and grandmothers, I want to thank you for teaching me be a person with kind heart, and be grateful to everyone in my life. You always care about my mental and physical health, my diet and my safety, and think about me every seconds when I'm studying abroad. I want to thank my husband for your enormous support and help. You always make me laugh when I'm down and have faith in everything I do. Thank you for leading me to the right direction when it comes to life as well as the research, life is not always easy, but it's great to have you by my side. Last but not the least, I want to thank my two important friends Guandong (Elliot) Yang and Tianyu Gao (TG), who brough me endless love, support, and laugh. Thanks Elliot for traveling across the country to attend my wedding and always be there for me when I need someone to talk. Thanks TG for endlessly sharing funny stories and be on the phone for me to cheer me up. I could not ask for more to have two such best friends.

Chapter 3, in full, has been submitted for publication as "Zhang, Yiran; Ronghui Xu. *cmprskcoxmsm: Use IPW to Estimate Treatment Effect under Competing Risks*, submitted to

CRAN Package". The dissertation author was the primary author on this paper.

Chapter 4, in full, has been submitted for publication as "Zhang, Yiran; Ronghui Xu. *Marginal Structural Illness-Death Models for Semi-Competing Risks Data*, submitted to Statistics in Medicine". The dissertation author was the primary author on this paper.

Data collection and analysis for all projects were obtained from the Honolulu Heart Program (HHP) and Honolulu-Asia Aging Study (HAAS), funded by the National Institutes of Health grant NIH R03 AG062432.

VITA

2016	Bachelor of Science in Applied Mathematics. South China University of Technology, Guangzhou, Guangdong, China
2018	Master of Science in Public Health in Biostatistics. Emory University, Atlanta, GA, USA
2022	Doctor of Philosophy in Biostatistics. University of California San Diego, San Diego, CA, USA

PUBLICATIONS

- **Zhang, Y.** and Xu, R. *Marginal Structural Illness-Death Models for Semi-Competing Risks Data*. 2022, arXiv:2204.10426.
- **Zhang, Y.** and Xu, R. *semicmprskcoxmsm: Use inverse probability weighting to estimate treatment effect for semi competing risks data*. 2022, R package version 0.1.0.
- Lukin, D., Faleck, D., Xu, R., **Zhang, Y.**, Weiss, A., Aniwan, S., ... and Dulai, P. S. *Comparative safety and effectiveness of vedolizumab to tumor necrosis factor antagonist therapy for ulcerative colitis*. Clinical Gastroenterology and Hepatology, 20(1), 126-135, 2022.
- **Zhang, Y.** and R. Xu. *cmprskcoxmsm: Use IPW to Estimate Treatment Effect under Competing Risks*. 2021, R package version 0.2.1.
- Infante, M. A., **Zhang, Y.**, Brumback, T., Brown, S. A., Colrain, I. M., Baker, F. C., ... and Thompson, W. K. *Adolescent Binge Drinking Is Associated With Accelerated Decline of Gray Matter Volume*. Cerebral Cortex, bhab368, 2021.
- Cederquist, L., LaBuzetta, J. N., Cachay, E., Friedman, L., Yi, C., Dibsie, L., and **Zhang, Y.** *Identifying disincentives to ethics consultation requests among physicians, advance practice providers, and nurses: a quality improvement all staff survey at a tertiary academic medical center*. BMC Medical Ethics, 22(1), 1-8, 2021.
- Bohm, M., Xu, R., **Zhang, Y**, Varma, S., Fischer, M., Kochhar, G., ... and Chablaney, S. *Comparative safety and effectiveness of vedolizumab to tumour necrosis factor antagonist therapy for Crohn's disease*. Alimentary pharmacology & therapeutics, 52(4), 669-681, 2020.
- **Zhang, Y.**, Hedo, R., Rivera, A., Rull, R., Richardson, S., and Tu, X. M. *Post hoc power analysis: is it an informative and meaningful analysis?*. General psychiatry, 32(4), 2019.

- Li, X., **Zhang, Y.**, Meisel, J., Jiang, R., Behera, M., and Peng, L. *Validation of the newly proposed American Joint Committee on Cancer (AJCC) breast cancer prognostic staging group and proposing a new staging system using the National Cancer Database. Breast Cancer Research and Treatment*. 171(2), 303-313, 2018.

- Colasanti, J., Sumitani, J., Mehta, C. C., **Zhang, Y.**, Nguyen, M. L., Del Rio, C., and Armstrong, W. S. *Implementation of a rapid entry program decreases time to viral suppression among vulnerable persons living with HIV in the Southern United States*. In Open forum infectious diseases (Vol. 5, No. 6, p. ofy104). US: Oxford University Press 2018.

ABSTRACT OF THE DISSERTATION

Causal Inference for Competing Risks and Semi-competing Risks Data

by

Yiran Zhang

Doctor of Philosophy in Biostatistics

University of California San Diego, 2022

Professor Ronghui Xu, Chair

In this dissertation, we utilize the novel statistical methods for obtaining causal effect under competing risks and semi-competing risks data in survival analysis. This dissertation is comprised of three main settings. In the first setting, we aim to assess the causal effect of mid-life alcohol exposure to the late life cognitive score which is related to Alzheimer's disease (AD) using a large scale longitudinal data. We applied the marginal structural model (MSM) with inverse probability weighted (IPW) to adjust for time-varying confounding. We found that there is a significant decline in cognitive scores among heavy drinkers compared always light drinker. However, since the cognitive scores also changes over time, learning the relationship of alcohol exposure and time to cognitive impairment is also worth to explore.

In the second setting, we are interested in mid-life alcohol exposure to late life time to cognitive impairment which is also related to AD. Under this setting, as people are in their late-life stage, death prevents us from observing cognitive impairment. In survival analysis, death is considering as competing event. To estimate the causal effect of point treatment to time to event with the existence of competing event, we applied the MSM Cox proportional hazards model with IPW. Since hazard ratio is hard to interpret in medical research, we proposed predicted risk contrasts formula under the MSM Cox model.

Observing the trend that people die quickly after experiencing cognitive impairment, in the third settings, we proposed a MSM illness-death to assess the causal effect for alcohol exposure to time to cognitive impairment, death and death after cognitive impairment. We considered two specific such models, the usual Markov illness-death structural model and the general Markov illness-death structural model which incorporates a frailty term. For interpretation purposes, risk contrasts under the structural models are defined. To accommodate the possibility of misspecification of propensity score model, we also derived the augmented IPW estimator under MSM illness-death usual Markov model.

Chapter 1

Introduction and Data Resource

1.1 Introduction

More than 25 million people worldwide are affected by dementia, most suffering from Alzheimer's disease (AD) or Alzheimer's disease related disease (ADRD). According to NIH, the number of affected people is expected to increase three-fold by 2050, from the currently approximate 5 million (10% of persons aged 65 and older). The risk factors, including many exposures or lifestyle, other than age or AD-related genotype, are not well understood. Without further understanding, effective prevention strategies remain a challenge.

Compared with abstinence, moderate alcohol intake is often considered to be associated with health benefits ([30]; [21]). Such benefits have also been reported to extend to cognitive health in late life in several studies ([44]; [36]), while other studies found the opposite ([27]). As It is now generally accepted that the pathogeneses of AD probably begin decades earlier and involve a progressive accumulation of tissue injury and cellular loss with disruptions of physiologic or metabolic systems accruing over decades. Therefore, it is crucial to have available reliable longitudinal measures for alcohol consumption and many other relevant variables, as well as to carefully examine the myriad associated factors that may confound or otherwise interfere with a confident assessment of the independent contributions of alcohol. Many events and factors from early in life including genetic and environmental exposures may contribute to dementia later in life. For this paper, We will focus on mid-life, represent important candidate determinants

of disease processes leading to late-life Alzheimer's or an allied dementing condition.

This dissertation contains both applying contemporary causal inference methods and developing the novel statistical methods for investigating highly complex relationships between mid-life exposures and late-life cognitive outcomes related to Alzheimer's disease. Our projects focus on using the rich data resources of the Honolulu Heart Program (HHP) and the Honolulu-Asia Aging Study (HAAS), linked epidemiologic projects that collected longitudinal alcohol, physical and cognitive measures from mid-to late life. Our projects also reflect an urgent need for more effective analytic methods to disentangle causal factors from allied confounders that have likely contributed to previous inconsistencies using traditional statistical approaches.

Our first project focus on adapting the novel causal inference methods in survival analysis to analyze the HAAS data and estimate the causal effect of mid-life alcohol exposure to the late life cognitive impairment with existence of death as competing event. Estimating the causal effect of a treatment or exposure is not straightforward for an observational study. In the observational study, the assumption of no confounders for the exposure or treatment of interest is violated. We considered the marginal structural model (MSM) with inverse probability weighting (IPW) ([46]). For survival outcomes, [16] proposed the marginal structural Cox proportional hazards model to estimate the averaged treatment effect. In failure-time settings, a competing event is any event that makes it impossible for the event of interest to occur. In our case, some participant dead before the cognitive impairment occur. The main contribution of the first project is, we expanded estimating the averaged treatment effect (ATE) that was defined as hazard ratio by modeling the marginal structural cause-specific hazards models to estimate the cumulative incidence functions (i.e. risks) as ATE, and we also provided inference on the risk difference or risk ratio at any given time. We adapted our model to the HHP and HAAS study data, details were shown in chapter one.

The second project is the extension from the first project. In competing risk setting, not only we considered time to moderate impairment or time to death as event of interest, but also were interested in time to death following moderate impairment. We referred this situation

as semi-competing risk where a subject can experience both non-terminal event (moderate impairment) and terminal event (death). To modeled the semi-competing risk data, [67] proposed an illness–death compartment model with a shared gamma frailty. In the illness–death model, a subject can either transit directly to the terminal event or first to the nonterminal event and then to the terminal event. For the frailty model, [59] proposed a general proportional hazards model with normal frailty. We extended the above two models in the causal inference setting, we developed a marginal structural illness-death proportional hazard mixed model (PHMM) with IPW, and used nonparametric maximum likelihood estimation (NPMLE) for estimating the ATE, variance components was estimated by bayesian bootstrap.

In the longitudinal data analysis, we usually have two strategies analyzing the clustered data. The first is using the conditional model, i.e.: mixed effect model. The other is using the marginal model, i.e.: generalized estimating equation (GEE). Our third project is inspired from the nature of analyzing longitudinal data, and extends it to causal survival analysis. Also, variance estimation using bootstrap for the second project was time-consuming. We proposed to fit the illness–death model by modeling marginal distributions ([62]) where no specific structure of dependence among the distinct failure times on each subject is imposed, which is called usual Markov model in the illness-death model setting. Each marginal distribution of the failure times is formulated by a MSM Cox proportional hazards model with IPW, and we estimated the regression parameters in the Cox models by maximizing the failure-specific partial likelihoods. Under this setting, variance can be estimated by using robust sandwich estimator.

Since the IPW estimator is biased if the propensity score model is misspecified, an augmented IPW (AIPW) estimator with doubly robust properties can protect against such model misspecification. It would also allow us to apply machine learning or nonparametric methods to the propensity score model. [43] and [55] have already developed the AIPW estimator for the marginal structural Cox model, and it is nature to extend their work for the illness-death model setting.

1.2 The Longitudinal Study of Honolulu Heart Program and Honolulu-Asia Aging Study

Data Resource The Honolulu Heart Program (HHP) was established in 1965 with an NHLBI, NIH research contract as a longitudinal epidemiologic study of rates and risk factors for heart disease and stroke in men of Japanese ancestry living on Oahu and born 1900 through 1919. Of the approximately 12,000 Japanese-American men then living on Oahu, 8,006 participated in the initial examination and interview (1965, then aged 45-65 years). Subsequent HHP examinations/interviews occurred in 1968- 71 (exam 2, n=7,498), and 1971-74 (exam 3, n=6,860). A follow-up HHP questionnaire/telephone interview (mailout, 1986-89, n=4,655) was focused on diet, use of supplemental vitamins, and general health. Survival, vital status, and cardiovascular illnesses were continuously monitored by community surveillance and repeated contracts with participants and family members.

The Honolulu-Asia Aging Study (HAAS) was established in 1991 (HHP exam 4, n=3,734) as a continuation of the HHP with a shift in focus to brain aging, AD, vascular dementia, other causes of cognitive and motor impairment, stroke, and the common chronic conditions of late-life. Eight further HAAS exams were done at 2-3 year intervals until 2012, except for exam 8 and 9 (See Appendix A Figure A.1 for the date distribution at each exam), when only about 500 of the original 8,006 were still alive. Neuropsychologic screening was done at all 9 HAAS examinations, with persons suspected of cognitive or motor impairment receiving full neurologic and neuropsychologic diagnostic evaluations. Blood testing for Apolipoprotein E genotype (APOE), hormone levels, hematologic levels, HDL and LDL cholesterol, C reactive protein, fibrinogen, fasting glucose and insulin levels, and other factors were done at the baseline HAAS examination. Subjective and objective assessments of physical function (grip strength, chair stands, timed walk, balance), were done at 5 HAAS examinations. Comprehensive research-protocol brain autopsies were completed for 852 HAAS decedents between 1992 - 2011. Below is the date distribution for each visit:

Table 1.1. Range for CASI visit Date

	First Visit Date for exam	Last Visit Date for exam
Exam 4	1991-02-22	1993-10-22
Exam 5	1994-03-31	1996-04-30
Exam 6	1997-10-09	1999-02-12
Exam 7	1999-07-06	2000-10-16
Exam 8	2001-12-03	2004-02-20
Exam 9	2003-11-17	2005-09-13
Exam 10	2007-04-04	2009-03-03
Exam 11	2009-04-09	2010-09-30
Exam 12	2011-01-17	2011-08-29

Mid-life alcohol exposure: Estimates of the total ethanol intake from reported alcohol exposure patterns were calculated as ounces of ethanol per month for beer, liquor, wine and sake using algorithms based on average unit sizes and usual percentages of alcohol. Consistent interview and recording methods were employed at exam 1 & 3 of HHP. Alcohol intake was not considered for exam 2, considering exam 2 (1968-71) occurred right after exam 1 (1965). We also excluded the alcohol consumption for exam 4 (1986-89) to avoid unreliable recall at the time of the HAAS examination, especially for those with impaired cognition. The local culture of these men reflects an unusual honesty because alcohol exposure is neither admired nor discouraged.

For statistical analysis, binary or categorical exposures usually have better epidemiology interpretation, we thus dichotomized the alcohol exposure by more than 14 drinks per week as heavy drinker and less equal than 14 drinks per week as light drinker. Here, we defined 0.5 oz pure alcohol as one drink, and 14 drinks per week would be 7 oz a week, thus 30.1 oz per month. For alcohol exposure history, participants were classified into four alcohol exposure categories, based on intake estimations at both exam 1 & 3: men who were (1) both light drinkers at exam 1 & 3 (light - light), (2) light drinkers at exam 1 then moderate drinkers at exam 3 (light - heavy), (2) heavy drinkers at exam 1 then light drinkers at exam 3 (heavy - light) and (4) both moderate drinkers at exam 1 & 3 (heavy - heavy). It has been showed ([4]) that light to moderate alcohol drinkers or non-drinkers in the mid-life have reduced risk of dementia. Thus, for the survival

analysis, we will focus on comparing light-light (both light drinkers at exam 1 & 3) versus the rest three groups. For the data analysis in Chapter 3 and Chapter 4, we define the exposed group as heavy drinker: who at least be a heavy drinker once at exam 1 & 3, i.e: light - heavy, heavy - light, heavy - heavy drinkers, and define the control group as light drinker who is both light drinkers at exam 1 & 3.

Cognitive outcomes: The central purpose of the HAAS was to employ precise, standardized, consistent methods to estimate rates and risk factors for AD, vascular dementia, and related conditions. For our analyses, cognitive impairment will be based on sequential scores on the Cognitive Assessment and Screening Instrument (CASI), which was first assessed at HHP exam 4. The CASI includes tests of attention, concentration, orientation, short- and long-term memory, language, visual construction, list generating, fluency, abstraction, and judgment; it has a score range of 0 to 100. The distribution of CASI was highly left-skewed, with most of the scores on the high end. Despite the CASI scores, we are also interested in years (days) to moderate cognitive impairment. A CASI score below 74 is considered as moderate cognitive impairment, and below 60 is considered severe cognitive impairment. Numerous data plots show that CASI scores below 60 rarely stabilize or improve thereafter, we thus focus on time to moderate impairment as out time-to-event outcome.

Competing event and censoring: Since HAAS is a longitudinal observational study, it is very hard to follow up all the participants until the end of the study or the death (distribution of number of participant across each visit is showed in Appendix A: Table A.1). Meanwhile, we identified that some participants dead long after they dropped from the study, and the death dates were extracted from the death certificant: see Appendix A Figure A.2 for last visit age and death age distribution across each visit. Under this data structure, for participants who have death date in record, we need to identify whether death is the event that happened in HAAS study.

In HAAS study, exam visits were conducted in every 3-4 year. We can then assign each event: moderate cognitive impairment, death and loss to follow up to the proper exam visits, based on the dates that those events happened. For each event, we assign the exam visit based on

visit date distribution Table 1.1, and claim that if the participant died after the first visit date and before the last visit date of an exam, then that exam number will be assigned as the death exam visit number. Following are some special occasions in the data and solutions to it:

- The last visit date and death date are in the same exam: the death will be assigned to the next exam visit.
- The death occurs between two visits: then death will be assigned to the later exam visit.
- The exam visit that is assigned to death date is greater than last visit number plus one (i.e. participant died more than two years after the last visit), then it will be considered as loss to follow up (i.e. censoring in our case).
- The death occurs after first date of exam 9 and before last date of exam 8: if the participant's last visit is exam 7, then death will be assigned to exam 8; if the participant's last visit is exam 8, then death will be assigned to exam 9 (all of the participants in this case are whose event happened before death or loss to follow up).

After assigned the death exam number to each participant, Table 1.2 shows the range of the death date for each visit:

Table 1.2. Range for Death Date across each visit

	First Death Date for exam	Last Death Date for exam
Exam 4		
Exam 5	1991-09-04	1996-04-09
Exam 6	1994-09-13	1999-02-11
Exam 7	1998-03-07	2000-09-26
Exam 8	2000-01-27	2004-02-16
Exam 9	2002-06-19	2005-08-21
Exam 10	2004-03-31	2009-01-02
Exam 11	2007-09-30	2010-05-04
Exam 12	2009-10-14	2011-08-02
Exam 13	2011-09-22	2012-07-19

Confounders: According to [65], systolic blood pressure and heart rate were associated with cognitive impairment, and [53] show that regular drinking users tend to have high blood

pressure. Meanwhile, [34] showed that Apolipoprotein E genotype (APOE) is strongly associated with Alzheimer's disease. In epidemiology study, education years and baseline age are considered as important baseline demographic data to represent the social economics status, we thus also included baseline age and education years as confounder. Based on all the information, for all the analysis, we considered systolic blood pressure, heart rate, baseline age, education years, APOE, as confounders.

1.3 Aims and Organization of this Dissertation

This dissertation contains both applying contemporary causal inference methods and developing the novel statistical methods for investigating highly complex relationships between mid-life exposures and late-life cognitive outcomes related to Alzheimer's disease. Our projects focus on using the rich data resources of the Honolulu Heart Program (HHP) and the Honolulu-Asia Aging Study (HAAS), linked epidemiologic projects that collected longitudinal alcohol, physical and cognitive measures from mid-to late life. Our projects also reflect an urgent need for more effective analytic methods to disentangle causal factors from allied confounders that have likely contributed to previous inconsistencies using traditional statistical approaches.

For the chapter one, we first gave a literature review on scientific background of the dissertation as well as the introduction of the study that inspired for the later work. Because of the complexity of the data structure, we also introduced the detailed steps for cleaning the data.

For the chapter two, we first gave a literature review on existing statistical methods in causal inference for longitudinal data with continuous outcome. We then set up the potential outcome framework and applied the novel MSM model with IPW for the scientific question that inspired from the HHP and HAAS study. Details analysis step and theoretical background were provided, and we perform the analysis on accessing the causal effect of mid-life alcohol exposure to late-life cognitive functions.

For the chapter three, we first gave a literature review on existing statistical methods in

survival analysis with competing risks data, and then we set up the potential outcome framework in causal survival analysis. To estimate the causal effect of mid-life alcohol exposure to time to moderate cognitive impairment with death as competing event, we adapted the MSM cause-specific Cox proportional hazards model with IPW. We also provided the formula for cumulative incidence functions (i.e. risks) from predicted MSM model. An R package *cmprskcoxmsm* was developed for estimating point treatment causal effect for time to event data with the existence of competing events.

For the chapter four, we inspired from the previous chapter and treated death as semi-competing event, as we are also interested in how the mid-life alcohol exposure affects the time to moderate cognitive impairment, time to death as well as time to death after cognitive impairment. We applied the three-state illness-death model to observational data using the potential outcomes framework. Inverse probability of treatment weighting is used to fit these structural models. Under the Cox model formulation, typical software used to fit the Cox regression model can be used to fit the usual Markov model in the absence of frailty. With the frailty term under the general Markov model, a weighted EM algorithm is developed and its convergence property studied. The intense simulation studies showed the good performance of our proposed methods.

For the chapter five, since the IPW estimator is biased if the propensity score model is misspecified, an augmented IPW (AIPW) estimator with doubly robust properties can protect against such model misspecification. Three AIPW estimators under MSM usual Markov model were derived, and It would also allow us to apply machine learning or nonparametric methods to the propensity score model. This work is still under investigation.

Chapter 2

Assess the Causal Effect of Time Varying Alcohol Exposure on Cognitive Assessment Score at the Start of HAAS Study

2.1 Abstract

Objectives: To assess the causal effect of mid-life alcohol exposure on late-life cognitive score that relates to Alzheimer's disease (AD) in a 2628 Japanese-American men from the population-based Honolulu Heart Program (HHP) and The Honolulu-Asia Aging Study (HAAS).

Methods: Three on-site exams were taken since 1965, while drinking consumption was assessed at Exam 1 and 3 before the Cognitive Abilities Screening Instrument (CASI) was examed(exam 4). All the exams were conducted at Kuakini Medical Center, Honolulu, Hawaii. Based on whether they consume greater 30.1 ounces per month, we stratified those participants in heavy and light drinker at each exam. The Marginal structure model with inverse propensity score weighting (IPW) was used to assess the causal effect of drinking status to CASI controlling for time-varying confounders: Systolic blood press and Heart rate, and baseline confounders: baseline age, education, APOE genotype.

Main Results: The difference of mean CASI was compared between 4 different alcohol exposure groups. We found that participants who are heavy drinker at both two exam visits have 1.02 ($p = 0.582$) lower CASI score than light drinker at two exam visits. There is no significant

difference in CASI score for any pairwise comparisons.

Conclusions: Mid-life alcohol exposure does not have causal effect on decreasing or increasing for late-life cognitive functions. More research is needed, and we should limit the recommendation of heavy alcohol consumption for all older people.

2.2 Introduction

More than 25 million people worldwide are affected by dementia, most suffering from Alzheimer's disease (AD). According to NIH, the number of affected people is expected to increase three-fold by 2050, from the currently approximate 5 million (10% of persons aged 65 and older). The risk factors, including many exposures or lifestyle, other than age or AD-related genotype, are not well understood. Without further understanding, effective prevention strategies remain a challenge.

Compared with abstinence, light alcohol intake is often considered to be associated with health benefits ([28, 30, 21]). Such benefits have also been reported to extend to cognitive health in late life in several studies ([42, 44]), while other studies found the opposite ([27]). It is now generally accepted that the pathogeneses of AD probably begin decades earlier and involve a progressive accumulation of tissue injury and cellular loss with disruptions of physiologic or metabolic systems accruing over decades. Therefore, it is crucial to have available reliable longitudinal measures for alcohol consumption and many other relevant variables, as well as to carefully examine the myriad associated factors that may confound or otherwise interfere with a confident assessment of the independent contributions of alcohol.

The alcohol consumption is easy to regulate or interfere, such as, The Center for Disease Control and Health Promotion (CDC) has many suggestions and regulations on alcohol consumption. Therefore, to investigate the causal effect of alcohol use on AD or ADRD will be necessary such that we can interfere with one's alcohol usage or make recommendations on alcohol consumption.

In randomized clinical trial (RCT), explaining the causal effect of an exposure or a treatment is simple, because randomization ensures that there are no confounding factors for the exposure or treatment of interest. However, sometimes RCT is unfeasible or unethical, and a well-designed observational study is also useful for medical research, policy study etc. The biggest problem of observational studies is that the assumption of no confounders for the

exposure or treatment of interest is violated (which prevented by randomization and blinding in RCT), but many advanced statistical methods are developed to control over confounders. Donald Rubin ([50]) first proposed a Neyman–Rubin causal model to estimate causal effect on the framework of potential outcomes and th tink about causation in both observational and experimental studies. [46] then proposed the Marginal structural model (MSM) with inverse probability weighting (IPW) methods for epidemiology studies, where "structural" means modeling the probabilities of counterfactual outcome, and "marginal" means it models the marginal distribution of the counterfactual outcome rather than the joint distribution. MSM is a powerful tool for confounding control in observational studies especially for longitudinal study with the existence of time-varying exposure, outcome and confounders. For longitudinal data analysis, time-varying confounders is an inevitable feature. Also, in some longitudinal studies, time-varying confounders affect and can be affected by time-varying treatment assignment, for example: In HIV study, CD4 count will affect the initialization of zidovudine (AZT) treatment, and meanwhile AZT will decrease the CD4 count. In this case, stratified analysis or standard parametric regression will cause the bias estimation of causal effect. While MSM is a widely use methodology for handling time-varying confounders, parametric g-formula is also a widely use methods. Parametric g-formula use the standardization techniques to decompose the averaged treatment effect (ATE) and assign the parametric model for each identifiable term in the formula. [64] explained how to use parametric g-formula in the regression analysis, while [63] extended it to the survival outcome.

For this paper, we will focus on mid-life alcohol consumption, represent important candidate determinants of disease processes leading to AD or ADRD. Our paper will be accomplished using the rich data resources of the HHP and the HAAS, linked epidemiologic projects that collected longitudinal alcohol, physical and cognitive measures from mid-to late life. We will develop and implement novel causal inference approaches and statistical models to examine the relationship between time-varying moderate alcohol consumption and cognitive impairment while adjusting the time-varying exposures. Our main interest is on studying the causal relation-

ship between mid-life alcohol exposure and the late life cognitive functions. For the HHP study, we observed the time-varying alcohol exposure, as well as time-varying confounders: blood pressure (BP) and heart rate (HR). We first plan to study the causal effect of time-varying mid-life alcohol exposure on late-life cognitive outcomes. Figure 2.1 shows the directed acyclic graph (DAG) depicting the causal relations between the baseline confounders Z_0, the time-varying mid-life alcohol exposures A_1 & A_3, the time-varying confounders Z_1 and Z_3: BP and HR, and the outcome Y (CASI score). The time invariant confounder Z_0 will be: education years, baseline age and APOE positive.

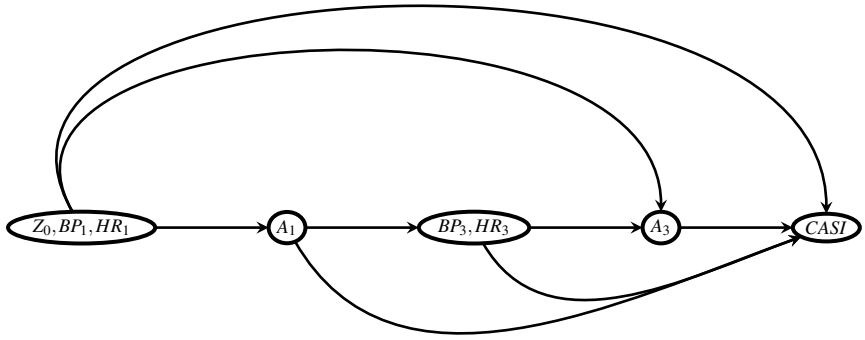

Figure 2.1. Directed acyclic graph (DAG) for the causal relations between the baseline and time-varying confounders Z_0 and Z_1, Z_3, time-varying exposures A, and outcome Y

2.3 Methods

2.3.1 Data Resource

Data: Here we consider the data from the Honolulu Hearth Program project (HHP, 1965 -1974), in order to study the effects of mid-life alcohol exposure collected during HHP on early-late-life cognitive function scores (CASI), full description of the data are shown in Chapter 1. HHP followed a cohort of Japanese men born between 1900 - 1920 and lived in Hawaii, and demographic data, vital data and life-style questionnaire are obtained in all four exam visits in HHP.

Cognitive outcomes: We are interested in CASI score (detailed introduction of CASI can be find in Chapter 1 Section 1.2) at exam 4. The CASI score has a range of 0 to 100.15, and the distribution of CASI was highly left-skewed, with most of the scores on the high end. We, therefore, used the log(101-CASI) transformation for all the analyses, which resulted in an approximately normal distribution.

Time-varying confounders and time-invariant confounders: Time-invariant confounders are discussed in Chapter 1 Section 1.2). From Figure 2.1, we can tell that the systolic blood pressure and ventricular hear rate at exam 3 will affected by the the alcohol exposure at exam 1, thus the systolic blood pressure and ventricular hear rate were treated as time-varying confounders that change over each exam visit, and were affected by the previous exam's exposure data.

Sample Attrition and Exclusions: Until exam 3 of HHP, a total of 6860 men participated. Of this total, 3674 (54%) continue participating in the HAAS examination, while 2530 (37%) had died by the time of the HAAS examination and 603 (9%) were alive but did not participate. Of those 3674 participants, 2654 were included in the analysis. 1020 participants were excluded because of incomplete alcohol intake data and insufficient information on covariates.

2.3.2 Statistical approaches

Due to the non-random nature of this observational longitudinal study, and also the existence of time-varying confounders, we used inverse probability weighting (IPW) of PS to estimate the mean treatment effect (ATE) of heavy drinking versus light drinking among the study population. PS provides an estimate of an individual's probability of receiving an intervention at each visit, based on information available to the individual's historical data, and PS can help to account for selection bias inherent in routine observational data. .

Following [46], let $Y^{a_1 a_3}$ be the value of Y that would have been observed had all subjects received dose history (a_1, a_3) rather than their observed alcohol exposure history (A_1, A_3). Note that $A_k, k = 1, 3$ is dichotomous on each exam visit, we will have $2^2 = 4$ possible counterfactuals. We also make the following assumptions that are commonly used in causal inference:

(I) Sequential exchangeability: $Y^{a_1 a_3} \perp A_k \mid A_{k-1}, Z_0, Z_k, k = 1, 2$

(II) Sequential positivity: $P(A_k = a_k \mid a_{k-1}, z_0, z_{k-1}) > 0, k = 1, 2$

(III) Consistency: If $A_1 = a_1, A_3 = a_3$, then $Y^{a_1 a_3} = Y$

The sequential exchangeability means that the probability of being exposed at each time depends on exposed and confounders history and, conditional on this history, does not depend on any unmeasured causes of the outcome. The sequential positivity means that the probability of being exposed at each time k will never be 0 or 1, no matter the past confounders and exposure history. Consistency means that if $A = a$ for a given individual, then $Y^a = Y$ for that individual. Meanwhile, the distribution of CASI was highly left-skewed, with most of the scores on the high end. We, therefore, used the log(101-CASI) transformation for all the analyses, which resulted in an approximately normal distribution, see Appendix A.

Under the identifiability condition, the mean of potential outcome $E(Y^{a_1a_3})$ can be fit with a saturated linear MSM ([46]):

$$E(Y^{a_1a_3}) = \beta_0 + \beta_1 * a_{01} + \beta_2 * a_{10} + \beta_3 * a_{11} \tag{2.1}$$

where $a_{01} = 0,1, k = 1,2$ represent whether the subject is in Light - Heavy drinker group. Similar interpretation can also be used for a_{00}, a_{01}, a_{10}.

