2021

Statistical methods for the restricted mean survival time and lifetime risk

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

STATISTICAL METHODS FOR THE RESTRICTED MEAN SURVIVAL TIME AND LIFETIME RISK

by

SARAH C. CONNER

B.S., George Washington University, 2011
M.P.H., Columbia University, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader

Ludovic Trinquart, PhD
Associate Professor of Biostatistics

Second Reader

Martin G. Larson, ScD
Research Professor of Biostatistics

Third Reader

Alexa Beiser, PhD
Professor of Biostatistics

Fourth Reader

Michael P. LaValley, PhD
Professor of Biostatistics

ACKNOWLEDGMENTS

First, I would like to sincerely thank my advisor, Ludovic Trinquart, for his many years of dedicated mentorship and support. Without your encouragement, I would have never pursued this PhD or made it as far as I have today. I am still in awe at the journey we have had from Cochrane (cpt fich) to BU, and at all we have accomplished together in this dissertation. Your unrelenting guidance has made me a better writer and independent researcher, and built my confidence as a statistician. I would also like to thank Alexa Beiser for her mentorship at FHS, teaching me to be a biostatistician, and continuing to support me on my dissertation committee. Thank you to my committee members for making time to provide insightful feedback for this research: Martin Larson, Michael LaValley, and Michael Pencina. Thank you to Emelia Benjamin for assisting my professional development, helping me see the application beyond the data, and supporting my F31 application.

Second, I am so grateful for the friends I made at BU. Alicia, I don't think a day goes by where we do not talk. I cannot express how thankful I am for our friendship and that we had each other along the way, from our first semester classes to our defenses and job interviews - we did it! Isabelle, my RMST-confidant: I could not have embarked on this dissertation phase without you and our cubicle, tea kettles, and life discussions. Thank you for always being there for me when I am over-thinking and encouraging me to hit the "submit" button. Kendra and Adrienne, I am so happy we got to work together as FHS RAs, but even happier that after many SAS macros and cookie trips later, we continue to be friends. And thank you to all of my BU classmates, especially Thomas, Elise, Ariel, Adlin, Yineng, Zach, Qiuxi, Ben, Sarah, Taylor, and Katie.

Finally, I am tremendously thankful to my family and friends outside of school.

Thank you to my parents, Patti and Kip, and my sister Lindsay, for teaching me hard work and patience. Thank you to my husband, Kevin, for joining me on this journey in Boston and grounding me. I cannot tell you how appreciative I am for our home together with Shelby.

Additional Acknowledgments

I would like to extend my sincere gratitude to Katia Bulekova for her continued support with the Shared Computing Cluster and patiently helping me improve the efficiency of my R code. Thank you to the Framingham Heart Study participants, who made these analyses possible.

This work was funded by the National Heart, Lung, and Blood Institute (NHLBI, F31HL145904) and the National Institute of General Medical Sciences (NIGMS, T32GM74905-14). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

STATISTICAL METHODS FOR THE RESTRICTED MEAN SURVIVAL TIME AND LIFETIME RISK

SARAH C. CONNER

Boston University, Graduate School of Arts and Sciences, 2021

Major Professor: Ludovic Trinquart, PhD

ABSTRACT

In observational studies or non-randomized studies with censored data, associations are commonly measured with adjusted hazards ratios from multivariable proportional hazards models. The difference in restricted mean survival times (RMST) up to a pre-specified time point is an alternative measure that offers a clinically meaningful interpretation. However, existing methods for the difference in RMST do not address statistical challenges common in observational research, such as time-varying confounding or competing risks. Several regression-based methods exist to estimate an adjusted difference in RMSTs, but they digress from the model-free method of taking the area under the survival function. The lifetime risk is another widely reported metric for disease incidence in the presence of censoring and the competing risk of death. The lifetime risk is usually estimated by the Aalen-Johansen estimator allowing delayed entry, but there are no methods to directly model the lifetime risk. These gaps in statistical methodology may limit the applications of the difference in RMST and lifetime risk. By addressing these methodological gaps, we aim to promote the reporting of the RMST and lifetime risk in observational and non-randomized studies.

In this dissertation, we present novel statistical methods for the difference in

RMST, difference in restricted mean time lost (RMTL), and lifetime risk. First, we introduce an estimator and the associated variance for the adjusted difference in RMST using inverse probability weighting (IPW). Second, we demonstrate how to estimate the adjusted difference in RMST accounting for time-varying confounding with the parametric g-formula. We also demonstrate how to address missing data challenges with sequential multiple imputation. Third, we introduce an IPW-based estimator and the associated variance for the adjusted difference in RMTL in the presence of competing risks. We also propose a regression model of the RMTL conditional on covariates with inverse probability of censoring weighting. Finally, we present a new regression model for the lifetime risk using pseudo-observations of the Aalen-Johansen estimator. In simulation studies, we demonstrate our proposed estimators are unbiased and perform well under various settings. We illustrate the methods for incident coronary heart disease and atrial fibrillation in the Framingham Heart Study.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AF	Atrial Fibrillation
AIC	Akaike's Information Criterion
BMI	Body Mass Index
CHD	Coronary Heart Disease
CI	Confidence Interval
HR	Hazards Ratio
IPW	Inverse Probability Weighting
IPCW	Inverse Probability of Censoring Weighting
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
RMST	Restricted Mean Survival Time
RMTL	Restricted Mean Time Lost

CHAPTER 1

Introduction

1.1 RESTRICTED MEAN LIFETIME MEASURES

In observational studies and non-randomized clinical trials, the endpoint of interest is frequently the time to event, such as survival time or time of disease incidence, denoted as T. The mean survival time, $E(T)$, is not estimable because of right-censoring on T. However, it is possible to estimate the restricted mean survival time (RMST) up to time τ,

$$\mu(\tau) = E(T \wedge \tau) = \int_0^\tau S(t)dt$$

where $S(t) = P(T > t)$ is the survival function. If death is the outcome, the RMST is interpreted as the life expectancy until time τ. This interpretation is clinically meaningful and may be advantageous when communicating results.(Trinquart et al., 2016; Weir et al., 2019) The RMST can be estimated non-parametrically by taking the area under the Kaplan-Meier estimator of the survival function up to τ,

$$\hat{\mu}(\tau) = \int_0^\tau \hat{S}(t)dt$$

where $\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{\Delta N_\cdot(t_j)}{Y(t_j)}\right)$, $t_j : t_1, \ldots, t_D$ are the ordered, unique event times prior to τ, $N_\cdot(t)$ is the number of events by t, $\Delta N_\cdot(t)$ is the increment $N_\cdot(t) - N_\cdot(t-)$, and $Y(t)$ is the number at risk at t.(Kaplan & Meier, 1958) The difference in RMST between groups is obtained by subtracting subgroup-specific RMSTs.

In an observational or non-randomized clinical trial setting, covariate adjust-

ment is necessary. To adjust for confounders, the RMST can be modeled conditional on covariates. Karrison (1987), Zucker (1998) and Chen & Tsiatis (2001) proposed variations of proportional hazards models stratified by exposure group. (Andersen & Pohar Perme, 2010) and (Tian et al., 2014) proposed generalized linear models for the RMST, estimated with pseudo-observation and inverse probability of censoring weighting (IPCW), respectively. Royston et al. (2015) proposed a flexible parametric model for the log cumulative hazard with restricted cubic splines. However, regression methods are inconsistent with the original approach of taking the area under the survival curve. In Chapter 2 of this dissertation, I develop an adjusted estimator of the difference in RMST to account for confounding using inverse probability weighting (IPW). This approach is consistent with the original idea of taking the area under the Kaplan-Meier curve.

Furthermore, values of covariates may be measured at multiple timepoints and vary over time. One can incorporate time-varying covariates in a Cox proportional hazards model or using the g-methods, i.e. the parametric g-formula, inverse probability weighting of marginal structural models, and g-estimation.(Robins, 1986; Witteman et al., 1998; Taubman et al., 2009; Danaei et al., 2016; Naimi et al., 2017) However, statistical methods to estimate the RMST while accounting for time-varying confounding are limited.(Hagiwara et al., 2019) In Chapter 3 of this dissertation, I demonstrate how to derive the difference in RMST with the parametric g-formula to account for time-varying confounding. I also demonstrate how to address missing data challenges, which arise with repeated examinations, using sequential multiple imputation.

The restricted mean time lost (RMTL) is a closely related measure to the RMST. In the competing risk setting with K causes of failure $\epsilon \in (1, \ldots, K)$, the cause-

specific RMTL is the area under the cause-specific cumulative incidence function until time τ,

$$\nu_k(\tau) = \int_0^\tau F_k(t)dt.$$

The RMTL gives the lost life expectancy due to cause k until time τ, or in the disease incidence setting, the lost disease-free time until τ. In the absence of competing risks, $L(\tau) = \tau - \mu(\tau)$. The cause-specific RMTL can be estimated by taking the area under the Aalen-Johansen estimator of the cause-specific cumulative incidence function until time τ,

$$\hat{\nu}_k(\tau) = \int_0^\tau \hat{F}_k(t)dt$$

where $N_k(t)$ denotes events of cause k, $Y(t)$ is the number at risk, and $\hat{S}(t)$ is the Kaplan-Meier estimator of being alive and event-free at t. The difference in RMTL between groups is obtained by subtracting subgroup-specific RMTLs.

The RMTL can also be modeled conditional on covariates to adjust for confounders. Similar to the RMST, modeling approaches are inconsistent with the original approach of taking the area under the cumulative incidence curve.(Andersen, 2013) Despite the high prevalence of competing risks, most statistical development has focused on the RMST instead of the cause-specific RMTL. In Chapter 4 of this dissertation, I develop an adjusted estimator of the RMTL to account for confounding using IPW. I also develop a regression model for the RMTL conditional on covariates using inverse probability of censoring weighting (IPCW).

1.2 LIFETIME RISK

Another measure commonly reported in competing risks data is the residual lifetime risk. The residual lifetime risk is the long-term, cumulative risk for developing a disease after a certain index age, τ_0.(Beiser et al., 2000) Many clinical applications from longitudinal cohort studies have examined the lifetime risk of various diseases.(Feuer et al., 1993; Lloyd-Jones et al., 2004, 2006; Seshadri et al., 2006; Berry et al., 2012; Staerk et al., 2018; Brookmeyer & Abdalla, 2019) Risk measures over longer time frames have been shown to motivate lifestyle behavior, particularly among younger individuals or those who may not be at short-term risk of disease.(Lloyd-Jones et al., 2004) Thus, correctly quantifying the lifetime risk is essential.

The lifetime risk is estimated by using age as the time scale and by accounting for the competing risk of death, in which an individual may die before experiencing disease. The lifetime risk is estimated at a fixed point, as the cumulative incidence function for the disease estimated at an advanced age, e.g. 95 years. The residual lifetime risk can be estimated non-parametrically by the Aalen-Johansen estimator,

$$\hat{F}_1(\tau|T > \tau_0) = \sum_{\tau_0 < t_j \leq \tau} \hat{S}(t_{j-1}) \frac{\Delta N_1(t_j)}{Y(t_j)}$$

where $\tau_0 < t_j \leq \tau$ are the unique ages at failure and the risk set $Y(t)$ is modified to allow delayed entries, $\hat{S}(t) = \prod_{\tau_0 < t_j \leq t} \left(1 - \frac{\Delta N(t_j)}{Y(t_j)}\right)$. Beiser et al. (2000); Gaynor et al. (1993); Du (2010); Andersen et al. (1993)

The difference in residual lifetime risk can be obtained by subtracting subgroup-specific estimates of the lifetime risk. However, multivariable prediction of lifetime risk has yet to be demonstrated.(Beiser et al., 2000; Allignol et al.,

2011) In Chapter 5 of this dissertation I address this gap by developing a regression model for the lifetime risk conditional on covariates using the pseudo-observation approach.

1.3 NEED FOR STATISTICAL DEVELOPMENT BEYOND THE HAZARDS RATIO

The objectives of this dissertation are to express associations between risk factors and the different in RMST, RMTL, or lifetime risk. These measures of association can complement measures of association which are more frequently rep rorted in analyses of lifetime data, such as the hazards ratio (HR) or subdistribution HR. Furthermore, we can make individualized predictions based on these alternative measures of association.

Analyses of lifetime data typically involve the hazard function $\lambda(t)$, which is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P\{T \in (t, t+\Delta t) | T \geq t\}}{\Delta t}.$$

This is because the hazard is still valid in the presence of right censoring.(Beyersmann et al., 2009) Consequently, associations between covariates and time-to-event outcomes are commonly expressed with HRs estimated by Cox proportional hazards models. In the case of competing risks, there are two hazard frameworks to express associations. The first is the subdistribution HR, where the subdistribution hazard for cause k, $\lambda_k(t)$, is defined as

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P\{T \in (t, t+\Delta t), \epsilon = k | T \geq t \cup T < t, \epsilon \neq k\}}{\Delta t}.$$

Subdistribution HRs are commonly estimated by Fine-Gray proportional hazards

models. The second framework is the cause-specific HR, where the cause-specific hazard for cause k, $\alpha_k(t)$, is defined as

$$\alpha_k(t) = \lim_{\Delta t \to 0} \frac{P\{T \in (t, t+\Delta t), \epsilon = k | T \geq t\}}{\Delta t}.$$

The cause-specific HR is commonly estimated with a Cox proportional hazards model for event k, while treating other event types as censored.