Since $Y^{a_1a_3}$ is counterfactual outcome and cannot be fully observed, we thus use time-varying IPW to create a pseudo population, where in the pseudo population, the counterfactual mean $E(Y^{a_1a_3})$ is the $E(Y|A_1 = a_1, A_3 = a_3)$. We define the IPW as following:

$$W^{a_1a_3} = \frac{1}{Pr(A_1 = 1|Z_0, Z_1)} * \frac{1}{Pr(A_3 = 1|Z_3, A_1, Z_0, Z_1)}$$

and the stabilized inverse probability weights are:

$$SW^{a_1a_3} = \frac{Pr(A_1 = 1)}{Pr(A_1 = 1|Z_0, Z_1)} * \frac{Pr(A_3 = 1|A_1)}{Pr(A_3 = 1|Z_3, A_1, Z_0, Z_1)}$$

For the weights at exam 3, we have unstabilized and stabilized weights. For stabilized weights:

$$SW = \frac{Pr(A_3 = 1|A_1)}{Pr(A_3 = 1|Z_3, A_1, Z_0, Z_1)}$$

Where we estimate the $Pr(A_3 = 1|A_1)$ by the marginal proportion of 4 different alcohol exposure class.

PS were calculated using R package "*twang*" (Toolkit for Weighting and Analysis of Nonequivalent Groups). Instead of using parametric logistic regression model, "*twang*" use the nonparametric boosted regression as the predicted probability of heavy drinking vs light drinking, conditioned on the measured historical variables thought to be confounders or predictors for the outcome of interest, as well as the alcohol exposure. Notice that PS model for exam 3 should

include baseline confounders, the alcohol exposure status, systolic blood pressure and ventricular hear rate at exam 1, and systolic blood pressure and ventricular hear rate at exam 3.

2.4 Results

The demographic characteristics of the participants at baseline (N=2654) are presented in Table 2.1. The baseline mean of length of education was slightly shorter in Heavy + Heavy groups (9.87 years), and the Light + Light group tends to have a smaller proportion of E4Positive (17.8%).

Table 2.1. Demographic data between four alcohol exposure groups

	Light - Light ($n = 1908$)	Light - Heavy ($n = 300$)	Heavy- Light ($n = 134$)	Heavy - Heavy ($n = 312$)
Age	52.85 (4.55)	53.69 (4.78)	52.00 (3.99)	52.58 (4.41)
Education (Years)	10.85 (3.25)	9.92 (3.16)	10.27 (2.58)	9.87 (2.97)
APOE Positive (Yes)	343 (17.8%)	63 (21.2%)	31 (23.1%)	60 (18.7%)
Baseline BP	128.75 (17.84)	131.18 (18.43)	135.14 (18.38)	134.76 (19.33)
Baseline HR	76.27 (11.91)	75.35 (11.25)	78.98 (11.10)	78.44 (12.27)

The distribution of CASI and log(101-CASI) for four alcohol exposure schemes is shown in Figure 2.2, and more distribution transformation plots are shown in the Appendix A. We can tell we have most of the participants (N=1474) who are heavy drinkers at both exam 1 and 3, and the participants tend to have a lower CASI, which means impairment cognitive performance.

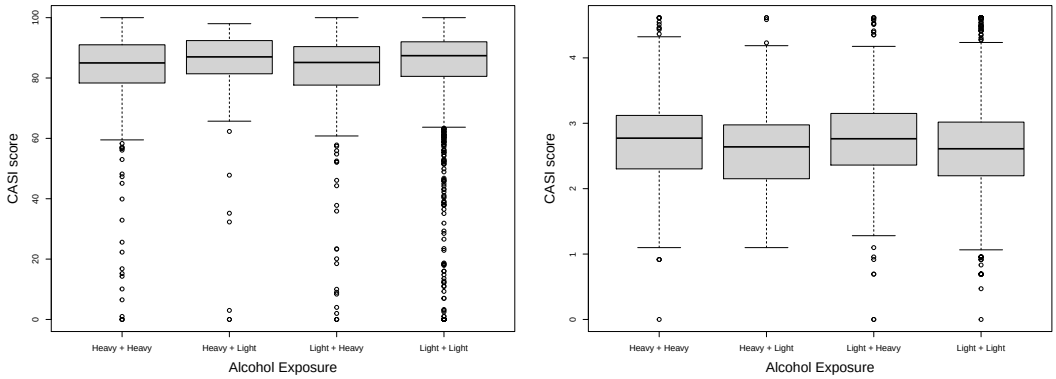

Figure 2.2. Distribution of CASI (left) and log(101-CASI) (right) across four alcohol exposure groups

The distribution of propensity score (probability of being heavy drinker) at exam1 & 3 are shown in in the Figure 2.3, where we found that the sequential positivity is not violated. The standardized mean difference plots and more PS distribution plots across the two exams visit are shown in the Appendix A.

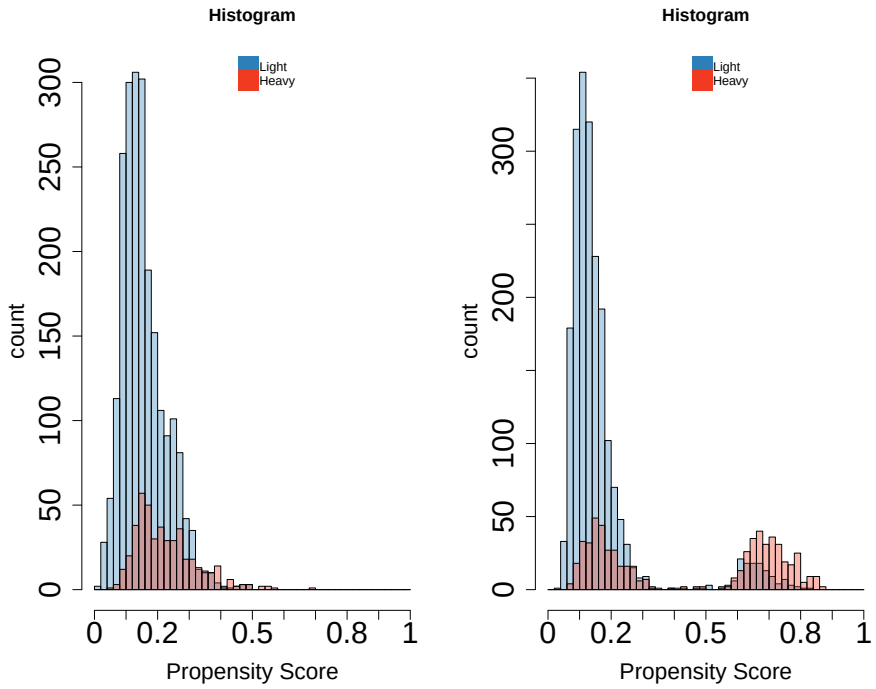

Figure 2.3. PS distribution for exam 1 (left) and exam 3 (right)

For the regression results, we provide three methods to estimate the standard error. The first one **SE.r** is the output generated by r summary table, and they use the formula:

$$SE.r = \sqrt{(X'WX)^{-1} * Var(\varepsilon)}$$

The second one **SE.WLS** is using sandwich formula for weighted least square:

$$SE.WLS = \sqrt{[(X'WX)^{-1}X'W] * [(X'WX)^{-1}X'W]' * Var(Y)}$$

were we estimate the variance of Y by the sample variance of CASI at time 4.

The third one **SE.Boot** is using bootstrap, where we generated 1000 bootstrap samples, each formed by resampling pairs $(x_i, y_i, w_i$ with replacement from the original data. Then, we run a weighted linear regression on each of the bootstrap samples of size n and extract the coefficient

of interest. We can estimated the variance-covariance matrix by taking variance and covariance for our bootstrap samples. The p-value based on different standard errors are also provided.

There are 4 possible combinations for the alcohol exposure class, then we will have 6 pairwise comparisons for estimated CASI in total. The results for the 6 comparisons are presented below, 3 possible standard error estimates are also provided for each comparison:

Table 2.2. MSM regression results for six pairwise comparison, where 00 represent Light - Light group, 01 represent Light - Heavy group, 10 represent Heavy - Light group and 11 represent Heavy - Heavy group

	Estimate	SE.r	SE.WLS	SE.Boot
01 vs. 00	0.069	0.045 (p=0.127)	0.046 (p=0.135)	0.045 (p=0.131)
10 vs. 00	-0.022	0.069 (p=0.750)	0.068 (p=0.748)	0.058 (p=0.739)
11 vs. 00	0.073	0.044 (p=0.095)	0.047 (p=0.118)	0.045 (p=0.111)
10 vs. 01	-0.091	0.085 (p=0.287)	0.086 (p=0.288)	0.077 (p=0.286)
11 vs. 01	-0.004	0.067 (p=0.952)	0.070 (p=0.954)	0.067 (p=0.954)
11 vs. 10	0.094	0.085 (p=0.262)	0.086 (p=0.269)	0.078 (p=0.272)

The results of MSM with IPW were presented at Table 2.2. The estimation what were obtained by fitting the model (1) were β_1=0.069 (standard errors se=0.046), β_2=-0.022 (se=0.069), β_3=0.073 (se=0.044) which can be interpreted as the difference of mean log(101-CASI) between the Light + Heavy, Heavy + Light, Heavy+ Heavy and Light + Light (Table 2.2). As can be seen from the standard errors, all coefficients were not statistically significantly different from 0. The other 3 pairwise differences were shown in the Table 2.2, there were also no statistically significant differences.

We also provided the weighted mean of CASI for 4 alcohol exposure groups with 95% confidence interval:

Table 2.3. Weighted mean of CASI for 4 alcohol exposure groups with 95% confidence interval

	Estimate	SE.r	SE.WLS	SE.Boot
Light - Light	87.29	(86.86, 87.71)	(86.85, 87.72)	(86.88, 87.75)
Light - Heavy	86.31	(85.04, 87.48)	(85.01, 87.50)	(85.00, 87.51)
Heavt - Light	87.59	(85.70, 89.24)	(85.72, 89.26)	(85.75, 89.22)
Heavy - Heavy	86.25	(85.02, 87.38)	(84.93, 87.46)	(84.90, 87.43)

2.5 Discussion

Alcohol intake is frequent among elder people living in the United States. The results of the present study suggest the non-significant causal relationship between moderate alcohol intake and late-age cognitive functions among these Japanese-American men. The non-significant causal relationship suggested moderate alcohol intake may not have side effects on the late-life cognitive functions. While these results may increase the attention for the benefit of moderate alcohol consumption on the cognitive outcomes, our study still has evidence showed that the men who are both moderate drinkers at both exam 1 and 3 have poor cognitive performance than other groups. More research is needed, and we should limit the recommendation of moderated drinking of alcohol for all older people.

Our study is the first study to apply the novel causal inference model to examine the causation of moderate alcohol consumption on cognitively functions in the sizeable mid-age sample. Prior investigations used alternative approaches to study associations with healthy longevity in women. Results from the Nurses' Health Study, showed an association between moderate, regular alcohol consumption at midlife and successful ageing defined as living to age 70 without physical or cognitive impairment2. Although our outcome differs from this previous studies, we all in agreement concerning the potential benefits of alcohol consumption for healthy brain function. Our study extends the findings to men and to a large population longitudinal study.

There are several limitations to this study. We examined the cognitive function at only one time point, so there is no comparison between the baseline cognitive function to the end of follow-up. Meanwhile, the assumption of MSM is that there is no unmeasured confounders over time other than blood pressure and heart rate, which is implausible for the medical research, sensitivity analysis should be performed to test the existence of unmeasured confounders. There were several strengths to our study. The extensive data collected on this cohort allowed for control of many potential confounders such as education, gene related to AD. Further, by using the

marginal structural model, we can control the time-varying confounders such as blood pressure and heart rate, to assess the unbiased estimator for the drinking effect.

For the future direction: there is abundant evidence showing that blood pressure may be associated with cognitive impairment. In the meantime, plenty of medication was designed to control blood pressure, and it is easy for us to control the blood pressure on the population. Thus, developing and applying novel causal inference methods to detect the role of blood pressure on cognitive outcomes is necessary. At the same time, we do not make full use of the HAAS longitudinal study, where CASI score changes over each visit. From this data structure, instead of comparing CASI score directly between alcohol exposure group, learning the relationship of alcohol exposure and time to moderate impairment is also an interesting area to explore.

Chapter 3

Assess the Causal Effect of Mid-life Alcohol Exposure on Time to Moderate Cognitive Impairment with Death as Competing Risk

3.1 Introduction

MSM has been proved as a useful tool for estimating causal effect in longitudinal study, and [16] later extended the MSM to the survival setting, specifically to the Cox proportional hazard model by modeling the counterfactual event time. Based on the idea proposed by [16], later, MSM has been extended for many different settings in survival analysis. For example: [38] proposed a MSM that permits estimation of cause-specific hazards in situations where more than one cause of death is of interest, also [6] extended MSM for survival analysis while handling the informative censoring. [49] also proposed a Bayesian MSM for survival outcomes. In particular, they take a Bayesian nonparametric approach, using a combination of a dependent Dirichlet process and Gaussian process to model the observed data. While MSM is a popular methods in causal survival analysis, other statistical models are also developed for survival analysis. For example: [2] talked about comparing G-formula and IPW in survival analysis and proposed a new pseudo-observations method for estimating the survival probability, and [7] presented a one-step Targeted Maximum Likelihood Estimator (TMLE) for estimating the counterfactual

average survival curve. Instrumental variable analysis [56] and mediation analysis [61] are also been widely discussed nowadays. Some scholar also focused on developing the the doubly robust efficient estimator for the causal effect, for example: [10] derived the doubly robust efficient estimator for semi-parametric additive hazards models, and [19] established the efficient and doubly robust estimator for learning treatment effects under additive hazards models with high-dimensional confounding.

In general, for survival analysis, we assumed that there is only one event of interest which means only one survival endpoint. However, in many contexts such as drug study, one can experience difference types of event, for example: disease relapse, death, or serious adverse event etc, which are all the event of interest for us. Meanwhile, the occurrence of one type of event may prevent us from observing other events, and those events are called competing risk in survival analysis. In survival analysis literatures, two types of model are widely use for compering risk analysis. The first one is called cause-specific cox model, which we model the cause-specific hazard for each event type, and the overall hazard for all the event is the the sum of cause-specific hazard over the all the event types. The second one is called sub-distribution model proposed by [11], it is a semi-parametric proportional hazards model for the sub-distribution, and we can use the model to estimate the cumulative incidence function (CIF) which is defined as the probability of an event in the setting where other competing risks are acknowledged to exist. Competing risk data is a very common data in survival analysis, thus extending the causal inference methods to the competing risk data setting is essential. In the recent years, there are tremendous literatures discussing causal inference in competing risk setting. Some literatures focused on using the parametric g-formula methods to estimate the ATE. For example: [70] claimed that counterfactual hazard contrasts cannot generally be interpreted as causal effects, and proposed two parametric g-formula methods for competing risk which competing events are depicted as time-varying covariates. Followed by Young's setting, [54] proposed the idea of separable effects to study the causal effect of a treatment on an event of interest, and defined the separable direct effect as treatment effect on the event of interest not mediated by its effect

on the competing event while the indirect effect as treatment effect on the event of interest only through its effect on the competing event.

The objective of this first project is to adapt the MSM with IPW to the cause-specific hazard model to the HAAS study data, and proposed a new predicted counterfactual risk based on the results of cause-specific hazard model. For the HAAS study data, we first studied the data distribution, make the definition of the event, competing event and censoring.

3.2 Methods

3.2.1 Data Resource

Data: Here we consider the data from the the Honolulu Heart Program (HHP) and the Honolulu-Asia Aging Study (HAAS), in order to study the effects of mid-life alcohol exposure collected during HHP on late life time to moderate cognitive impairment as well as time to death, with considering the other as competing event, full description of two studies are shown in Chapter 1. Mid-life alcohol exposure and baseline confounders are obtained from HHP study, while the time to event data are obtained from HHP study. Both studies followed a cohort of Japanese men born between 1900 - 1920 and lived in Hawaii. Here we take the first exam visit at HAAS study, also the exam 4 at HHP study as the well-defined time zero.

Outcomes: Moderate cognitive impairment was defined based on CASI score less or equal than 74 (more description of CASI see in Chapter 1 Section 1.2). Since HAAS was conducted during the late-life of the participants, death were observed during the study and prevented us observing the moderate cognitive impairment. More details about defining the competing events in HAAS data can be also found in Chapter 1 Section 1.2.

Exposure: The mid-life alcohol exposure was collected during the HHP study between 1965-73. The Heavy Drinking group consisted of individuals who had heavy drinking at one point during mid-life, and the Light Drinking those who never had heavy drinking during mid-life.

Confounders: The confounders were decided by literature review and clinical experi-

ences, as well as availability of the data. Demographic data such as baseline age, APOE are considered as confounders by recent clinical papers. Systolic blood pressure and ventricular heart rate at exam 4 (baseline of HAAS study) are also considered as the confounders that is related to the outcomes. Since CASI score are different among participants, we also considered baseline CASI score (CASI score at exam 4) as the baseline confounder.

Sample Attrition and Exclusions: Until exam 4 of HHP, a total of 2654 men participated. Of this total, we only include participants with normal cognitive function (CASI ≥ 74) at baseline, and after excluding missing values for exposure and confounders, we have 1881 participants in total. Of those participants, 1390 (74%) participants have been considered as Light Drinking and 491 (26%) are considered as Heavy drinking.

3.2.2 Statistical Methods

Following by survival notation with presence of possible right censoring and competing event, we define T as time to event, and and J as the event type indicator (more specifically, $J = 1$ is the event I, $J = 2$ is event II). We define A as the exposure indicator. Let T^a be the potential time to event, and J^a be the analogous potential event-type indicator. It is potential because not all the subjects are exposed to $A = a$, then we only observe part of the variables from population. We defined Z as the baseline confounders. The causal relationship can be shown from the below DAG:

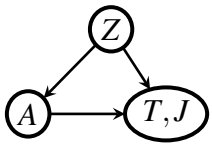

Figure 3.1. The causal directed acyclic graph

We make the following assumptions that are commonly used in causal inference:

(I) Exchangeability: $(T^a, J^a) \perp A \mid Z$

(II) Positivity: $P(A = a \mid Z) > 0$

(III) Consistency: If A=a, then $T^a = T$, $J^a = J$

The exchangeability means that within levels of Z, all other predictors of the outcome are equally distributed between the every treated groups. We can also say that the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates Z. In observational study, it also called ignobility, meaning that there is no unmeasured confounders in the study. The positivity assumption means that the probability of receiving every value of treatment conditional on Z is greater than zero, i.e., positive. Meaning, in the study, each subject could be assigned to either one of treatment. The consistency assumption means that the values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data, it also means we observe one of the potential outcomes at a time.

Followed by Hernan (2001), we also make the assumption on noninformative censoring:

(IV) Noninformative censoring: $(T^a, J^a) \perp C^a \mid Z$

We then focus on defining average treatment effect (ATE). The ATE is the average effect in the population level, it projects to the entire population. In the continuous outcome setting, ATE is defined as $E(Y^{A=a_1} - Y^{A=a_0})$, where Y is an arbitrary continuous outcome. For the survival analysis setting, we first specify the marginal structural cause-specific cox proportional hazards model for event $j = 1, 2$:

$$\lambda_{T^a, J^a = j}(t) = \lambda_{0j}(t) e^{\beta_j * a} \tag{3.1}$$

which is the cause-specific hazard of T^a at t under treatment a, $\lambda_{0j}(t)$ is the unspecified baseline

cause-specific hazards for event j. Then $\exp(\beta_j)$ is the ATE, which is the causal hazard ratio for event j.

The potential outcome, by definition, is not fully observed. Following [46] and [16], to estimate the parameters in model (3.1), we use the inverse probability weights (IPW) to create a pseudo population, where in the pseudo population, the confounder Z get controlled. The inverse probability weights for treatment $A = a$ can be defined as:

$$W(a) = \frac{1}{P(A = a \mid Z)}$$

and the stabilized inverse probability weights for treatment $A = a$ will be:

$$SW(a) = \frac{P(A = a)}{P(A = a \mid Z)} \tag{3.2}$$

After we estimating the cause specific hazard: λ_j^a using IPW, we could estimate the corresponding CIF. Following [18], we could use :

$$\hat{P}(T^a < t, J^a = j) = \int_0^t \hat{S}^a(u) d\hat{\Lambda}_j^a(u),$$

where $\hat{S}^a(u)$ is the estimated over all survival function for T^a, $\hat{S}^a(u) = e^{-\hat{\Lambda}_1^a(u) - \hat{\Lambda}_2^a(u)}$, $\hat{\Lambda}_j^a(u) = \hat{\Lambda}_{0j}(u) e^{\hat{\beta}_j * a}$, and $\hat{\Lambda}_{0j}(u)$ is a Breslow-type estimator of the baseline cumulative hazard (for j=1,2).

For this analysis, we fit the saturated MSM cause-specific Cox proportional hazards model 3.1. Where $a = 0, 1$ represent whether the subject is in light drinker group, or in rest of three groups (called heavy drinker group). The propensity score: probability of being heavy drinker was calculated for each subjects based on selected confounders (Chapter 1 Section 1.2). We also calculated the predicted CIF for time to moderate impairment and time to death, confidence band is provided by using bootstrap. All the analysis were conducted using R package *cmprskcoxmsm* that was developed based on this project. For the package, we used the *twang*

package to generate the propensity score which is based on generalized boosted tree (GBM).

3.3 Results

The demographic characteristics of the participants at baseline (N=1881) are presented in Table 3.1, 29 participants are removed from analysis because of missing baseline confounders. The baseline mean of length of education was slightly higher in Light-Light groups (11.22 years), and the Non Light-Light group tends to have a higher proportion of APOE Positive (21.4%), higher systolic blood pressure (151.40) and lower baseline CASI score (87.38).

Table 3.1. Baseline demographic of two alcohol exposure groups

	Light Drinking ($n = 1390$)	Heavy Drinking ($n = 491$)
Systolic BP	148.76 (21.26)	151.40 (21.95)
Baseline Age	77.05 (3.80)	77.12 (3.75)
Education (Years)	11.22 (3.12)	10.42 (2.97)
APOE Positive (Yes)	254 (18.3%)	105 (21.4%)
Heart Rate	31.22 (4.62)	31.88 (4.83)
Baseline CASI	88.53 (6.00)	87.38 (6.13)

We first need to generate the IPW weights for the data, below is the distribution of propensity score (check positivity assumption) and the standard mean difference (SMD) plot before and after weighting, we can tell that the positivity assumption is not violated under this case, and by weighting, the confounders are controlled well.

After having the IPW, the next step is to fit the marginal structural cause-specific Cox proportional hazard model for estimating alcohol exposure effects: we fit the model using the stabilized weight:

Table 3.2. Regression results for comparing non Light-Light vs Light-Light drinkers for years to moderate impairment

	Estimate	Robust SE	z-value	p-value	Hazard Ratio	95% CI
Heavy Drinking	0.202	0.08	2.518	0.012$*$	1.224	(1.046, 1.432)

From the results, we can see that the estimated hazard ratio is 1.224 with 95% CI (1.046, 1.432), so there is significant mid-life alcohol exposure effect on time to late life moderate

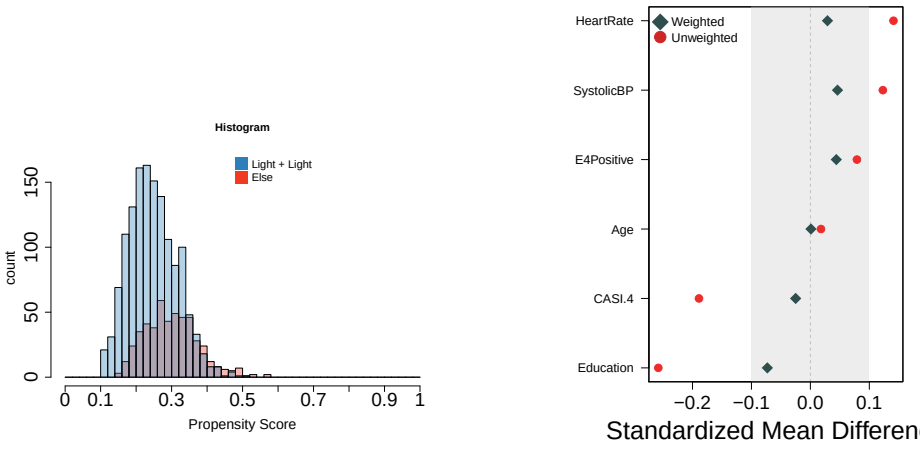

Figure 3.2. Propensity score distribution (left) and SMD before and after weighting

cognitive impairment.

We can also estimate the CIF from the data for the different event types, the 95% confidence interval of the CIF is calculated by bootstrap. We estimate the CIF for the moderate impairment (left) and death (right), we can tell that even though the risk of moderate impairment and risk of death are higher in the Heavy drinking group, but the confidence band are overlapped, so the difference is not statistically significant.

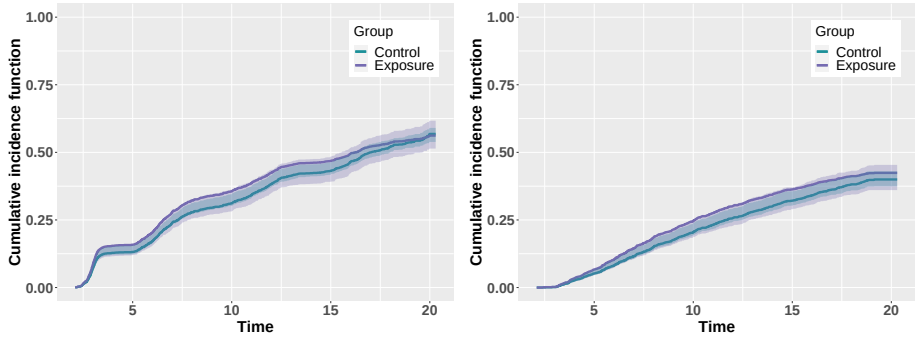

Figure 3.3. CIF for moderate impairment (left) and death (right) with 95% confidence band

3.4 Discussion

Alcohol intake is frequent among elder people living in the United States. The results of the present study suggest the significant causal relationship between alcohol intake and late-age

cognitive functions among these Japanese-American men wgucg suggested having heavy alcohol intake during mid-life may have side effects on the late-life cognitive functions. More research is needed, and we should limit the recommendation of moderated drinking of alcohol for all older people.

Our study is the first study to apply the novel causal inference model to examine the causation of alcohol consumption on cognitively functions in the sizable mid-age sample. Prior investigations used alternative approaches to study associations with healthy longevity in women. Results from the Nurses' Health Study, showed an association between heavy, irregular alcohol consumption at midlife and successful aging defined as living to age 75 without physical or cognitive impairment ([42]). Although our outcome differs from this previous studies, we all in agreement concerning the potential harm of alcohol consumption for unhealthy brain function. Our study extends the findings to men and to a large population longitudinal study.

There are several limitations to this study. We examined all the confounders at only one time point, and we did not consider time varying confounders in our analysis because of large percentage of missing value. Meanwhile, the assumption of MSM is that there is no unmeasured confounders over time other than blood pressure and heart rate, which is implausible for the medical research, sensitivity analysis should be performed to test the existence of unmeasured confounders. Meanwhile, there were several strengths to our study. The extensive data collected on this cohort allowed for control of many potential confounders such as education, gene related to Alzheimer's disease. Further, by using the marginal structural model, we can control the existing confounders such as blood pressure and heart rate, to assess the unbiased estimator for the alcohol exposure effect.

Acknowledgement

Chapter 3, in full, has been submitted for publication as "Zhang, Yiran; Ronghui Xu. *cmprskcoxmsm: Use IPW to Estimate Treatment Effect under Competing Risks*, submitted to

CRAN Package". The dissertation author was the primary author on this paper.

Chapter 4

Marginal Structural Illness-Death Models for Semi-Competing Risks Data

4.1 Abstract

The three-state illness–death model has been established as a general approach for regression analysis of semi-competing risks data. In this paper we apply it to a class of marginal structural models for observational data. We consider two specific such models, the usual Markov illness–death structural model and the general Markov illness–death structural model which incorporates a frailty term. For interpretation purposes, risk contrasts under the structural models are defined. Inference under the usual Markov model can be carried out using estimating equations with inverse probability weighting, while inference under the general Markov model requires a weighted EM algorithm. We study the inference procedures under both models using extensive simulations, and apply them to the analysis of mid-life alcohol exposure on late life cognitive impairment as well as mortality using the Honolulu-Asia Aging Study data set. The R codes developed in this work have been implemented in the R package *semicmprskcoxmsm* that is publicly available on CRAN.

4.2 Introduction

Our work was motivated by the longitudinal epidemiologic Honolulu-Asia Aging Study (HAAS). The HAAS cohort is comprised of the surviving participants from the Honolulu Heart Program (HHP), a prospective, community-based cohort study of heart disease and stroke established in 1965 with about 8,000 men of Japanese ancestry living on the island of Oahu, who were born between 1900-1919. HAAS was established in 1991 and was brought to closure in 2012 with the goal of determining the prevalence, incidence, and risk factors for Alzheimer's disease (AD) and brain aging. Demographic data, vital status and diet data were collected every 2-3 years during the HHP period, and neuropsychologic assessment were performed every 2-3 years during the HAAS. Our goal is to assess the causal effect of mid-life alcohol exposure captured during HHP on late life outcomes collected in HAAS. In particular, a subject may develop cognitive impairment, then die, or die without cognitive impairment. These are referred to as semi-competing risks where there are non-terminal events (cognitive impairment) and terminal events (death). As outcomes we are interested in time to non-terminal event and time to terminal event, as well as time to the terminal event following the non-terminal event.

The above semi-competing risks setting is the same as the three-states illness-death model depicted in Figure 4.1,[67] which was first introduced by [13]. We assume that a subject starts in the "healthy" state (state 0), then transition into the cognitive impairment (state 1) or death state (state 2), which are also referred to as the intermediate or non-terminal, and the terminal state, respectively. The corresponding transition events are then the non-terminal event and the terminal event, respectively. [67] discussed extensively the illness-death model for semi-competing risks data, and also incorporated a shared frailty term in the illness-death model that encompasses previous works such as the copula model of [12]. The illness-death model with shared frailty has been extended to different situations including in the presence of left truncation,[29] or for a nested case-control study.[22] [31] extended this model to the Bayesian paradigm. [1] developed an R package to analyze semi-competing risks data under the illness-death model

using parametric models and the Bayesian method, but not for the semiparametric Cox model formulation.

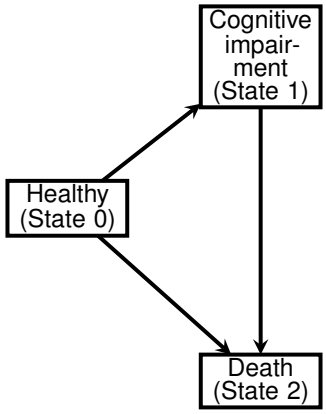

Figure 4.1. Three-state illness-death model

For observational data, marginal structural models (MSM) have been established as a valuable tool for identifying causal effects, which can be consistently estimated using the inverse-probability-of-treatment weighting (IPTW).[46, 16] In this paper we consider a class of marginal structural illness–death models, with and without a shared frailty term. For the former an EM type iterative algorithm is needed in order to estimate the parameters. The structural models give rise to interpretable causal quantities such as different types of risk contrasts in the multi-state setting.[37] The remainder of this article is organized as follows. In the next section we introduce the structural models and assumptions. In Section 3 we discuss inference under the usual Markov illness-death structural model and Section 4 the general Markov illness-death structural model, where a weighted EM algorithm is developed and studied. In Section 5 we carry out extensive simulation studies to assess the performance under the two models including when either one of the model is valid while the other is not. We apply the approaches to the HAAS data set described above in Section 6 and conclude with more discussion in the last section.