Proportional hazards and subdistribution hazards models permit adjustment for covariates, and an array of statistical developments have addressed other challenges including time-varying covariates and delayed entry. For this reason, proportional hazards and subdistribution models are frequently the default choice for analyses of lifetime data in observational and nonrandomized clinical trials. However, the HR depends on follow-up duration if the proportional hazards assumption is violated. In the case of nonproportional hazards, reporting only the HR may result in incorrect conclusions. Furthermore, the interpretation of survival benefit using HRs is challenging.(Blagoev et al., 2012; Trinquart et al., 2016; Weir et al., 2019) The HR is a relative measure and does not communicate any information about the absolute effect, including the reference group.(Trinquart et al., 2017) Thus, conventional methods for analyzing lifetime data may limit risk communication.

Furthermore, the HR is averaged over time, which induces differential selection bias even for randomized treatments.(Hernán, 2010) Consequently, the HR cannot offer a causal interpretation even when adjusted for confounders. In contrast, the difference in RMST or RMTL estimated with IPW and the parametric g-formula can offer causal interpretations under certain identifiability assumptions. Thus, our approaches for the difference in RMST and RMTL using IPW and the

parametric g-formula can be used for causal inference with certain assumptions. However, we do not make these causal assumptions in the illustrative examples in this dissertation.

While the difference in RMST is gaining popularity and the lifetime risk of disease is frequently reported, the HR remains the default measure of association for lifetime data. The development of new statistical methods for the RMST and lifetime risk will allow greater flexibility to address these common statistical challenges. In order to promote the application and reporting of the RMST and lifetime risk, we are developing the tools to address statistical challenges that are already addressed for the HR and Cox model.

1.4 INCIDENT ATRIAL FIBRILLATION

Our statistical developments are motivated by studying incident atrial fibrillation (AF) in the Framingham Heart Study. AF is the most common arrhythmia and affects about 3-6 million Americans, a number that is expected to rise to 12-15 million by 2050.(Misayaka et al., 2006; Benjamin et al., 2017) AF is associated with increased risks of myocardial infarction, heart failure, stroke, dementia, and death.(Benjamin et al., 2017) Primary and secondary prevention of AF are key. The use of metrics such as the RMST and lifetime risk may help clinicians and patients to communicate and understand risk. These metrics may also be helpful to clinical practice guidelines.(Hindricks et al., 2020)

AF is subject to the competing risk of death, in which the occurrence of death precludes observing incident AF. The overall objective of this dissertation is to improve statistical methods for estimating the RMST and lifetime risk, and improve our understanding of AF epidemiology with these new methods.

1.5 PROPOSED METHODS

In Chapter 2, we present an estimator for the adjusted restricted mean survival time using inverse probability weighting (IPW). In Chapter 3, we show how to estimate the restricted mean survival time while accounting for time-varying confounders using the parametric g-formula. In Chapter 4, we present an estimator for the adjusted restricted mean time lost in the presence of competing risks using IPW, and a regression model estimated with IPCW. In Chapter 5, we present a regression model for the lifetime risk using pseudo-observations.

CHAPTER 2

Adjusted restricted mean survival times

2.1 BACKGROUND

In observational studies with censored data, exposure-outcome associations are commonly measured with adjusted HRs from multivariable Cox proportional hazards models. The difference in RMSTs up to a pre-specified time point, τ, is an alternative measure that offers a clinically meaningful interpretation.(Royston & Parmar, 2011, 2013; Uno et al., 2014; Zhao et al., 2016) In the observational setting, several methods have been proposed to adjust the difference in RMST for potential confounders. The methods of Karrison (1987), Zucker (1998), and Chen & Tsiatis (2001) invoke proportional hazards models stratified by exposure group. These methods assume proportional hazards for the covariate effects. Other regression-based approaches, such as Royston & Parmar (2002)'s flexible parametric model, Andersen & Pohar Perme (2010)'s pseudo-observation model, Tian et al. (2014)'s IPCW or ANCOVA-type model, or Díaz et al. (2018)'s targeted minimum loss based estimation rely on regression-based models, which is inconsistent with the model-free method of estimating RMST in the unadjusted setting, i.e. taking the area under the survival function.

Inverse probability weighting has been previously used when estimating the adjusted RMST. Wei (2008) first applied inverse probability weighting to the Nelson-Aalen estimator via stratified Cox models. Schaubel & Wei (2011) and Zhang & Schaubel (2011, 2012a,b) later advanced these methods to be doubly robust and incorporate dependent censoring. Although the Nelson-Aalen and Kaplan-Meier estimators are asymptotically equivalent, we focus on the Kaplan-

Meier estimator in this chapter. Unlike many previous methods, our proposed method does not rely on regression models to estimate the adjusted RMST; rather, it relies on the Kaplan-Meier estimator, and is more similar to estimating RMST in the unadjusted setting by integrating the survival function.

We introduce a new estimator of the adjusted RMST, in which we integrate an adjusted Kaplan-Meier estimator with inverse probability weighting (IPW).(Conner et al., 2019b) Then, the adjusted difference in RMSTs is the area between the two IPW-adjusted survival functions. We also provide the variance of the IPW adjusted difference in RMSTs.

2.2 INVERSE PROBABILITY WEIGHTED ADJUSTED DIFFERENCE IN RESTRICTED MEAN SURVIVAL TIMES

Let T_i denote the time to event, C_i denote the time to censoring, and $X_i = T_i \wedge C_i$. Let $(X_i; \delta_i; A_i; Z_i)$, $i = 1, \cdots, n$, denote a sample of right-censored survival data, in which δ_i is the event indicator (which takes value 1 if the observed time corresponds to an event time and 0 if right-censored), A_i is the exposure group taking values $a = 1, \cdots, A^*$ for A^* exposure groups, and Z_i is a vector of covariates of dimension $p \times 1$. At time t_j, there are d_j^a events out of Y_j^a individuals at risk in group a.

We weight subjects according to the inverse of the probability of being in their observed exposure group.(Xie & Liu, 2005) Each subject is assigned a weight, $w_i = \sum_{a=1}^{A^*} \frac{1(A_i=a)}{p_i^a}$, for which 1 is the indicator function, and $p_i^a = P(A_i = a|Z_i)$ can be deterministic (i.e., survey weights) or estimated (i.e. with a logistic model). Our methods consider the weights as fixed, but in practice we estimate weights based on the data. Xie & Liu (2005); Fitzmaurice et al. (2012); Robins et al. (2000) The

weights are then applied to the number of events and number at risk to obtain an adjusted Kaplan-Meier estimator and variance which incorporates the weights. At time t_j, the weighted d_j^a and Y_j^a in group a are given by $\tilde{d}_j^a = \sum_{i:X_i=t_j} w_i \delta_i 1(A_i = a)$ and $\tilde{Y}_j^a = \sum_{i:X_i \geq t_j} w_i 1(A_i = a)$. Then, the adjusted Kaplan-Meier estimate at time t in group a is given by $\hat{S}_a(t) = \prod_{t_j \leq t} \left[1 - \frac{\tilde{d}_j^a}{\tilde{Y}_j^a}\right]$. Xie and Liu have shown that the variance can be estimated by $\widehat{var}\left[\hat{S}_a(t)\right] = [\hat{S}_a(t)]^2 \sum_{t_j \leq t} \frac{\tilde{d}_j^a}{M_j^a(\tilde{Y}_j^a - \tilde{d}_j^a)}$, for which $M_j^a = \frac{\left[\sum_{i:T_i \geq t_j} w_i 1(A_i=a)\right]^2}{\sum_{i:T_i \geq t_j} w_i^2 1(A_i=a)}$.(Xie & Liu, 2005) We show that the IPW-adjusted RMST up to the time point τ in group a is

$$\hat{\mu}_a(\tau) = \int_0^\tau \hat{S}_{adj}^a(t)\, \mathrm{d}t, \tag{2.1}$$

and the variance of the adjusted RMST can be estimated by

$$\widehat{var}(\hat{\mu}_a(\tau)) = \sum_{j:t_j \leq \tau} \left[\sum_{i=j}^{\tau} \hat{S}_a(t_i)(t_{i+1} - t_i)\right]^2 \frac{\tilde{d}_j}{M_j(\tilde{Y}_j - \tilde{d}_j)}. \tag{2.2}$$

Proof.

The variance of the IPW adjusted RMST, $\hat{\mu}_a(\tau)$, can be written as

$$\begin{aligned}
\mathrm{var}\{\hat{\mu}_a(\tau)\} &= E\left[\hat{\mu}_a(\tau)^2\right] - \left(E\left[\hat{\mu}_a(\tau)\right]\right)^2 \\
&= E\left(\int_0^\tau \int_0^\tau \hat{S}_a(u)\hat{S}_a(v)dudv\right) - E\left(\int_0^\tau \hat{S}_a(u)du\right)E\left(\int_0^\tau \hat{S}_a(v)dv\right) \\
&= \int_0^\tau \int_0^\tau \left(E\left[\hat{S}_a(u)\hat{S}_a(v)\right] - S_a(u)S_a(v)\right)dudv.
\end{aligned} \tag{2.3}$$

Following the notation of Xie and Liu Xie & Liu (2005) we introduce $s_j = \frac{S_a(t_j)}{S_a(t_{j-1})}$

and $\hat{s}_j = 1 - \frac{\tilde{d}_j}{\tilde{Y}_j}$. It follows that

$$\begin{aligned} E\left[\hat{S}_a(u)\hat{S}_a(v)\right] &= E\left[\prod_{j:t_j\le u}\hat{s}_j \prod_{j:t_j\le v}\hat{s}_j\right] \\ &= E\left[\prod_{j:t_j\le u}\hat{s}_j^2 \prod_{j:u<t_j\le v}\hat{s}_j\right] \quad \text{if } u < v \\ &= \prod_{j:t_j\le u} E_j\left[\hat{s}_j^2\right] \prod_{j:u<t_j\le v} E_j\left[\hat{s}_j\right] \\ &= \prod_{j:t_j\le u}\left(s_j^2 + var_j(\hat{s}_j)\right) \prod_{j:u<t_j\le v} s_j \end{aligned}$$

in which E_j denotes the conditional expectation given information up to time t_j, and var_j denotes the conditional variance given information up to time t_j. We let $u = t_h$ and $v = t_k$, and have

$$\prod_{j:u<t_j\le v} s_j = s_{h+1}\cdot s_{h+2}\cdots s_k = \frac{S_a(t_{h+1})}{S_a(t_h)}\cdots\frac{S_a(t_k)}{S_a(t_{k-1})} = \frac{S_a(t_k)}{S_a(t_h)} = \frac{S_a(v)}{S_a(u)}.$$

Moreover, Xie & Liu (2005) showed that

$$\prod_{j:t_j\le u}\left(s_j^2 + var_j(\hat{s}_j)\right) = (S_a(u))^2 \prod_{j:t_j\le u}\left(1 + \frac{1-s_j}{M_j s_j}\right),$$

in which $M_j = \frac{\left[\sum_{i:T_i\ge t_j} w_i\right]^2}{\sum_{i:T_i\ge t_j} w_i^2}$. Thus,

$$\begin{aligned} E\left[\hat{S}_a(u)\hat{S}_a(v)\right] &= \left[\{S_a(u)\}^2 \prod_{j:t_j\le u}\left(1 + \frac{1-s_j}{M_j s_j}\right)\right]\frac{S_a(v)}{S_a(u)} \\ &= S_a(u)S_a(v)\prod_{j:t_j\le u}\left(1 + \frac{1-s_j}{M_j s_j}\right). \end{aligned}$$

Following the result of Xie & Liu (2005), $\frac{1}{M_j} \to 0$ under the condition that $\frac{\max_{i:T_i \geq t_j} w_i}{\sum_{i:T_i \geq t_j} w_i} \to 0$, allowing us to ignore $\frac{1}{M_j}$ terms of the second order. We then obtain the following approximation

$$
\begin{aligned}
E\left[\hat{S}_a(u)\hat{S}_a(v)\right] - S_a(u)S_a(v) &= S_a(u)S_a(v)\left(\prod_{j:t_j \leq u}\left(1 + \frac{1-s_j}{M_j s_j}\right) - 1\right) \\
&\doteq S_a(u)S_a(v)\left(\sum_{j:t_j \leq u}\frac{1-s_j}{M_j s_j}\right) \qquad (2.4)
\end{aligned}
$$

in which $\doteq$ denotes approximate equality. We plug 2.4 into 2.3 to obtain the variance of the IPW-adjusted RMST as

$$
\begin{aligned}
\text{var}\{\hat{\mu}_a(\tau)\} &= \int_0^\tau \int_0^\tau \left(E\left[\hat{S}_a(u)\hat{S}_a(v)\right] - S_a(u)S_a(v)\right) du dv \\
&\doteq \sum_{i=0}^{\tau}\sum_{l=0}^{\tau}\left[S_a(u_i)S_a(v_l)\left(\sum_{j:t_j \leq u}\frac{1-s_j}{M_j s_j}\right)(u_{i+1}-u_i)(v_{l+1}-v_l)\right] \\
&= \sum_{j:t_j \leq \tau}\left[\sum_{i=j}^{\tau} S_a(u_i)(u_{i+1}-u_i)\sum_{l=j}^{\tau} S_a(v_l)(v_{l+1}-v_l)\right]\left(\frac{1-s_j}{M_j s_j}\right) \\
&= \sum_{j:t_j \leq \tau}\left[\sum_{i=j}^{\tau} S_a(t_i)(t_{i+1}-t_i)\right]^2\left(\frac{1-s_j}{M_j s_j}\right). \qquad (2.5)
\end{aligned}
$$

Finally, we estimate Equation 2.5 by

$$
\begin{aligned}
\hat{\text{var}}\{\hat{\mu}_a(\tau)\} &= \sum_{j:t_j \leq \tau}\left[\sum_{i=j}^{\tau}\hat{S}_a(t_i)(t_{i+1}-t_i)\right]^2 \frac{1-\hat{s}_j}{M_j \hat{s}_j} \\
&= \sum_{j:t_j \leq \tau}\left[\sum_{i=j}^{\tau}\hat{S}_a(t_i)(t_{i+1}-t_i)\right]^2 \frac{1-(1-\frac{\tilde{d}_j}{\tilde{Y}_j})}{M_j(1-\frac{\tilde{d}_j}{\tilde{Y}_j})}
\end{aligned}
$$

$$= \sum_{j:t_j \leq \tau} \left[\sum_{i=j}^{\tau} \hat{S}_a(t_i)(t_{i+1} - t_i) \right]^2 \frac{\tilde{d}_j}{M_j(\tilde{Y}_j - \tilde{d}_j)}.$$

which is 2.5.

2.3 REGRESSION-BASED METHODS

We describe two regression-based methods proposed previously to estimate the difference in RMST adjusted for potential confounders: Tian et al. (2014)'s IPCW model and Andersen & Pohar Perme (2010)'s pseudo-observation method.

2.3.1 Inverse probability of censoring weighted model

Tian et al. (2014) have proposed an IPCW adjusted analysis that relates the RMST directly to an exposure of interest Z and additional covariates X, also referred to as an ANCOVA-type model. The model is given by $\hat{\mu}(\tau) = \beta' W_i = \alpha + \beta_z Z_i + \beta_x' X_i$, for which $\beta' = (\alpha, \beta_z, \beta_x')$ and $W' = (1, Z, X')$. In the case of two independent groups, exposed ($Z = 1$) and unexposed ($Z = 0$), the regression coefficient β_z is the adjusted difference in RMST between the two groups.

An estimate of β is obtained with estimating equations for RMST based on the IPCW technique to handle censored data: $S(\beta) = n^{-1} \sum_i \frac{\tilde{\Delta}_i}{\hat{G}(Y_i)} W_i (Y_i - \beta' W_i) = 0$, with $Y_i = \min(T_i, \tau)$, $\tilde{\Delta}_i = 1(Y_i \leq C_i)$, and $\hat{G}()$ the Kaplan-Meier estimator of the censoring distribution. The asymptotic variance of $\hat{\beta}$ is described in the supplementary material of Tian et al. (2014).

2.3.2 Pseudo-observation model

Andersen & Pohar Perme (2010) have proposed a pseudo-observation regression model to assess the effects of covariates on the RMST.(Andersen et al., 2004; An-

dersen & Perme, 2010; Andersen, 2013) Up to a time point τ, $\hat{\mu}(\tau)$ is the estimator of the RMST from the integrated Kaplan-Meier estimator based on all observations; $\hat{\mu}(\tau)^{(-i)}$ is the leave-one-out estimator for μ_τ based on all observations except i^{th} observation. The ith pseudo-observation is defined as $\hat{\mu}_i(\tau) = n \cdot \hat{\mu}(\tau) - (n-1) \cdot \hat{\mu}(\tau)^{(-i)}$. The mean of the pseudo-observations is an estimate of the RMST at time τ for the complete sample.Andersen & Pohar Perme (2010)

To examine the effect of an exposure Z on the RMST, while controlling for additional covariates X, we use an uncensored generalized linear model for the pseudo-observations given by $\mu_i(\tau) = \beta' W_i = \alpha + \beta_z Z_i + \beta_x' X_i$.(Andersen et al., 2004) Again, the regression coefficient β_z is the adjusted difference in RMST between groups. An estimate of β can be obtained using generalized estimating equations: $\sum_i \left(\frac{\partial}{\partial \beta}(\beta' W_i)\right)' V_i^{-1} (\mu_i(\tau) - \beta' W_i) = 0$. The covariance matrix V_i for the solution $\hat{\beta}$ can be estimated by a sandwich estimator described by Andersen et al. (2004) and Klein et al. (2008).

2.4 SIMULATION STUDY

We conducted Monte Carlo simulation studies to examine the statistical performance of the proposed method in estimating the marginal difference in adjusted RMST. We compared its statistical performance to the IPCW and pseudo-observation methods using four criteria: relative bias, mean squared error, empirical coverage rate, and the relative error of the model-based standard errors with respect to the empirical standard errors, as defined by Morris et al. (2019).

2.4.1 Data generation process

We simulated 1,000 datasets for each scenario. We adapted a previously described data-generating process.(Austin, 2009) For each subject, we first simulated 10 covariates (x_1 to x_{10}) from independent standard normal distributions. We then randomly generated an exposure status, where $Z = 1$ denotes exposed and $Z = 0$ denotes unexposed, from a Bernoulli distribution with parameter $P(Z = 1)$ defined by a logistic model, $\text{logit}(P(Z = 1)) = \beta_0 + \beta_w x_1 + \beta_w x_2 + \beta_m x_3 + \beta_m x_4 + \beta_s x_5 + \beta_s x_6 + \beta_{vs} x_7$, and we considered a range of subjects exposed and varying strengths of association between exposure and outcome. We set the intercept β_0 to generate a desired proportion of exposed subjects. We set $\beta_w = \log(1.25)$, $\beta_m = \log(1.5)$, $\beta_s = \log(1.75)$, and $\beta_{vs} = \log(2)$ to denote weak, moderate, strong, and very strong associations, respectively. Finally, we generated the time to event measured in years by using a stratified Weibull regression model.(Bender et al., 2005) We first generated a linear predictor, $\text{LP} = \beta_W x_2 + \beta_m x_4 + \beta_s x_6 + \beta_{vs} x_7 + \beta_w x_8 + \beta_m x_9 + \beta_s x_{10}$. Based on a random number u drawn from a uniform distribution on the interval 0 to 1, we then generated the times to event as $t = t_1 z + t_0(1 - z)$, with $t_0 = \left(-\frac{\log u}{\lambda_0 \exp \text{LP}}\right)^{1/\nu_0}$, and $t_1 = \left(-\frac{\log u}{\lambda_1 \exp(\beta_z + \text{LP})}\right)^{1/\nu_1}$. We set the shape and scale parameters (ν_1, λ_1) and (ν_0, λ_0) in the exposed and unexposed groups, respectively, to obtain specific time-to-event patterns. We also set β_z to reflect a desired strength of association between the exposure and the hazard of event. In this design, (x_2, x_4, x_6, x_7) are associated with both the exposure and outcome and thus are confounders. (x_1, x_3, x_5) are associated with the exposure only, and (x_8, x_9, x_{10}) are associated with the outcome only.

2.4.2 Scenarios

We allowed the following factors to vary: the sample size (250, 500, 1,000), the proportion of exposed subjects (5%, 10%, 20%, 30%, 40%, 50%), and the exposure effect $\exp(\beta_z)$ (1.25, 1.5, 2). In addition, we defined three settings for the time-to-event patterns: proportional hazards, non-proportional hazards with crossing survival curves, and non-proportional hazards with an early survival difference (Table 2.1). In the early survival difference scenarios, the HR is initially large but approaches the null value of 1 over time. The resulting survival functions are visualized in Figure A.2, and the corresponding hazard functions are available in Figure A.1. We thus examined 54 scenarios with 1000 simulated datasets of size n=1,000, n=500, and n=250 each.

2.4.3 Statistical analysis in simulated datasets

We pre-specified the time point $\tau = 10$ years for all analyses using the simulated datasets. In each simulated dataset, we estimated the difference in adjusted RMST, θ_i, between exposed and unexposed subjects. For the proposed method, we estimated the probability of being exposed using a logistic regression model of the exposure z as a function of the 7 covariates $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ that affect the exposure. This allowed us to obtain adjusted Kaplan-Meier estimates of the survival function in the exposed and unexposed groups, respectively. Each adjusted RMST was estimated by integration of the corresponding survival function up to τ. The standard error of the difference in adjusted RMST was estimated as described in Section 2.2.

We also used the IPCW and pseudo-observation methods as comparators. Both RMST models included the exposure variable z and all covariates associated with

the time to event ($x_2, x_4, x_6, x_7, x_8, x_9, x_{10}$) through an identity link function. A key difference between our proposed method and the regression-based methods is that the regression-based methods can identify covariates predictive of the outcome.

2.4.4 Assessment of statistical performance

We assessed the statistical performance of each method of estimating the marginal effect in terms relative bias, mean squared error, empirical coverage rate, and relative error of the model-based standard errors with respect to the empirical standard errors.(Morris et al., 2019) To determine the marginal difference in RMST, θ, between exposed and unexposed subjects, we generated a dataset consisting of 1,000,000 subjects. We generated the covariates (x_1 to x_{10}) and exposure z as described previously. We generated times to event t_0 first assuming that all subjects were unexposed. We then simulated times to event t_1 assuming that all subjects were exposed. For the exposed and unexposed populations separately, we estimated the RMST by integrating the Kaplan-Meier estimate of the survival function up to τ and we calculated the difference in the resulting RMSTs to obtain the marginal effect θ.

We estimated the mean difference in adjusted RMST as $\hat{\theta} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\theta}_i$. We examined four performance criteria in estimating the true marginal difference in RMST: the mean relative bias, $(\frac{\hat{\theta}-\theta}{\theta})$; the mean squared error of the estimated effect, $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_i - \theta)^2$; the empirical coverage rate, estimated as the proportion of 95% confidence intervals that covered θ; and the relative error of the model-based standard errors (ModSE) $\sqrt{\frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}\hat{V}(\hat{\theta}_i)}$ with respect to the empirical standard errors (EmpSE), $\sqrt{\frac{1}{n_{sim}-1}\sum_{i=1}^{n_{sim}}(\hat{\theta}_i - \bar{\theta})}$, defined as $\frac{ModSE-EmpSE}{EmpSE}$.(Morris et al., 2019).

2.4.5 Results

In the proportional hazard setting, our proposed method using inverse probability weighting performed consistently across settings, with negligible bias. The generated marginal effects are presented in Table 2.2. In the proportional hazard and non-proportional hazard with early survival difference settings, the difference in restricted mean survival time increases as the marginal effect of exposure increases. However, in the non-proportional hazard setting where survival curves cross, the difference in restricted mean survival time decreases as the marginal effect of exposure increases. Figure A.2 gives the survival curves under each setting.

Table 2.1: Simulation scenarios

	Unexposed		**Exposed**	
Setting	ν_0	λ_0	ν_1	λ_1
PH	2	0.0083	2	0.0100
Non-PH, survival curves cross	1	0.0667	15	0.0003
Non-PH, early survival difference	1	0.0100	10	0.2207

PH: proportional hazards.
ν and λ denote the shape and scale parameters of the Weibull distributions.

Table 2.2: Marginal differences in restricted mean survival times for each simulation scenario

	Exposure effect		
Setting	Weak β_z=log(1.25)	Moderate β_z=log(1.5)	Very Strong β_z=log(2.0)
PH	-0.65 yrs.	-0.96 yrs.	-1.45 yrs.
Non-PH, survival curves cross	0.97 yrs.	0.74 yrs.	0.36 yrs.
Non-PH, early survival difference	-1.36 yrs.	-1.76 yrs.	-2.36 yrs.

PH: proportional hazards.
β_z denotes the strength of the association between the exposure and outcome.

With sample size n=1,000, the relative bias of the pseudo-observation and ANCOVA methods increased as the proportion of exposed decreased, reaching 32%

under weak (β_z=log(1.25)) effects, 23% under moderate (β_z=log(1.5)) effects, and 15% under very strong (β_z=log(2.0)) effects at 5% exposure (Table 2.3, Figure A.3). In the non-proportional hazard setting where survival curves cross, the trend in relative bias using the pseudo-observation and ANCOVA methods was more dramatic: the relative bias reached 38% with weak association between exposure and outcome, 46% with moderate association, and 88% with very strong association at 5% exposure. In the non-proportional hazard setting with early survival difference, all methods performed similarly in terms of relative bias.