4.3 Three-State Illness-Death model

4.3.1 Definitions and assumptions

For our setup, assume a well-defined time zero, and let random variables T_1 and T_2 denote time to the non-terminal and the terminal event since time zero, respectively. If a subject does not experience the non-terminal event before the terminal event, we define $T_1 = +\infty$.[67, 12] Denote the joint density of T_1 and T_2 as $f(t_1,t_2)$ in the upper wedge $0 < t_1 \leq t_2$, and the density of T_2 along the line $t_1 = +\infty$ as $f_\infty(t_2)$ for $t_2 > 0$. Note that for semi-competing risks data, we do not observe any data in the lower wedge $0 < t_2 < t_1 < +\infty$; see Figure 4.2. We also denote the bivariate survival function of T_1 and T_2 in the upper wedge as $S(t_1,t_2)$.

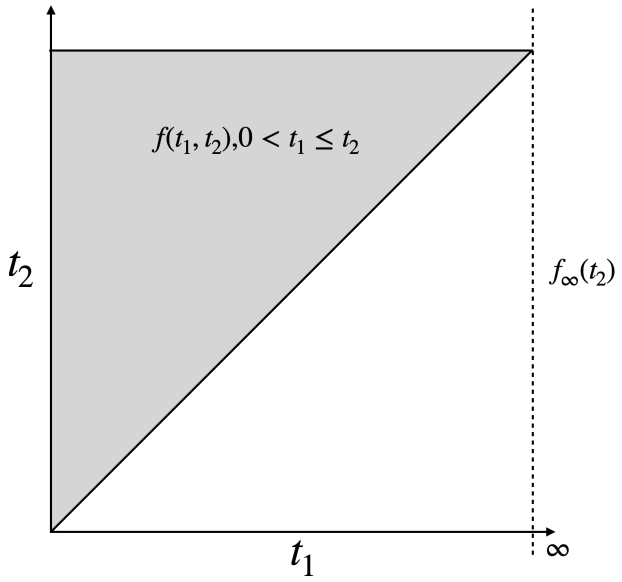

Figure 4.2. Joint density function of T_1 and T_2

The multi-state model quantifies event rates and event risks based on the history of events, and is completely specified by the three transition intensities below, also referred to as transition rates in the literature. Let $\lambda_1(t_1)$ and $\lambda_2(t_2)$ be the transition rates from the initial healthy state to the non-terminal, and the terminal state, respectively, and $\lambda_{12}(t_2 \mid t_1)$ the transition rate from the

non-terminal state to the terminal state. That is,

$$\lambda_1(t_1) = \lim_{\Delta \to 0^+} \frac{P\left(T_1 \in [t_1, t_1+\Delta) \mid T_1 \geq t_1, T_2 \geq t_1\right)}{\Delta}, \tag{4.1}$$

$$\lambda_2(t_2) = \lim_{\Delta \to 0^+} \frac{P\left(T_2 \in [t_2, t_2+\Delta) \mid T_1 \geq t_2, T_2 \geq t_2\right)}{\Delta}, \tag{4.2}$$

$$\lambda_{12}(t_2 \mid t_1) = \lim_{\Delta \to 0^+} \frac{P\left(T_2 \in [t_2, t_2+\Delta) \mid T_1 = t_1, T_2 \geq t_2\right)}{\Delta}. \tag{4.3}$$

Note that (4.1) and (4.2) are in fact the cause-specific hazards in the usual competing risks setting, for time to the non-terminal event and time to the terminal event without non-terminal event, respectively. In general, $\lambda_{12}(t_2 \mid t_1)$ can depend on both t_1 and t_2. In the following we consider the commonly used Markov assumption: $\lambda_{12}(t_2 \mid t_1) = \lambda_{12}(t_2)$, i.e. the transition rate from non-terminal to terminal state does not depend on what value T_1 takes.

While the transition rates in (4.1) - (4.3) completely specifies the three-state illness-death model, for interpretation purposes various risk type quantities can be of interest in practice. Cumulative incidence function (CIF) are commonly used for competing risks,[24] that is, for the non-terminal event, denoted by $F_1(t_1)$ below, and for the terminal event without the non-terminal event, denoted by $F_2(t_2)$ below. In addition, we may also consider a third CIF, denoted by $F_{12}(t_1, t_2)$, for the terminal event following the non-terminal event.[37] We have

$$F_1(t_1) = P(T_1 \leq t_1, \delta_1 = 1) = \int_0^{t_1} S(u)\lambda_1(u)du, \tag{4.4}$$

$$F_2(t_2) = P(T_2 \leq t_2, \delta_2 = 1, \delta_1 = 0) = \int_0^{t_2} S(u)\lambda_2(u)du, \tag{4.5}$$

$$F_{12}(t_1, t_2) = P(T_2 \leq t_2 \mid T_1 \leq t_1, T_2 \geq t_1) = 1 - \exp\left\{-\int_{t_1}^{t_2} \lambda_{12}(u)du\right\}, \tag{4.6}$$

where $S(t) = \exp\left[-\int_0^t \{\lambda_1(u) + \lambda_2(u)\}\, du\right]$.

In the presence of right censoring, such as lost to follow-up or administrative censoring, let C be the time to right censoring since time zero. Denote $X_1 = \min(T_1, T_2, C)$, $X_2 = \min(T_2, C)$, and the event indicators $\delta_1 = I\{X_1 = T_1\}$, $\delta_2 = I\{X_2 = T_2\}$, where $I(\cdot)$ is the indicator function.

Let $A = \{0,1\}$ be a binary treatment assignment, possibly not randomized. Following [40] and [51] framework of potential outcomes, we denote T_1^a, T_2^a, C^a as potential time to the non-terminal event, terminal event and censoring under treatment $a = 0,1$. And X_1^a, X_2^a, δ_1^a and δ_2^a are similarly defined. Let Z be a p-dimensional vector of covariates. Denote $\pi(Z) = P(A = 1 \mid Z)$, often referred to as the propensity score. The causal relationship of the variables defined above can be depicted in a graphical display called a chain graph as in Figure 4.3,[57] where the undirected line indicates correlation. A chain graph without undirected edges is known as a causal directed acyclic graphs (DAG).

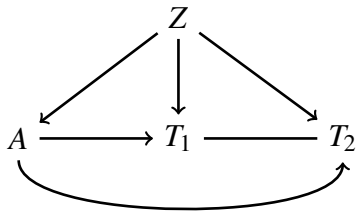

Figure 4.3. Causal chain graph representation of semi-competing risks data

We assume the following, which are commonly used in order to identify the causal estimands to be specified later:

(I) Stable unit treatment value assumption (SUTVA): there is only one version of the treatment and that there is no interference between subjects.

(II) Exchangeability: $(T_1^a, T_2^a) \perp A \mid Z$.

(III) Positivity: $\pi(Z) > 0$.

(IV) Consistency: If $A = a$, then $T_1^a = T_1$, $T_2^a = T_2$, $C^a = C$.

Exchangeability implies that within levels of the variable Z, the potential event times (T_1^a, T_2^a) and the treatment assignment A are independent. It is also called (conditional) ignobility, and that there are no unmeasured confounders. The positivity assumption requires that the probability of receiving either treatment ($A = 1$) or control ($A = 0$) is positive for any given value of Z. The consistency assumption here links the potential outcomes with the observed outcomes. For more discussion on these assumptions, please see [17].

We also assume:

(IV) Non-informative censoring: $(T_1^a, T_2^a) \perp C^a \mid Z$.

4.3.2 The structural models

Let $\lambda_1(t_1;a)$, $\lambda_2(t_2;a)$ and $\lambda_{12}(t_2|t_1;a)$ be the transition rates corresponding to the counterfactual states under the three-state model, $a = 0, 1$. [3] discussed about modeling each transition intensity by a Cox type proportional intensities regression model. Following the same idea, we can postulate the semi-parametric Cox models for these transition rates, which are also hazard functions.[67, 3] In particular, we consider the following *usual Markov illness-death structural model*:[67]

$$\lambda_1(t_1;a) = \lambda_{01}(t_1)e^{\beta_1 a},\ t_1 > 0; \tag{4.7}$$

$$\lambda_2(t_2;a) = \lambda_{02}(t_2)e^{\beta_2 a},\ t_2 > 0; \tag{4.8}$$

$$\lambda_{12}(t_2|t_1;a) = \lambda_{03}(t_2)e^{\beta_3 a},\ 0 < t_1 < t_2. \tag{4.9}$$

The joint distribution of T_1 and T_2 under model (5.4) - (5.6) will be given as a special case below.

The usual Markov illness-death model can be extended by incorporating a frailty term, to the *general Markov illness-death structural model*. The frailty term induces further correlation between T_1 and T_2, beyond what is already contained in the joint distribution of T_1 and T_2 above. It also models unobserved heterogeneity among individuals.[26, 41] Following [59] we consider the log-normal distribution for the frailty, and we have

$$\lambda_1(t_1|b;a) = \lambda_{01}(t_1)e^{\beta_1 a+b},\ t_1 > 0; \tag{4.10}$$

$$\lambda_2(t_2|b;a) = \lambda_{02}(t_2)e^{\beta_2 a+b},\ t_2 > 0; \tag{4.11}$$

$$\lambda_{12}(t_2|t_1,b;a) = \lambda_{03}(t_2)e^{\beta_3 a+b},\ 0 < t_1 < t_2, \tag{4.12}$$

where $b \sim N(0,\sigma^2)$. Obviously model (5.4) - (5.6) is a special case of (4.10) - (4.12) by setting

$b = 0$.

Recall the joint density $f(t_1,t_2)$ and the bivariate survival function $S(t_1,t_2)$ previously defined in the upper wedge $t_1 \leq t_2$, and the density function $f_\infty(t_2)$ along the line $t_1 = +\infty$. In the Supplementary Materials we show that these quantities can be derived as functions of the transition rates (4.1) - (4.3). With the models specified in (4.10) - (4.12) we then have the following quantities that will be used later:

$$f(t_1,t_2;a) = \lambda_{01}(t_1)\lambda_{03}(t_2)e^{\beta_1 a+b+\beta_3 a+b}\exp\left(-\Lambda_{01}(t_1)e^{\beta_1 a+b} - \Lambda_{02}(t_1)e^{\beta_1 a+b}\right) \times \exp\left(-\Lambda_{03}(t_1,t_2)e^{\beta_3 a+b}\right), \tag{4.13}$$

$$f_\infty(t_2;a) = \lambda_{02}(t_2)e^{\beta_2 a+b}\exp\left(-\Lambda_{01}(t_2)e^{\beta_1 a+b} - \Lambda_{02}(t_2)e^{\beta_2 a+b}\right), \tag{4.14}$$

$$S(t,t;a) = \exp\left(-\Lambda_{01}(t)e^{\beta_1 a+b} - \Lambda_{02}(t)e^{\beta_2 a+b}\right), \tag{4.15}$$

where $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(u)du$ for $j = 1,2$, and $\Lambda_{03}(t_1,t_2) = \Lambda_{03}(t_2) - \Lambda_{03}(t_1)$ with $\Lambda_{03}(t) = \int_0^t \lambda_{03}(u)du$.

4.3.3 Likelihood

In this subsection we assume that the treatment A is randomized so that we can write down the relevant probabilities for the four scenarios below. We will then use inverse probability weighting (IPW) to create a pseudo-randomized sample. Denote $O_i = (X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i}, A_i)$ the observed data for subject i, and L_c the likelihood conditional on the random effect b. We have the following four different scenarios:

(i) Non-terminal event then censored prior to terminal event: $X_{1i} = T_{1i}, X_{2i} = C_i, \delta_{1i} =$

$1, \delta_{2i} = 0,$

$$
\begin{aligned}
L_c(O_i \mid b_i) &= \int_{X_{2i}}^{+\infty} f(X_{1i}, t_2) dt_2 \\
&= \lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} \\
&\quad \times \exp\left\{ -\Lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} - \Lambda_{02}(X_{1i}) e^{\beta_2 A_i + b_i} - \Lambda_{03}(X_{1i}, X_{2i}) e^{\beta_3 A_i + b_i} \right\};
\end{aligned}
$$

(ii) Non-terminal event and then terminal event: $X_{1i} = T_{1i}, X_{2i} = T_{2i}, \delta_{1i} = 1, \delta_{2i} = 1,$

$$
\begin{aligned}
L_c(O_i \mid b_i) &= f(X_{1i}, X_{2i}) \\
&= \lambda_{01}(X_{1i}) \lambda_{03}(X_{2i}) e^{\beta_1 A_i + b_i + \beta_3 A_i + b_i} \\
&\quad \times \exp\left\{ -\Lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} - \Lambda_{02}(X_{1i}) e^{\beta_1 A_i + b_i} - \Lambda_{03}(X_{1i}, X_{2i}) e^{\beta_3 A_i + b_i} \right\};
\end{aligned}
$$

(iii) Terminal event without non-terminal event: $X_{1i} = T_{2i}, X_{2i} = T_{2i}, \delta_{1i} = 0, \delta_{2i} = 1,$

$$
\begin{aligned}
L_c(O_i \mid b_i) &= f_\infty(X_{2i}) \\
&= \lambda_{02}(X_{2i}) e^{\beta_2 A_i + b_i} \exp\left\{ -\Lambda_{01}(X_{2i}) e^{\beta_1 A_i + b_i} - \Lambda_{02}(X_{2i}) e^{\beta_2 A_i + b_i} \right\};
\end{aligned}
$$

(iv) Censored before any event: $X_{1i} = X_{2i} = C_i, \delta_{1i} = 0, \delta_{2i} = 0,$

$$
L_c(O_i \mid b_i) = S(X_{1i}, X_{2i}) = \exp\left\{ -\Lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} - \Lambda_{02}(X_{2i}) e^{\beta_2 A_i + b_i} \right\}.
$$

Combining the above four scenarios, we have

$$
\begin{aligned}
L_c(O_i \mid b_i) = & \left\{ \lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} \right\}^{\delta_{1i}} \exp\left\{ -\Lambda_{01}(X_{1i}) e^{\beta_1 A_i + b_i} \right\} \\
& \cdot \left\{ \lambda_{02}(X_{2i}) e^{\beta_2 A_i + b_i} \right\}^{\delta_{2i}(1-\delta_{1i})} \exp\left\{ -\Lambda_{02}(X_{1i}) e^{\beta_2 A_i + b_i} \right\} \\
& \cdot \left\{ \lambda_{03}(X_{2i}) e^{\beta_3 A_i + b_i} \right\}^{\delta_{2i}\delta_{1i}} \exp\left\{ -\Lambda_{03}(X_{1i}, X_{2i}) e^{\beta_3 A_i + b_i} \right\}. \qquad (4.16)
\end{aligned}
$$

4.4 The Usual Markov Structural Model

In the absence of randomization, denote $w_i = A_i/\hat{\pi}(\mathbf{Z}) + (1-A_i)/\{1-\hat{\pi}(\mathbf{Z})\}$ as the IP weight for subject i. In practice, $\pi(\cdot)$ is unknown and can be estimated from the data by either specifying a parametric model such as the logistic regression,[46] or use nonparametric methods such as boosted trees.[35]

For the usual Markov illness-death model, with $b_i = 0$ in (4.16), we have the weighted log-likelihood

$$\begin{aligned}\log L_w = &\sum_i w_i \left[\delta_{1i}\{\beta_1 A_i + \log(\lambda_{01}(X_{1i}))\} - \Lambda_{01}(X_{1i})e^{\beta_1 A_i}\right] \\ &+ \sum_i w_i \left[\delta_{2i}(1-\delta_{1i})\{\beta_2 A_i + \log(\lambda_{02}(X_{2i}))\} - \Lambda_{02}(X_{1i})e^{\beta_2 A_i}\right] \\ &+ \sum_i w_i \left[\delta_{2i}\delta_{1i}\{\beta_3 A_i + \log(\lambda_{03}(X_{2i}))\} - \Lambda_{03}(X_{1i}, X_{2i})e^{\beta_3 A_i}\right]. \qquad (4.17)\end{aligned}$$

It can be seen that the parameters for the three transition rates (β_j, Λ_{0j}), $j = 1,2,3$, are variationally independent in the above likelihood and therefore can be estimated separately. Note that the semiparametric approach under the Cox type models discretizes the baselines hazards $\lambda_{0j}(\cdot)$ into point masses at the observed event times and estimates the cumulative $\Lambda_{0j}(\cdot)$ as step functions. It can be verified that maximizing (4.17) is equivalent to maximizing the following three weighted Cox regression model likelihoods: 1) treating the non-terminal event as the event of interest, and terminal event without non-terminal or originally censored as 'censored'; 2) treating the terminal event without non-terminal as the event of interest, and non-terminal event or originally censored as 'censored'; 3) treating the terminal event following the non-terminal as the event of interest, left truncated at the time of the non-terminal event (so only those who had the non-terminal event are included), and originally censored as 'censored'. Then the standard software (e.g. *coxph()* in R package 'survival') can be used to obtain the estimates $(\hat{\beta}_j, \hat{\Lambda}_{0j})$, $j = 1,2,3$.

In order to obtain the variance of the estimates, if we assume the estimated weights

in (4.17) as known, then the robust sandwich variance estimator in standard software such as *coxph()* can be used to obtain the estimated variance for $\hat{\beta}_j, j = 1,2,3$. In the Supplementary Materials we provide the formulas for estimating the covariances between $\beta_j, j = 1,2,3$. In addition, we may also use the bootstrap variance estimator which accounts for the uncertainty in estimating the weights.

For causal interpretation, we may define the risk contrasts as the difference or the ratio between the CIF's under the structural models with $a = 1$ and $a = 0$. In particular,

$$F_1(t_1;a) = \exp(\beta_1 a) \int_0^{t_1} S(u;a)\lambda_{01}(u)du, \tag{4.18}$$

$$F_2(t_2;a) = \exp(\beta_2 a) \int_0^{t_2} S(u;a)\lambda_{02}(u)du, \tag{4.19}$$

$$F_{12}(t_1,t_2;a) = 1 - \exp\left\{-e^{\beta_3 a}\int_{t_1}^{t_2} \lambda_{03}(u)du\right\}, \tag{4.20}$$

where $S(t;a) = \exp\left[-\int_0^t \left\{\lambda_{01}(u)e^{\beta_1 a} + \lambda_{02}(u)e^{\beta_2 a}\right\} du\right]$. We estimate the contrasts by plugging in the parameter estimates, and obtain their 95% confidence intervals (CI) using bootstrap. We note that for simple competing risk data under the marginal structural Cox model, such risk contrasts are available in the R package 'cmprskcoxmsm'.[72]

4.5 The General Markov Structural Model

Under the general Markov illness-death model (4.10) - (4.12) where $b \sim N(0,\sigma^2)$, let $\theta = (\beta_1,\beta_2,\beta_3,\Lambda_{01},\Lambda_{02},\Lambda_{03},\sigma^2)$. Denote $O = \{O_i\}_{i=1}^n$. The weighted observed data likelihood is:

$$L_w(\theta;O) \;=\; \prod_i \left\{\int L(\theta;O_i \mid b_i)\cdot f(\theta;b_i)db_i\right\}^{w_i}, \tag{4.21}$$

where $f(\theta;b_i)$ is the normal density function. Then the estimate $\hat{\theta}$ can be obtained by maximizing (4.21).

We introduce below an EM type algorithm in order to maximize (4.21). Denote $Q(\theta,\tilde{\theta})$ the expectation of the weighted log-likelihood of the augmented data (y_i,b_i), $i=1,...,n$, conditional on the observed data and the current parameter value $\tilde{\theta}$:

$$Q(\theta,\tilde{\theta})=\sum_i \mathbb{E}\left[w_i\cdot l\left(\theta_i;O_i|b_i\right)\mid O,\tilde{\theta}\right]+\sum_i \mathbb{E}\left[w_i\cdot \log f\left(\theta;b_i\right)\mid O,\tilde{\theta}\right], \tag{4.22}$$

where

$$\begin{aligned} l\left(\theta;O\mid b\right)=\Big[&\delta_1\left\{b+\beta_1 A+\log\left(\lambda_{01}(X_1)\right)\right\}\\ &+\delta_2(1-\delta_1)\left\{b+\beta_2 A+\log\left(\lambda_{02}(X_2)\right)\right\}\\ &+\delta_2\delta_1\left\{b+\beta_3 A+\log\left(\lambda_{03}(X_2)\right)\right\}\\ &-\Lambda_{01}(X_1)e^{\beta_1 A+b}-\Lambda_{02}(X_1)e^{\beta_2 A+b}-\Lambda_{03}(X_1,X_2)e^{\beta_3 A+b}\Big]. \end{aligned} \tag{4.23}$$

Then $Q=Q_1+Q_2+Q_3+Q_4$, where

$$\begin{aligned} Q_1(\beta_1,\lambda_{01})=\sum_i w_i\Big[&\delta_{1i}\big\{\,\mathbb{E}(b_i)+\beta_1 A_i+\log\left(\lambda_{01}(X_{1i})\right)\big\}\\ &-\Lambda_{01}(X_{1i})\exp\{\beta_1 A_i+\log\mathbb{E}(e^{b_i})\}\Big], \end{aligned} \tag{4.24}$$

$$\begin{aligned} Q_2(\beta_2,\lambda_{02})=\sum_i w_i\Big[&\delta_{2i}(1-\delta_{1i})\big\{\,\mathbb{E}(b_i)+\beta_2 A_i+\log\left(\lambda_{02}(X_{2i})\right)\big\}\\ &-\Lambda_{02}(X_{1i})\exp\{\beta_2 A_i+\log\mathbb{E}(e^{b_i})\}\Big], \end{aligned} \tag{4.25}$$

$$\begin{aligned} Q_3(\beta_3,\lambda_{03})=\sum_i w_i\Big[&\delta_{2i}\delta_{1i}\big\{\,\mathbb{E}(b_i)+\beta_3 A_i+\log\left(\lambda_{03}(X_{2i})\right)\big\}\\ &-\Lambda_{03}(X_{1i},X_{2i})\exp\{\beta_3 A_i+\log\mathbb{E}(e^{b_i})\}\Big], \end{aligned} \tag{4.26}$$

$$Q_4(\sigma^2)=\sum_i w_i\Big\{-\frac{1}{2}\big(\log 2\pi+\log\sigma^2\big)-\frac{1}{2\sigma^2}\,\mathbb{E}(b_i^2)\Big\}, \tag{4.27}$$

where $E\{h(b_i)\}=E\{h(b_i)\mid O_i,\tilde{\theta}\}$ is shorthand for a function $h(\cdot)$ of b_i. Analogous to the EM

algorithm, we iterate between the E-steps and the M-steps described below until convergence.

E-step

The conditional expectations in (4.24) - (4.27) are all in form of $E\{h(b_i) \mid O_i, \tilde{\theta}\} = \int h(b_i) f(b_i \mid O_i, \tilde{\theta}) db_i$, where $h(b_i) = e^{b_i}$ in (4.24) - (4.26) and $h(b_i) = b_i^2$ in (4.27). These two expectations are not in closed form; however, we can approximate these integrals by numerical methods, specifically by (adaptive) Gaussian quadrature.[14, 45] Details of computation are shown in the Supplement Materials.

M-step

The M-step conveniently separates the update of β_j and Λ_{0j} for $j = 1,2,3$ from that of the variance component σ^2. For Q_1 - Q_3, similar to Section 4.4, (4.24) - (4.26) are equivalent to the weighted log-likelihood functions in a Cox regression with additional known offsets $\mu_i = \log E(e^{b_i} \mid O, \tilde{\theta})$. In order to maximize Q_4, we set

$$\frac{\partial Q_4}{\partial \sigma^2} = \sum_i w_i \Big\{ -\frac{1}{2\sigma^2} + \frac{\mathbb{E}(b_i^2 \mid O, \tilde{\theta})}{2\sigma^4} \Big\} = 0,$$

leading to

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n w_i \mathbb{E}(b_i^2 \mid O, \tilde{\theta})}{\sum_{i=1}^n w_i}, \tag{4.28}$$

In the lemma below, we establish the following property of the above weighted EM algorithm, which is similar to that of the EM algorithm.

lemma 1. *Suppose $L_w(\theta; O)$ is the weighted observed data likelihood. At step k of the algorithm denote $\theta^{(k)}$ the current value, and $\theta^{(k+1)}$ the value that maximizes $Q(\theta, \theta^{(k)})$. Then:*

$$L_w(\theta^{(k+1)}; O) \geq L_w(\theta^{(k)}; O). \tag{4.29}$$

The proof of the lemma is given in the Supplement Materials. Following [66] or Theorem 4.12 in [32], since $Q(\theta;\tilde{\theta})$ is continuous in both θ and $\tilde{\theta}$, then all limit points of the weighted EM sequence $\{\theta^{(k)}\}$ are stationary points of $L_w(\theta;O)$, and $L_w(\theta^{(k)};O)$ converges monotonically to $L_w(\theta^*;O)$ for some stationary point θ^*. In addition, for existence of such limit point(s) [58] proposed a condition for the usual unweighted EM algorithm: as long as the maximizer in the M-step is unique. We can show that this result extends immediately to our weighted EM algorithm. And finally, our M-step satisfies this condition, i.e. the maximizer in the M-step is unique.

As initial values we use for β_j and Λ_{0j}, $j = 1,2,3$, the estimates from weighted Cox regression without the offsets, i.e. from the usual Markov model of the previous section; and $\sigma^2 = 1$. The stop criteria we use in this paper are convergence in the log-likelihood as well as in parameters of interest: $|\log L_w(\theta^{(k+1)};y) - \log L_w(\theta^{(k)};y)| \leq 10^{-5}$, $|\beta_j^{(k+1)} - \beta_j^{(k)}| \leq 10^{-3}$, $j = 1,2,3$ and $|\sigma^{2(k+1)} - \sigma^{2(k)}| \leq 10^{-3}$.

Variance estimate

The variance of the parameter estimates following a typical EM algorithm can be estimated by the inverse of a (discrete) observed information matrix calculated using Louis' formula, including for the nonparametric maximum likelihood estimator (NPMLE) under, for example, the semiparametric proportional hazards mixed models.[59] For observational data, however, inference using the weighted NPMLE under semiparametric models requires the derivation of efficient influence functions,[5] and is generally non-trivial under the normal frailty construct.[39, 33] In the following we use bootstrap to obtain the variance estimator for $\hat{\theta}$.

Risk contrasts

Similar to what we proposed under the usual Markov model, we also can define the risk contrasts under the general Markov model. Since the general Markov models are conditional on

the random effect b, we have the following conditional risk:

$$F_1(t_1 \mid b;a) = \exp(\beta_1 a + b) \int_0^{t_1} S(u \mid b;a)\lambda_{01}(u)du, \tag{4.30}$$

$$F_2(t_2 \mid b;a) = \exp(\beta_2 a + b) \int_0^{t_2} S(u \mid b;a)\lambda_{02}(u)du, \tag{4.31}$$

$$F_{12}(t_1,t_2 \mid b;a) = 1 - \exp\left\{-e^{\beta_3 a+b} \int_{t_1}^{t_2} \lambda_{03}(u)du\right\}, \tag{4.32}$$

where

$$S(t \mid b;a) = \exp\left[-\int_0^t \left\{\lambda_{01}(u)e^{\beta_1 a+b} + \lambda_{02}(u)e^{\beta_2 a+b}\right\} du\right] \tag{4.33}$$

$$= \exp\left\{-e^{\beta_1 a+b}\Lambda_{01}(t) - e^{\beta_2 a+b}\Lambda_{02}(t)\right\}. \tag{4.34}$$

As discussed earlier the frailty term, or equivalently, the random effect b represents the unobserved heterogeneity among the individuals. As such, the above conditional risk represents individual risk, and the risk contrasts the individual risk contrasts. We therefore have the individual risk difference (IRD) and the individual risk ratio (IRR). Under the random effects model, for $i = 1,2,...,n$, the predicted random effect is $\hat{b}_i = \mathbb{E}(b_i \mid O_i, \hat{\theta})$.[59] We then obtain the predicted IRD and the predicted IRR. For inference on these individual risk contrasts, Bayesian bootstrap[25] may be used which, unlike the usual resampling with replacement, preserves each individual i in the original data set. Details of the Bayesian bootstrap are provided in the Supplementary Materials. Note that because b is random, the common terminology in the literature is 'predicted' instead of 'estimated', and 'prediction interval (PI)' instead of CI.

4.6 Simulation

We carry out extensive Monte Carlo simulation studies in order to assess the performance of the estimation procedure described above. We use the idea from [15] to simulate data under the marginal structural model (4.10) - (4.12). We also adapt the method from [23], originally

designed for simulating semi-competing risk data with gamma frailty. Very briefly the following steps are used to to generate the data; more details are provided in the Supplementary Materials.

- Generate $U_1 \sim U(0,1)$ and $U_2 \sim U(0,1)$;
- Generate confounder $Z = (Z_1, Z_2, Z_3)^\intercal$, with $Z_j = U_1 + U_2 + \varepsilon_j, j = 1,2,3$, where $\varepsilon_1 \sim N(0,1)$, $\varepsilon_2 \sim N(0,1.5)$ and $\varepsilon_3 \sim N(0,1.8)$;
- Generate $A \sim$ Bernoulli(p_A), where $p_A = \text{logit}^{-1}(\alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3)$, with $\alpha_0 = 0.5, \alpha_1 = 0.1, \alpha_2 = -0.1, \alpha_3 = -0.2$;
- Let $\lambda_{01}(t) = \lambda_{02}(t) = 2e^{-t} I(0 \le t \le 3) + 2e^{-3} I(t > 3)$ and $\lambda_{03}(t) = 2\lambda_{01}(t)$. Then with probability $P(T_1 = \infty)$ given in the Supplementary Materials,

$$T_2 = \Lambda_{01}^{-1}\left(-\frac{\log(U_1)}{\exp(\beta_1 A + b) + \exp(\beta_2 A + b)}\right);$$

and with probability $1 - P(T_1 = \infty)$,

$$T_1 = \Lambda_{01}^{-1}\left(-\frac{\log(U_1)}{\exp(\beta_1 A + b) + \exp(\beta_2 A + b)}\right),$$

$$T_2 = \Lambda_{01}^{-1}\left(-\frac{\log(U_2)}{2\exp(\beta_3 A + b)} + \Lambda_{01}(t_1)\right).$$

- Generate Censoring time $C \sim U(0.4, 0.5)$, which leads to an average censoring rate around 20%.

We set $\beta_1 = \beta_2 = 1$, $\beta_3 = 0.5$. Weights are calculated by fitting the logistic regression with Z_1, Z_2, Z_3 as covariates. We run 500 simulations for each case. Table 4.1 and 4.2 report, for sample size n=250 and n=500, respectively, the estimate, the empirical standard deviation (SD), the mean of estimated standard errors (SE), and the coverage probability (CP) of the nominal 95% confidence intervals. Under the usual Markov model, we estimate the asymptotical variance

of β_j, $j = 1,2,3$ using both the model-based formulas, which ignores the uncertainty in the estimation of the weights, and bootstrap.

Table 4.1. Simulation results with n = 250; $\beta_1 = \beta_2 = 1$ and $\beta_3 = 0.5$. The true value for $\Lambda_{01}(1)$ = $\Lambda_{02}(1)$ = 1.264, and $\Lambda_{03}(1)$ = 2.528.

	Usual Markov Model					General Markov Model				
σ^2	Par	Estimate	SD	model/boot SE	model/boot CP	Par	Estimate	SD	SE	CP
0	β_1	0.995	0.211	0.214 / 0.219	95.6% / 95.4%	β_1	1.063	0.197	0.201	94.8%
	β_2	1.005	0.206	0.203 / 0.210	94.8% / 95.1%	β_2	1.042	0.201	0.203	94.8%
	β_3	0.503	0.268	0.263 / 0.260	94.6% / 94.8%	β_3	0.497	0.213	0.211	95.3%
	$\Lambda_{01}(1)$	1.219	0.264	0.259	94.8%	$\Lambda_{01}(1)$	1.323	0.275	0.280	94.6%
	$\Lambda_{02}(1)$	1.206	0.285	0.281	94.8%	$\Lambda_{02}(1)$	1.315	0.293	0.289	95.3%
	$\Lambda_{03}(1)$	2.470	0.484	0.491	96.1%	$\Lambda_{03}(1)$	2.472	0.367	0.365	95.6%
						σ^2	0.038	0.018	0.030	98.0%
0.5	β_1	0.778	0.198	0.196 / 0.199	80.9% / 81.3%	β_1	1.011	0.258	0.267	96.1%
	β_2	0.782	0.204	0.209 / 0.204	82.4% / 81.8%	β_2	1.005	0.261	0.267	96.1%
	β_3	0.215	0.218	0.218 / 0.213	79.6% / 78.9%	β_3	0.509	0.269	0.275	94.9%
	$\Lambda_{01}(1)$	1.096	0.168	0.166	77.7%	$\Lambda_{01}(1)$	1.292	0.367	0.364	94.9%
	$\Lambda_{02}(1)$	1.036	0.193	0.200	78.3%	$\Lambda_{02}(1)$	1.315	0.362	0.368	95.1%
	$\Lambda_{03}(1)$	2.749	0.406	0.403	83.5%	$\Lambda_{03}(1)$	2.460	0.518	0.521	95.6%
						σ^2	0.572	0.199	0.193	92.9%
1	β_1	0.670	0.210	0.202 / 0.205	66.2% / 65.9%	β_1	0.993	0.258	0.270	95.5%
	β_2	0.679	0.198	0.201 / 0.195	68.6% / 69.1%	β_2	0.992	0.272	0.262	94.9%
	β_3	0.104	0.243	0.239 / 0.240	60.0% / 60.4%	β_3	0.492	0.316	0.309	93.8%
	$\Lambda_{01}(1)$	0.984	0.172	0.177	69.1%	$\Lambda_{01}(1)$	1.290	0.395	0.394	94.5%
	$\Lambda_{02}(1)$	0.987	0.147	0.145	67.5%	$\Lambda_{02}(1)$	1.295	0.396	0.402	94.1%
	$\Lambda_{03}(1)$	3.010	0.548	0.549	71.8%	$\Lambda_{03}(1)$	2.459	0.603	0.595	95.7%
						σ^2	1.089	0.270	0.275	93.6%
2	β_1	0.561	0.201	0.205 / 0.202	41.8% / 41.7%	β_1	0.985	0.301	0.291	95.9%
	β_2	0.555	0.209	0.202 / 0.211	40.4% / 39.6%	β_2	0.989	0.303	0.295	95.7%
	β_3	0.003	0.233	0.226 / 0.229	33.2% / 34.0%	β_3	0.488	0.368	0.359	94.8%
	$\Lambda_{01}(1)$	0.920	0.134	0.128	19.4%	$\Lambda_{01}(1)$	1.233	0.330	0.333	94.3%
	$\Lambda_{02}(1)$	0.923	0.146	0.151	21.8%	$\Lambda_{02}(1)$	1.246	0.329	0.335	93.8%
	$\Lambda_{03}(1)$	3.785	0.615	0.610	11.5%	$\Lambda_{03}(1)$	2.513	0.583	0.590	96.6%
						σ^2	1.912	0.318	0.326	93.1%

Table 4.2. Simulation results with n = 500; $\beta_1 = \beta_2 = 1$ and $\beta_3 = 0.5$. The true value for $\Lambda_{01}(1) = \Lambda_{02}(1) = 1.264$, and $\Lambda_{03}(1) = 2.528$.

	Usual Markov Model					General Markov Model				
σ^2	Par	Estimate	SD	model/boot SE	model/boot CP	Par	Estimate	SD	SE	CP
0	β_1	1.003	0.147	0.147 / 0.146	95.0% / 96.0%	β_1	1.031	0.147	0.146	94.2%
	β_2	1.000	0.141	0.137 / 0.145	94.8% / 95.5%	β_2	1.040	0.145	0.147	95.5%
	β_3	0.499	0.149	0.153 / 0.151	94.6% / 95.2%	β_3	0.542	0.157	0.161	95.5%
	$\Lambda_{01}(1)$	1.233	0.210	0.202	95.4%	$\Lambda_{01}(1)$	1.226	0.200	0.194	94.1%
	$\Lambda_{02}(1)$	1.254	0.204	0.198	94.8%	$\Lambda_{02}(1)$	1.214	0.232	0.202	93.9%
	$\Lambda_{03}(1)$	2.465	0.344	0.336	94.5%	$\Lambda_{03}(1)$	2.544	0.331	0.339	94.4%
						σ^2	0.029	0.011	0.023	98.6%
0.5	β_1	0.762	0.141	0.143 / 0.141	71.2% / 70.0%	β_1	1.006	0.227	0.230	95.8%
	β_2	0.775	0.151	0.148 / 0.146	75.4% / 73.9%	β_2	0.997	0.229	0.233	96.1%
	β_3	0.219	0.158	0.160 / 0.158	68.0% / 66.8%	β_3	0.496	0.211	0.202	94.4%
	$\Lambda_{01}(1)$	1.183	0.138	0.130	69.4%	$\Lambda_{01}(1)$	1.252	0.302	0.293	94.6%
	$\Lambda_{02}(1)$	1.178	0.146	0.139	68.6%	$\Lambda_{02}(1)$	1.249	0.295	0.292	94.8%
	$\Lambda_{03}(1)$	2.734	0.361	0.356	72.1%	$\Lambda_{03}(1)$	2.485	0.501	0.489	95.2%
						σ^2	0.566	0.179	0.186	93.3%
1	β_1	0.667	0.146	0.137 / 0.143	55.2% / 56.4%	β_1	1.000	0.209	0.202	94.4%
	β_2	0.661	0.142	0.150 / 0.143	59.4% / 56.3%	β_2	0.998	0.211	0.202	95.2%
	β_3	0.105	0.153	0.154 / 0.153	47.2% / 49.4%	β_3	0.498	0.223	0.216	94.8%
	$\Lambda_{01}(1)$	1.018	0.124	0.123	56.7%	$\Lambda_{01}(1)$	1.283	0.273	0.278	96.1%
	$\Lambda_{02}(1)$	1.035	0.126	0.125	52.8%	$\Lambda_{02}(1)$	1.289	0.269	0.275	95.5%
	$\Lambda_{03}(1)$	2.868	0.441	0.435	62.8%	$\Lambda_{03}(1)$	2.475	0.511	0.499	94.7%
						σ^2	1.063	0.189	0.184	93.9%
2	β_1	0.563	0.149	0.142 / 0.144	33.8% / 35.2%	β_1	1.009	0.268	0.273	95.6%
	β_2	0.550	0.149	0.147 / 0.144	34.2% / 34.4%	β_2	1.007	0.271	0.276	95.6%
	β_3	0.005	0.165	0.167 / 0.159	14.6% / 13.8%	β_3	0.492	0.291	0.303	94.4%
	$\Lambda_{01}(1)$	0.920	0.104	0.099	10.8%	$\Lambda_{01}(1)$	1.244	0.302	0.300	94.6%
	$\Lambda_{02}(1)$	0.933	0.111	0.108	12.4%	$\Lambda_{02}(1)$	1.250	0.306	0.301	95.2%
	$\Lambda_{03}(1)$	3.721	0.557	0.551	9.3%	$\Lambda_{03}(1)$	2.479	0.499	0.506	94.9%
						σ^2	1.924	0.255	0.252	93.0%

When $\sigma^2 = 0$, we see that the estimation under the usual Markov model is nearly unbiased, in particular for the larger sample size $n = 500$, and the coverage of the confidence intervals (CI) based on the normal approximation is very close to the nominal level. We note that the margin of error using 500 simulation runs to estimate the coverage of 95% CI's is 0.019, so that the range of coverage probability (CP) should be mostly within 93.1% to 96.9%. We also see that when $\sigma^2 = 0$, the estimation under the general Markov mode performed well for β_j and $\Lambda_{0j}(01)$, $j = 1,2,3$. However, the mean of the estimated standard error of σ^2 is much higher than the empirical standard deviation, and the CI overcovers. We note that this is the boundary cases considered in [68], where the asymptotical distribution is no longer normal.

When $\sigma^2 > 0$, we see that our estimator under the general Markov model is quite accurate for even the smaller sample size $n = 250$, the SEs are close to the sample SD and the coverage probabilities are good. The estimates under the usual Markov model is obviously biased with poor coverage of the CI's when $\sigma^2 > 0$.

Finally, we note that the variances of the estimators are generally larger under the general Markov, as more parameter is estimated.

4.7 Application to HAAS study

For this analysis, we are interested in the effect of mid-life alcohol exposure on cognitive impairment as well as death, which are semi-competing risks. In the HHP-HAAS study, alcohol consumption was assessed by self-report and translated into units of drinks per month. Estimates of the total ethanol intake from reported drinking patterns were calculated as ounces per month for beer, liquor, wine, and sake using algorithms based on average unit sizes and usual alcohol percentages. The alcohol consumption was then dichotomized into light drinking ($\leq$30.1 oz/month) vs heavy drinking ($>$30.1 oz/month). The "mid-life" alcohol exposure was collected during the HHP study between 1965-73. The Heavy Drinking group consisted of individuals who had heavy drinking at one point during mid-life, and the Light Drinking those

who never had heavy drinking during mid-life. Cognitive impairment was based on scores from the Cognitive Assessment and Screening Instrument (CASI), where a score below 74 was considered a moderate impairment (MI).

The confounders were decided by literature review and clinical experiences, as well as availability of the data. Literatures show that vital data such as blood pressure and heart rate are associated with drinking habits, as well as the cognitive health. Meanwhile, demographic data such as age, years of education, are also related to cognitive impairment and drinking habits. The Apolipoprotein E is the first identified genetic susceptibility factor for sporadic AD. Towards understanding determinants of cognitive impairment and factors associated with drinking habits, the final set of baseline confounders are baseline CASI score, systolic blood pressure, heart rate, Apolipoprotein E genotype positive, years of education and baseline age. We only include participants with normal cognitive function (CASI $\geq$ 74) at baseline, and after excluding missing values for exposure and confounders, we have 1881 participants in total.

Table 4.3. Event counts by heavy versus light alcohol drinking in the HAAS data

	Heavy Drinking ($n = 491$)	Light Drinking ($n = 1390$)	Overall ($n = 1881$)
Event			
censor	84 (17.1%)	273 (20.9%)	357 (19.0%)
death without moderate impairment	163 (33.2%)	474 (34.1%)	637 (33.9%)
moderate impairment then censor	57 (11.6%)	204 (14.7%)	261 (13.9%)
moderate impairment then death	187 (38.1%)	439 (31.6%)	626 (33.3%)

Since HAAS is a long-term epidemiology study, lost to follow-up occurs at every exam visit. On the other hand, death certificates were obtained for many participants, even after lost to follow-up. For this reason, we needed to properly define the death for the semi-competing risks data. If the death date is after the participant's recorded last visit date from the study, we consider this participant lost to follow-up. More details of data pre-processing can be found in [71].

Propensity scores (PS) were calculated using R package *twang* (Toolkit for Weighting and Analysis of Nonequivalent Groups), which estimates the PS using boosted regression as

the predicted probability of being heavy versus light drinking, conditional on the measured baseline confounders. Before applying the IPW approach to the multi-state model, we obtained stabilized weights and trimmed them within (0.1, 10). In Supplementary Materials we show the PS histograms in the heavy and light drinking groups as a check of the positivity assumption, where the PS distributions are seen to be bounded away from zero and one. We also plot the standardized mean difference (SMD) to check the balance of each confounder before and after weighting, where the SMD's of all the confounders are within the interval [-0.1, 0.1] after weighting.

We apply our proposed methods to the HAAS data. We first fit the usual Markov structural model and the results are in the top half of Table 4.4. We see that the transition rates to moderate impairment or death without moderate impairment are significantly higher in the heavy drinking group compared to the light drinking group. But we don't see a significant difference in the transition rates to death after moderate impairment.

We then fit the general Markov structural model and the results are in the bottom half of Table 4.4. The convergence plot of the parameters and the likelihood during the weighted EM algorithm are provided in the Supplement Materials, where we stopped at 168 EM steps for the final results. Compared to the results under the usual Markov model, the magnitude of all three estimated effects are further away from the null, and all three transition rates are significantly higher in the heavy drinking group than the light drinking group. The phenomenon of more significant and away-from-the-null regression effects after accounting for the frailty is known in the literature under the Cox model.[8]

Finally, we estimate the causal risk contrasts under the structural models. For illustration purposes we fix $t_1 = 8$ years in $F_{12}(t_1, t_2; a)$ and $F_{12}(t_1, t_2|b; a)$; that is, the cumulative incidence rate of death following MI by 8 years. We show the estimated risk curves in Figure 4.4 first row under the usual Markov model, and the risk contrasts in Table 4.5 for heavy versus light drinking. It is seen that the risk contrasts for the two competing events, MI and death without MI, are significantly different from the null at 5 and 10 years, but not so at 15 and 20 years. The

Table 4.4. Parameter estimates of heavy ($a = 1$) versus light ($a = 0$) drinking using the HAAS data

	Estimate	SE	Hazard Ratio (HR)	95% CI of HR
The usual Markov model				
moderate impairment	0.202	0.079	1.224	[1.047, 1.431]*
death without moderate impairment	0.285	0.094	1.331	[1.105, 1.603]*
death after moderate impairment	0.152	0.089	1.164	[0.975, 1.388]
The general Markov model				
moderate impairment	0.264	0.072	1.302	[1.131, 1.499]*
death without moderate impairment	0.359	0.103	1.431	[1.170, 1.752]*
death after moderate impairment	0.274	0.109	1.315	[1.062, 1.628]*
σ^2	0.752	0.107	-	[0.542, 0.962]

* indicates statistical significance at $\alpha = 0.05$ two-sided

risk contrasts for death following MI by 8 years are not significantly different from the null at 10, 15 or 20 years under the usual Markov model.

We also show the predicted conditional risk curves at different b values $(0, \pm\hat{\sigma}, \pm 2\hat{\sigma})$ in Figure 4.4, rows 2-6. In Figure 4.5 we plot the IRD and IRR at 10 years with 95% PI's of 100 participants from every percentile of the predicted b values. We note the different significance results for IRD and IRR: the IRD tends to be significantly different from the null for b values closer to zero, while the IRR tends to be significantly different from the null for negative b values. This appears to be generally the case for all three outcomes: MI, death without MI, and death following MI by 8 years. More discussion will follow in the next section.

Table 4.5. Estimated risk difference (RD) and risk ratio (RR) under the usual Markov model for moderate impairment (MI), death, and death following MI by $t_1 = 8$ years.

	Time	RD (95% CI)	RR (95% CI)
MI	5	0.026 (0.009, 0.043)*	1.203 (1.073, 1.364)*
	10	0.044 (0.010, 0.080)*	1.142 (1.031, 1.265)*
	15	0.036 (−0.003, 0.078)	1.085 (0.991, 1.189)
	20	−0.006 (−0.053, 0.047)	0.989 (0.909, 1.086)
Death	5	0.014 (0.004, 0.026)*	1.280 (1.071, 1.522)*
	10	0.042 (0.008, 0.077)*	1.203 (1.033, 1.396)*
	15	0.042 (−0.005, 0.084)	1.130 (0.986, 1.279)
	20	0.024 (−0.028, 0.073)	1.061 (0.931, 1.189)
Death after MI	10	0.036 (−0.007, 0.081)	1.136 (0.973, 1.304)
	15	0.052 (−0.011, 0.107)	1.071 (0.985, 1.151)
	20	0.014 (−0.002, 0.030)	1.014 (0.998, 1.032)

* indicates statistical significance at $\alpha = 0.05$ two-sided.

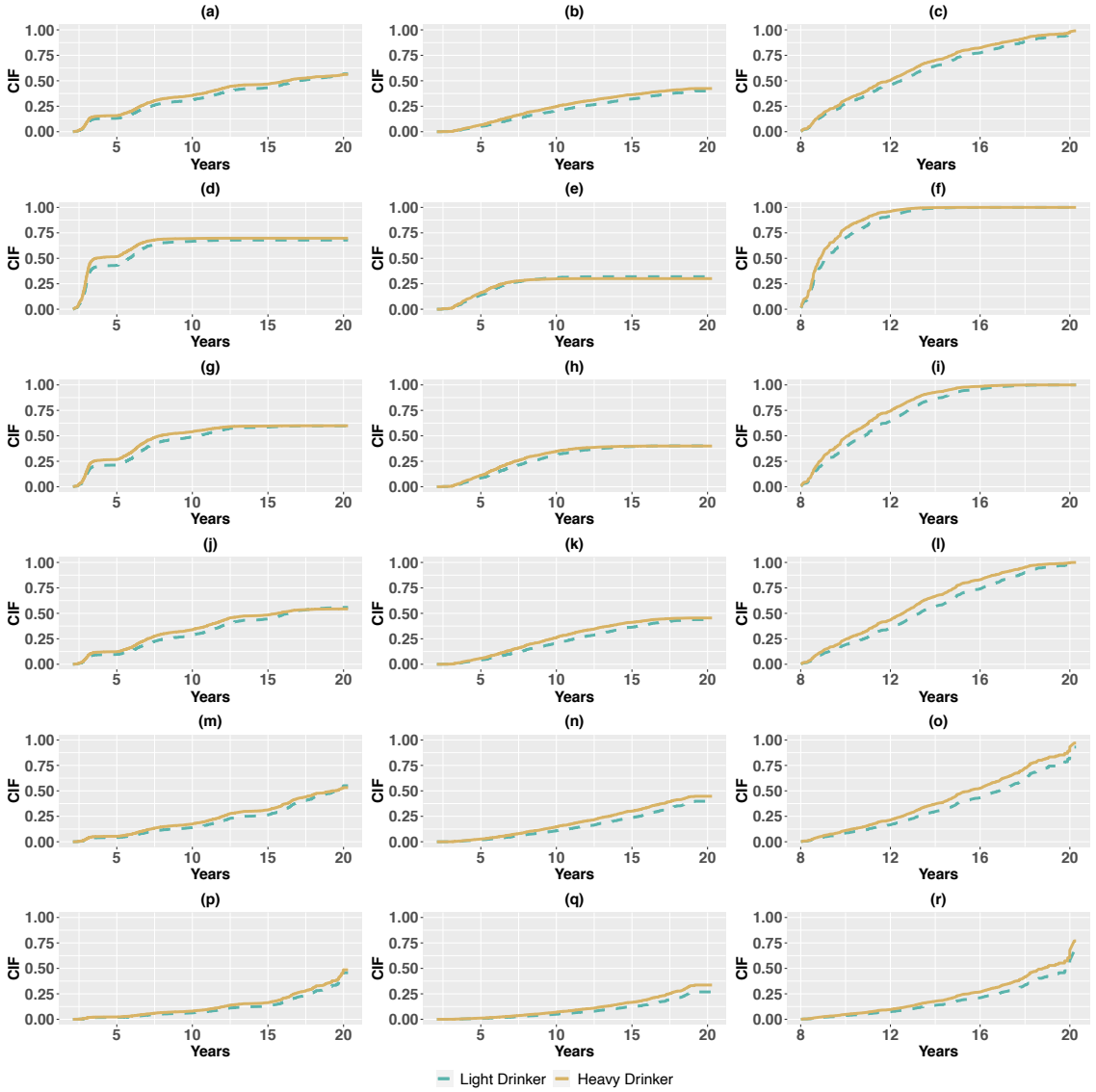

Figure 4.4. Risk plots for HAAS data under the usual Markov model, row 1; and conditional risk plots under the general Markov model, rows 2-6, for $b = 2\hat{\sigma}(1.734)$, $\hat{\sigma}(0.867)$, 0, $-\hat{\sigma}(-0.867)$ and $-2\hat{\sigma}(-1.734)$, respectively. The columns from left to right are: moderate impairment (MI), death without MI, and death following MI by $t_1 = 8$ years.

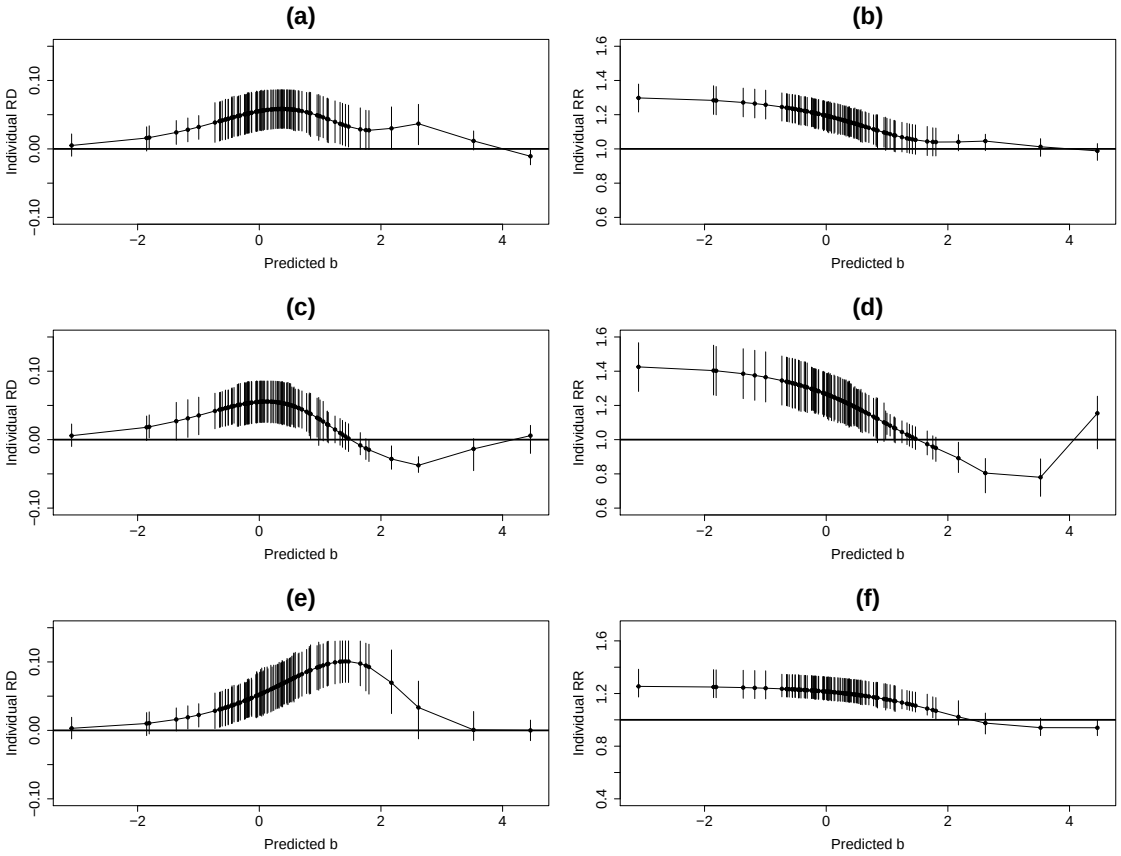

Figure 4.5. Individual risk difference (left) and individual risk ratio (right) at 10 years with 95% prediction intervals, for 100 participants of the HAAS study at every percentile of the predicted b's. Top row: moderate impairment (MI); middle row: death without MI; bottom row: death following MI by $t_1 = 8$ years.

4.8 Discussion

In this paper we applied the three-state illness-death model to observational data using the potential outcomes framework. Inverse probability of treatment weighting is used to fit these structural models. Under the Cox model formulation, typical software used to fit the Cox regression model can be used to fit the usual Markov model in the absence of frailty. With the frailty term under the general Markov model, a weighted EM algorithm is developed and its convergence property studied. The simulation studies showed the good performance of our proposed methods.

For applications in practice, we have defined cumulative risk based causal contrasts and illustrated their use. Under the general Markov model with frailty, these give rise to individual risk contrasts IRD and IRR. This is consistent with the random effects modeling formulation, where individual trajectories, for example, from longitudinal data can be estimated and predicted. We have extended this feature to the causal inference setting, when the individual heterogeneity is modeled using random effects. It might also be of some interest to compare the IRD and IRR to the RD and RR under the usual Markov model without frailty, and note some similarity between the first and the fourth row of Figure 4.4, where the random effect b is set to its mean value of zero. We note that these two sets of contrasts are not the same, especially since the Cox model is not collapsible; and the interpretations are different for these two sets of contrasts.

Semi-competing risks data have recently been considered under the mediation setup with the non-terminal event as a mediator. [20, 69] Our multi-state structural models instead consider the total effect of the exposure on all three outcomes: non-terminal event, and terminal event with and without non-terminal event.

For future work, since the IPW estimator is biased if the propensity score model is misspecified, an augmented IPW (AIPW) estimator with doubly robust properties can protect against such model misspecification. It would also allow us to apply machine learning or nonparametric methods to the propensity score model. [43] and [55] have already developed

the AIPW estimator for the marginal structural Cox model, and it is nature to extend their work for the models in this paper. This is currently under investigation. Another future direction is to develop sensitivity analysis approaches for various assumptions including unmeasured confounding as well as modeling assumptions that are used.

The R codes developed in this work have been implemented in the R package *semicmprskcoxmsm* that is publicly available on CRAN.

Acknowledgement

This research was partially supported by NIH/NIA grant R03 AG062432. We thank Dr. Andrew Ying for discussion regarding Lemma 1.

Chapter 4, in full, has been submitted for publication as "Zhang, Yiran; Ronghui Xu. *Marginal Structural Illness-Death Models for Semi-Competing Risks Data*, submitted to Statistics in Medicine". The dissertation author was the primary author on this paper.

Chapter 5

Augmented IPW for Illness-Death Usual Markov Models

5.1 Introduction

MSM with IPW has been proved as a useful tool for estimating causal effect in causal inference for survival data [16]. However, the effect will be biased if the propensity scores (PS) are not correctly specified. In the observational studies, the relationship between confounders are complicated, and involves with lots of interactions. Under this circumstances, simply put the parametric model such as logistic regression model on PS will cause bias for estimating causal effect. To protect the possibility of misspecification of PS mode, the augmented IPW estimator were developed. Nowadays, as IPW estimator are widely used under different setting, AIPW estimator are also derived. For example, [47] proposed an AIPW estimator under the setting for point exposure and continuous outcome, they formed the problem for estimating causal effect using potential outcome framework as the missing data problem, and propose the locally and globally adaptive semiparametric efficient estimators. [48] further extended to the survival analysis with informative censoring, and proposed the augmented inverse probability of censoring weighted (AIPCW) estimator. [43] and [55] also derived the AIPW estimator under Cox MSM proportional hazards model, and used machine learning models for both PS model and outcome model. The AIPW estimators come with the IPW part and augmentation part, which generally need us to put the model on both the PS model and the conditional outcome

regression model. It has the doubly robust property that the ATE is unbiased estimated if either the propensity score model is correctly specified or the outcome regression is correctly specified, or both are correctly specified. Many papers have conducted intensive simulation studies to show the plausible properties of AIPW estimators. Moreover, as recent developing of machine learning models, and we can put the machine learning model to both parts of AIPW estimator, then the high-dimensional confounder space in some large-scale observational study can be handled.

From previous project, we have derived the IPW estimator under MSM for competing risks and semi-competing risks setting. However, because of the complexity of semi-competing risks data structure, few studies extends to the AIPW estimator. However, semi-competing risk data becomes more and more common in survival analysis, especially in Alzheimer's or HIV study, thus extending current causal inference methodology to semi-competing risks data is important, and will benefit for epidemiologists or clinical researchers. We have discussed literatures for causal analysis with semi-competing risks data in the previous Chapter, most of them focus on principle stratification or mediation analysis. For this paper, we extended the easy to understand illness-model, and derive the AIPW estimators under the MSM illness-death usual Markov setting. We gave the detailed derivation by starting from IPW estimating equation, and show how to derive the augmentation based on IPW scores. We discussed the possibilities and difficulties of putting the conditional model on outcome, and the next step.

5.2 Marginal Structure usual Markov illness-death model

Following [73], let A be the binary treatment assignment, i.e. $A \in 0,1$. Let C^a, T_1^a and T_2^a be the counterfactural censoring, non-terminal event time, and terminal event time under treatment a. If the subject fails before the non-terminal event occurs, then we define $T_1^a = \infty$. Let Z be the p-dimensional vector of covariates. We also let: $X_2^a = \min(T_2^a, C^a)$, $X_1^a = \min(T_1^a, \min(T_2^a, C^a))$ and the event indicator $\delta_1^a = I\left\{X_1^a = T_1^a\right\}$, $\delta_2^a = I\left\{X_2^a = T_2^a\right\}$. We also denote $\pi(Z) = P(A = 1|Z)$, often referred to as the propensity score.

5.2.1 The Usual Markov Illness–Death Model

Let $\lambda_1(t_1;a)$, $\lambda_2(t_2;a)$ and $\lambda_{12}(t_2|t_1;a)$ be the transition rates corresponding to the counterfactual states under the three-state model, $a = 0, 1$. That is,

$$\lambda_1(t_1) = \lim_{\Delta\to 0^+} \frac{P\left(T_1^a \in [t_1, t_1+\Delta) \mid T_1^a \geq t_1, T_2^a \geq t_1\right)}{\Delta}, \tag{5.1}$$

$$\lambda_2(t_2) = \lim_{\Delta\to 0^+} \frac{P\left(T_2^a \in [t_2, t_2+\Delta) \mid T_1^a \geq t_2, T_2^a \geq t_2\right)}{\Delta}, \tag{5.2}$$

$$\lambda_{12}(t_2 \mid t_1) = \lim_{\Delta\to 0^+} \frac{P\left(T_2^a \in [t_2, t_2+\Delta) \mid T_1^a = t_1, T_2^a \geq t_2\right)}{\Delta}. \tag{5.3}$$

[3] discussed about modeling each transition intensity by a Cox type proportional intensities regression model. Following the same idea, we can postulate the semi-parametric Cox models for these transition rates, which are also hazard functions [67, 3]. In particular, we consider the following *usual Markov illness-death structural model* from [67].

$$\lambda_1(t_1;a) = \lambda_{01}(t_1)e^{\beta_1 a},\ t_1 > 0; \tag{5.4}$$

$$\lambda_2(t_2;a) = \lambda_{02}(t_2)e^{\beta_2 a},\ t_2 > 0; \tag{5.5}$$

$$\lambda_{12}(t_2|t_1;a) = \lambda_{03}(t_2)e^{\beta_3 a},\ 0 < t_1 < t_2. \tag{5.6}$$

subsectionAssumptions

We assume the following, which are commonly used assumptions in order to identify the causal estimands to be defined later:

(I) Stable unit treatment value assumption (SUTVA): there is only one version of the potential outcomes and that there is no interference between subjects.

(II) Exchangeability: $(T_1^a, T_2^a) \perp A|Z$.

(III) Positivity: $\pi(\mathbf{Z}) > 0$.

(IV) Consistency: If $A = a$, then $T_1^a = T_1$, $T_2^a = T_2$, $C^a = C$.

(V) Non-informative censoring: $(T_1^a, T_2^a, Z) \perp C^a$, $C^a \perp Z$.