Table 2.3: Relative bias in simulation study, sample size n=1000

		Proportional hazards			**Non-proportional hazards, curves cross**			**Non-proportional hazards, early survival difference**		
Association	**Exposed, %**	IPW	Pseudo	IPCW	IPW	Pseudo	IPCW	IPW	Pseudo	IPCW
Weak,	0.05	0.070	0.317	0.317	-0.036	0.386	0.386	0.023	0.072	0.072
β_z=log(1.25)	0.10	0.027	0.273	0.273	-0.027	0.282	0.282	0.011	0.090	0.090
	0.20	0.026	0.219	0.219	0.001	0.182	0.182	0.008	0.091	0.091
	0.30	0.005	0.166	0.166	-0.006	0.089	0.089	-0.002	0.091	0.091
	0.40	0.009	0.116	0.116	-0.015	0.015	0.015	-0.008	0.073	0.073
	0.50	-0.009	0.040	0.040	-0.012	-0.054	-0.054	0.009	0.067	0.067
Moderate,	0.05	0.073	0.233	0.233	-0.033	0.461	0.461	0.038	0.038	0.038
β_z=log(1.5)	0.10	0.020	0.217	0.217	-0.020	0.356	0.356	0.007	0.044	0.044
	0.20	0.011	0.174	0.174	-0.003	0.204	0.204	0.002	0.069	0.069
	0.30	0.002	0.124	0.124	0.002	0.096	0.096	0.010	0.075	0.075
	0.40	0.010	0.090	0.090	-0.009	0.011	0.011	-0.002	0.067	0.067
	0.50	0.015	0.046	0.046	-0.004	-0.072	-0.072	0.009	0.061	0.061
Strong,	0.05	0.045	0.150	0.150	-0.044	0.877	0.877	0.022	-0.029	-0.029
β_z=log(2.0)	0.10	0.011	0.146	0.146	-0.056	0.614	0.614	0.020	0.008	0.008
	0.20	0.006	0.126	0.126	0.011	0.350	0.350	-0.001	0.036	0.036
	0.30	0.005	0.100	0.100	-0.032	0.120	0.120	0.010	0.050	0.050
	0.40	-0.001	0.076	0.076	-0.011	-0.026	-0.026	0.004	0.055	0.055
	0.50	-0.006	0.041	0.041	-0.005	-0.188	-0.188	0.000	0.053	0.053

IPW: inverse probability weighted Kaplan-Meier estimator of difference in restricted mean survival times, pseudo: pseudo-observation model of Andersen & Pohar Perme (2010)
IPCW: inverse probability of censoring weighted model of Tian et al. (2014)

The mean squared error was similar using all methods at 50% exposure (equal size groups) (Table 2.4, Figure A.4). However, the mean squared error increased as the proportion of exposed decreased across all settings and methods. The mean squared error was larger with the proposed method, as compared to the pseudo-observation and ANCOVA methods in the non-proportional hazards with early survival difference and proportional hazards settings, reaching 0.83 and 0.4 respectively at 5% exposure. All methods performed well in terms of MSE in the non-proportional hazards settings where survival curves cross.

Table 2.4: Mean squared error in simulation study, sample size n=1000

		Proportional hazards			**Non-proportional hazards, curves cross**			**Non-proportional hazards, early survival difference**		
Association	**Exposed, %**	IPW	Pseudo	IPCW	IPW	Pseudo	IPCW	IPW	Pseudo	IPCW
Weak	0.05	0.375	0.116	0.116	0.176	0.199	0.199	0.753	0.124	0.124
β_z=log(1.25)	0.10	0.176	0.084	0.084	0.111	0.119	0.119	0.369	0.096	0.096
	0.20	0.092	0.058	0.058	0.056	0.065	0.065	0.191	0.075	0.075
	0.30	0.058	0.042	0.042	0.045	0.038	0.038	0.114	0.064	0.064
	0.40	0.049	0.032	0.032	0.048	0.032	0.032	0.093	0.059	0.059
	0.50	0.052	0.030	0.030	0.053	0.033	0.033	0.078	0.056	0.056
Moderate	0.05	0.376	0.128	0.128	0.211	0.180	0.180	0.826	0.110	0.110
β_z=log(1.5)	0.10	0.182	0.093	0.093	0.109	0.110	0.110	0.363	0.073	0.073
	0.20	0.090	0.061	0.061	0.058	0.053	0.053	0.183	0.065	0.065
	0.30	0.063	0.044	0.044	0.051	0.036	0.036	0.134	0.068	0.068
	0.40	0.049	0.037	0.037	0.047	0.033	0.033	0.089	0.058	0.058
	0.50	0.050	0.031	0.031	0.057	0.034	0.034	0.081	0.055	0.055
Strong	0.05	0.405	0.113	0.113	0.238	0.159	0.159	0.765	0.100	0.100
β_z=log(2.0)	0.10	0.216	0.090	0.090	0.114	0.087	0.087	0.408	0.060	0.060
	0.20	0.108	0.070	0.070	0.066	0.049	0.049	0.193	0.056	0.056
	0.30	0.071	0.051	0.051	0.052	0.033	0.033	0.123	0.057	0.057
	0.40	0.049	0.039	0.039	0.045	0.028	0.028	0.093	0.061	0.061
	0.50	0.054	0.031	0.031	0.063	0.036	0.036	0.083	0.060	0.060

IPW: inverse probability weighted Kaplan-Meier estimator of difference in restricted mean survival times, pseudo: pseudo-observation model of Andersen & Pohar Perme (2010)
IPCW: inverse probability of censoring weighted model of Tian et al. (2014)

In most settings, the coverage of our method slightly surpassed 95% (Figure A.5). In the non-proportional hazards settings where survival curves cross, the coverage of the pseudo-observation and IPCW methods decreased as the proportion of exposed decreased, reaching 67% under weak and moderate effects. These results were similar in the proportional hazards settings.

The patterns of the relative errors for our proposed method were consistent with under or over coverage, where a positive relative error indicates overcoverage and a negative relative error indicates undercoverage (Figure A.6).Morris et al. (2019) Morris et al suggested bias as another possible reason for under or overcoverage, however our proposed method displayed little bias.Morris et al. (2019) For the pseudo-observation and IPCW methods, coverage appears to improve as bias decreases. However, patterns between relative errors and coverage did not hold for the pseudo-observation and IPCW methods. We note that our proposed method demonstrated higher relative error than the pseudo-observation and IPCW methods in many scenarios, sometimes reaching 20%.

Finally, results followed a similar pattern when the sample size was n=500 and n=250. The mean squared error of our proposed method increased as sample size decreased and the proportion of exposed decreased. Figures for all results are available in the appendix (Figures A.3-A.14).

2.5 ILLUSTRATIVE EXAMPLE: FRAMINGHAM HEART STUDY

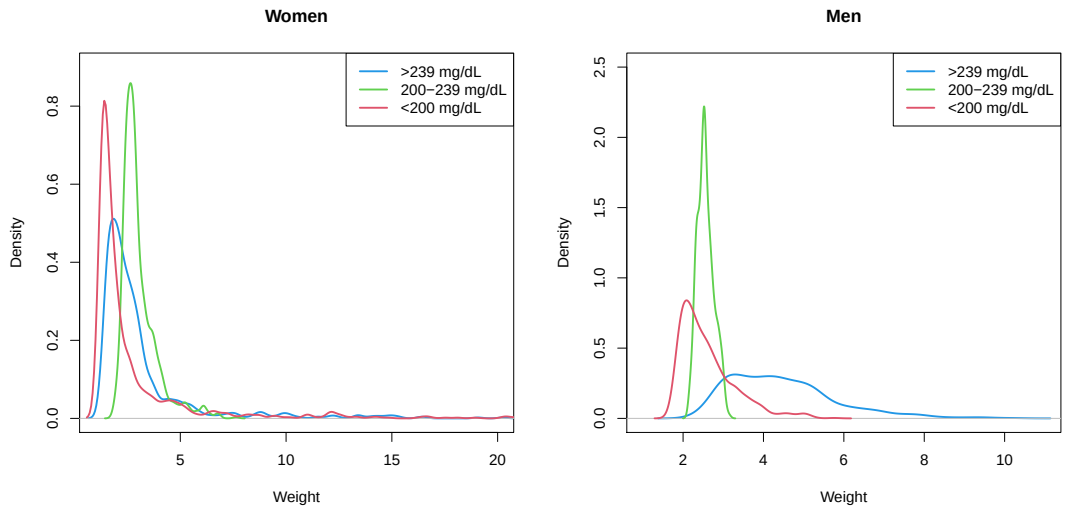
2.5.1 Methods

We illustrate the proposed method with the Framingham coronary heart disease (CHD) 10-year risk score model. The Framingham Heart Study is a long-term prospective study of the etiology of cardiovascular disease in the community of

Framingham, Massachusetts. Wilson et al. (1998) examined the association between total cholesterol and 10-year risk of CHD in the Original and Offspring Cohorts. Participants attended either the 11th examination in the Original cohort or the first examination in the Offspring cohort, and were free of CHD at baseline. Participants were followed for up to 12 years for the incidence of CHD. The exposure of interest was total cholesterol categorized as low (< 200 mg/dL), moderate ($200 - 239$ mg/dL), and high (≥ 240 mg/dL). The model was adjusted for age, hypertension, smoking status, diabetes, and HDL-cholesterol. Hypertension was categorized into four groups based on systolic and diastolic blood pressure, consistent with JNC-V definitions: normal including optimal (systolic <130 mmHg and diastolic <85 mmHg), high normal (systolic 130-139 mmHg or diastolic 85-89 mmHg), hypertension stage I (systolic 140-159 mmHg or diastolic 90-99 mmHg), and hypertension stages II-IV (systolic $\geq$ 160 mmHg or diastolic $\geq$ 100 mmHg).(Gifford Jr, 1993) Analyses were performed in men and women separately without formally testing for interaction. Risk factors were considered significant at a 5% two-sided level of significance. We fit the same multivariable Cox regression model for CHD, and used the same predictors to estimate the adjusted difference in RMSTs for CHD between total cholesterol groups.

We obtained adjusted HRs for high versus low total cholesterol and moderate versus low total cholesterol. We further assessed the proportional hazards assumption by the Grambsch-Thernau test at $\alpha = 0.10$.(Grambsch & Therneau, 1994) We also estimated differences in adjusted RMST between the cholesterol groups. We defined τ as 10 years for RMST measures, as Wilson et al predicted the 10-year risk of CHD. We applied our proposed method, the pseudo-observation approach, and the IPCW model to obtain the adjusted RMST. For our method, we obtained

Figure 2.1: Examination of inverse probability weights in the Framingham Heart Study



weights by fitting a multinomial logistic model to determine the individual predicted probabilities of being in each total cholesterol group according to the participant profile of all other covariates.

2.5.2 Results

First, we examine the distribution of the IP weights in Figure 2.1. We do not observe any extremely large weights, and the distributions display reasonable overlap. The IPW adjusted Kaplan-Meier cumulative risk curves and adjusted RMSTs by total cholesterol level obtained from our proposed method are presented in Figures 2.2 and 2.3.

The adjusted differences in RMSTs and HRs are presented in Table 2.5. In both men and women, the adjusted HRs show that higher versus lower total cholesterol level is associated with CHD. The HRs of moderate versus low total cholesterol and high versus low total cholesterol were 1.3 [95% CI:1.0, 1.7] and 2.0 [95% CI:1.5,

Figure 2.2: Adjusted Kaplan-Meier curves for coronary heart disease by total cholesterol level in men from the Framingham Heart Study

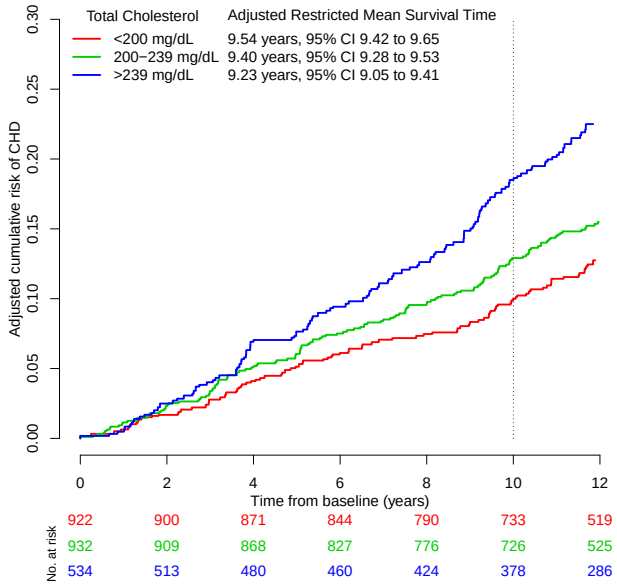
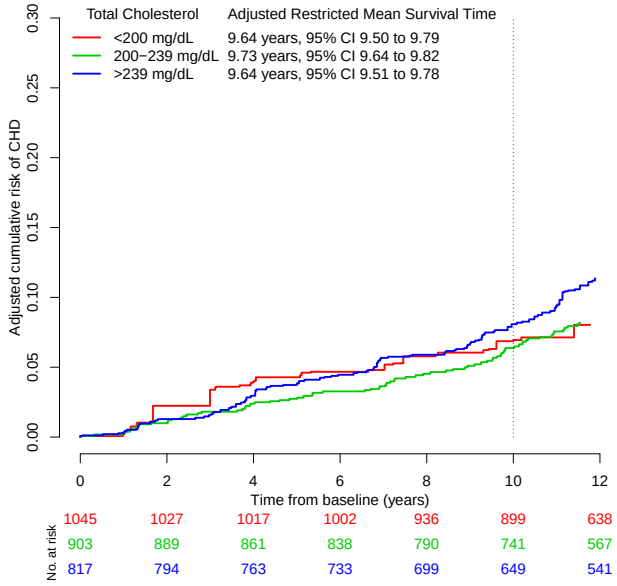


Figure 2.3: Adjusted Kaplan-Meier curves for coronary heart disease by total cholesterol level in women from the Framingham Heart Study



2.6] in men and 1.7 [95% CI:1.1, 2.6] and 2.1 [95% CI:1.4, 3.2] in women.

Table 2.5: Adjusted hazards ratios and differences in restricted mean survival times between total cholesterol levels in the Framingham Heart Study

	Men			Women		
Method	**Est.**	**95% CI**	**p**	**Est.**	**95% CI**	**p**
Total cholesterol, 200-239 mg/dL vs. <200 mg/dL						
HR	1.34	(1.03, 1.74)	0.031	1.68	(1.09, 2.57)	0.018
IPW difference in RMSTs	-1.58 mo.	(-3.58, 0.42)	0.118	1.02 mo.	(-0.98, 3.02)	0.316
Pseudo-observation difference in RMSTs	-1.57 mo.	(-3.43, 0.29)	0.097	-0.10 mo.	(-1.51, 1.32)	0.888
IPCW difference in RMSTs	-1.36 mo.	(-3.22, 0.49)	0.151	0.00 mo.	(-1.42, 1.40)	0.998
Total cholesterol, >239 mg/dL vs. <200 mg/dL						
HR	1.95	(1.49, 2.56)	<.001	2.06	(1.35, 3.15)	0.001
IPW difference in RMSTs	-3.66 mo.	(-6.22, -1.10)	0.005	-0.01 mo.	(-2.36, 2.34)	0.994
Pseudo-observation difference in RMSTs	-3.46 mo.	(-5.95, -0.96)	0.006	-1.09 mo.	(-2.93, 0.74)	0.244
IPCW difference in RMSTs	-3.07 mo.	(-5.56, -0.58)	0.016	-0.95 mo.	(-2.78, 0.88)	0.309

HR: hazards ratio, IPW: inverse probability weighting, RMST: restricted mean survival time

For both moderate versus low total cholesterol and high versus low total cholesterol, the proportional hazards assumption was met among men (p=0.41 and p=0.68), but not among women (p=0.04 and p=0.04). This suggests that the adjusted HRs for women should be interpreted with caution.

With our proposed method, the adjusted mean times to CHD (RMST) for low, moderate, and high total cholesterol were 9.5, 9.4, and 9.2 years in men over a 10 year time span. In women, the adjusted mean times to CHD were 9.7, 9.7, and 9.6 years over a 10 year time span.

In men, the difference in RMSTs between moderate and low total cholesterol was similar across the three methods suggesting a small decrease in the mean time to CHD with an increase in total cholesterol: -1.6 months [95% CI: -3.6, 0.4] using our proposed method, -1.6 months [95% CI: -3.4, 0.3] with the pseudo-observation model, and -1.4 months [95% CI: -3.2, 0.5] with the IPCW model. In women, the difference in RMSTs was slightly different between methods but small to negligible: 1.0 months [95% CI: -1.0, 3.0] using our proposed method, -0.1 months [95% CI: -1.5, 1.3] with the pseudo-observation model, and 0 months [95% CI: -1.4, 1.4] with the IPCW model.

As for high versus low total cholesterol, the difference in RMSTs in men was -3.7 months [95% CI: -6.2, -1.1] using our proposed method, -3.5 months [95% CI: -6.0, -1.0] with the pseudo-observation method, and -3.1 months [95% CI: -5.6, -0.6] with the IPCW model. In women, the difference in RMSTs was -0.01 months [95% CI: -2.4, 2.3] using our proposed method, -1.1 months [95% CI: -2.9, 0.7] with the pseudo-observation method, and -1.0 months [95% CI: -2.8, 0.9] with the IPCW model.

Among both men and women, the differences in RMSTs between moderate and

low total cholesterol were not significant. As for high versus low total cholesterol, the differences in RMSTs were significant among men but not in women. However, all adjusted HRs were significant. Whereas the adjusted HRs convey significant relative effects, examination of the absolute effects via difference in RMSTs gives a different picture: the mean difference in time to CHD is approximately 6 weeks in men and negligible in women, comparing moderate to low total cholesterol. The HRs and difference in RMSTs are consistent among men, but it appears that the significant HRs among women may not be clinically meaningful.

2.6 DISCUSSION

We developed a method to derive adjusted RMSTs and their differences, which can be used to measure the effect of exposures or treatments in observational studies. In our simulation study, the proposed method had similar statistical performance as compared to regression-based methods previously described; all methods performed well in terms of relative bias, mean squared error, and coverage, but their advantages differed by scenario. For example, our method had lower relative bias compared to the regression-based methods in the non-proportional hazards setting where survival curves cross. However, the regression-based methods had lower mean squared error than our method in the non-proportional hazards with early survival difference setting, particularly for proportions of exposure below 30%. In our Framingham Heart Study example, all methods produced consistent RMST-based measures in men, whereas our method differed slightly from the regression-based methods in women.

Because observational studies must adjust for confounding factors, epidemiologists frequently use the Cox proportional hazards model or related approaches

and typically report only adjusted hazards ratios.(Hernán, 2010) However, measures of effect based on adjusted RMST offer a different perspective. We submit that the RMST should be reported systematically alongside hazards ratios. The adjusted RMST is on the time scale, which provides clinicians and patients with background information (the adjusted RMST in the unexposed group) as well as the absolute effect of the exposure (the difference in adjusted RMST between the exposed and unexposed groups). In the Framingham Heart Study example, the adjusted HRs suggested that higher total cholesterol had a significant impact on the time to CHD, with HRs ranging from 1.3 to 2.1. In contrast, the differences in RMST between those with high versus low total cholesterol were less than 4 months in men and less than 1 month in women, over a 10 year time span. We illustrate how the difference in RMSTs can provide a different interpretation of the effect of cholesterol level on CHD risk. This is consistent with the recent work of Finegold et al. (2016), which found interventions such as statins had a mean lifespan gain of 7 months, and 95% of participants had no gain in lifespan. We emphasize that the difference in RMSTs is truncated to a 10 year time span, and is not to be interpreted as a gain or loss over the lifespan. Additionally, the adjusted Kaplan-Meier curves indicate that the difference in RMSTs would continue to increase with a larger time span. If the RMSTs are reported as complementary to the HR, they would offer a tool to compare the potential of different exposures in similar target populations.(Weir et al., 2019) This will assist researchers and clinicians in interpreting the impact of such exposures on outcomes.

In contrast with regression-based methods, our method is consistent with the typical RMST approach in the unadjusted framework: calculating the area below the curve. This is congruent with the visualization offered by the adjusted Kaplan-

Meier curves. In randomized trials, the Kaplan-Meier graph frequently complements the reported HR due to the shortcomings of the HR's interpretation and assumptions.(Hernán, 2010) Our approach and its implementation in R allow producing the adjusted Kaplan Meier curves. By reporting both the adjusted Kaplan-Meier curve and RMST, it is easier to gain a sense of an effect's intensity because of the time domain. Adjusted RMST-based measures bring value to interpreting time-to-event outcomes, especially in the field of epidemiology. Additionally, randomized trials commonly report unadjusted Kaplan-Meier curves alongside unadjusted hazards ratios, yet observational studies do not report adjusted Kaplan-Meier curves. An advantage of our method is the visualization of the adjusted RMST with the adjusted Kaplan-Meier curves.

In general, Kaplan-Meier curves are not feasible with continuous covariates. Since we used adjusted Kaplan-Meier curves to calculate RMST, our method does not allow the estimation of the effect of a continuous covariate on the RMST. If the objective is to express the effect for a continuous covariate, one can use the pseudo-observation method, IPCW model, or other regression-based methods.

A limitation of adjusting Kaplan-Meier curves with inverse probability weights or propensity scores is the potential for very large or small weights.(Jackson et al., 2017) Extreme weights typically occur when there is a rare patient profile that is frequent in the adjusting population; a large weight will be assigned to this subgroup. In consequence, the survival curve will have a large variance. Therefore, the adjusted RMST also will have a large standard error. There are solutions to address extreme weights. One can stabilize weights via trimming or truncation.(Austin, 2014; Robins et al., 2000) Additionally, one can potentially improve the adequacy of the model by using interaction terms or exploring other model types. In our anal-

ysis, we used logistic models to estimate the propensity scores. However, there are other many options to estimate weights. For instance, models can be fit with a probit or log link rather than a logit link, or with random forest methods. However, the chosen derivation of the weights may make model-based assumptions.

A second limitation of our estimator is that we consider the weights to be fixed when deriving the variance, but in practicality, the weights may be estimated from the data. Estimating weights will introduce sampling variability, which is not accounted for in the variance. To consider the weights as random and account for the the weights being estimated from data, it is common to use adjusted sandwich variance estimators or the bootstrap.(Fitzmaurice et al., 2012; Austin, 2016) Typically, the sandwich estimator adjusted for estimation of weights will give smaller standard errors.(Fitzmaurice et al., 2012) Finally, Xie and Liu compared the estimated variance using estimated weights with the variance estimated from Monte Carlo simulation, and found they were comparable under various censoring scenarios.(Xie & Liu, 2005)

Our simulation study compared three methods of obtaining the adjusted difference in RMST, however other methods exist which were not included in our simulation study. For example, it is possible to use piecewise exponential models, proportional hazards models, or augmented inverse-weighted Nelson-Aalen estimators.(Karrison, 1987; Zucker, 1998; Chen & Tsiatis, 2001; Wei, 2008; Zhang & Schaubel, 2011; Schaubel & Wei, 2011; Zhang & Schaubel, 2012a,b) Although some approaches use proportional hazards models, they do not all assume proportional hazards of the exposure of interest, i.e. via stratification.(Zucker, 1998; Chen & Tsiatis, 2001; Wei, 2008; Zhang & Schaubel, 2011; Schaubel & Wei, 2011; Zhang & Schaubel, 2012a,b) The IPCW and pseudo-observation methods do not uti-

lize Cox proportional hazards models. Another approach is the Royston-Parmar model that relates the log cumulative hazard function to the covariates, and in which the log baseline cumulative hazard is modeled as a cubic spline function of log time.(Royston & Parmar, 2013) This model can be further extended for non-proportional hazards, by including an interaction term between the spline function and the covariates. Future work will entail comparing these methods as well.

Finally, we note that the Kaplan-Meier estimator is asymptotically equivalent to the Nelson-Aalen estimator. Previous works have derived the restricted mean survival time by applying inverse probability weighting to the Nelson-Aalen estimator with augmentation terms to improve efficiency. Therefore, the asymptotic results of our proposed method are a special case of previous methods using inverse-weighted Nelson-Aalen estimators if the augmentation terms are set to zero.(Schaubel & Wei, 2011; Zhang & Schaubel, 2011, 2012a,b)

In conclusion, our proposed method for RMST using adjusted Kaplan-Meier curves produces results consistent with existing methods for adjusted RMST. In addition, it is congruent with the visualization of the exposure's effect with adjusted Kaplan-Meier curves. Due to the challenging interpretation of the HR and proportional hazards assumption, we submit that the RMST should be reported systematically alongside hazards ratios. We advocate the use of RMST-based measures because of their clinically relevant interpretations. Our method offers both easily-interpreted summary measures and visualization of absolute effects through adjusted Kaplan-Meier curves, which are missing from the observational literature. It also does not rely on model assumptions, such as proportional hazards.

CHAPTER 3

Adjusted restricted mean survival time with time-varying confounding

3.1 BACKGROUND

Confounders that are also intermediate variables cannot be accounted for by using traditional statistical methods, such as the Cox proportional hazards model with time-varying covariates. For example, body mass index (BMI) is a known, modifiable risk factor of AF.(Alonso et al., 2013; Berkovitch et al., 2016) Most previous studies have analyzed the association between atrial fibrillation (AF) and BMI at cohort entry, without accounting for changes in risk factors over time. Berkovitch et al. (2016) examined change in BMI as a time-varying covariate. However, changes in BMI may also affect other vascular risk factors via cardiac remodeling. For instance, increasing BMI increases the risk of developing hypertension and heart failure, both associated with increased risk of AF.(Litwin, 2010; Mahajan et al., 2018) Thus, vascular risk factors create time-varying confounding which depends on past BMI; they are said to be intermediate variables.

Unlike standard statistical methods, the g-methods, i.e. the parametric g-formula, inverse probability weighting of marginal structural models, and g-estimation, can account for the fact that BMI history is associated with other time-varying confounders.(Danaei et al., 2016; Taubman et al., 2009; Witteman et al., 1998; Naimi et al., 2017; Robins, 1986) Since the Cox model cannot address this, previous results on the association between BMI and AF may be biased. The parametric g-formula also allows estimation of the effect of interventions based on real-life scenarios but applied to all individuals: for example, the difference in AF risk had everyone been non-obese versus everyone had been obese. G-methods have

not been applied in AF research.