5.2.2 Martingale

Following [52], we define the counterfactual counting process: $N^a_{k\ell}(t)$: counting from state k to state ℓ and $Y^a_j(t)$: counterfactual at-risk process in state j under treatment a,

$$N^a_{01}(t) = \mathbb{1}\left(X^a_1 \leq t, \delta^a_1 = 1\right), \tag{5.7}$$

$$N^a_{02}(t) = \mathbb{1}\left(X^a_2 \leq t, \delta^a_1 = 0, \delta^a_2 = 1\right), \tag{5.8}$$

$$N^a_{12}(t) = \mathbb{1}\left(X^a_2 \leq t, \delta^a_1 = 1, \delta^a_2 = 1\right). \tag{5.9}$$

We also have counterfactual at risk process in state 0 and state 1,

$$Y^a_0(t) = \mathbb{1}\left(X^a_1 \geq t\right), \tag{5.10}$$

$$Y^a_1(t) = \mathbb{1}\left(X^a_2 \geq t \geq X^a_1\right). \tag{5.11}$$

Based on standard counting process theory, under non-informative censoring assumption, the counterfactual martingales are defined as following,

$$M^a_{01}(t) = N^a_{01}(t) - \int_0^t Y^a_0(u)\lambda_{01}(u)e^{\beta_1 a}du, \tag{5.12}$$

$$M^a_{02}(t) = N^a_{02}(t) - \int_0^t Y^a_0(u)\lambda_{02}(u)e^{\beta_2 a}du, \tag{5.13}$$

$$M^a_{12}(t) = N^a_{12}(t) - \int_0^t Y^a_1(u)\lambda_{03}(u)e^{\beta_3 a}du, \tag{5.14}$$

where $M_{k\ell}(t)$ are orthogonal local square integrable martingales with predictable variation processed given by $< M_{k\ell} > (t) = \int_0^t Y^a_k(u)\,\lambda_{0\ell}(u)\,e^{\beta_\ell a}du$, for $k\ell = \{01, 02, 12\}$.

For the observed data, assume a well-defined time zero, we let random variables T_1 and T_2 denote observed time to the non-terminal and the terminal event since time zero, respectively. If a subject does not experience the non-terminal event before the terminal event, we define $T_1 = +\infty$, [67, 12, 73]. If we denote C as observed censoring time, we then have the observed time

$X_2 = \min(T_2, C)$, $X_1 = \min(T_1, \min(T_2, C))$ and the observed event indicator $\delta_1 = I\{X_1 = T_1\}$, $\delta_2 = I\{X_2 = T_2\}$. We also denote the joint density of T_1 and T_2 as $f(t_1, t_2)$ in the upper wedge $0 < t_1 \leq t_2$, and the density of T_2 along the line $t_1 = +\infty$ as $f_\infty(t_2)$ for $t_2 > 0$. We also denote the bivariate survival function of T_1 and T_2 in the upper wedge as $\tilde{S}(t_1, t_2)$. Based on the observed data, we further define the observed counting process,

$$N_{01}(t) = \mathbb{1}(X_1 \leq t, \delta_1 = 1), \tag{5.15}$$

$$N_{02}(t) = \mathbb{1}(X_2 \leq t, \delta_1 = 0, \delta_2 = 1), \tag{5.16}$$

$$N_{12}(t) = \mathbb{1}(X_2 \leq t, \delta_1 = 1, \delta_2 = 1), \tag{5.17}$$

and

$$Y_0(t) = \mathbb{1}(X_1 \geq t), \tag{5.18}$$

$$Y_1(t) = \mathbb{1}(X_2 \geq t \geq X_1). \tag{5.19}$$

5.3 Augmented IPW Scores

We proved from [73] that β_j, $j = 1, 2, 3$ are variationally independent, and can be estimated separately by fitting three Cox models. Following [9], and motivating by the fact that $M_{k\ell}(t)$ is a zero-mean martingale process, we then obtain the full data scores,

$$U^F_{k\ell 1} = \sum_{a=0,1} dM^a_{k\ell}(u), \tag{5.20}$$

$$U^F_{k\ell 2} = \sum_{a=0,1} \int_0^\tau a dM^a_{k\ell}(u), \tag{5.21}$$

where $k\ell = \{01, 02, 12\}$.

To estimate β_j, $j = 1, 2, 3$, we use the Inverse Probability Weighting (IPW) approach to

create a pseudo-randomized sample. We denote $w = A/P(A=1|Z) + (1-A)/\{1-P(A=1|Z)\}$ as the IP weight, and obtain the IP weighted full-data scores:

$$U_{k\ell 1}^{IPW} = wdM_{k\ell}(t), \tag{5.22}$$

$$U_{k\ell 2}^{IPW} = \int_0^{\tau} wAdM_{k\ell}(t), \tag{5.23}$$

where $M_{k\ell}(t) = AM_{k\ell}^1(t) + (1-A)M_{k\ell}^0(t)$, $k\ell = \{01, 02, 12\}$. Noticed that because of consistency, we have

$$M_{01}(t) = N_{01}(t) - \int_0^t Y_0(u)\,\lambda_{01}(u)\,e^{\beta_1 A}du, \tag{5.24}$$

$$M_{02}(t) = N_{02}(t) - \int_0^t Y_0(u)\,\lambda_{02}(u)\,e^{\beta_2 A}du, \tag{5.25}$$

$$M_{12}(t) = N_{12}(t) - \int_0^t Y_1(u)\,\lambda_{03}(u)\,e^{\beta_3 A}du. \tag{5.26}$$

In [73], we studied the performance of IPW illness-death usual Markov models for estimating β_j, $j = 1,2,3$ when the weights w are known or consistently estimated. However, the propensity score model might be misspecified, and the IPW approach will be biased. To protect against possible misspecification of the propensity score model, we now augment the IPW scores to obtain a doubly robust score $U_{k\ell s}$, $s = 1,2$. Following [60] and [43] we obtain the following augmented IPW score:

$$U = U^{IPW} - \Pi\left(U^{IPW}|\Gamma_2\right), \tag{5.27}$$

where Γ_2 is the propensity score tangent space (also is called as augmentation space), and we denote $\Pi(q(\cdot)|\Gamma_2)$ as the projection of a function $q(\cdot)$ onto Γ_2 in the Hilbert space equipped with

covariance inner product. We have the following equations to describe $\Pi\left(U_{k\ell}^{IPW}|\Gamma_2\right)$:

$$\begin{aligned}\Pi\left(U_{k\ell 1}^{IPW}|\Gamma_2\right) &= wdM_{k\ell}(t) - wdE\{M_{k\ell}(t)|A,Z\} \\ &\quad + dE\{M_{k\ell}(t)|A=1,Z\} + dE\{M_{k\ell}(t)|A=0,Z\}, \qquad (5.28)\\ \Pi\left(U_{k\ell 2}^{IPW}|\Gamma_2\right) &= \int_0^{\tau} \left[wAdE\{M_{k\ell}(t)|A,Z\} - dE\{M_{k\ell}(t)|A=1,Z\}\right]. \qquad (5.29)\end{aligned}$$

We further plug (5.28) and (5.29) into (5.27), we obtain the **AIPW scores**:

$$\begin{aligned}U_{k\ell 1} &= wdM_{k\ell}(t) - wdE\{M_{k\ell}(t)|A,Z\} \\ &\quad + dE\{M_{k\ell}(t)|A=1,Z\} + dE\{M_{k\ell}(t)|A=0,Z\}, \qquad (5.30)\\ U_{k\ell 2} &= \int_0^{\tau} \left[wAdM_{k\ell}(t) - wAdE\{M_{k\ell}(t)|A,Z\} + dE\{M_{k\ell}(t)|A=1,Z\}\right], \qquad (5.31)\end{aligned}$$

where $k\ell = 01, 02, 12$.

Let $\lambda_1(t_1|A,Z)$, $\lambda_{12}(t_2|A,Z)$, $\lambda_2(t_2|A,Z)$ be the conditional transition hazards, conditional on A and Z, for non-terminal event, terminal event without non-terminal and terminal event following non-terminal event respectively. We then can define $\Lambda_1(t_1|A,Z)$, $\Lambda_2(t_2|A,Z)$, $\Lambda_{12}(t_2|A,Z)$ as the conditional transition cumulative hazards. We further denote $G(t|A,Z)$ as the conditional survivorship for censoring C, $S(t|A,Z)$ as the conditional overall survival function for T_1, T_2, and $S(t|A,Z) = \exp\{-\Lambda_1(t|A,Z) - \Lambda_2(t|A,Z)\}$. We also recall that $\Lambda_{0j}(t)$, $j = 1,2,3$ is the cumulative baseline hazard function from structural model (5.4) - (5.6). After the derivation as

given in the Appendix, we obtain

$$E\{M_{01}(t)|A,Z\} = \int_0^t G(u|A)S(u|A,Z)d\Lambda_1(u|A,Z) - \int_0^t G(u|A)S(u|A,Z)e^{\beta_1 A}d\Lambda_{01}(u), \tag{5.32}$$

$$E\{M_{02}(t)|A,Z\} = \int_0^t G(u|A)S(u|A,Z)d\Lambda_2(u|A,Z) - \int_0^t G(u|A)S(u|A,Z)e^{\beta_2 A}d\Lambda_{02}(u), \tag{5.33}$$

$$E\{M_{03}(t)|A,Z\} = \int_0^t G(u|A)dF_{12}(u|A,Z) - \int_0^t G(u|A)\left\{\tilde{S}(0,u|A,Z) - S(u|A,Z)\right\}e^{\beta_3 A}d\Lambda_{03}(u). \tag{5.34}$$

Recall we defined the propensity score $\pi(Z)$. We also denote

$$U_{01} = U_{01}(\beta_1, \Lambda_{01}; \pi, \Lambda_1, \Lambda_2, G),$$

$$U_{02} = U_{02}(\beta_2, \Lambda_{02}; \pi, \Lambda_1, \Lambda_2, G),$$

$$U_{12} = U_{12}(\beta_3, \Lambda_{03}; \pi, \Lambda_1, \Lambda_2, \Lambda_{12}, G),$$

where $\pi, \Lambda_1, \Lambda_2, \Lambda_{12}, G$ are the nuisance parameters in the model.

We also denote $\pi^o, \Lambda_1^o, \Lambda_2^o, \Lambda_{12}^o, G^o$ as the true value of those quantities, we then have the following doubly robust property for score functions $U_{k\ell} = [U_{k\ell 1}, U_{k\ell 2}]^\intercal$, $k\ell \in \{01, 02, 12\}$ with respect to π and $\Lambda_1, \Lambda_2, \Lambda_{12}$.

Theorem 1. ***Doubly robust property****: Under Assumption (I) - (V):*

(1) $E\{U_{01}(\beta_1^o, \Lambda_{01}^o; \pi, \Lambda_1, \Lambda_2, G^o)\} = 0$, *if* $\pi = \pi^o$ *or* $\Lambda_1 = \Lambda_1^o, \Lambda_2 = \Lambda_2^o$;

(2) $E\{U_{02}(\beta_2^o, \Lambda_{02}^o; \pi, \Lambda_1, \Lambda_2, G^o)\} = 0$, *if* $\pi = \pi^o$ *or* $\Lambda_1 = \Lambda_1^o, \Lambda_2 = \Lambda_2^o$;

(3) $E\{U_{12}(\beta_3^o, \Lambda_{03}^o; \pi, \Lambda_1, \Lambda_2, \Lambda_{12}, G^o)\} = 0$, *if* $\pi = \pi^o$ *or*
$\Lambda_1 = \Lambda_1^o, \Lambda_2 = \Lambda_2^o, \Lambda_{12} = \Lambda_{12}^o$.

Details of proof are shown in the Appendix.

Given the i.i.d observed data $(X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i}, A_i, Z_i)$, $i = 1, 2, ..., n$, we solve

$$\frac{1}{n}\sum_{i=1}^{n} U_{kl1,i} = 0, \tag{5.35}$$

$$\frac{1}{n}\sum_{i=1}^{n} U_{kl2,i} = 0, \tag{5.36}$$

where $k\ell = 01, 02, 12$, to estimate β_j and Λ_{0j}, $j = 1, 2, 3$, respectively.

Following [43], we denote $h(\cdot|A = a, Z_i) = h_i(\cdot, a)$ for any function with the form $h(\cdot|A = a, Z)$. For example, we have: $S_i(t, a) = S(t|A = a, Z_i)$. We also denote $R_{0i}(t, S, G) = Y_{0i}(t) - G_i(t, A_i)S_i(t, A_i)$ and $R_{1i}(t, S, G) = Y_{1i}(t) - G_i(t, A_i)\{S_i(0, t, A_i) - S_i(t, A_i)\}$. Then plugging (5.32) - (5.34) into (5.30) and (5.31), we have the following AIPW scores

$$\begin{aligned} U_{011,i} = {} & w\{dN_{01}(u) - G(u,A)S(u,A)d\Lambda_1(u,A)\} + \sum_{a=0,1} G(u,a)S(u,a)d\Lambda_1(u,a) \\ & - d\left\{ we^{\beta_1 A}Y_0(u) - wG(u,A)S(u,A)e^{\beta_1 A} + \sum_{a=0,1} G(u,a)S(u,a)e^{\beta_1 a} \right\}\Lambda_{01}(t), \end{aligned} \tag{5.37}$$

$$\begin{aligned} U_{012,i} = \int_0^t \Bigg[& wA\{dN_{01}(u) - G(u,A)S(u,A)d\Lambda_1(u,A)\} + G(u,1)S(u,1)d\Lambda_1(u,1) \\ & - d\left\{ wAe^{\beta_1 A}Y_0(u) - wAG(u,A)S(u,A)e^{\beta_1 A} + G(u,1)S(u,1)e^{\beta_1} \right\}\Lambda_{01}(t) \Bigg] \end{aligned} \tag{5.38}$$

to estimate β_1 and Λ_{01};

$$\begin{aligned} U_{021,i} = {} & w\{dN_{02}(u) - G(u,A)S(u,A)d\Lambda_2(u,A)\} + \sum_{a=0,1} G(u,a)S(u,a)d\Lambda_2(u,a) \\ & - d\left\{ we^{\beta_2 A}Y_0(u) - wG(u,A)S(u,A)e^{\beta_2 A} + \sum_{a=0,1} G(u,a)S(u,a)e^{\beta_2 a} \right\}\Lambda_{02}(t), \end{aligned} \tag{5.39}$$

$$\begin{aligned} U_{022,i} = \int_0^t \Bigg[& wA\{dN_{02}(u) - G(u,A)S(u,A)d\Lambda_2(u,A)\} + G(u,1)S(u,1)d\Lambda_2(u,1) \\ & - d\left\{ wAe^{\beta_2 A}Y_0(u) - wAG(u,A)S(u,A)e^{\beta_2 A} + G(u,1)S(u,1)e^{\beta_2} \right\}\Lambda_{02}(t) \Bigg] \end{aligned} \tag{5.40}$$

to estimate β_2 and Λ_{02}; and

$$\begin{aligned} U_{121,i} = w\{dN_{12}(u) - G(u,A)dS_{12}(u,A)\} + \sum_{a=0,1} G(u,a)dS_{12}(u,a) \\ - d\Bigg[we^{\beta_3 A}Y_1(u) - wG(u,A)\left\{\tilde{S}(0,u,A) - S(u,A)\right\}e^{\beta_3 A} \\ + \sum_{a=0,1} G(u,a)\left\{\tilde{S}(0,u,A) - S(u,A)\right\}e^{\beta_2 a}\Bigg]\Lambda_{02}(t), \end{aligned} \tag{5.41}$$

$$\begin{aligned} U_{122,i} = \int_0^t \Bigg(wA\{dN_{12}(u) - G(u,A)dS_{12}(u,A)\} + G(u,1)dS_{12}(u,A) \\ - d\Bigg[wAe^{\beta_3 A}Y_1(u) - wAG(u,A)\left\{\tilde{S}(0,u,A) - S(u,A)\right\}e^{\beta_3 A} \\ + G(u,1)\left\{\tilde{S}(0,u,1) - S(u,1)\right\}e^{\beta_3}\Bigg]\Lambda_{03}(t)\Bigg) \end{aligned} \tag{5.42}$$

to estimate β_3 and Λ_{03}.

For $\ell = 0, 1$, let

$$\begin{aligned} \mathcal{S}_{01}^{(\ell)}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G) = \frac{1}{n}\sum_{i=1}^{n}\{\, w_i \exp(\beta_1 A_i)A_i^{\ell}R_{0i}(t,S,G) \\ + \sum_{a=0,1} a^{\ell}\exp(\beta_1 a)G_i(t,a)S_i(t,a)\,\}, \end{aligned} \tag{5.43}$$

$$\begin{aligned} \mathcal{S}_{02}^{(\ell)}(t;\beta_2,\pi,\Lambda_1,\Lambda_2,G) = \frac{1}{n}\sum_{i=1}^{n}\{\, w_i \exp(\beta_2 A_i)A_i^{\ell}R_{0i}(t,S,G) \\ + \sum_{a=0,1} a^{\ell}\exp(\beta_2 a)G_i(t,a)S_i(t,a)\,\}, \end{aligned} \tag{5.44}$$

$$\begin{aligned} \mathcal{S}_{12}^{(\ell)}(t;\beta_3,\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[w_i \exp(\beta_3 A_i)A_i^{\ell}R_{1i}(t,S,G) \\ + \sum_{a=0,1} a^{\ell}\exp(\beta_3 a)G_i(t,a)\left\{\tilde{S}_i(0,t,a) - S_i(t,a)\right\}\Bigg]. \end{aligned} \tag{5.45}$$

We first solve (5.35) to obtain $\Lambda_{0j},\ j=1,2,3$. For any $t\in[0,\tau]$, we have the following:

$$\begin{aligned}\Lambda_{01}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G) = \int_0^t \sum_{i=1}^n [\, w_i \{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\ - \sum_{a=0,1} G(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\,] \\ \times \frac{1}{n\cdot \mathcal{S}_{01}^{(0)}(u;\beta_1,\pi,\Lambda_1,\Lambda_2,G)}, \qquad (5.46)\end{aligned}$$

$$\begin{aligned}\Lambda_{02}(t;\beta_2,\pi,\Lambda_1,\Lambda_2,G) = \int_0^t \sum_{i=1}^n [\, w_i \{dN_{02i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{2i}(u,A_i)\} \\ - \sum_{a=0,1} G(u,a)S_i(u,a)d\Lambda_{2i}(u,a)\,] \\ \times \frac{1}{n\cdot \mathcal{S}_{02}^{(0)}(u;\beta_1,\pi,\Lambda_1,\Lambda_2,G)}, \qquad (5.47)\end{aligned}$$

$$\begin{aligned}\Lambda_{03}(t;\beta_3,\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G) = \int_0^t \sum_{i=1}^n [\, w_i \{dN_{12i}(u) - G_i(u,A_i)dS_{12i}(u,A_i)\} \\ - \sum_{a=0,1} G(u,a)dS_{12i}(u,a)\,] \\ \times \times \frac{1}{n\cdot \mathcal{S}_{12}^{(0)}(t;\beta_3,\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G)}. \qquad (5.48)\end{aligned}$$

Plugging $\Lambda_{0j},\ j=1,2,3$, back into (5.36), we obtain the following AIPW scores for

estimating β_j, $j = 1,2,3$:

$$U_{012}^{(n)}(\beta_1;\pi,\Lambda_1,\Lambda_2,G) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^\tau w_i\left\{A_i - \bar{\mathcal{A}}_{01}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G)\right\} \times \{dN_{01i}(t) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} + \int_0^\tau \sum_{a=0,1}\left\{a - \bar{\mathcal{A}}_{01}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G)\right\} \times G_i(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\Bigg], \tag{5.49}$$

$$U_{022}^{(n)}(\beta_2;\pi,\Lambda_1,\Lambda_2,G) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^\tau w_i\left\{A_i - \bar{\mathcal{A}}_{02}(t;\beta_2,\pi,\Lambda_1,\Lambda_2,G)\right\} \times \{dN_{02i}(t) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{2i}(u,A_i)\} + \int_0^\tau \sum_{a=0,1}\left\{a - \bar{\mathcal{A}}_{02}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G)\right\} \times G_i(u,a)S_i(u,a)d\Lambda_{2i}(u,a)\Bigg], \tag{5.50}$$

$$U_{122}^{(n)}(\beta_3;\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^\tau w_i\left\{A_i - \bar{\mathcal{A}}_{12}(t;\beta_3,\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G)\right\} \{dN_{12i}(t) - G_i(u,A_i)dS_{12i}(u,A_i)\} + \int_0^\tau \sum_{a=0,1}\left\{a - \bar{\mathcal{A}}_{12}(t;\beta_3,\pi,\Lambda_1,\Lambda_2,\Lambda_{12},G)\right\} \times G_i(u,a)dS_{12i}(u,a)\Bigg], \tag{5.51}$$

where $\bar{\mathcal{A}}_{kl} = \mathcal{S}_{kl}^{(1)}/\mathcal{S}_{kl}^{(0)}$, for $kl \in \{01,02,12\}$. Details of the derivation are shown in Appendix.

To estimate β_j, $j = 1,2,3$ using (5.49) - (5.51) respectively, we first need to estimate π, Λ_1, Λ_2, Λ_{12}, and G in above equations. We can estimate G using nonparametric Kaplan-Meier estimator. In practice, the propensity score $\pi(\cdot)$ can be estimated from the data by either specifying a parametric model such as the logistic regression or using nonparametric methods such as boosted trees. For estimating $\Lambda_1(t|A,Z)$, $\Lambda_2(t|A,Z)$ and $\Lambda_{12}(t|A,Z)$, a model compatible

with the MSM (5.4) - (5.6) need to be used. For $\Lambda_1(t|A,Z)$ and $\Lambda_2(t|A,Z)$, [67] and [73] discussed that those are cause-specific cumulative hazard functions, nonparametric methods for cause-specific function under competing risks setting can be applied. [18] discussed using the LASSO under the cause-specific Cox proportional hazards model, and [18] discussed using random forest to estimate the cumulative incidence function (CIF), which can be used to estimate cause-specific cumulative hazard functions.

5.4 Simulation

We followed the simulation steps from [73] to generate confounding under a MSM usual Markov model:

- Generate $U_1 \sim U(0,1)$ and $U_2 \sim U(0,1)$;
- Generate confounder $Z = (Z_1, ..., Z_k)^\intercal$ using U_1 and U_2;
- Generate $A \sim$ Bernoulli(p_A), where p_A are generated using Z;
- Let $\lambda_{01}(t) = \lambda_{02}(t) = 2e^{-t}I(0 \leq t \leq 3) + 2e^{-3}I(t > 3)$ and $\lambda_{03}(t) = 2\lambda_{01}(t)$. Then with probability $P(T_1 = \infty)$

$$T_2 = \Lambda_{01}^{-1}\left\{-\frac{\log(U_1)}{\exp(\beta_1 A) + \exp(\beta_2 A)}\right\};$$

and with probability $1 - P(T_1 = \infty)$,

$$T_1 = \Lambda_{01}^{-1}\left\{-\frac{\log(U_1)}{\exp(\beta_1 A) + \exp(\beta_2 A)}\right\},\ T_2 = \Lambda_{01}^{-1}\left\{-\frac{\log(U_2)}{2\exp(\beta_3 A)} + \Lambda_{01}(t_1)\right\}.$$

- Generate Censoring time $C(a) \sim 0.3\exp(0.2a + 0.2)$, which leads to an average censoring rate around 15%.

We consider the following two scenarios:

Table 5.1. Scenarios 1 and 2 of the simulation to generate confounders and treatment assignment.

Scenario	Confounders and Treatment generating mechanism
1 PS: Logistic	$Z_j = U_1 + U_2 + \varepsilon_j$, where $\varepsilon_j \sim N(0,1)$, for $j = 1,2,3$ $p_A = \text{logit}^{-1}(0.5 + 0.1Z_1 - 0.1Z_2 - 0.2Z_3)$
2 PS: Soft Partition	$Z_j = (U_1 + U_2 + \varepsilon_j)/\sqrt{2}$, where $\varepsilon_j \sim N(0,1)$, for $j = 1,2,..,6$ $p_A = 0.8 * I(\sum_{i=1}^{6} Z_i^2 < \chi_{0.5,6}) + 0.2 * I(\sum_{i=1}^{6} Z_i^2 \geq \chi_{0.5,6})$

We fix $\beta_1 = \beta_2 = 1$, $\beta_3 = 0.5$. In the simulation, the propensity score is estimated using logistic regression and generalized boosted model (R package *twang*). Conditional hazard $\Lambda_1(t|A,Z)$, $\Lambda_2(t|A,Z)$ and $\Lambda_{12}(t|A,Z)$ are estimated using semi-parametric Cox proportional hazards model with confounders Z and treatment A as covariates, without interaction terms, as well as using LASSO Cox proportional hazards model (R package *glmnet*) with covariates that are generate from Z with quadratic and cubic terms, as well as two-way and three-way interactions, and the treatment A. We used the nonparametric Kaplan-Meier estimator for the censoring distribution $G(t|A)$. To solve the equation (5.49) and (5.50), we use grid search and search between $(0.8, 1.3)$ with every 0.001. To solve equation (5.51), we use grid search to search between $(0.3, 0.8)$ with every 0.001. Standard error are estimated by sandwich formula for IPW estimator and bootstraps for AIPW estimators.

We run 500 simulations with a sample size of 500 each, and compared the IPW with AIPW methods.

Table 5.2. Scenario 1, Treatments are generated from logistic regression without interaction. The margin of error for 95% CI with 500 runs is: 0.019, which the range of coverage probability (CP) should be within 93.1% to 96.9%

		IPW					AIPW			
		Estimate	SD	SE	CP		Estimate	SD	SE	CP
PS	**Parameter**					**Outcome**				
Logistic	β_1	1.003	0.154	0.152	95.4%	Cox	0.998	0.135	0.131	95.8%
						Cox Lasso	1.002	0.142	0.149	95.7%
	β_2	1.000	0.145	0.139	94.8%	Cox	0.994	0.128	0.130	95.8%
						Cox Lasso	0.995	0.133	0.137	94.9%
	β_3	0.499	0.151	0.149	95.2%	Cox	0.502	0.145	0.139	94.7%
twang	β_1	0.998	0.148	0.150	95.6%	Cox	0.996	0.141	0.147	94.8%
						Cox Lasso	0.993	0.153	0.150	95.5%
	β_2	0.998	0.139	0.136	94.6%	Cox	1.000	0.135	0.140	95.6%
						Cox Lasso	1.004	0.138	0.147	94.9%
	β_3	0.503	0.150	0.153	95.4%	Cox	0.495	0.156	0.162	96.2%

Table 5.3. Scenario 2, Treatments are generated from binary indicator functions that correlated with Z. The margin of error for 95% CI with 500 runs is: 0.019, which the range of coverage probability (CP) should be within 93.1% to 96.9%

		IPW					AIPW			
		Estimate	SD	SE	CP		Estimate	SD	SE	CP
PS	**Parameter**					**Outcome**				
Logistic	β_1	1.093	0.160	0.167	95.8%	Cox	0.988	0.124	0.119	94.1%
						Cox Lasso	0.991	0.133	0.129	94.5%
	β_2	1.087	0.153	0.167	96.2%	Cox	0.979	0.129	0.124	94.1%
						Cox Lasso	0.983	0.136	0.133	94.5%
	β_3	0.483	0.180	0.188	96.9%	Cox	0.511	0.137	0.141	93.9%
						Cox Lasso	0.512	0.151	0.148	93.4%
twang	β_1	0.788	0.149	0.148	34.8%	Cox	0.958	0.146	0.147	94.8%
						Cox Lasso	0.885	0.158	0.150	91.5%
	β_2	0.785	0.150	0.148	34.1%	Cox	0.969	0.142	0.140	95.6%
						Cox Lasso	0.871	0.154	0.147	92.1%
	β_3	0.265	0.183	0.186	38.3%	Cox	0.477	0.188	0.182	95.4%
						Cox Lasso	0.411	0.191	0.186	90.3%

Results from Scenario 1 show that when propensity score model are correctly specified, both IPW and AIPW estimators are performed well. On the other hand, results from Scenario 2 show that when PS model is not correctly specified, the IPW estimators still perform relatively good when using logistic regression for PS model and the performance further get improved in AIPW estimators. The performance of using *twang* for PS model generally is bad. However, when using semi-parametric Cox proportional hazards model for conditional hazards, the performance

of AIPW estimators get improved compare to IPW estimators. Simulation studies for using *twang* as well as LASSO Cox are still under investigation, and more simulation studies will be performed in the future.

Meanwhile, [18] discussed using random forest to estimate the cumulative incidence function (CIF). By definition of CIF, we have,

$$F_1(t|A,Z) = \int_0^t S(u|A,Z) d\Lambda_1(u|A,Z), \tag{5.52}$$

$$F_2(t|A,Z) = \int_0^t S(u|A,Z) d\Lambda_2(u|A,Z). \tag{5.53}$$

Then

$$S(t|A,Z) = 1 - F_1(t|A,Z) - F_2(t|A,Z),$$

and

$$\Lambda_1(t|A,Z) = \int_0^t \frac{1}{S(u|A,Z)} dF_1(u|A,Z),$$

$$\Lambda_2(t|A,Z) = \int_0^t \frac{1}{S(u|A,Z)} dF_2(u|A,Z),$$

where $F_1(t|A,Z)$ and $F_2(t|A,Z)$ can be estimated non-parametrically through random forest. This model is also under investigation in simulation studies.