Previous applications of g-methods on topics outside of AF research have reported relative risks, risk differences, or hazard ratios to measure associations.(Taubman et al., 2009; Witteman et al., 1998; Collaboration et al., 2018; Lodi et al., 2015; Shakiba et al., 2018; Keil et al., 2014) In this chapter, we demonstrate the derivation of RMST when using the parametric g-formula to account for time-varying confounding.(Conner et al., 2019a) We use the parametric g-formula to estimate the association between hypothetical BMI interventions and the risk of AF, as well as the difference in RMSTs (AF-free time) while accounting for time-varying confounding in data from the Framingham Heart Study. We considered two time horizons, $\tau = 10$ years and $\tau = 20$ years.

3.2 MOTIVATING EXAMPLE: INCIDENT ATRIAL FIBRILLATION IN THE FRAMINGHAM HEART STUDY

We included participants from the FHS Original and Offspring cohorts who were AF-free and attended an examination at age 50 years, plus or minus five years. Follow up began at the examination closest to age 50 years. For the Original cohort, the earliest exam to enter was Examination 11 (1968-1971) due to a shift in adjudication of AF around 1970.(Kannel et al., 1961; Staerk et al., 2017, 2018) However, we used covariate information from Examinations 9 and 10 to include covariate history with a lag of 3 years, as described in Section 3.4.2. For the Offspring cohort, the earliest exam to enter was Examination 2 (1979-1983) in order to incorporate covariate history with a lag of 3 years from Examination 1, which we describe in more detail in Section 3.4.2.(Feinleib et al., 1975) AF diagnosis, including atrial flutter, was adjudicated by two cardiologists using examination records, medical

records, electrocardiograms, and hospital contacts. In our analyses, follow-up continued until the earliest of first diagnosed AF, death, last FHS examination or medical contact, or end of follow-up (December 31st, 2015 or 10 and 20 years since age 50 years, plus or minus five years). Our final sample consisted of 4,392 AF-free participants.

The exposure of interest was BMI over time. We adjusted analyses on sex, baseline age, and on the following time-varying covariates: smoking status (current vs. former/never), systolic blood pressure (SBP), diastolic blood pressure (DBP), antihypertensive treatment, history of diabetes, heart failure, and myocardial infarction. We selected these covariates in alignment with the CHARGE AF simple risk score, and we added sex.(Alonso et al., 2013; Schnabel et al., 2009) We also conducted analyses stratified by sex.

3.3 CONVENTIONAL APPROACHES

We first fit Cox proportional hazards models for 10 and 20 years of follow up. We estimated hazard ratios for BMI at baseline adjusting for the aforementioned covariates at baseline, time-varying BMI adjusting for covariates at baseline, and time-varying BMI adjusting for time-varying covariates. BMI associations were calculated for a 5 kg/m^2 decrease in BMI. Similarly, we repeated these models with BMI dichotomized as non-obese (BMI<30 kg/m^2) and obese (BMI30kg/m^2).

3.4 PARAMETRIC G-FORMULA FOR THE DIFFERENCE IN RESTRICTED MEAN SURVIVAL TIMES

To account for time-varying covariates which depend on previous BMI, we estimated the associations of BMI interventions and incident AF using the paramet-

ric g-formula method.(Witteman et al., 1998; Collaboration et al., 2018; Lodi et al., 2015; Shakiba et al., 2018; Keil et al., 2014) This method adjusts for time-varying covariates which depend on previous exposure by leveraging the past value of covariates. The method works in two steps. First, we estimated the joint density of time-varying covariates given the covariate history through parametric models. Second, we conducted Monte Carlo simulation to estimate the risk of incident AF under a given BMI intervention, defined by specific BMI profiles. For example, we simulated participants maintaining BMI below 30 kg/m^2 and maintaining BMI at least 30 kg/m^2 over a pre-specified time period. By contrasting the counterfactual outcomes, we estimated the association for obesity and the risk of incident AF.

3.4.1 BMI interventions of interest

We examined BMI interventions with regard to incident AF via the following comparisons: 1) maintaining BMI between 18.5 kg/m^2 and 29.9 kg/m^2 at all years (non-obese) versus maintaining BMI between 30 kg/m^2 and 41 kg/m^2 at all years (obese); 2) maintaining BMI between 18.5 kg/m^2 and 29.9 kg/m^2 at all years (non-obese) versus the natural course, or the observed empirical distribution of time-varying BMI (Logan et al., 2018); and, 3) a 10% decrease in BMI each year among individuals with 25.0 kg/m^2 until BMI reaches 25.0 kg/m^2 versus the natural course (Table 3.1). These interventions and comparisons are fashioned to emulate real-life weight management strategies, such as bariatric surgery or more gradual weight management.(Abed et al., 2013; Pathak et al., 2014; Edwards et al., 2016) The interventions are maintained until the pre-specified time period, AF, or death.

Table 3.1: BMI interventions and comparisons assessed with the g-formula

Intervention	Reference
Non-obese at all times, BMI [18.5, 29.9]	Obese at all times, BMI [30, 40]
Non-obese at all times, BMI [18.5, 29.9]	Natural course
10% decrease in BMI per year when BMI>=25	Natural course

3.4.2 Causal structure and covariate history

In Figure 3.1, we present a directed acyclic graph to illustrate the pathways between BMI, other covariates, and AF over time. Directed acyclic graphs are a popular tool in epidemiology to visualize confounding and causal relationships. Figure 3.1 illustrates the causal structure between time-varying BMI, time-varying confounder L, and the risk of AF. Time goes from left to right, and thus BMI and L are prior to AF; measures at time $k-1$ are prior to those at time k.

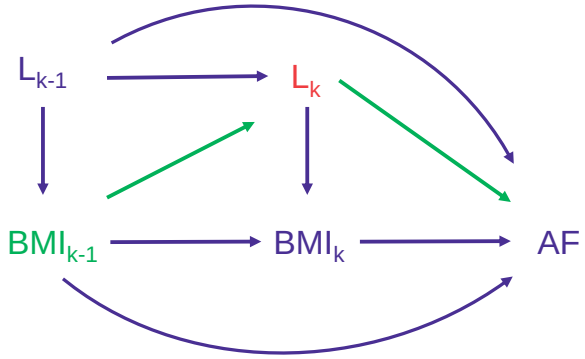


Figure 3.1: Directed acyclic graph of body mass index, other time-varying covariates, and atrial fibrillation

The directed acyclic graph displays repeated measures at years $k-1$ and k. BMI_k denotes the exposure, body mass index (BMI), at year k. L_k denotes confounders at year k (for example, systolic blood pressure). AF denotes the outcome, new-onset atrial fibrillation (AF). Arrows indicate associations (for example, the association of BMI and incident AF). Adjustment for intermediate variables L_k (red) in a Cox model will block the path between BMI_{k-1} and AF (green), which prevents us from observing the full association. However, g-methods can accommodate this scenario. If BMI_{k-1} did not cause AF through L_k (the green arrows were not present), then L_k would not be an intermediate variable and adjustment for L_k would not block the association of BMI_{k-1} and AF.

Arrows from one variable to another indicate that we make the assumption of a direct causal effect from the first variable to the second (not mediated by other variables in the graph). Adjustment for L_k as a time-varying covariate in a Cox proportional hazards model would block the path between BMI_{k-1} and subsequent AF, preventing one from observing the full association between BMI_{k-1} and AF. However, g-methods such as the parametric g-formula can accommodate this causal structure.

In our analyses, we incorporated the covariate history with a lag of 3 years. At year k, the associations between BMI history (BMI_{k-1}) and L_k indicate how a participants other risk factors may be affected by his or her previous BMI at years $k-3$, $k-2$ and $k-1$. The risk of incident AF at year k may be influenced by their

current BMI (BMI_k), their BMI history $(BMI_{k-3}, BMI_{k-2}, BMI_{k-1})$, and their risk factors $(L_{k-3}, L_{k-2}, L_{k-1}, L_k)$.

3.4.3 Monte Carlo simulation for association of BMI interventions

For each BMI intervention described in Section 3.4.1, we used the previously estimated parametric models to generate a pseudo-population of size n=10,000 participants. Beginning with the observed covariate values, we generated covariates at each year using the estimated regression coefficients from the covariate models. Throughout the process, generated BMI values were modified to match an assigned intervention and used in generating subsequent covariate values. For example, if participants maintain a BMI below 30 kg/m^2 over time, any simulated BMI value at or above 30.0 kg/m^2 is updated to 29.9 kg/m^2. We give an example of this process in Figure B.1. The risk at each time point conditional on the covariate history was also estimated using the estimated regression coefficients from a discrete hazard model, or pooled logistic model, for AF described in Section 3.2. Using the g-formula, we first estimated hazard ratio and risk ratio at 10 and 20 years. We then derived the absolute risk difference and difference in RMST.

3.4.4 Difference in restricted mean survival times

In the absence of repeated examinations, the g-formula involves standardization of the risk of AF for a given BMI intervention across confounder profiles $\boldsymbol{X_i}$,

$$\begin{aligned} E\{I(AF) = 1|BMI\} &= \sum_x E\{I(AF_i) = 1|BMI_i, \boldsymbol{X_i}\} \cdot f_{\boldsymbol{X}}(x) \\ &= \sum_x P(AF_i = 1|BMI_i, \boldsymbol{X_i}) \cdot f_{\boldsymbol{X}}(x) \end{aligned}$$

However, the covariates are not fixed over time but change at each year k. Therefore, we index on year k and must account for time-varying confounding.

To account for time-varying confounding, we first estimated the hazard of AF each year conditional on the covariate history. We define the conditional probability of AF for individual i at year k under a given BMI intervention,

$$P(AF_{i,k+1} = 1 | AF_{i,k} = 0, \boldsymbol{BMI}_{i,k}, \boldsymbol{L}_{i,k}, \boldsymbol{X_i})$$

where $\boldsymbol{BMI}_{ik}$ denotes the time-varying BMI values corresponding to the assigned BMI intervention, $\boldsymbol{L}_i$ denotes the vector of time-varying risk factors with lags, and $\boldsymbol{X_i}$ denotes time-invariant covariates (sex, age). We fit a pooled logistic model for incident AF,

$$\begin{aligned} &\text{logit}\{P(AF_{i,k+1} = 1 | AF_{i,k} = 0, \boldsymbol{BMI}_{i,k}, \boldsymbol{L}_{i,k}, \boldsymbol{X_i})\} \\ &= \beta_0 + (\boldsymbol{BMI}_{i,k})^T \boldsymbol{\beta} + \boldsymbol{L}_{i,k}^T \boldsymbol{\gamma} + \boldsymbol{X_i}^T \boldsymbol{\alpha} + r(t_k)^T \boldsymbol{\theta} \end{aligned} \tag{3.1}$$

in which $\boldsymbol{L}_{i,k}$ denotes all observed time-varying covariates at year k and $r(t_k)$ denotes a restricted cubic spline function for time.(Green & Symons, 1983; Cupples et al., 1988; D'Agostino et al., 1990; Ngwa et al., 2016)

In each simulated pseudo-population, we use the estimated coefficients from 3.1 to obtain predicted conditional probabilities of AF at each time point for each individual in the pseudo-population based on their simulated covariates,

$$\hat{P}(AF_{i,k+1} = 1 | AF_{i,k} = 0, \boldsymbol{BMI}_{i,k}, \boldsymbol{L}_{i,k}, \boldsymbol{X_i})$$

Next, we estimate the conditional cumulative probability of being AF-free at year

k for each individual in the pseudo-population,

$$\begin{aligned}\hat{S}(t_k|\boldsymbol{BMI}_{i,k}, \boldsymbol{L}_{i,k}, \boldsymbol{X_i}) &= 1 - \hat{P}(AF_{i,k} = 1|\boldsymbol{BMI}_{i,k}, \boldsymbol{L}_i, \boldsymbol{X_i}) \\ &= \prod_{j:t_j \leq t_k} \{1 - \hat{P}(AF_{i,j+1} = 1|AF_{i,j} = 0, \boldsymbol{BMI}_{i,j}, \boldsymbol{L}_{i,j}, \boldsymbol{X_i})\}.\end{aligned}$$

Then, we estimate the marginal probability of being AF-free at year k under a given BMI intervention by averaging over all individuals in the pseudo-population,

$$\hat{S}(t_k|\boldsymbol{BMI}_k) = \frac{1}{10,000} \sum_{i=1}^{10,000} \hat{S}(t_k|\boldsymbol{BMI}_{i,k}, \boldsymbol{L}_{i,k}, \boldsymbol{X_i}).$$

To estimate the adjusted RMST at τ under a given BMI intervention, $\mu(\tau|\boldsymbol{BMI}_k)$, we used rectangular area approximation with equally spaced intervals by summing the survival probabilities at each year,

$$\begin{aligned}\hat{\mu}(\tau|\boldsymbol{BMI}_k) &= \sum_{k=0}^{K} \hat{S}(t_k|\boldsymbol{BMI}_k)(t_{k+1} - t_k) \\ &= \sum_{k=1}^{K} \hat{S}(t_k|\boldsymbol{BMI}_k) \qquad (3.2)\end{aligned}$$

where $\hat{S}(t_0|\boldsymbol{BMI}_k) = 1$ and $t_{K+1} = \tau$. We use equally spaced intervals since the parametric g-formula assumes a discrete-time framework. Since we assumes equally spaced intervals of one unit (year), this is simply the sum of $\hat{S}(t_k|\boldsymbol{BMI}_k)$.