Appendix A

HHP and HAAS Data Distribution

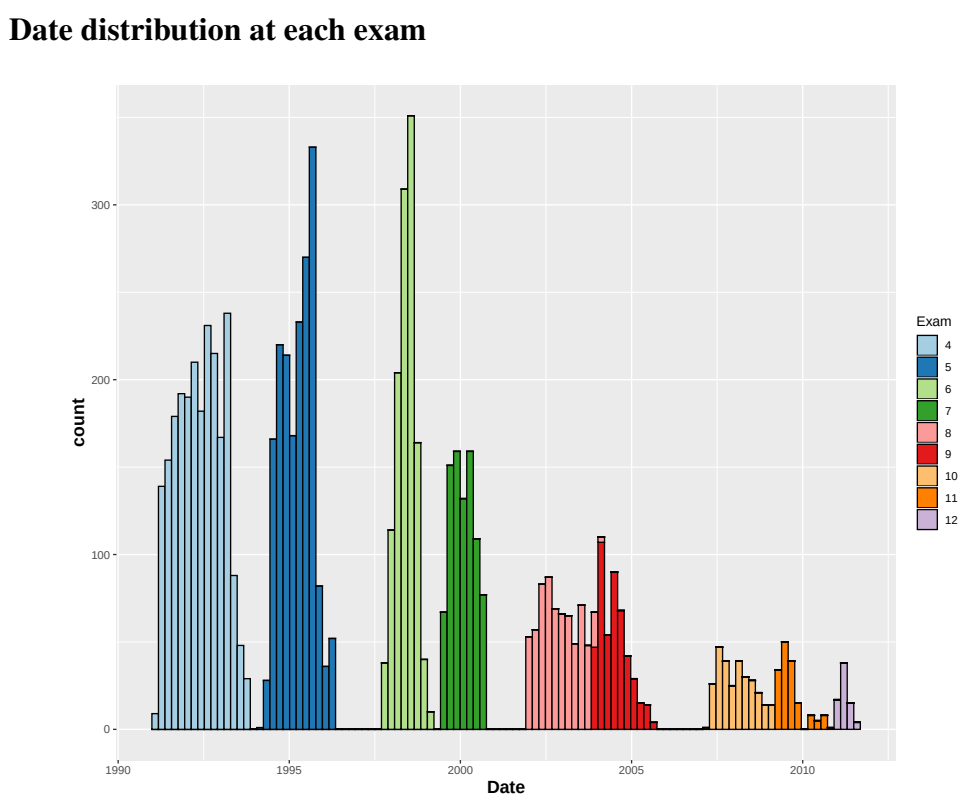

Figure A.1. Date distribution at each exam

Number of participant for each exam

Table A.1. Number of participant for each exam

	Number of participant
Exam 4	1910
Exam 5	1826
Exam 6	1791
Exam 7	1359
Exam 8	1168
Exam 9	853
Exam 10	688
Exam 11	411
Exam 12	278

Comparison for the last visit age for those have death date and their age of death

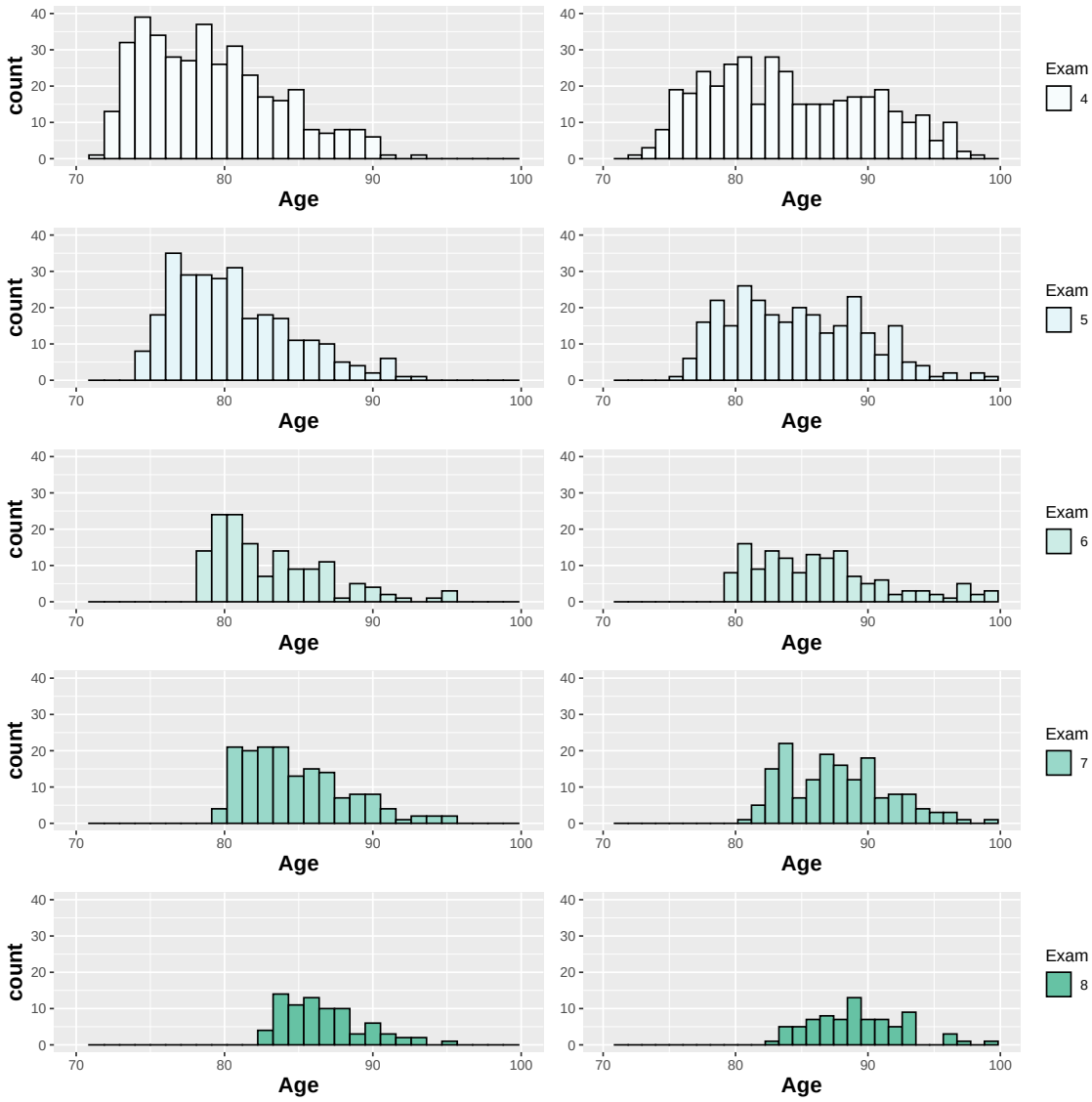

Figure A.2. Last Age for those who died (left), Death Age(right)

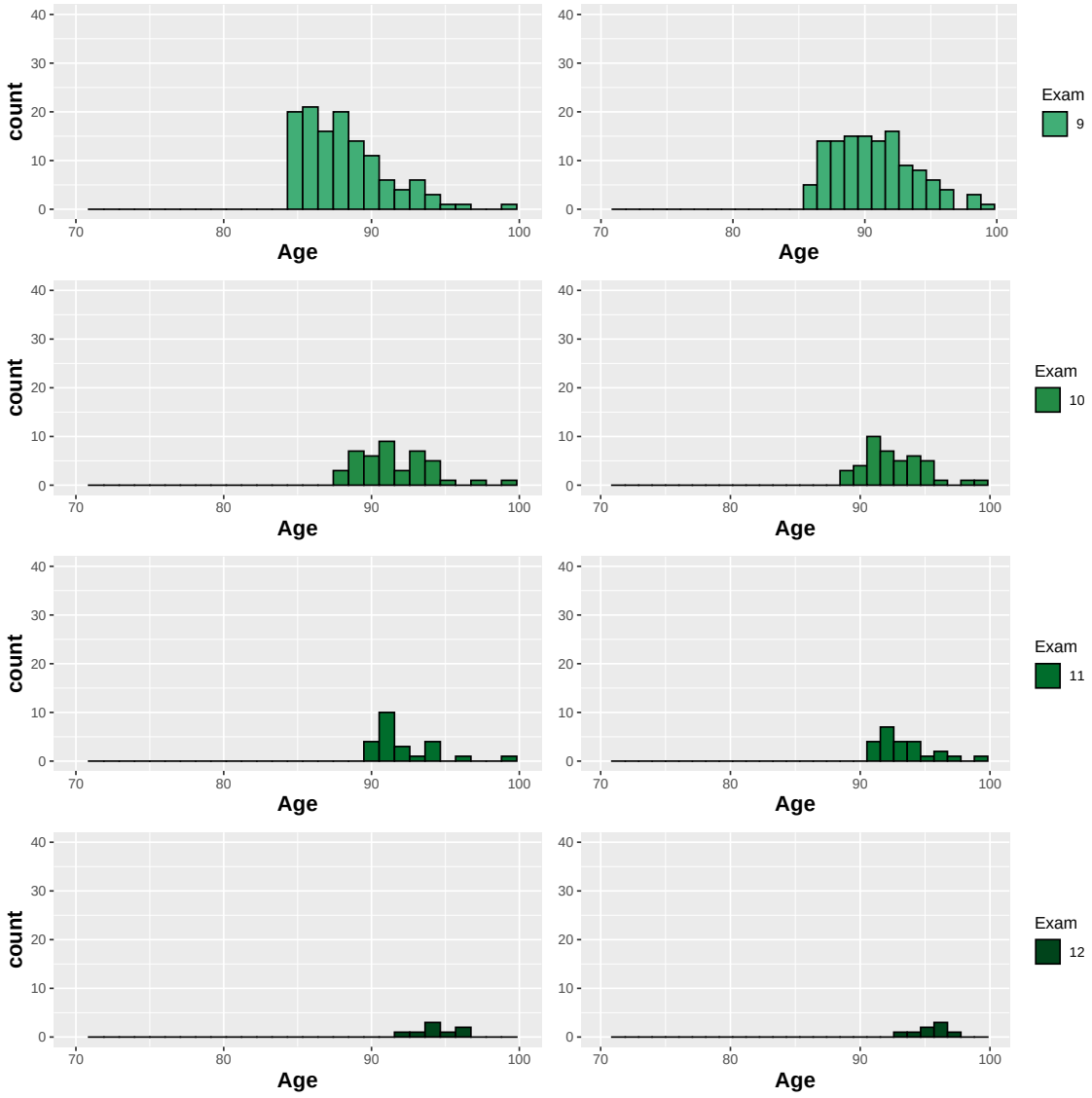

Figure A.3. Last Age for those who died (left), Death Age(right)

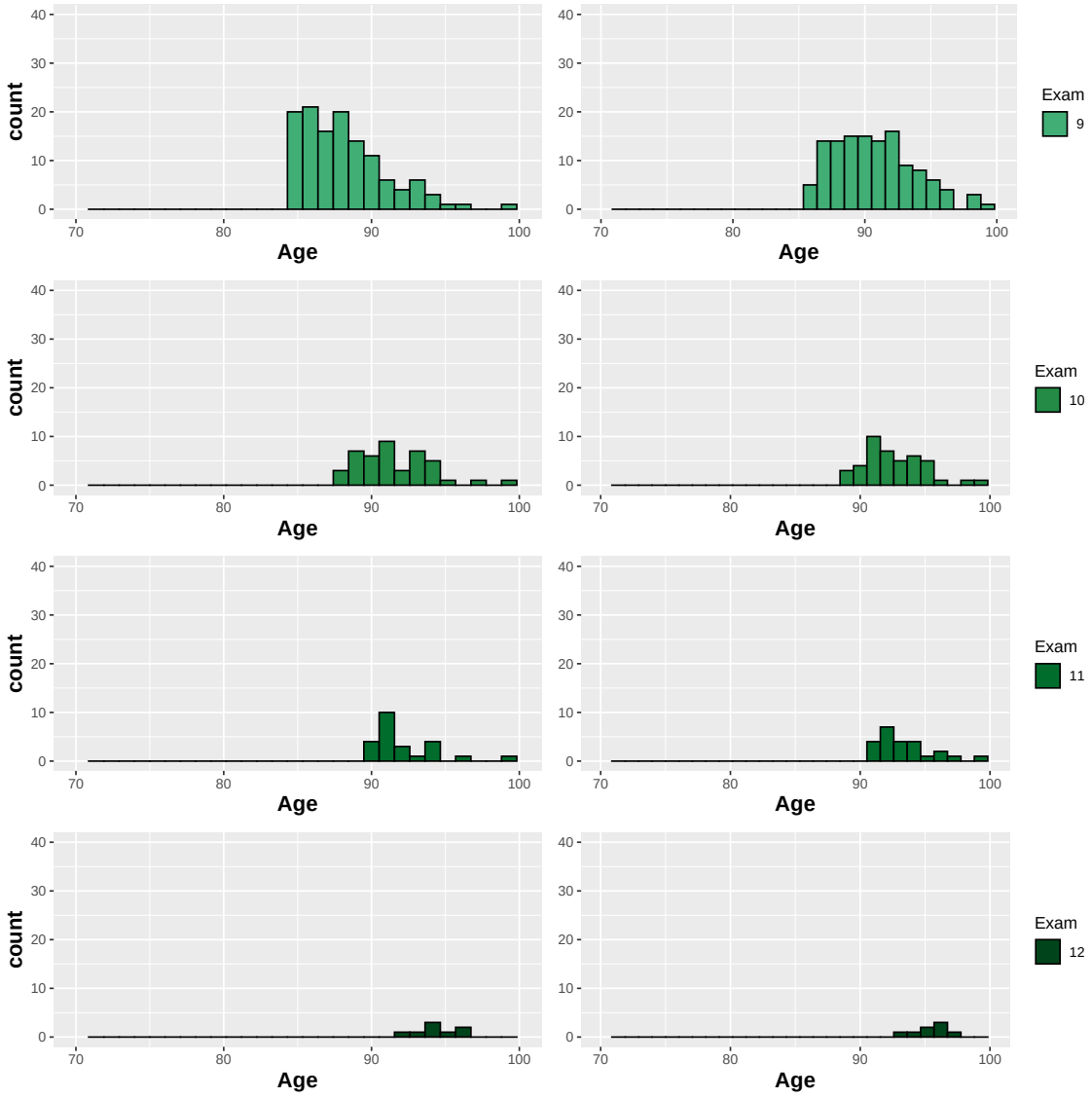

Figure A.4. Year difference of last age for those who died and their death age

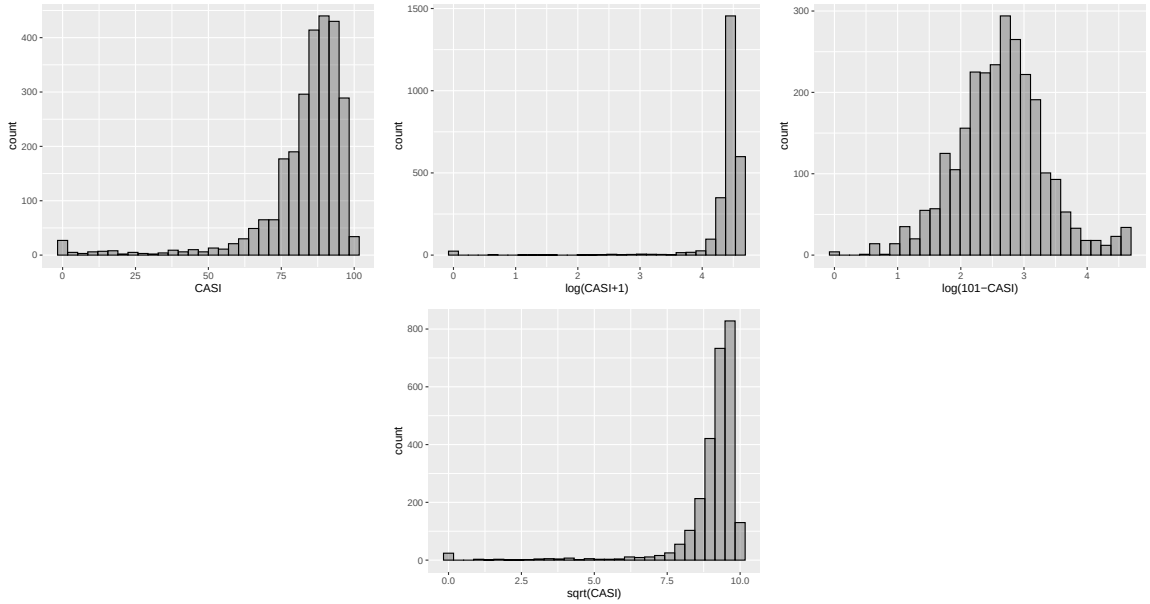

Figure A.5. Histogram of different transformation of CASI

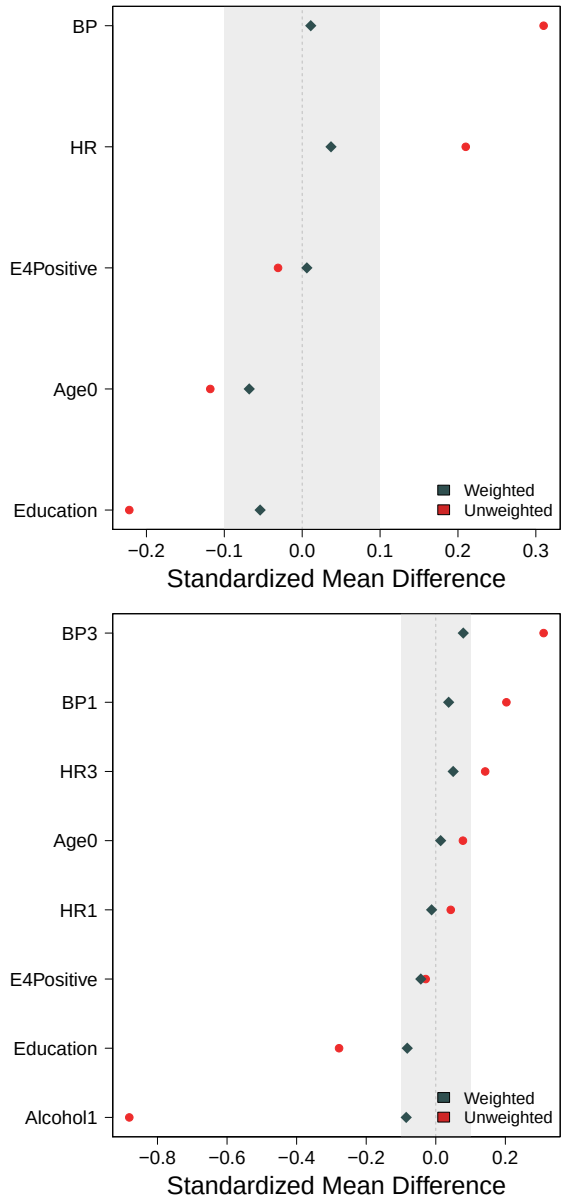

Figure A.6. SMD plot before and after weighting for exam 1 (left) and exam 3 (right)

Spaghetti plots for semi-competing risk data

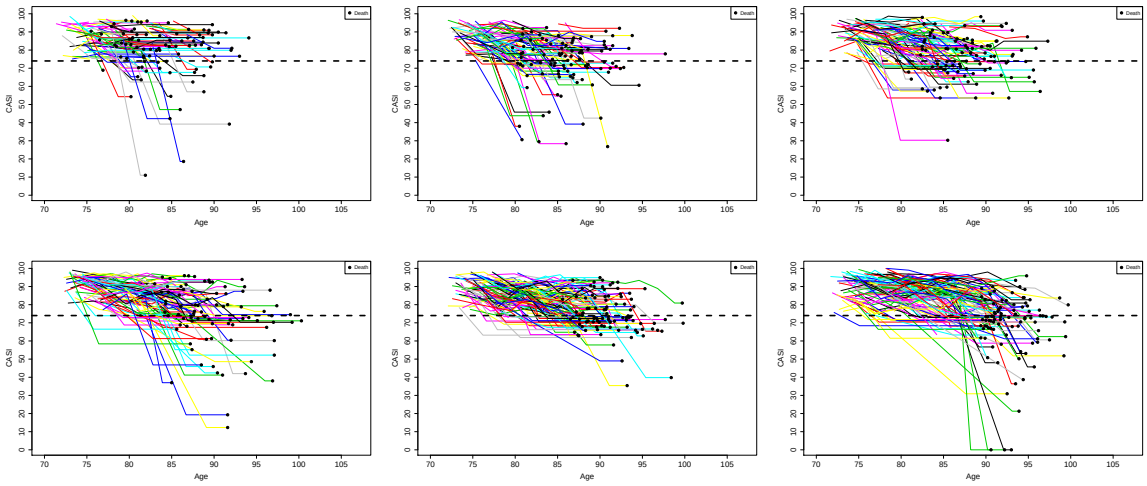

Figure A.7. Spaghetti plots for randomly selected subjects that died in different exam

Appendix B

Supplementary materials for Chapter 4

Derivation of $f(t_1,t_2)$, $f_\infty(t_2)$ and $S(t,t)$

$$
\begin{aligned}
f_\infty(t_2) &= \lim_{\Delta\to 0} \frac{P(T_1 \geq t_2, T_2 \in [t_2, t_2+\Delta))}{\Delta} \\
&= \lim_{\Delta\to 0} \frac{P(T_1 \geq t_2, T_2 \in [t_2, t_2+\Delta))}{P(T_1 \geq t_2, T_2 \geq t_2)\,\Delta} \times P(T_1 \geq t_2, T_2 \geq t_2) \\
&= \lambda_2(t_2) S(t_2, t_2)
\end{aligned}
$$

We also have:

$$
\begin{aligned}
f(t_1,t_2) &= \lim_{\Delta\to 0}\lim_{\delta\to 0}\frac{P(T_1\in[t_1,t_1+\delta),T_2\in[t_2,t_2+\Delta))}{\Delta\delta} \\
&= \lim_{\Delta\to 0}\lim_{\delta\to 0} P(T_1\geq t_1,T_2\geq t_1)\times\frac{P(T_1\in[t_1,t_1+\delta),T_2\geq t_1)}{P(T_1\geq t_1,T_2\geq t_1)\delta} \\
&\times\frac{P(T_2\in[t_2,t_2+\Delta)\mid T_1\in[t_1,t_1+\delta),T_2\geq t_1)}{\Delta} \\
&= \lim_{\Delta\to 0}\lim_{\delta\to 0} P(T_1\geq t_1,T_2\geq t_1)\times\frac{P(T_1\in[t_1,t_1+\delta),T_2\geq t_1)}{P(T_1\geq t_1,T_2\geq t_1)\delta} \\
&\times\frac{P(T_2\in[t_2,t_2+\Delta),T_2\geq t_1\mid T_1\in[t_1,t_1+\delta))}{P(T_2\geq t_1\mid T_1\in[t_1,t_1+\delta))\Delta} \\
&= \lim_{\Delta\to 0}\lim_{\delta\to 0} P(T_1\geq t_1,T_2\geq t_1)\times\frac{P(T_1\in[t_1,t_1+\delta),T_2\geq t_1)}{P(T_1\geq t_1,T_2\geq t_1)\delta} \\
&\times\frac{P(T_2\in[t_2,t_2+\Delta)\mid T_1\in[t_1,t_1+\delta))}{P(T_2\geq t_2\mid T_1\in[t_1,t_1+\delta))\Delta}\times\frac{P(T_2\geq t_2\mid T_1\in[t_1,t_1+\delta))}{P(T_2\geq t_1\mid T_1\in[t_1,t_1+\delta))} \\
&= S(t_1,t_1)\lambda_1(t_1)\lambda_{12}(t_2\mid t_1)\exp\left\{-\int_{t_1}^{t_2}\lambda_{12}(u|t_1)du\right\}
\end{aligned}
$$

We further have:

$$
\begin{aligned}
\lambda_1(t_1) &= \lim_{\Delta\to 0^+}\frac{P(T_1\in[t_1,t_1+\Delta)\mid T_1\geq t_1,T_2\geq t_1)}{\Delta} \\
&= \frac{\lim_{\Delta\to 0^+}P(T_1\in[t_1,t_1+\Delta),T_1\geq t_1,T_2\geq t_1)/\Delta}{P(T_1\geq t_1,T_2\geq t_1)} \\
&= \frac{\lim_{\Delta\to 0^+}P(T_1\in[t_1,t_1+\Delta),T_2\geq t_1)/\Delta}{P(T_1\geq t_1,T_2\geq t_1)} \\
&= \frac{\int_{t_1}^{+\infty}f(t_1,u)du}{P(T_1\geq t_1,T_2\geq t_1)} \\
&= \frac{-\frac{\partial}{\partial t_1}S(t_1,t_2)|_{t_2=t_1}}{S(t_1,t_1)} \\
&= -\frac{\partial}{\partial t_1}\log S(t_1,t_2)|_{t_2=t_1}
\end{aligned}
\tag{B.1}
$$

Similar derivation can be applied to obtain

$$
\lambda_2(t_2)=\lim_{\Delta\to 0^+}P(T_2\in[t_2,t_2+\Delta)\mid T_1\geq t_2,T_2\geq t_2)/\Delta=-\partial\log S(t_1,t_2)/\partial t_2|_{t_1=t_2}.
$$

By solving the partial derivative equations with the initial condition $S(0,0) = 1$, we have

$$S(t,t) = e^{-(\Lambda_1(t)+\Lambda_2(t))}.$$

We then have (4.13) - (4.15) in the main text.

Variance-covariance under the usual Markov model

For the ith individual, let the at-risk process for non-terminal event, terminal event without non-terminal event, and terminal event following non-terminal event as $Y_{1i}(t) = I(X_{1i} \geq t)$, $Y_{2i}(t) = I(X_{2i} \geq t, X_{1i} \geq t)$, and $Y_{3i}(t) = I(X_{2i} \geq t \geq X_{1i})$. It is also convenient to introduce the following notation:

$$\begin{aligned}
S_{1w}^{(1)}(\hat{\beta}_1;t) &= \sum_{\ell=1}^{n} w_\ell Y_{1\ell}(t) A_\ell \exp(\hat{\beta}_1 A_\ell),\ S_{1w}^{(0)}(\hat{\beta}_1;t) = \sum_{\ell=1}^{n} w_\ell Y_{1\ell}(t) \exp(\hat{\beta}_1 A_\ell);\\
S_{2w}^{(1)}(\hat{\beta}_2;t) &= \sum_{\ell=1}^{n} w_\ell Y_{2\ell}(t) A_\ell \exp(\hat{\beta}_2 A_\ell),\ S_{2w}^{(0)}(\hat{\beta}_2;t) = \sum_{\ell=1}^{n} w_\ell Y_{2\ell}(t) \exp(\hat{\beta}_2 A_\ell);\\
S_{3w}^{(1)}(\hat{\beta}_3;t) &= \sum_{\ell=1}^{n} w_\ell Y_{3\ell}(t) A_\ell \exp(\hat{\beta}_3 A_\ell),\ S_{3w}^{(0)}(\hat{\beta}_3;t) = \sum_{\ell=1}^{n} w_\ell Y_{3\ell}(t) \exp(\hat{\beta}_3 A_\ell).
\end{aligned}$$

Then the robust sandwich variance estimator is given by $V(\hat{\beta}) = B(\hat{\beta})M(\hat{\beta})B(\hat{\beta})$, where $B(\hat{\beta}) = -\partial^2 \log L_w(\beta)/\partial\beta^2|_{\beta=\hat{\beta}}/n = [b_{jj}]_{j=1,2,3}$ is a diagonal matrix,

$$\begin{aligned}
b_{11} &= -\frac{1}{n}\sum_{i=1}^{n} w_i \delta_{i1} \left\{ A_i - \frac{S_{1w}^{(1)}(\hat{\beta}_1;X_{1i})}{S_{1w}^{(0)}(\hat{\beta}_1;X_{1i})} \right\},\\
b_{22} &= -\frac{1}{n}\sum_{i=1}^{n} w_i (1-\delta_{i1}) \delta_{i2} \left\{ A_i - \frac{S_{2w}^{(1)}(\hat{\beta}_2;X_{1i})}{S_{2w}^{(0)}(\hat{\beta}_2;X_{1i})} \right\},\\
b_{33} &= -\frac{1}{n}\sum_{i=1}^{n} w_i \delta_{i1} \delta_{i2} \left\{ A_i - \frac{S_{3w}^{(1)}(\hat{\beta}_3;X_{2i})}{S_{3w}^{(0)}(\hat{\beta}_3;X_{2i})} \right\};
\end{aligned}$$

and $M(\hat{\beta}) = \sum_{i=1}^{n} \hat{U}^{(i)}(\hat{\beta})\hat{U}^{(i)}(\hat{\beta})'/n$ with

$$
\begin{aligned}
U_1^{(i)}(\hat{\beta}_1) =& w_i \delta_{1i} \left\{ A_i - \frac{S_{1w}^{(1)}(\hat{\beta}_1; X_{1i})}{S_{1w}^{(0)}(\hat{\beta}_1; X_{1i})} \right\} \\
& - w_i \cdot \sum_{\ell=1}^{n} w_\ell \delta_{1\ell} \frac{Y_{1i}(X_{1\ell}) \exp(\beta_1 A_i)}{S_{1w}^{(0)}(\hat{\beta}_1; X_{1\ell})} \left\{ A_i - \frac{S_{1w}^{(1)}(\hat{\beta}_1; X_{1\ell})}{S_{1w}^{(0)}(\hat{\beta}_1; X_{1\ell})} \right\}, \\
U_2^{(i)}(\hat{\beta}_2) =& w_i (1 - \delta_{1i}) \delta_{2i} \left\{ A_i - \frac{S_{2w}^{(1)}(\hat{\beta}_2; X_{1i})}{S_{2w}^{(0)}(\hat{\beta}_2; X_{1i})} \right\} \\
& - w_i \cdot \sum_{\ell=1}^{n} w_\ell (1 - \delta_{1\ell}) \delta_{2\ell} \frac{Y_{2i}(X_{1\ell}) \exp(\beta_2 A_i)}{S_{2w}^{(0)}(\hat{\beta}_2; X_{1\ell})} \left\{ A_i - \frac{S_{2w}^{(1)}(\hat{\beta}_2; X_{1\ell})}{S_{2w}^{(0)}(\hat{\beta}_2; X_{1\ell})} \right\}, \\
U_3^{(i)}(\hat{\beta}_3) =& w_i \delta_{1i} \delta_{2i} \left\{ A_i - \frac{S_{3w}^{(1)}(\hat{\beta}_3; X_{2i})}{S_{3w}^{(0)}(\hat{\beta}_3; X_{2i})} \right\} \\
& - w_i \cdot \sum_{\ell=1}^{n} w_\ell \delta_{1\ell} \delta_{2\ell} \frac{Y_{3i}(X_{2\ell}) \exp(\beta_3 A_i)}{S_{3w}^{(0)}(\hat{\beta}_3; X_{2\ell})} \left\{ A_i - \frac{S_{3w}^{(1)}(\hat{\beta}_3; X_{2\ell})}{S_{3w}^{(0)}(\hat{\beta}_3; X_{2\ell})} \right\}.
\end{aligned}
$$

Proof of Lemma 1

Proof. From (4.21) in the main text, we have:

$$
\begin{aligned}
l_w(\theta;O) =& \log L_w(\theta;O) \\
=& \log\left\{\prod_i \left(\int L(\theta;O_i \mid b_i)\cdot f(\theta;b_i)db_i\right)^{w_i}\right\} \\
=& \sum_i w_i \log \int \frac{L(\theta;O_i \mid b_i)\cdot f(\theta;b_i)}{f(b_i|\theta^{(k)},O_i)} f(b_i|\theta^{(k)},O_i)db_i \\
=& \sum_i w_i \log \mathbb{E}\left[\left.\frac{L(\theta;O_i \mid b_i)\cdot f(\theta;b_i)}{f(b_i|\theta^{(k)},0_i)}\right| \theta^{(k)},O_i\right] \qquad \text{(B.2)} \\
\geq& \sum_i w_i \mathbb{E}\left[\log\left(\frac{L(\theta;O_i \mid b_i)\cdot f(\theta;b_i)}{f(b_i|\theta^{(k)},O_i)}\right) \mid \theta^{(k)},O_i\right] \qquad \text{(B.3)} \\
=& \sum_i \mathbb{E}_{\theta^{(k)}}\left[w_i \cdot l(\theta;O_i \mid b_i) \mid O_i\right] + \mathbb{E}\left[w_i \cdot \log f(b_i;\theta)) \mid \theta^{(k)},O_i\right] \\
& - \mathbb{E}\left[w_i \cdot \log f(b_i \mid \theta^{(k)},O_i) \mid \theta^{(k)},O_i\right] \\
=& Q(\theta;\theta^{(k)}) - \sum_i w_i \mathbb{E}\left[\log f(b_i \mid \theta^{(k)},O_i) \mid \theta^{(k)},O_i\right],
\end{aligned}
$$

where the inequality above comes from Jensen's inequality. If $\theta = \theta^{(k)}$, (B.2) becomes

$$
\begin{aligned}
& \sum_i w_i \log \mathbb{E}\left[\frac{L(\theta^{(k)};O_i \mid b_i)\cdot f(\theta^{(k)};b_i)}{f(b_i|\theta^{(k)},O_i)} \mid \theta^{(k)},O_i\right] \\
=& \sum_i w_i \log \mathbb{E}\left[f(O_i \mid \theta^{(k)}) \mid \theta^{(k)},O_i\right] \\
=& \sum_i w_i \log f(O_i \mid \theta^{(k)}) \\
=& \sum_i w_i \mathbb{E}\left[\log f(O_i \mid \theta^{(k)}) \mid \theta^{(k)},O_i\right],
\end{aligned}
$$

which equals (B.3).

Then we have $l_w(\theta^{(k)};O) = Q(\theta^{(k)};\theta^{(k)}) - \sum_i w_i \mathbb{E}\left[\log f(b_i \mid \theta^{(k)},O_i) \mid \theta^{(k)},O_i\right]$.

Therefore,

$$
\begin{aligned}
& l_w(\theta^{(k+1)};O) - l_w(\theta^{(k)};O) \\
\geq & Q(\theta^{(k+1)};\theta^{(k)}) - Q(\theta^{(k)};\theta^{(k)}) - \Bigg(\sum_i w_i \mathbb{E}\left[\log f(b_i \mid \theta^{(k)}, O_i) \mid \theta^{(k)}, O_i \right] \\
& - \sum_i w_i \mathbb{E}\left[\log f(b_i \mid \theta^{(k)}, O_i) \mid \theta^{(k)}, O_i \right] \Bigg) \\
= & Q(\theta^{(k+1)};\theta^{(k)}) - Q(\theta^{(k)};\theta^{(k)}).
\end{aligned}
$$

Since $\theta^{(k+1)}$ maximizes $Q(\theta, \theta^{(k)})$, $Q(\theta^{(k+1)};\theta^{(k)}) - Q(\theta^{(k)};\theta^{(k)}) \geq 0$.

Therefore $l_w(\theta^{(k+1)};O) \geq l_w(\theta^{(k)};O)$, and $L_w(\theta^{(k+1)};O) \geq L_w(\theta^{(k)};O)$. □

Detailed calculation of $E(h(b_i)|O_i,\tilde{\theta})$

We have

$$\begin{aligned}
E(h(b_i)|O_i;\tilde{\theta}) &= \int h(b_i)\cdot f(b_i \mid O_i;\tilde{\theta})db_i \\
&= \int h(b_i)\cdot \frac{f(O_i,b_i;\tilde{\theta})}{f(O_i;\tilde{\theta})}db_i \\
&= \int h(b_i)\cdot \frac{f(O_i \mid b_i;\tilde{\theta})f(b_i;\tilde{\theta})}{f(O_i;\tilde{\theta})}db_i,
\end{aligned}$$

where

$$\begin{aligned}
f(O_i;\tilde{\theta}) &= \int f(O_i,b_i;\tilde{\theta})db_i \\
&= \int f(O_i \mid b_i;\tilde{\theta})\cdot f(b_i;\tilde{\theta})db_i.
\end{aligned}$$

After plugging in model based quantities, we have

$$\begin{aligned}
f(O_i;\tilde{\theta}) = \int \Bigg[& \{\tilde{\lambda}_{01}(X_{1i}) \\
& \exp(\tilde{\beta}_1 A_i + b_i)\}^{\delta_{1i}} \exp\{-\tilde{\Lambda}_{01}(X_{1i})\exp(\tilde{\beta}_1 A_i + b_i)\} \\
\cdot & \{\tilde{\lambda}_{02}(X_{2i}) \\
& \exp(\tilde{\beta}_2 A_i + b_i)\}^{\delta_{2i}(1-\delta_{1i})} \exp\{-\tilde{\Lambda}_{02}(X_{1i})\exp(\tilde{\beta}_2 A_i + b_i)\} \\
\cdot & \{\tilde{\lambda}_{03}(X_{2i}) \\
& \exp(\tilde{\beta}_3 A_i + b_i)\}^{\delta_{2i}\delta_{1i}} \exp\{-\tilde{\Lambda}_{03}(X_{1i},X_{2i})\exp(\tilde{\beta}_3 A_i + b_i)\}\Bigg] \\
\cdot & \left[\frac{\exp(-\frac{b_i^2}{2\tilde{\sigma}^2})}{\sqrt{2\pi\tilde{\sigma}^2}}\right] db_i.
\end{aligned}$$

Then we have

$$
\begin{aligned}
E(h(b_i)|O_i;\tilde{\theta}) = \int \frac{h(b_i)}{f(O_i;\tilde{\theta})} \\
&\cdot \Bigg[\Big\{ \tilde{\lambda}_{01}(X_{1i}) \\
&\qquad \cdot \exp(\tilde{\beta}_1 A_i + b_i) \Big\}^{\delta_{1i}} \exp\{-\tilde{\Lambda}_{01}(X_{1i}) \exp(\tilde{\beta}_1 A_i + b_i)\} \\
&\cdot \Big\{ \tilde{\lambda}_{02}(X_{2i}) \\
&\qquad \exp(\tilde{\beta}_2 A_i + b_i) \Big\}^{\delta_{2i}(1-\delta_{1i})} \exp\{-\tilde{\Lambda}_{02}(X_{1i}) \exp(\tilde{\beta}_2 A_i + b_i)\} \\
&\cdot \Big\{ \tilde{\lambda}_{03}(X_{2i}) \\
&\qquad \exp(\tilde{\beta}_3 A_i + b_i) \Big\}^{\delta_{2i}\delta_{1i}} \exp\Big\{ -\tilde{\Lambda}_{03}(X_{1i}, X_{2i}) \exp(\tilde{\beta}_3 A_i + b_i) \Big\} \Bigg] \\
&\cdot \left[\frac{\exp(-\frac{b_i^2}{2\tilde{\sigma}^2})}{\sqrt{2\pi\tilde{\sigma}^2}} \right] db_i.
\end{aligned}
$$

Numerical methods such as adaptive Gaussian quadrature can be used to calculate the integral, which is what we use in this paper.

Details for M-step

To maximize the Q_1, we apply the profile likelihood method where we fix β_1 first. If we denote the entry time for i as V_i, then the formula (4.24) can be written as:

$$\begin{aligned} Q_1(\beta_1, \lambda_{01}) &= \sum_i w_i \Big\{ \delta_{1i} \log \lambda_{01}(X_{1i}) + \delta_{1i}\beta_1 A_i + \delta_{1i}\mathbb{E}(b_i) - e^{\beta_1 A_i + \mu_i} \int_{V_i}^{X_{1i}} \lambda_{01}(s)ds \Big\} \\ &= \sum_i w_i \Big\{ \delta_{1i} \log \lambda_{01}(X_{1i}) + \delta_{1i}\beta_1 A_i + \delta_{1i}\mathbb{E}(b_i) - e^{\beta_1 A_i + \mu_i} \sum_{k: V_i \le t_{1k} \le X_{1i}} \lambda_{01}(t_{1k}) \Big\} \end{aligned}$$

the second equality is because of the nonparametric likelihood discretizes λ_{01} to mass points at the observed non-terminal event times $0 < t_{11} < ... < t_{1k_1}$. For the given β_1, we want to maximize the Q_1 over $\lambda_{01} = (\lambda_{11}, ..., \lambda_{1k_1})$, then for k, we have:

$$\frac{\partial Q_1(\beta_1, \lambda_{01})}{\partial \lambda_{1k}} = \frac{w_k}{\lambda_{1k}} - \sum_{i: V_i \le t_{1k} \le X_{1i}} w_i e^{\beta_1 A_i + \mu_i}$$

and:

$$\hat{\lambda}_{1k} = \frac{w_k}{\sum_{i: V_i \le t_{1k} \le X_{1i}} w_i e^{\beta_1 A_i + \mu_i}} \tag{B.4}$$

for $k = 1, 2, ..., k_1$. Substituting (B.4) into Q_1, we will have:

$$
\begin{aligned}
Q_1 = &\sum_i w_i \Bigg\{ \delta_{1i} \log \frac{w_i}{\sum\limits_{j \in R_i} w_j e^{\beta_1 A_j + \mu_j}} + \delta_{1i} \beta_1 A_i + \delta_{1i} \mathbb{E}(b_i) \\
&\quad - \sum_{k: V_i \le t_{1k} \le X_{1i}} \frac{w_k \delta_{1k} e^{\beta_1 A_k + \mu_k}}{\sum\limits_{j \in R_k} w_j e^{\beta_1 A_j + \mu_j}} \Bigg\} \\
= &\sum_i w_i \left\{ \delta_{1i} \log \frac{e^{\beta_1 A_i}}{\sum\limits_{j \in R_i} w_j e^{\beta_1 A_j + \mu_j}} \right\} \\
&\quad - \sum_i \sum_{k: V_i \le t_{1k} \le X_{1i}} w_i \cdot \frac{w_k e^{\beta_1 A_i + \mu_i}}{\sum\limits_{j \in R_k} w_j e^{\beta_1 A_j + \mu_j}} + s_1 + s_2 \\
= &\sum_i w_i \left\{ \delta_{1i} \log \frac{e^{\beta_1 A_i}}{\sum\limits_{j \in R_i} w_j e^{\beta_1 A_j + \mu_j}} \right\} - \sum_{k=1}^{k_1} \sum_{i: X_{1i} \ge t_{1k}} w_k \cdot \frac{w_i e^{\beta_1 A_i + \mu_i}}{\sum\limits_{j \in R_k} w_j e^{\beta_1 A_j + \mu_j}} + s_1 + s_2 \\
= &\sum_i w_i \left\{ \delta_{1i} \log \frac{e^{\beta_1 A_i}}{\sum\limits_{j \in R_i} w_j e^{\beta_1 A_j + \mu_j}} \right\} - \sum_{k=1}^{k_1} w_k \frac{\sum_{i: X_{1i} \ge t_{1k}} w_i e^{\beta_1 A_i + \mu_i}}{\sum\limits_{j \in R_k} w_j e^{\beta_1 A_j + \mu_j}} + s_1 + s_2 \\
= &\sum_i w_i \left\{ \delta_{1i} \big(\beta_1 A_i - \log \sum_{j \in R_i} w_j e^{\beta_1 A_j + \mu_j} \big) \right\} + \sum_{k=1}^{k_1} w_k + s_1 + s_2
\end{aligned} \tag{B.5}
$$

To maximize Q_2 and Q_3, we cannot use the profile likelihood approach. Let $0 < t_{21} < ... < t_{2k_2}$ be the distinct terminal event times without non-terminal event, and $0 < t_{31} < ... < t_{3k_3}$ be the distinct terminal event times following non-terminal event. We also let $\lambda_{02} = (\lambda_{21}, ..., \lambda_{2k_2})$ and $\lambda_{03} = (\lambda_{31}, ..., \lambda_{3k_3})$. Same as the Q_1, we can write Q_2, Q_3 as:

$$
\begin{aligned}
Q_2(\beta_2, \lambda_{02}) = &\sum_i w_i \Bigg\{ \delta_{2i}(1 - \delta_{1i}) \log \lambda_{02}(X_{2i}) + \delta_{2i}(1 - \delta_{1i}) \beta_2 A_i + \delta_{2i}(1 - \delta_{1i}) \mathbb{E}(b_i) \\
&\qquad - e^{\beta_2 A_i + \mu_i} \sum_{j: 0 \le t_{2j} \le X_{1i}} \lambda_{2j} \Bigg\} \\
Q_3(\beta_3, \lambda_{03}) = &\sum_i w_i \Bigg\{ \delta_{2i} \delta_{1i} \log \lambda_{03}(X_{2i}) \\
&\qquad + \delta_{2i} \delta_{1i} \beta_3 A_i + \delta_{2i} \delta_{1i} \mathbb{E}(b_i) - e^{\beta_3 A_i + \mu_i} \sum_{j: X_{1i} < t_{3j} \le X_{2i}} \lambda_{3j} \Bigg\}
\end{aligned}
$$

We denote the score:

$$U_2 = \frac{\partial Q_2}{\partial \beta_2} = \sum_i w_i A_i \left\{ \delta_{2i}(1-\delta_{1i}) - \Lambda_{02}(X_{1i}) e^{\beta_2 A_i + \mu_i} \right\} = 0 \tag{B.6}$$

$$U_3 = \frac{\partial Q_3}{\partial \beta_3} = \sum_i w_i A_i \left\{ \delta_{2i}\delta_{1i} - \Lambda_{03}(X_{1i}, X_{2i}) e^{\beta_3 A_i + \mu_i} \right\} = 0 \tag{B.7}$$

$$U_{4j} = \frac{\partial Q_2}{\partial \lambda_{2j}} = \frac{w_j}{\lambda_{2j}} - \sum_i \mathbb{1}(t_{2j} \le X_{1i}) w_i \cdot e^{\beta_2 A_i + \mu_i} = 0 \tag{B.8}$$

$$U_{5j} = \frac{\partial Q_3}{\partial \lambda_{3j}} = \frac{w_j}{\lambda_{3j}} - \sum_i \mathbb{1}(X_{1i} < t_{3j} \le X_{2i}) w_i \cdot e^{\beta_3 A_i + \mu_i} = 0 \tag{B.9}$$

Assuming no ties, from (B.6) and (B.8), we will get:

$$\hat{\beta}_2 = \log\left(\frac{\sum\limits_{i:A_i=1} w_i \delta_{2i}(1-\delta_{1i})}{\sum\limits_{i:A_i=1} w_i \Lambda_{02}(X_{1i}) e^{\mu_i}} \right), \tag{B.10}$$

$$\hat{\lambda}_{2j} = \frac{w_j}{\sum\limits_i \mathbb{1}(t_{2j} \le X_{1i}) w_i \cdot e^{\beta_2 A_i + \mu_i}}; \tag{B.11}$$

and from (B.7) and (B.9), we will get:

$$\hat{\beta}_3 = \log\left(\frac{\sum\limits_{i:A_i=1} w_i \delta_{2i}\delta_{1i}}{\sum\limits_{i:A_i=1} w_i \Lambda_{03}(X_{1i}, X_{2i}) e^{\mu_i}} \right), \tag{B.12}$$

$$\hat{\lambda}_{3j} = \frac{w_j}{\sum\limits_i \mathbb{1}(X_{1i} < t_{3j} \le X_{2i}) w_i \cdot e^{\beta_3 A_i + \mu_i}}. \tag{B.13}$$

Bayesian bootstrap

For each bootstrap sample:

- Generate n standard exponential (mean and variance 1) random variates : $u_1, u_2, ..., u_n$;
- The weights for the Bayesian bootstrap are: $w_i^{boot} = u_i/\bar{u},\ i = 1, 2, ..., n$, where $\bar{u} = n^{-1}\sum_{i=1}^{n} u_i$;
- Calculate the propensity score and IP weights w_i^{IPW} based on Bayesian bootstrap weighted data, and assigned the weights for fitting the MSM general Markov model as $w_i = w_i^{boot} * w_i^{IPW}$.
- After obtaining $\hat{\theta}$ and $\hat{b}_i$, for each individual i, calculate the IRR and IRD by plugging $\hat{\theta}, \hat{b}_i$ and $a = 0, a = 1$ separately into (4.30) - (4.32) from main text at time t: $\hat{F}_{1i}(t \mid b_i; 1) - \hat{F}_{1i}(t \mid b_i; 0)$, $\hat{F}_{2i}(t \mid b_i; 1) - \hat{F}_{2i}(t \mid b_i; 0)$ and $\hat{F_{12i}}(t_1, t \mid b_i; 1) - \hat{F_{12i}}(t_1, t \mid b_i; 0)$, etc..

The 95% prediction intervals (PI) are obtained by the normal approximation using bootstrap standard error.

Details for the simulation steps

Following [23], from (4.14) in the main text and $\lambda_{01}(t) = \lambda_{02}(t) = 2\exp(-t)I(0 \le t \le 3) + 2\exp(-3)I(t \ge 3)$ and $\lambda_{03}(t) = 2\lambda_{01}(t)$, we have

$$\begin{aligned} P(T_1 = \infty) &= \int_0^{+\infty} f_\infty(t \mid b)dt \\ &= \int_0^{+\infty} e^{\beta_{2z}+b}\lambda_{02}(t)\exp\big\{-e^{\beta_{1z}+b}\Lambda_{01}(t) - e^{\beta_{2z}+b}\Lambda_{02}(t)\big\}dt \\ &= \frac{e^{\beta_{2z}}}{e^{\beta_{1z}} + e^{\beta_{2z}}}. \end{aligned} \tag{B.14}$$

We can also derive the conditional marginal density of T_1 when $T_1 < \infty$ from $f(t_1, t_2 \mid b)$ as:

$$\begin{aligned}
f(t_1 \mid b) &= \int_{t_1}^{+\infty} f(t_1, t \mid b) dt \\
&= \int_{t_1}^{+\infty} e^{\beta_1 z + \beta_3 z + 2b} \lambda_{01}(t_1) \lambda_{03}(t) \\
&\quad \cdot \exp\Big\{ - e^{\beta_1 z + b} \Lambda_{01}(t_1) - e^{\beta_2 z + b} \Lambda_{02}(t_1) - e^{\beta_3 z + b} \Lambda_{03}(t_1, t) \Big\} dt \\
&= e^{\beta_1 z + b} \lambda_{01}(t_1) \exp\Big\{ - e^{\beta_1 z + b} \Lambda_{01}(t_1) - e^{\beta_2 z + b} \Lambda_{02}(t_1) \Big\} \\
&\quad \cdot \int_{t_1}^{\infty} \exp\big(- e^{\beta_3 z + b} \Lambda_{03}(t_1, t) \big) d\big\{ e^{\beta_3 z + b} \Lambda_{03}(t) \big\} \\
&= e^{\beta_1 z + b} \lambda_{01}(t_1) \exp\Big\{ - e^{\beta_1 z + b} \Lambda_{01}(t_1) - e^{\beta_2 z + b} \Lambda_{02}(t_1) \Big\} \\
&= e^{\beta_1 z + b} \lambda_{01}(t_1) \exp\Big\{ - e^{\beta_1 z + b} \Lambda_{01}(t_1) - e^{\beta_2 z + b} \Lambda_{01}(t_1) \Big\}. \qquad \text{(B.15)}
\end{aligned}$$

Therefore the conditional survival functions of T_1 conditional on b are

$$\begin{aligned}
S_1(t_1 \mid b) &= P(t_1 \le T_1 < \infty) + P(T_1 = \infty) \\
&= \int_{t_1}^{+\infty} f(t \mid b) dt + Pr(T_1 = \infty) \\
&= \frac{e^{\beta_1 z}}{e^{\beta_1 z} + e^{\beta_2 z}} \exp\Big\{ - (e^{\beta_1 z + b} + e^{\beta_2 z + b}) \Lambda_{01}(t_1) \Big\} + \frac{e^{\beta_2 z}}{e^{\beta_1 z} + e^{\beta_2 z}}, \qquad \text{(B.16)}
\end{aligned}$$

and

$$\begin{aligned}
S_1(t_1 \mid T_1 < \infty, b) &= \frac{S_1(t_1, T_1 < \infty \mid b)}{1 - Pr(T_1 = \infty)} \\
&= \exp\Big\{ - (e^{\beta_1 z + b} + e^{\beta_2 z + b}) \Lambda_{01}(t_1) \Big\}. \qquad \text{(B.17)}
\end{aligned}$$

We also need the conditional joint probability $P(T_2 > t_2, T_1 \in [t_1, t_1+\Delta t] \mid b)$, $t_1 < t_2 < \infty$:

$$
\begin{aligned}
&P(T_2 > t_2, T_1 \in [t_1, t_1+\Delta t] \mid b) = \int_{t_2}^{+\infty} f(t_1, t \mid b) dt \\
=&e^{\beta_1 z+b}\lambda_{01}(t_1) \cdot \exp\big[-e^{\beta_1 z+b}\Lambda_{01}(t_1) - e^{\beta_2 z+b}\Lambda_{02}(t_1) - e^{\beta_3 z+b}(\Lambda_{03}(t_2) - \Lambda_{03}(t_1))\big] \\
=&e^{\beta_1 z+b}\lambda_{01}(t_1) \cdot \exp\big[-e^{\beta_1 z+b}\Lambda_{01}(t_1) - e^{\beta_2 z+b}\Lambda_{01}(t_1) - 2e^{\beta_3 z+b}(\Lambda_{01}(t_2) - \Lambda_{01}(t_1))\big]. \quad \text{(B.18)}
\end{aligned}
$$

Therefore, the conditional survival function for T_2 given $T_1 = t_1 < \infty$ and b is:

$$
\begin{aligned}
S_{21}(t_2 \mid t_1, b) = P(T_2 > t_2 \mid T_1 = t_1, b) &= \frac{P(T_2 > t_2, T_1 \in [t_1, t_1+\Delta t] \mid b)}{f(t_1 \mid b)} \\
&= \exp\big(-2e^{\beta_3 z+b}\{\Lambda_{01}(t_2) - \Lambda_{01}(t_1)\}\big), \quad \text{(B.19)}
\end{aligned}
$$

and the conditional survival function for T_2 given $T_1 = \infty$ and b is

$$
\begin{aligned}
S_{21}(t_2 \mid T_1 = \infty, b) = P(T_2 > t_2 \mid T_1 = \infty, b) &= \frac{P(T_2 > t_2, T_1 = \infty \mid b)}{Pr(T_1 = \infty)} \\
&= \frac{\int_{t_2}^{+\infty} f_\infty(t \mid b) dt}{Pr(T_1 = \infty)} \\
&= \exp\Big\{-(e^{\beta_1 z+b} + e^{\beta_2 z+b})\Lambda_{01}(t_2)\Big\}. \quad \text{(B.20)}
\end{aligned}
$$

Based on the above, we can generate the event time T_1, T_2: with probability $P(T_1 = \infty)$, we can generate T_2 from $S_{21}(t_2 \mid T_1 = \infty, b)$, and with probability $1 - P(T_1 = \infty)$, we can generate T_1 from $S_1(t_1 \mid T_1 < \infty, b)$, then generate T_2 from $S_{21}(t_2 \mid t_1, b)$ conditioning on the observed value of $T_1 = t_1$.

HAAS data analysis

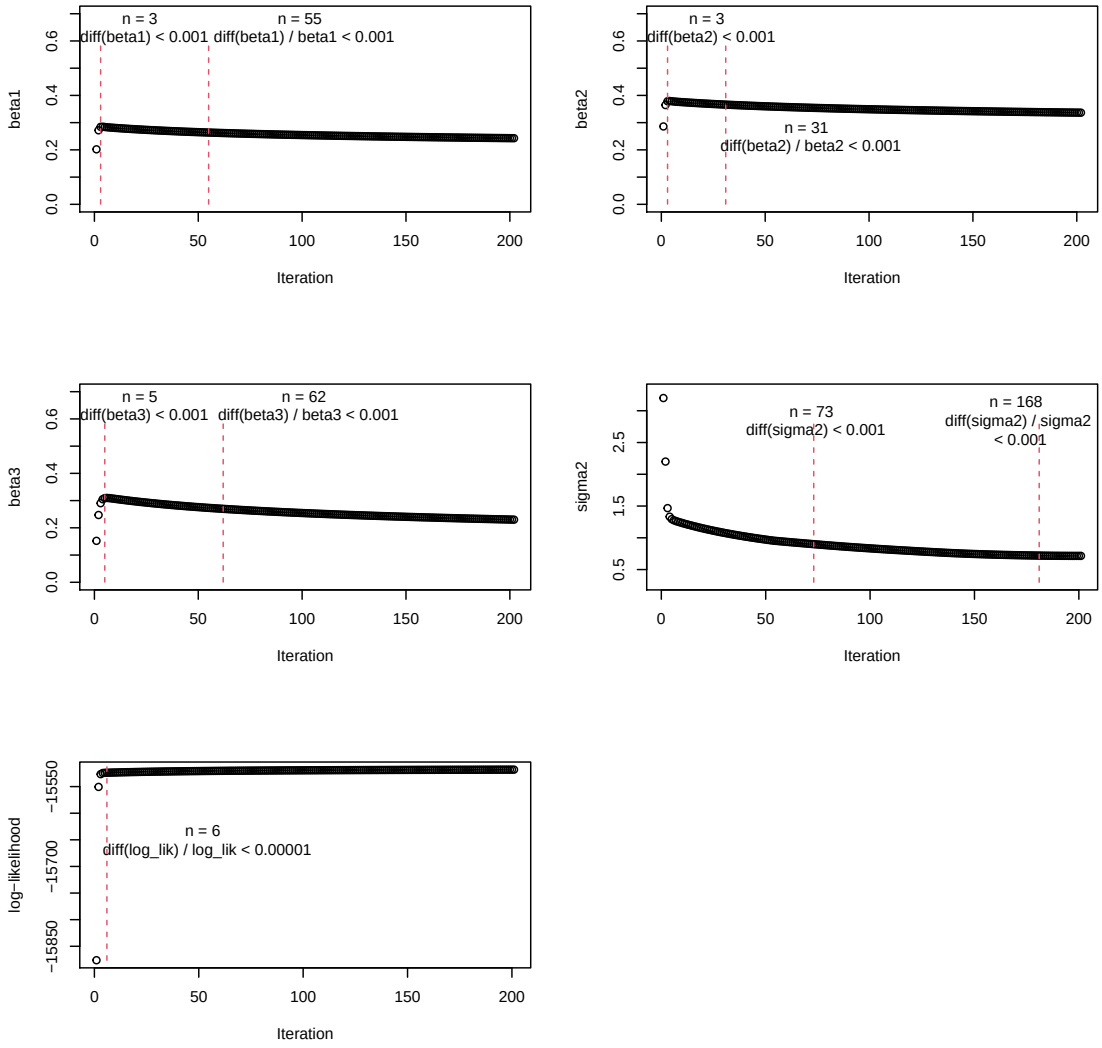

Figure B.1. Convergence plots for the HAAS data analysis under the general Markov model

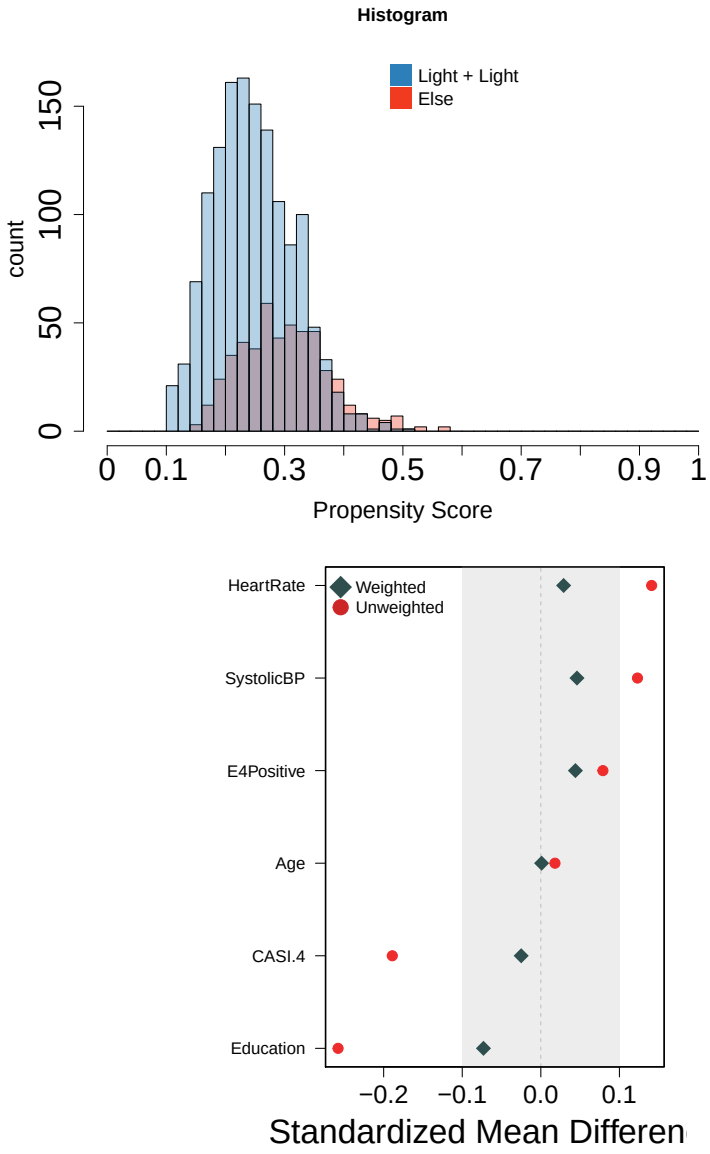

Figure B.2. Distribution of PS (top) and SMD plot (bottom)

Appendix C

Supplementary materials for Chapter 5

Relationships between transitional hazards and joint density of T_1, T_2

Following the supplementary materials from [73], we obtain the relationship between three transitional hazards and joint density of T_1, T_2. We have,

$$\lambda_1(t_1) = \lim_{\delta \to 0^+} \frac{P(T_1 \in [t_1, t_1+\delta), T_2 \geq t_1)}{P(T_1 \geq t_1, T_2 \geq t_1)\delta} \tag{C.1}$$

$$= \frac{\int_{t_1}^{+\infty} f(t_1, s)ds}{S(t_1, t_1)}, \tag{C.2}$$

$$\lambda_2(t_2) = \lim_{\delta \to 0^+} \frac{P(T_2 \in [t_2, t_2+\delta), T_1 \geq t_2)}{P(T_1 \geq t_2, T_2 \geq t_2)\delta} \tag{C.3}$$

$$= \frac{f_\infty(t_2)}{S(t_2, t_2)}, \tag{C.4}$$

and,

$$\lambda_{12}(t_2) = \lim_{\Delta \to 0^+} \frac{P\left(T_2 \in [t_2, t_2+\Delta) \mid T_1 = t_1, T_2 \geq t_2\right)}{\Delta} \tag{C.5}$$

$$= \lim_{\delta \to 0^+, \Delta \to 0^+} \frac{P\left(T_2 \in [t_2, t_2+\Delta), T_1 \in [t_1, t_1+\delta)\right)}{P\left(T_2 \geq t_2, T_1 \in [t_1, t_1+\delta)\right) \delta\Delta} \tag{C.6}$$

$$= \frac{f(t_1, t_2)}{\int_{t_2}^{+\infty} f(t_1, s) ds}. \tag{C.7}$$

Details for calculating $E\{M_{kl}(t) \mid A,Z\}$

Since $E\{M_{k\ell}(t)|A,Z\} = E\{N_{k\ell}(t)|A,Z\} - \int_0^t E\{Y_k(u)|A,Z\}\exp(\beta_\ell A)d\Lambda_{0\ell}(u|A,Z)$, we first show the detail calculations for $E\{N_{k\ell}(t)|A,Z\}$, $k\ell = 01,02,12$.

From [73], we have

$$f(t_1,t_2|A,Z) = S(t_1|A,Z)\lambda_1(t_1|A,Z)\lambda_{12}(t_2|A,Z)e^{-\{\Lambda_{12}(t_2|A,Z)-\Lambda_{12}(t_1|A,Z)\}}, \tag{C.8}$$

$$f_\infty(t_2|A,Z) = S(t_2|A,Z)\lambda_2(t_2|A,Z), \tag{C.9}$$

where $S(t|A,Z) = \exp\{-\Lambda_1(t|A,Z) - \Lambda_2(t|A,Z)\}$ is called conditional overall survival function. We also note that $S(t|A,Z) = \tilde{S}(t,t|A,Z)$, where $\tilde{S}(t,t|A,Z)$ is the conditional joint survival for T_1 and T_2 as we discussed in Section 1.1 and in [73].

We then have,

$$\begin{aligned}
E\{N_{01}(t)|A,Z\} &= E\{I(X_1 \le t, \delta_1 = 1)|A,Z\} && \text{(C.10)}\\
&= P(T_1 \le t, T_1 \le T_2, T_1 \le C|A,Z) && \text{(C.11)}\\
&= \int_0^t \int_{t_1}^{+\infty}\int_{t_1}^{+\infty} f_c(c|A)f(t_1,t_2|A,Z)dcdt_2dt_1 && \text{(C.12)}\\
&= \int_0^t G(t_1|A)\int_{t_1}^{+\infty} f(t_1,t_2|A,Z)dt_2dt_1 && \text{(C.13)}\\
&= \int_0^t G(t_1|A)S(t_1|A,Z)\lambda_1(t_1|A,Z)\\
&\qquad \times \int_{t_1}^{+\infty}\lambda_{12}(t_2|A,Z)e^{-\{\Lambda_{12}(t_2|A,Z)-\Lambda_{12}(t_1|A,Z)\}}dt_2dt_1 && \text{(C.14)}\\
&= \int_0^t G(t_1|A)S(t_1|A,Z)d\Lambda_1(t_1|A,Z), && \text{(C.15)}
\end{aligned}$$

because $\int_{t_1}^{+\infty}\lambda_{12}(t_2|A,Z)e^{-\{\Lambda_{12}(t_2|A,Z)-\Lambda_{12}(t_1|A,Z)\}}dt_2 = 1$. Note also that if we denote $F_1(t|A,Z) = \int_0^t S(t_1|A,Z)d\Lambda_1(t_1|A,Z)$ as the CIF or subdistribution function of the non-terminal event, then the above can be more directly derived as $\int_0^t G(t_1|A)dF_1(t_1|A,Z)$.

$$E\{N_{02}(t)|A,Z\} = E\{I(X_2 \le t, \delta_1 = 0, \delta_2 = 1)|A,Z\} \tag{C.16}$$

$$= P(T_2 \le t, T_1 \ge T_2, T_2 \le C|A,Z) \tag{C.17}$$

$$= \int_0^t \int_{t_2}^{+\infty} f_c(c|A,Z) f_\infty(t_2|A,Z) dc dt_2 \tag{C.18}$$

$$= \int_0^t G(t_2|A) f_\infty(t_2|A,Z) dt_2 \tag{C.19}$$

$$= \int_0^t G(t_2|A) S(t_2|A,Z) d\Lambda_2(t_2|A,Z). \tag{C.20}$$

The above can also be more directly derived as $\int_0^t G(t_2|A) dF_2(t_2|A,Z)$, where $F_2(t|A,Z) = \int_0^t S(t_2|A,Z) d\Lambda_2(t_2|A,Z)$ is the CIF or subdistribution function of the terminal event without the non-terminal event.