We computed the survival probabilities at one year increments in simulated datasets without loss to follow-up. Since the examinations in Framingham Heart Study take place every two to six years, we chose to use years as the unit of time with one year increments; additional details on covariate measurements are explained in 3.4.5. Standard errors and 95% confidence intervals are obtained with

the bootstrap method. The difference in RMSTs is then given from the difference in RMSTs between under different BMI interventions, say always obese and always non-obese, is

$$\hat{\mu}(\tau|\boldsymbol{BMI}_k = 1) - \hat{\mu}(\tau|\boldsymbol{BMI}_k = 0). \tag{3.3}$$

We obtained standard errors using the non-parametric bootstrap with 500 samples. We repeated this process in 30 imputed datasets.

3.4.5 Multiple imputation of missing and annual-level data

Observational studies with varying exam cycles present a challenge when estimating the RMST with the g-formula. Framingham Heart Study participants have clinic exams every two to eight years, where exam attendance has ranged from 55% to 84%. In addition, some covariates may be missing at an attended exam. We used multiple imputation and linear interpolation to generate data with complete covariate information available each year (Figure 3.2). The interpolation allows estimating the survival probabilities at each year, and thus the RMSTs.

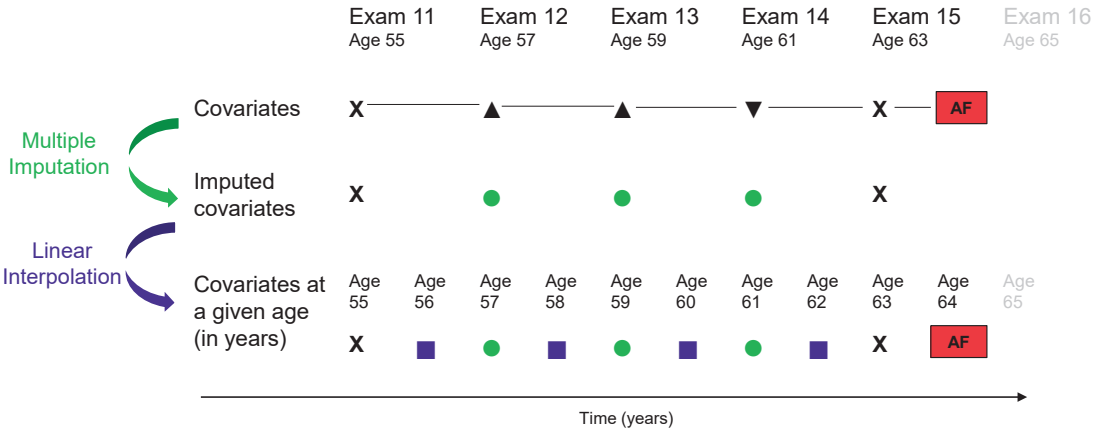


Figure 3.2: Multiple imputation and interpolation process

X completely measured, ▲ incomplete covariates, ▼ unattended exam, ● covariates multiply imputed, ■ covariates linearly interpolated. Examinations took place approximately every two years in the Original cohort and every four to eight years in the Offspring. Covariates of interest include body mass index (BMI), smoking, systolic blood pressure (SBP), diastolic blood pressure (DBP), antihypertensive treatment, diabetes, heart failure, and myocardial infarction.

First, we performed sequential multiple imputation to impute 30 datasets, accounting for temporality of covariates across successive exam cycles while removing participants from the imputation process at death.(Ning et al., 2013; Nevalainen et al., 2009; Schomaker & Heumann, 2018) We used data beginning at Examination 9 in the Original cohort and Examination 1 in the Offspring cohort. We imputed missing covariate values at entry examination and both missing covariate values and non-attended exams throughout the follow-up period. Imputed datasets were generated within each cohort and then combined. We repeated this process for 30 imputed datasets.

Second, we updated the yearly covariate values between exams. For continuous covariates, we filled in values using linear interpolation. For dichotomous covariates, we used midpoint interpolation to identify the year of the change in

covariate, if any. Among participants who experienced AF or died after their last in-clinic examination, we carried forward covariate values from the last exam (attended or imputed) until incident AF or death. We performed analyses in each of the 30 imputed datasets, and combined results according to Rubin's rule.(Schomaker & Heumann, 2018)

3.4.6 Sensitivity analyses

We performed sensitivity analyses under three different scenarios at 20 years of follow-up. Weight loss may be due to severe illness and associated with greater risk of morbidity, including AF.(Danaei et al., 2016; Taubman et al., 2009) Therefore, in our first sensitivity analysis we excluded participants diagnosed with cancer at entry exam and we censored participants upon cancer diagnosis.

In a second sensitivity analysis, we restricted to participants who attended at least three consecutive examinations with non-missing covariates in the Original cohort or two consecutive examinations with non-missing covariates in the Offspring cohort because of the need for lagged information (covariate history). We performed sequential multiple imputation for any missing covariate values during follow-up, as previously described.

Third, we considered a later entry age of 55 to 65 years. Finally, following a reviewers suggestion, we fit pooled negative binomial models for the hazard of AF instead of pooled logistic models.

3.4.7 Software

All analyses were performed in SAS 9.4. We used the GFORMULA SAS macros and made modifications to accommodate multiple imputation and a pooled nega-

tive binomial model.(Logan et al., 2018) To provide guidance on preparing the analytic dataset and using the parametric g-formula to estimate RMST, we share our SAS code and a working example. Our SAS program is available at github.com/sconner/afbmi-gformula.

3.5 RESULTS

We included 4,392 participants from the Framingham Heart Study who were AF-free at entry, which was the examination attended closest to 50 years. We followed participants for up to 20 years. We estimated HRs comparing time-varying non-obese vs. obese with traditional Cox models. We then used the parametric g-formula to compare non-obese vs. obese and 10% annual decrease in BMI (until normal weight is reached) vs. the natural course (no intervention). With the g-formula, we estimated HRs and differences in RMSTs, adjusted for sex, age, and time-varying risk factors.

3.5.1 Participant characteristics

Among the 4,392 AF-free participants, 53.4% were women (Table 3.2). The mean age at entry was 50.7 years. The mean BMI was 27.1 kg/m^2, and 20.4% were obese. The flow of participant selection is outlined in Figure B.2 in Appendix B. With a median follow-up of 23.8 years (Q1, Q3: 16.7, 30.2), there were 847 total AF events, with 489 in men and 358 in women. At 20 years, there were 389 total AF events, with 259 in men and 130 in women. For brevity, we present results for 20 years of follow-up. Results for 10 years of follow-up are available in Tables 3.3-3.5.

Table 3.2: Characteristics of Framingham Heart Study participants at entry (n=4,392)

	Overall	By sex		By cohort	
Risk factor	n=4,392	**Men** n=2,047	**Women** n=2,345	**Original** n=841	**Offspring** n=3,551
Age (years)	50.7 (2.2)	50.7 (2.2)	50.8 (2.2)	52.8 (1.7)	50.3 (2)
Women	2,345 (53.4)	-	-	468 (55.6)	1,877 (52.9)
BMI (kg/m2)	27.1 (5)	27.9 (4.2)	26.3 (5.6)	26.3 (4.3)	27.2 (5.1)
SBP (mm Hg)	125 (17)	128 (17)	123 (18)	132 (19)	124 (17)
DBP (mm Hg)	80 (10)	82 (10)	77 (10)	83 (11)	79 (10)
Current smoker	1,184 (29.5)	557 (29.8)	627 (29.3)	191 (41.4)	993 (28)
Use of hypertension medication	579 (13.2)	304 (14.9)	275 (11.7)	75 (8.9)	504 (14.2)
Diabetes	198 (5.1)	113 (6.1)	85 (4.1)	17 (2.8)	181 (5.5)
Heart failure	13 (0.3)	10 (0.5)	3 (0.1)	3 (0.4)	10 (0.3)
Myocardial infarction	95 (2.2)	81 (4)	14 (0.6)	25 (3.0)	70 (2.0)

Values are mean (SD) or n (%).

3.5.2 Relative measures of associations by conventional approaches and the g-formula

At 20 years of follow-up, Cox proportional hazards models with all covariates time-varying indicated non-obese participants had a 17% decreased hazard of AF compared to obese participants (HR: 0.83, 95% CI: 0.72, 0.97, last row in Table 3.3). When considering continuous BMI, on average the hazard of AF decreased by 12% per 5 kg/m^2 decrease in BMI (HR: 0.88, 95% CI: 0.82, 0.95, last row in Table 3.3). We clarify that this represents a 5 kg/m^2 shift in BMI, and not necessarily an individuals change over time. The hazard ratios comparing non-obese vs. obese changed slightly in magnitude when adjusting for all covariates at baseline, obesity as time-varying and other covariates at baseline, and both obesity and all covariates time-varying. However, the analogous hazard ratios for continuous BMI were nearly identical across models.

Using the g-formula, the hazard of AF was 27% lower had everyone been non-obese vs. obese (HR: 0.73, 95% CI: 0.58, 0.91, row 4 of Table 3.4). The risk ratio

Table 3.3: Hazard ratios of associations between body mass index and atrial fibrillation estimated with conventional Cox models

Model	Non-obese vs obese	5 kg/m^2 decrease in BMI
10 years		
All covariates at baseline	0.85 (0.72, 1.00)	0.93 (0.86, 1.00)
Time-varying obesity/BMI	0.83 (0.71, 0.97)	0.90 (0.84, 0.97)
All covariates time-varying	0.82 (0.70, 0.96)	0.90 (0.84, 0.96)
20 years		
All covariates at baseline	0.75 (0.63, 0.88)	0.88 (0.81, 0.95)
Time-varying obesity/BMI	0.82 (0.71, 0.95)	0.88 (0.82, 0.94)
All covariates time-varying	0.83 (0.72, 0.97)	0.88 (0.82, 0.95)

Data are adjusted hazard ratios and 95% confidence intervals. Cox models are adjusted for SBP, DBP, current smoking status, use of hypertension medication, diabetes status, history of heart failure, and history of myocardial infarction. We note that results for a 5 kg/m2 decrease in BMI represent average results for a shift in BMI, and not necessarily an individuals change over time. Results for time-varying obesity/BMI allowed obesity/BMI to vary over time, but all other covariates were measured at baseline.

was similar (RR: 0.75, 95% CI: 0.63, 0.89). Comparisons of BMI interventions with the natural course showed a small but not significant benefit in favor of the BMI intervention. When comparing non-obese to the natural course, the hazard and risk of AF both decreased by 8% (HR: 0.92, 95% CI: 0.78, 1.08 and RR: 0.92, 95% CI: 0.83, 1.02, row 5 of Table 3.4). When comparing the 10% BMI decrease per year intervention to the natural course, the hazard and risk of AF both decreased by 4% (HR: 0.96, 95% CI: 0.86, 1.08 and RR: 0.96, 95% CI: 0.92, 1.00, row 6 of Table 3.4).

In g-formula analyses performed in subgroups by sex, hazard ratios and risk ratios comparing non-obese and obese interventions were greater in magnitude for men than women, though most 95% confidence intervals contained the null value of 1 (Tables B.1 and B.2). Hazard ratios and risk ratios for the other comparisons were similar by sex.

Table 3.4: Relative measures of association between body mass index and atrial fibrillation estimated with the g-formula

	Hazard ratio	Risk ratio
10 years		
Non-obese vs. obese	0.77 (0.49, 1.21)	0.77 (0.51, 1.16)
Non-obese vs. natural course	0.98 (0.70, 1.39)	0.98 (0.74, 1.30)
10% decrease in BMI per year vs. natural course	1.00 (0.81, 1.25)	0.99 (0.94, 1.04)
20 years		
Non-obese vs. obese	0.73 (0.58, 0.91)	0.75 (0.63, 0.89)
Non-obese vs. natural course	0.92 (0.78, 1.08)	0.92 (0.83, 1.02)
10% decrease in BMI per year vs. natural course	0.96 (0.86, 1.08)	0.96 (0.92, 1.00)

Data are adjusted hazard ratios and 95% confidence intervals. Cox models are adjusted for SBP, DBP, current smoking status, use of hypertension medication, diabetes status, history of heart failure, and history of myocardial infarction. We note that results for a 5 kg/m2 decrease in BMI represent average results for a shift in BMI, and not necessarily an individuals change over time.

3.5.3 Absolute measures of associations by the g-formula

In Table 3.5, we present the absolute measures of associations by the g-formula. At 20 years of follow-up, the absolute risk of AF was 9.94% (95% CI: 8.48%, 11.40%) had everyone been non-obese and 13.27% had everyone been obese (95% CI: 11.26%, 15.28%), with a difference of -3.33% (95% CI: -5.48%, -1.18%). The mean AF-free lifetime was 19.22 years (95% CI: 19.01, 19.43) had everyone been non-obese and 19.03 years (95% CI: 18.86, 19.20) had everyone been obese over 20 years, with a difference in RMST of 2.25 months (95% CI: -0.66, 5.16). The risk difference was significant, whereas the difference in RMSTs was not. However, the directions of associations were consistent.