Recall that $S(t|A,Z) = \exp\{-\Lambda_1(t|A,Z) - \Lambda_2(t|A,Z)\}$, so it is a function of $\Lambda_1(t|A,Z)$ and $\Lambda_2(t|A,Z)$.

We also have,

$$E\{N_{12}(t)|A,Z\} = E\{I(X_2 \le t, \delta_1 = 1, \delta_2 = 1)|A,Z\} \tag{C.21}$$

$$= P(T_2 \le t, T_1 \le T_2, T_2 \le C|A,Z) \tag{C.22}$$

$$= \int_0^t \int_0^{t_2} \int_{t_2}^{+\infty} f_c(c|A,Z) f(t_1,t_2|A,Z) dc dt_1 dt_2 \tag{C.23}$$

$$= \int_0^t G(t_2|A) \int_0^{t_2} f(t_1,t_2|A,Z) dt_1 dt_2 \tag{C.24}$$

$$= \int_0^t G(t_2|A) f_{12}(t_2|A,Z) dt_2 = \int_0^t G(t_2|A) dF_{12}(t_2|A,Z), \tag{C.25}$$

if we denote $f_{12}(t_2|A,Z) = \int_0^{t_2} f(t_1,t_2|A,Z) dt_1$, and $F_{12}(t_2|A,Z) = \int_0^{t_2} f_{12}(u|A,Z) du$.

For implementation of the above, from (C.24) we need to calculate:

$$
\begin{aligned}
&\int_0^t G(t_2|A) \\
&\qquad \times \int_0^{t_2} S(t_1|A,Z)\lambda_1(t_1|A,Z)\lambda_{12}(t_2|A,Z)e^{-\{\Lambda_{12}(t_2|A,Z)-\Lambda_{12}(t_1|A,Z)\}}dt_1dt_2 \qquad \text{(C.26)} \\
&= \int_0^t G(t_2|A)e^{-\Lambda_{12}(t_2|A,Z)}\left\{\int_0^{t_2} S(t_1|A,Z)e^{\Lambda_{12}(t_1|A,Z)}d\Lambda_1(t_1|A,Z)\right\}d\Lambda_{12}(t_2|A,Z) \\
&= \int_0^t -G(t_2|A)\left\{\int_0^{t_2} S(t_1|A,Z)e^{\Lambda_{12}(t_1|A,Z)}d\Lambda_1(t_1|A,Z)\right\}d\{e^{-\Lambda_{12}(t_2|A,Z)}\} \\
&= \int_0^t G(t_2|A)K(t_2|A,Z)d\{e^{-\Lambda_{12}(t_2|A,Z)}\},
\end{aligned}
$$

where we denote $K(t_2|A,Z) = -\int_0^{t_2} S(t_1|A,Z)e^{\Lambda_{12}(t_1|A,Z)}d\Lambda_1(t_1|A,Z)$. The actual coding with estimators involves two summations.

We can similarly derive $E\{Y_k(u)|A,Z\}$, $k = 1,2$, and obtain

$$
\begin{aligned}
E\{Y_0(t)|A,Z\} &= E\{I(X_1 \geq t)|A,Z\} && \text{(C.27)} \\
&= P(T_1 \geq t, T_2 \geq t, C \geq t|A,Z) && \text{(C.28)} \\
&= \int_t^{+\infty} f_c(c|A,Z)dc \cdot \int_t^{+\infty}\int_t^{+\infty} f(t_1,t_2|A,Z)dt_2dt_1 && \text{(C.29)} \\
&= G(t|A)S(t|A,Z), && \text{(C.30)}
\end{aligned}
$$

and

$$
\begin{aligned}
E\{Y_1(t)|A,Z\} &= E\{I(X_2 \geq t > X_1)|A,Z\} && \text{(C.31)} \\
&= P(T_1 < t, T_2 \geq t, C \geq t|A,Z) && \text{(C.32)} \\
&= G(t|A)\left\{\tilde{S}(0,t|A,Z) - S(t|A,Z)\right\} && \text{(C.33)} \\
&= G(t|A)\int_0^t \int_t^{+\infty} f(t_1,t_2|A,Z)dt_2dt_1, && \text{(C.34)}
\end{aligned}
$$

where $\tilde{S}(0,t|A,Z) = P(T_1 \geq 0, T_2 \geq t)$ and $S(t|A,Z) = P(T_1 \geq t, T_2 \geq t)$.

For the implementation of the above, we have

$$\begin{aligned}
&\int_0^t \int_t^{+\infty} f(t_1,t_2|A,Z)dt_2dt_1 \\
&= \int_0^t S(t_1|A,Z)\lambda_1(t_1|A,Z)e^{\Lambda_{12}(t_1|A,Z)}dt_1 \cdot \int_t^{+\infty} \lambda_{12}(t_2|A,Z)e^{-\Lambda_{12}(t_2|A,Z)}dt_2 \\
&= e^{-\Lambda_{12}(t|A,Z)} \int_0^t S(t_1|A,Z)e^{\Lambda_{12}(t_1|A,Z)}d\Lambda_1(t_1|A,Z).
\end{aligned}$$

Proof for Theorem 1

We only take U_{01} as the example, the rest two can follow the proof. To prove

$$E\{U_{01}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1, \Lambda_2, G^*)\} = 0,$$

we need to prove

$$E\{U_{011}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1, \Lambda_2, G^*)\} = 0$$

and

$$E\{U_{012}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1, \Lambda_2, G^*)\} = 0,$$

if $\pi = \pi^*$ or $\Lambda_1 = \Lambda_1^*, \Lambda_1 = \Lambda_1^*$.

(i) When $\pi = \pi^*$, $E\{U_{01}^{AIPW}(\beta_1^*, \Lambda_{01}^*; \pi^*, \Lambda_1, \Lambda_2, G^*)\} = 0$:

$$\begin{aligned}
&E\{U_{012}(\beta_1^*, \Lambda_{01}^*; \pi^*, \Lambda_1, \Lambda_2, G^*)\} = \\
&= E\Bigg[\int_0^\tau wA dM_{01}(t) \\
&\quad - \int_0^\tau wA \left\{ G^*(t \mid A) S(t \mid A, Z) d\Lambda_1(t \mid A, Z) - G^*(t \mid A) S(t \mid A, Z) e^{\beta_2^* A} d\Lambda_{01}^*(t) \right\} \\
&\quad + \int_0^\tau \left\{ G^*(t \mid 1) S(t \mid 1, Z) d\Lambda_1(t \mid 1, Z) - G^*(t \mid 1) S(t \mid 1, Z) e^{\beta_2^*} d\Lambda_{01}^*(t) \right\} \Bigg]. \qquad \text{(C.35)}
\end{aligned}$$

By conditional expectation and $\pi = \pi^*$, let $DR1 = E\{U_{012}(\beta_1^*, \Lambda_{01}^*; \pi^*, \Lambda_1, \Lambda_2, G^*)\}$, we have,

$$\begin{aligned}
DR1 = &\int_0^\tau E\left[\frac{E\{A dM_{01}(t) \mid Z\}}{\pi^*} \right] \\
&- \int_0^\tau E\left[\frac{1}{\pi^*} E\left\{ A \left(G^*(t \mid A) S(t \mid A, Z) d\Lambda_1(t \mid A, Z) \right) \mid Z \right\} \right] \\
&+ \int_0^\tau E\left[\frac{1}{\pi^*} E\left\{ G^*(t \mid A) S(t \mid A, Z) e^{\beta_2^* A} d\Lambda_{01}^*(t) \mid Z \right\} \right] \\
&+ \int_0^\tau \left\{ G^*(t \mid 1) S(t \mid 1, Z) d\Lambda_1(t \mid 1, Z) - G^*(t \mid 1) S(t \mid 1, Z) e^{\beta_2^*} d\Lambda_{01}^*(t) \right\} \qquad \text{(C.36)}
\end{aligned}$$

Since A only takes value in 0 or 1, then based on the definition of propensity score, we have:

$$\begin{aligned}
DR1 = &\int_0^\tau E\left\{dM_{01}^1(t)\right\} \\
&- \int_0^\tau E\left\{G^*(t \mid 1)S(t \mid A,Z)d\Lambda_1(t \mid A,Z) - G^*(t \mid 1)S(t \mid 1,Z)e^{\beta_2^*}d\Lambda_{01}^*(t)\right\} \\
&+ \int_0^\tau E\left\{G^*(t \mid 1)S(t \mid 1,Z)d\Lambda_1(t \mid 1,Z) - G^*(t \mid 1)S(t \mid 1,Z)e^{\beta_2^*}d\Lambda_{01}^*(t)\right\} \\
= &\int_0^\tau E\left\{dM_{01}^1(t)\right\} = 0
\end{aligned} \tag{C.37}$$

Same algebra, for $t \in [0,\tau]$, we can have:

$$\begin{aligned}
&E\{U_{011}(\beta_1^*,\Lambda_{01}^*;\pi^*,\Lambda_1,\Lambda_2,G^*)\} \\
=&E\left[wE\left\{dM_{01}(t) \mid Z\right\}\right] \\
&- E\left[wE\left\{G^*(t \mid A)S(t \mid A,Z)d\Lambda_1(t \mid A,Z) - G^*(t \mid A)S(t \mid A,Z)e^{\beta_1^* A}d\Lambda_{01}^*(t) \mid Z\right\}\right] \\
&\quad + \left\{G^*(t \mid 1)S(t \mid 1,Z)d\Lambda_1(t \mid 1,Z) - G^*(t \mid 1)S(t \mid 1,Z)e^{\beta_1^*}d\Lambda_{01}^*(t)\right\} \\
&\quad + \left\{G^*(t \mid 0)S(t \mid 0,Z)d\Lambda_1(t \mid 0,Z) - G^*(t \mid 0)S(t \mid 0,Z)e^{\beta_2^*}d\Lambda_{01}^*(t)\right\}
\end{aligned} \tag{C.38}$$

Let $DR2 = E\{U_{011}(\beta_1^*,\Lambda_{01}^*;\pi^*,\Lambda_1,\Lambda_2,G^*)\}$, since $w = A/\pi + (1-A)/(1-\pi)$, and $\pi = E(A|Z)$ plug in and we have,

$$\begin{aligned}
DR2 =&E\left\{\frac{E(A \mid Z)}{\pi^*}dM_{01}^1(t)\right\} + E\left\{\frac{E(1-A \mid Z)}{1-\pi^*}dM_{01}^0(t)\right\} \\
&- \left\{G^*(t \mid 1)S(t \mid 1,Z)d\Lambda_1(t \mid 1,Z) - G^*(t \mid 1)S(t \mid 1,Z)e^{\beta_1^*}d\Lambda_{01}^*(t)\right\} \\
&- \left\{G^*(t \mid 0)S(t \mid 0,Z)d\Lambda_1(t \mid 0,Z) - G^*(t \mid 0)S(t \mid 0,Z)e^{\beta_1^*}d\Lambda_{01}^*(t)\right\} \\
&+ \left\{G^*(t \mid 1)S(t \mid 1,Z)d\Lambda_1(t \mid 1,Z) - G^*(t \mid 1)S(t \mid 1,Z)e^{\beta_1^*}d\Lambda_{01}^*(t)\right\} \\
&+ \left\{G^*(t \mid 0)S(t \mid 0,Z)d\Lambda_1(t \mid 0,Z) - G^*(t \mid 0)S(t \mid 0,Z)e^{\beta_1^*}d\Lambda_{01}^*(t)\right\} \\
=&E\left\{dM_{01}^1(t)\right\} + E\left\{dM_{01}^0(t)\right\} = 0
\end{aligned} \tag{C.39}$$

(ii) When $\Lambda_1 = \Lambda_1^*, \Lambda_2 = \Lambda_2^*$, $E\{U_{01}^{AIPW}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1^*, \Lambda_2^*, G^*)\} = 0$.

We first know that, when $\Lambda_1 = \Lambda_1^*$ and $\Lambda_2 = \Lambda_2^*$, then $S(t \mid A,Z) = S^*(t \mid A,Z)$. Plug in the true value of Λ_1^* and Λ_2^* to (5.32), we can have:

$$\begin{aligned} E\{M_{01}(t) \mid A,Z\} &= \int_0^t G^*(u \mid A) S^*(t \mid 1,Z) d\Lambda_1^*(t \mid 1,Z) \\ &\quad - \int_0^t G^*(u \mid A) S^*(u \mid A,Z) e^{\beta_1 * A} d\Lambda_{01}^*(u) \\ &= E\{M_{01}^*(t) \mid A,Z\} \end{aligned} \tag{C.40}$$

then,

$$\begin{aligned} E\{U_{012}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1^*, \Lambda_2^*, G^*)\} &= \int_0^\tau E\left[\frac{E\{A dM_{01}^*(t) \mid A,Z\}}{\pi}\right] \\ &\quad - \int_0^\tau E\left[\frac{E\{A dM_{01}^*(t) \mid A,Z\}}{\pi}\right] \\ &\quad + \int_0^\tau E\left[E\{dM_{01}^*(t) \mid A=1,Z\}\right] \\ &= \int_0^\tau E\left[dM_{01}^{1^*}(t)\right] = 0 \end{aligned} \tag{C.41}$$

Same idea, we can have:

$$\begin{aligned} &E\{U_{011}(\beta_1^*, \Lambda_{01}^*; \pi, \Lambda_1^*, \Lambda_2^*, G^*)\} \\ =&E\left[E\left\{w dM_{01}^*(t) \mid A,Z\right\} - E\left\{w dM_{01}^*(t) \mid A,Z\right\} + E\left\{dM_{01}^*(t) \mid 1,Z\right\} + E\left\{dM_{01}^*(t) \mid 0,Z\right\}\right] \\ =&E\left\{dM_{01}^{1^*}(t) \mid Z\right\} + E\left\{dM_{01}^{0^*}(t) \mid Z\right\} = 0 \end{aligned} \tag{C.42}$$

Derive the estimating equation for λ_{01} and β_1

We only take U_{01} as the example, the rest two can follow the derivation. We first have,

$$\begin{aligned}
U_{011}(t) = \int_0^t \Big[& wdN_{01}(u) - wde^{\beta_1 A} Y_0(u)\Lambda_{01}(u) \\
& - wG(u,A)S(u,A)d\Lambda_1(u,A) + wdG(u,A)S(u,A)e^{\beta_1 A}\Lambda_{01}(u) \\
& + G(u,1)S(u,1)d\Lambda_1(u,1) - dG(u,1)S(u,1)e^{\beta_1}\Lambda_{01}(u) \\
& + G(u,0)S(u,0)d\Lambda_1(u,0) - dG(u,0)S(u,0)\Lambda_{01}(u)\Big] \\
= \int_0^t \Bigg[& w\{dN_{01}(u) - G(u,A)S(u,A)d\Lambda_1(u,A)\} + \sum_{a=0,1} G(u,a)S(u,a)d\Lambda_1(u,a) \\
& - d\left\{ we^{\beta_1 A}Y_0(u) - wG(u,A)S(u,A)e^{\beta_1 A} + \sum_{a=0,1} G(u,a)S(u,a)e^{\beta_1 a} \right\}\Lambda_{01}(t)\Bigg]
\end{aligned} \tag{C.43}$$

Solving

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n U_{011,i}(t) = \frac{1}{n}\sum_{i=1}^n \Bigg[& \int_0^t w_i\{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\
& + \sum_{a=0,1} G_i(u,a)S_i(u,a)d\Lambda_{1i}(u,a) \\
& - d\{\, w_i e^{\beta_1 A}Y_{0i}(u) - w_i G_i(u,A_i)S_i(u,A_i)e^{\beta_1 A_i} \\
& + \sum_{a=0,1} G_i(u,a)S_i(u,a)e^{\beta_1 a}\,\}\Lambda_{01}(t)\Bigg] = 0
\end{aligned} \tag{C.44}$$

We then plug (5.43) to (C.44), we have,

$$\begin{aligned}
d\Lambda_{01}(t) = \sum_{i=1}^n [\, & w_i\{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\
& - \sum_{a=0,1} G(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\,] \times \frac{1}{n\cdot \mathcal{S}_{01}^{(0)}(u;\beta_1,\pi,\Lambda_1,\Lambda_2,G)},
\end{aligned} \tag{C.45}$$

then,

$$\begin{aligned}\Lambda_{01}(t) &= \int_0^t d\Lambda_{01}(t) \\ &= \int_0^t \sum_{i=1}^{n} [\, w_i \{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\ &\qquad - \sum_{a=0,1} G(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\,] \times \frac{1}{n \cdot \mathcal{S}_{01}^{(0)}(u;\beta_1,\pi,\Lambda_1,\Lambda_2,G)} \end{aligned} \tag{C.46}$$

We further have,

$$\begin{aligned}U_{012}(t) &= \int_0^{\tau} \Big[wAdN_{01}(u) - wAde^{\beta_1 A}Y_0(u)\Lambda_{01}(u) \\ &\qquad - wAG(u,A)S(u,A)d\Lambda_1(u,A) + wAdG(u,A)S(u,A)e^{\beta_1 A}\Lambda_{01}(u) \\ &\qquad + G(u,1)S(u,1)d\Lambda_1(u,1) - dG(u,1)S(u,1)e^{\beta_1}\Lambda_{01}(u)\Big] \\ &= \int_0^t \Big[wA\{dN_{01}(u) - G(u,A)S(u,A)d\Lambda_1(u,A)\} \\ &\qquad\quad + G(u,1)S(u,1)d\Lambda_1(u,1) \\ &\qquad\quad - d\{\, wAe^{\beta_1 A}Y_0(u) - wAG(u,A)S(u,A)e^{\beta_1 A} \\ &\qquad\qquad + G(u,1)S(u,1)e^{\beta_1}\,\}\Lambda_{01}(t)\Big] \end{aligned} \tag{C.47}$$

Solving the estimating equation and plug in (C.45), we have,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} U_{012,i}^{AIPW}(t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^t w_i A_i \{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\ + G_i(u,1)S_i(u,1)d\Lambda_{1i}(u,1)\Bigg] \\ - \frac{1}{n}\sum_{i=1}^{n}\Bigg(d\{\, w_i A_i e^{\beta_1 A_i} Y_{0i}(u) - w_i A_i G_i(u,A_i)S_i(u,A_i)e^{\beta_1 A_i} \\ + G_i(u,1)S_i(u,a)e^{\beta_1 a}\,\} \\ \times \sum_{i=1}^{n}[\, w_i\{dN_{01i}(u) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\ - \sum_{a=0,1} G(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\,] \\ \times \frac{1}{n\cdot \mathcal{S}_{01}^{(0)}(u;\beta_1,\pi,\Lambda_1,\Lambda_2,G)}\Bigg)\end{aligned} \tag{C.48}$$

Noticed that,

$$\begin{aligned}&\mathcal{S}_{01}^{(1)}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G) \\ &= \frac{1}{n}\sum_{i=1}^{n}\Big\{w_i \exp(\beta_1 A_i)A_i^{\ell} R_{0i}(t,S,G) + \exp(\beta_1)G_i(t,1)S_i(t,1)\Big\}\end{aligned} \tag{C.49}$$

then we have the estimating equation for β_1,

$$\begin{aligned}U_{012,n}(\beta_1;\pi,\Lambda_1,\Lambda_2,G) = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^{\tau} w_i\big\{A_i - \bar{\mathcal{A}}_{01}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G)\big\} \\ \times\{dN_{01i}(t) - G_i(u,A_i)S_i(u,A_i)d\Lambda_{1i}(u,A_i)\} \\ + \int_0^{\tau}\sum_{a=0,1}\big\{a - \bar{\mathcal{A}}_{01}(t;\beta_1,\pi,\Lambda_1,\Lambda_2,G)\big\} \\ \times G_i(u,a)S_i(u,a)d\Lambda_{1i}(u,a)\Bigg]\end{aligned} \tag{C.50}$$

Bibliography

[1] Danilo Alvares, Sebastien Haneuse, Catherine Lee, and Kyu Ha Lee. Semicomprisks: an R package for the analysis of independent and cluster correlated semi competing risks data. *The R journal*, 11(1):376, 2019.

[2] Per K Andersen, Elisavet Syriopoulou, and Erik T Parner. Causal inference in survival analysis using pseudo observations. *Statistics in medicine*, 36(17):2669–2681, 2017.

[3] Per Kragh Andersen, Lars Sommer Hansen, and Niels 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.

[4] Kaarin J Anstey, Holly A Mack, and Nicolas Cherbuin. Alcohol consumption as a risk factor for dementia and cognitive decline: meta analysis of prospective studies. *The American journal of Geriatric psychiatry*, 17(7):542–555, 2009.

[5] Norman E Breslow and Jon A Wellner. Weighted likelihood for semiparametric models and two phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics*, 34(1):86–102, 2007.

[6] Jenny Bryan, Zhuo Yu, and Mark J Van Der Laan. Analysis of longitudinal marginal structural models. *Biostatistics*, 5(3):361–380, 2004.

[7] Weixin Cai and Mark J van der Laan. One-step targeted maximum likelihood estimation for time to event outcomes. *Biometrics*, 76(3):722–733, 2020.

[8] Claude Chastang, David Byar, and Steven Piantadosi. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Statistics in Medicine*, 7(12):1243–1255, 1988.

[9] Kani Chen, Zhezhen Jin, and Zhiliang Ying. Semiparametric analysis of transformation models with censored data. *Biometrika*, 89(3):659–668, 2002.

[10] Oliver Dukes, Torben Martinussen, Eric J Tchetgen Tchetgen, and Stijn Vansteelandt. On doubly robust estimation of the hazard difference. *Biometrics*, 75(1):100–109, 2019.

[11] Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509, 1999.

[12] Jason P Fine, Hongyu Jiang, and Rick Chappell. On semi competing risks data. *Biometrika*, 88(4):907–919, 2001.

[13] Evelyn Fix and Jerzy Neyman. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 23(3):205–241, 1951.

[14] Walter Gander and Walter Gautschi. Adaptive quadrature—revisited. *BIT Numerical Mathematics*, 40(1):84–101, 2000.

[15] WG Havercroft and Vanessa Didelez. Simulating from marginal structural models with time dependent confounding. *Statistics in medicine*, 31(30):4190–4206, 2012.

[16] Miguel A Hernán, Babette Brumback, and James M Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448, 2001.

[17] Miguel A Hernán and James M Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2021.

[18] Jiayi Hou, Anthony Paravati, Jue Hou, Ronghui Xu, and James Murphy. High dimensional variable selection and prediction under competing risks with application to seer medicare linked data. *Statistics in medicine*, 37(24):3486–3502, 2018.

[19] Jue Hou, Jelena Bradic, and Ronghui Xu. Treatment effect estimation under additive hazards models with high dimensional confounding. *Journal of the American Statistical Association*, pages 1–16, 2021.

[20] Yen Tsung Huang. Causal mediation of semicompeting risks. *Biometrics*, 77(4):1143–1154, 2021.

[21] Harindra Jayasekara, Dallas R English, Robin Room, and Robert J MacInnis. Alcohol consumption over time and risk of death: a systematic review and meta analysis. *American journal of epidemiology*, 179(9):1049–1059, 2014.

[22] Ina Jazić, Stephanie Lee, and Sebastien Haneuse. Estimation and inference for semi competing risks based on data from a nested case control study. *Statistical methods in medical research*, 29(11):3326–3339, 2020.

[23] Fei Jiang and Sebastien Haneuse. Simulation of semicompeting risk survival data and estimation based on multistate frailty model. *Harvard University Biostatistics Working Paper Series*, 2015.

[24] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*. John Wiley & Sons, 2011.

[25] Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.

[26] Tony Lancaster and Stephen Nickell. The analysis of re employment probabilities for the unemployed. *Journal of the Royal Statistical Society: Series A (General)*, 143(2):141–152, 1980.

[27] Ellen Melbye Langballe, Helga Ask, Jostein Holmen, Eystein Stordal, Ingvild Saltvedt, Geir Selbæk, Arvid Fikseaunet, Sverre Bergh, Per Nafstad, and Kristian Tambs. Alcohol consumption and risk of dementia up to 27 years later in a large, population based sample: the hunt study, norway. *European journal of epidemiology*, 30(9):1049–1056, 2015.

[28] Cathy Lau Barraco and Michael E Dunn. Evaluation of a single session expectancy challenge intervention to reduce alcohol use among college students. *Psychology of Addictive Behaviors*, 22(2):168, 2008.

[29] Catherine Lee, Paola Gilsanz, and Sebastien Haneuse. Fitting a shared frailty illness death model to left-truncated semi competing risks data to examine the impact of education level on incident dementia. *BMC Medical Research Methodology*, 21(1):1–13, 2021.

[30] Christine M Lee, Melissa A Lewis, and Clayton Neighbors. Preliminary examination of spring break alcohol use and related consequences. *Psychology of Addictive Behaviors*, 23(4):689, 2009.

[31] Kyu Ha Lee, Sebastien Haneuse, Deborah Schrag, and Francesca Dominici. Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):253–273, 2015.

[32] Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

[33] Jerry J Maples, Susan A Murphy, and William G Axinn. Two level proportional hazards models. *Biometrics*, 58(4):754–763, 2002.

[34] Richard Mayeux, Ann M Saunders, Steven Shea, Suzanne Mirra, Denis Evans, Allen D Roses, Bradley T Hyman, Barbara Crain, Ming Xin Tang, and Creighton H Phelps. Utility of the apolipoprotein e genotype in the diagnosis of alzheimer's disease. *New England Journal of Medicine*, 338(8):506–511, 1998.

[35] Daniel F McCaffrey, Greg Ridgeway, and Andrew R Morral. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4):403, 2004.

[36] Jim McCambridge and Richard Saitz. Rethinking brief interventions for alcohol in general practice. *Bmj*, 356, 2017.

[37] Luís Meira Machado and Marta Sestelo. Estimation in the progressive illness death model: A nonexhaustive review. *Biometrical Journal*, 61(2):245–263, 2019.

[38] Erica EM Moodie, David A Stephens, and Marina B Klein. A marginal structural model for multiple outcome survival data: assessing the impact of injection drug use on several causes of death in the canadian co infection cohort. *Statistics in Medicine*, 33(8):1409–1425, 2014.

[39] Susan A Murphy and Aad W Van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465, 2000.

[40] Jerzy S Neyman. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Annals of Agricultural Sciences*, 10:1–51, 1923.

[41] Gert G Nielsen, Richard D Gill, Per Kragh Andersen, and Thorkild IA Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian journal of Statistics*, pages 25–43, 1992.

[42] Ruth Peters, Jean Peters, James Warner, Nigel Beckett, and Christopher Bulpitt. Alcohol, dementia and cognitive decline in the elderly: a systematic review. *Age and ageing*, 37(5):505–512, 2008.

[43] Denise Rava. *Survival Analysis and Causal Inference: from Marginal Structural Cox to Additive Hazards Model and beyond*. Ph.D. Thesis, University of California, San Diego, 2021.

[44] Emilie T Reas, Gail A Laughlin, Donna Kritz Silverstein, Elizabeth Barrett Connor, and Linda K McEvoy. Moderate, regular alcohol consumption is associated with higher cognitive function in older community dwelling adults. *The journal of prevention of Alzheimer's disease*, 3(2):105, 2016.

[45] John R Rice. A metalgorithm for adaptive quadrature. *Journal of the ACM (JACM)*, 22(1):61–82, 1975.

[46] James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[47] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

[48] Andrea Rotnitzky and James Robins. Inverse probability weighted estimation in survival analysis. *Encyclopedia of biostatistics*, 4:2619–2625, 2005.

[49] Jason Roy, Kirsten J Lum, and Michael J Daniels. A bayesian nonparametric approach to marginal structural models for point treatments and a continuous or survival outcome. *Biostatistics*, 18(1):32–47, 2017.

[50] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[51] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[52] Youyi Shu, John P Klein, and Mei Jie Zhang. Asymptotic theory for the Cox semi Markov illness death model. *Lifetime Data Analysis*, 13(1):91–117, 2007.

[53] Tom Skaria, Katharyn Jean Mitchell, Olga Vogel, Thomas Wälchli, Max Gassmann, and Johannes Vogel. Blood pressure normalization independent cardioprotective effects of endogenous, physical activity induced cgrp (calcitonin gene related peptide) in chronically hypertensive mice. *Circulation research*, 125(12):1124–1140, 2019.

[54] Mats J Stensrud, Jessica G Young, Vanessa Didelez, James M Robins, and Miguel A Hernán. Separable effects for causal inference in the presence of competing events. *Journal of the American Statistical Association*, pages 1–9, 2020.

[55] Eric J Tchetgen Tchetgen and James Robins. On parametrization, robustness and sensitivity analysis in a marginal structural cox proportional hazards model for point exposure. *Statistics & Probability Letters*, 82(5):907–915, 2012.

[56] Eric J Tchetgen Tchetgen, Stefan Walter, Stijn Vansteelandt, Torben Martinussen, and Maria Glymour. Instrumental variable estimation in a survival context. *Epidemiology (Cambridge, Mass.)*, 26(3):402, 2015.

[57] Eric J Tchetgen Tchetgen, Isabel R Fulcher, and Ilya Shpitser. Auto G computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021.

[58] Florin Vaida. Parameter convergence for EM and MM algorithms. *Statistica Sinica*, pages 831–840, 2005.

[59] Florin Vaida and Ronghui Xu. Proportional hazards model with random effects. *Statistics in medicine*, 19(24):3309–3324, 2000.

[60] Mark J Van der Laan and James M Robins. *Unified methods for censored longitudinal data and causality*, volume 5. Springer, 2003.

[61] Tyler J VanderWeele. Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)*, 22(4):582, 2011.

[62] Lee Jen Wei, Danyu Y Lin, and Lisa Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, 84(408):1065–1073, 1989.

[63] Lan Wen, Jessica G Young, James M Robins, and Miguel A Hernán. Parametric g formula implementations for causal survival analyses. *Biometrics*, 2021.

[64] Daniel Westreich, Stephen R Cole, Jessica G Young, Frank Palella, Phyllis C Tien, Lawrence Kingsley, Stephen J Gange, and Miguel A Hernán. The parametric g formula to estimate the effect of highly active antiretroviral therapy on incident aids or death. *Statistics in medicine*, 31(18):2000–2009, 2012.

[65] Jeff D Williamson, Nicholas M Pajewski, Alexander P Auchus, R Nick Bryan, Gordon Chelune, Alfred K Cheung, Maryjo L Cleveland, Laura H Coker, Michael G Crowe, William C Cushman, et al. Effect of intensive vs standard blood pressure control on probable dementia: a randomized clinical trial. *Jama*, 321(6):553–561, 2019.

[66] CF Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of statistics*, pages 95–103, 1983.

[67] Jinfeng Xu, John D Kalbfleisch, and Beechoo Tai. Statistical analysis of illness death processes and semicompeting risks data. *Biometrics*, 66(3):716–725, 2010.

[68] Ronghui Xu, Florin Vaida, and David P Harrington. Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. *Statistica Sinica*, 19(2):819, 2009.

[69] Yanxun Xu, Daniel Scharfstein, Peter Müller, and Michael Daniels. A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks. *Biostatistics*, 23(1):34–49, 2022.

[70] Jessica G Young, Mats J Stensrud, Eric J Tchetgen Tchetgen, and Miguel A Hernán. A causal framework for classical statistical estimands in failure time settings with competing events. *Statistics in medicine*, 39(8):1199–1236, 2020.

[71] Yiran Zhang. *Causal inference for competing risks and semi competing risks data*. Ph.D. Thesis, University of California, San Diego, 2022.

[72] Yiran Zhang and Ronghui Xu. *cmprskcoxmsm: Use IPW to Estimate Treatment Effect under Competing Risks*, 2021. R package version 0.2.1.

[73] Yiran Zhang and Ronghui Xu. Marginal structural illness death models for semi competing risks data. *arXiv preprint arXiv:2204.10426*, 2022.