When comparing non-obese and 10% decrease in BMI per year while overweight interventions to the natural course, differences favored the BMI intervention but were smaller in magnitude. The non-obese intervention decreased risk by 0.90 percentage points (95% CI: -1.97, 0.18) and increased mean AF-free lifetime by

Table 3.5: Absolute measures of association between body mass index and atrial fibrillation using parametric g-formula

10 years	Intervention	Comparator	Difference in risk (%) or RMST (months)
Non-obese vs. obese			
Risk, %	2.69 (1.35, 4.03)	3.49 (1.97, 5.02)	-0.81 (-2.21, 0.59)
RMST, years	9.88 (9.79, 9.96)	9.86 (9.78, 9.93)	0.24 (-0.94, 1.42)
Non-obese vs. natural course			
Risk, %	2.69 (1.34, 4.04)	2.74 (1.65, 3.83)	-0.05 (-0.92, 0.82)
RMST, years	9.88 (9.78, 9.97)	9.88 (9.83, 9.93)	-0.09 (-1.00, 0.82)
10% decrease in BMI per year vs. natural course			
Risk, %	2.71 (1.63, 3.79)	2.75 (1.66, 3.84)	-0.04 (-0.19, 0.11)
RMST, years	9.88 (9.83, 9.93)	9.88 (9.83, 9.93)	0.01 (-0.06, 0.07)
20 years	**Intervention**	**Comparator**	**Difference in risk (%) or RMST (months)**
Non-obese vs. obese			
Risk, %	9.94 (8.48, 11.40)	13.27 (11.26, 15.28)	-3.33 (-5.48, -1.18)
RMST, years	19.22 (19.01, 19.43)	19.03 (18.86, 19.20)	2.25 (-0.66, 5.16)
Non-obese vs. natural course			
Risk, %	9.95 (8.50, 11.40)	10.84 (9.73, 11.96)	-0.90 (-1.97, 0.18)
RMST, years	19.22 (19.01, 19.43)	19.20 (19.10, 19.29)	0.30 (-1.96, 2.55)
10% decrease in BMI per year vs. natural course			
Risk, %	10.45 (9.31, 11.60)	10.88 (9.75, 12.00)	-0.43 (-0.86, 0.00)
RMST, years	19.21 (19.12, 19.31)	19.19 (19.10, 19.29)	0.23 (-0.11, 0.57)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples. Analyses are adjusted for SBP, DBP, current smoking status, use of hypertension medication, diabetes status, history of heart failure, and history of myocardial infarction.

0.30 months (95% CI: -1.96, 2.55), whereas the 10% decrease in BMI per year while overweight intervention decreased risk by 0.43 percentage points (95% CI: -0.86, 0.00) and increased AF-free lifetime by 0.23 months (95% CI: -0.11, 0.57). Figure 3.3 depicts the g-formula Kaplan-Meier curves over 20 years for all intervention comparisons.

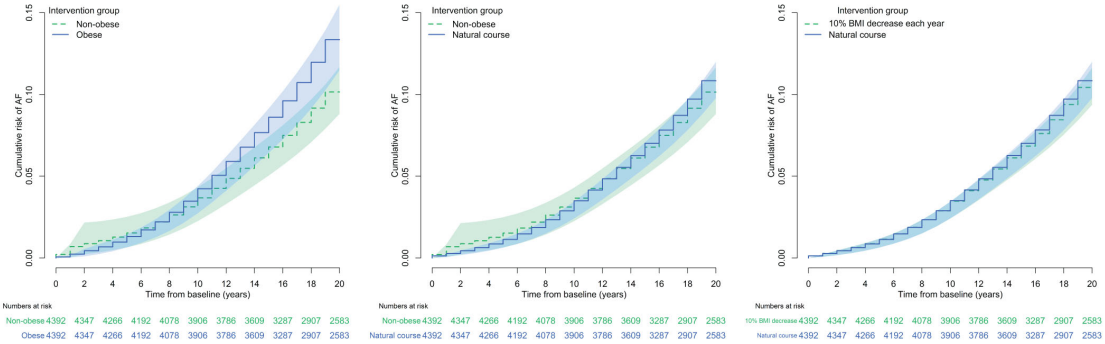


Figure 3.3: G-formula survival probabilities for cumulative risk of atrial fibrillation comparing simulated obese and non-obese populations

In analyses stratified by sex, the magnitude of association was again greater in men than women (Tables B.1 and B.2). The risk difference was -3.30% (95% CI: -6.89%, 0.30%) in men and -2.69% (95% CI: -5.37, -0.02%) in women, while the gain in mean AF-free lifetime over 20 years between non-obese and obese was 1.89 months (95% CI: -3.67, 7.44) in men and 1.19 months (95% CI: -2.34, 4.71) in women. Risks and mean AF-free lifetimes were similar between the non-obese intervention, 10% BMI decrease per year intervention, and the natural course. Though the measures of association were close to the null value, the direction of risk differences and differences in RMSTs disagreed when comparing non-obese to the natural course in both men and women. In Figure B.3 of the Appendix, we plot

the Kaplan-Meier curves for men and women to illustrate how measures of association may disagree when curves cross or overlap. In this case, the difference in cumulative risks is a point estimate at 20 years, while the difference in RMSTs reflects the entire time horizon from 0 to 20 years.

3.5.4 Sensitivity analyses

In additional sensitivity analyses, results were overall similar to our main findings. In cancer-free participants (1,825 men and 2,045 women), both relative and absolute associations were consistent with original analyses (Table B.3). For the complete-case setting (70.6% participants), 155 Original cohort participants (18.4%) and 2,947 Offspring cohort participants (83.0%) had complete information at entry and lagged exams. When comparing non-obese vs. obese, all associations at 20 years were consistent with our main approach but slightly greater in magnitude while comparisons of 10% decrease in BMI per year to the natural course were nearly identical (Table B.4). Non-obese vs. natural course associations disagreed in direction, though very small and close to the null. Our sensitivity analysis examining entry at age 60 years, plus or minus five years, included 3,350 women and 2,799 men. Both relative and absolute measures of association were similar to our main analyses but slightly smaller in magnitude (Table B.5). Under all interventions, AF risks were higher while mean AF-free lifetimes were lower (20.65% and 18.33 years under the natural course, Table B.5). Finally, when fitting pooled negative binomial models instead of pooled logistic models for the hazard of AF, results were nearly identical (Table B.6).

3.6 DISCUSSION

In summary, we found that decreased BMI was associated with lower rate of AF in men and women after accounting for time-varying covariates which depend on previous BMI when using varying measurse of association (RMST difference, risk difference, risk ratio, and hazards ratio). We demonstrated how g-methods can account for time-varying covariates which are also intermediate variables, unlike conventional statistical methods.

If a variable is associated with both the exposure and outcome, the association between exposure and outcome may be biased unless confounding is addressed. In longitudinal studies, confounding is commonly addressed by adjusting for time-varying covariates. However, if a time-varying covariate is also an intermediate variable between exposure and outcome, it gives rise to time-varying confounding. Adjustment for the time-varying covariate with traditional methods will preclude observing the full association between exposure and outcome, leading to biased estimates, and statistical methods like the g-methods are needed.

Time-varying confounding may commonly occur in longitudinal cardiovascular studies with repeated measurements of risk factors over time. For example, when studying the association of isolated systolic hypertension and cardiovascular death, one would account for confounding by adjusting for arterial rigidity.(Witteman et al., 1998) However, isolated systolic hypertension may lead to arterial rigidity, which in turn leads to negative health outcomes. The association between isolated systolic hypertension and cardiovascular death may be due to increased arterial rigidity that developed in between, induced by hypertension. Therefore, standard adjustment for arterial rigidity would block the association between isolated systolic hypertension and cardiovascular death. Instead, g-methods

like the g-formula studied here can address time-varying confounding.

In this paper, we quantified the association both on the relative scale and on the absolute scale. Similar to prior works, we observed large associations on the relative scale when comparing non-obese and obese interventions. However, associations on the absolute scale provide important context since the absolute risk of AF in individuals 45 to 55 years of age are relatively small at 10 and 20 years of follow-up (2.74% and 10.84% under the natural course, Table 3.5).

Using the g-formula, we observed a decrease in AF risk and a gain in mean AF-free lifetime if participants had been non-obese compared to obese, which is consistent with the literature.(Wang et al., 2004; Abed et al., 2013) The magnitude of this difference was larger at 20 years than 10 years. However, contrasts of non-obese and 10% decrease in BMI per year to the natural course yielded weaker associations. The smaller magnitude is somewhat expected due to the different reference group between comparisons: the first comparison contrasted maximally different groups, while the second and third comparisons contrasted groups that were relatively more similar.

While conventional Cox proportional hazards models can accommodate time-dependent covariates, they do not permit time-dependent confounders which depend on BMI history. Therefore, the parametric g-formula allows us to examine a different contrast of interest, i.e. the difference in mean-time free of AF had participants been obese vs. non-obese over 20 years. In fact, allowing covariates to vary over time when they are actually mediated by the exposure of interest can prevent observing the association between exposure and outcome (Figure 3.1). Overall, hazard ratios estimated with the g-formula comparing non-obese vs. obese were greater in magnitude than hazard ratios estimated with the Cox model. In fact, at

20 years, the Cox hazard ratio with all covariates and obesity measured at baseline was closer in magnitude to the g-formula estimate, while including covariates as time-varying reduced the magnitude. This is likely due to the covariates role as intermediate variables between obesity and incident AF: the causal pathway between obesity and AF is blocked when adjusting for the intermediate covariates (Figure 3.1). However, we note that the Cox model and g-formula hazard ratio do not measure the same thing: the Cox model compares non-obese vs. obese individuals, while the g-formula compares everyone had they been non-obese vs. everyone had they been obese. Additionally, we simulated non-obese as 18.5 to 25.0 in BMI and obese as 30 to 41 in BMI to avoid BMIs out of range in our sample.

Many studies have demonstrated that obesity is an AF risk factor.(Alonso et al., 2013; Berkovitch et al., 2016; Frost et al., 2014; Khan et al., 2018; Magnani et al., 2013; Norby et al., 2016; Rahman et al., 2016; Rosengren et al., 2009; Wang et al., 2004; Wong et al., 2015; Aune et al., 2017) We fit conventional Cox proportional hazards models, with and without time-varying covariates, to demonstrate that our data produces results consistent with the literature. Conventional hazard ratios estimated with Cox models with all risk factors measured at baseline were fairly consistent with the literature. In large meta-analyses of 51 and 25 studies, respectively, Wong et al. (2015) and Aune et al. (2017) demonstrated that increases in BMI were associated with increased risk of AF. Wong et al reported a 19% to 29% increase in AF odds per 5 kg/m^2 increase in BMI, while Aune et al reported a 28% increase in AF risk. Both of these results are greater in magnitude than the HRs we report at 20 years of follow-up (14% increase in hazard considering all risk factors at baseline, Table 3.3). However, our results were similar in magnitude to Schnabel et al. (2009)s findings in the Cardiovascular Health Study white sub-population

(14% increase in AF odds per 5 kg/m^2 increase in BMI), which may be comparable to FHS participants.

While obesity is a known risk factor for AF, our results were sometimes small on the absolute scale. Similar findings exist in the literature. For instance, in a large randomized trial of overweight and obese individuals with type 2 diabetes, the Look AHEAD (Action for Health in Diabetes) Study found rates of AF were not affected by a lifestyle intervention which included weight loss.(Alonso et al., 2015) However, the weight loss of the intervention was modest (6.0% mean weight loss from baseline). Additionally, in the Swedish Obese Subjects (SOS) matched cohort study, the risk of new-onset AF over a median follow-up of 19 years was 12.4% among those who underwent bariatric surgery vs. 16.8% among matched referents in usual care.(Jamaly et al., 2016) However, matching was imperfect in this study.

Differences in RMSTs have an appealing interpretation, and provide a magnitude of association which relative measures, such as the hazard ratio and risk ratio, cannot provide. An advantage of a measure in the time domain is the ease in interpretation. Unlike the hazard ratio, the difference in RMSTs does not depend on the proportional hazards assumption. The median survival time is an alternative measure, but cannot be estimated if the survival probability never decreases to 0.5 in the time horizon, which is common in cardiovascular outcomes. In our data, the median AF-free survival time cannot be estimated as the probability of AF-free survival never decreases to 0.5 over 20 years. However, the restricted mean survival time can always be estimated. To our knowledge, the difference in RMSTs using g-formula methods has only been reported once.(Lodi et al., 2015) Some previous applications of the parametric g-formula assessed examination cycles with

approximately equal intervals.(Danaei et al., 2016; Taubman et al., 2009; Witteman et al., 1998; Jain et al., 2016; Lajous et al., 2013; Ueda et al., 2018; Vangen-Lønne et al., 2018; Nascimento et al., 2017; Garcia-Aymerich et al., 2014; Danaei et al., 2013)

There are several limitations to our findings. First, we do not know reasons for weight loss in FHS participants and major weight loss is observed infrequently. Previous trajectory analyses in FHS and the Atherosclerosis Risk in Communities (ARIC) study have identified relatively flat longitudinal patterns in weight.(Norby et al., 2016; Rahman et al., 2016) Although we measure weight, variations in weight over time could be due to many reasons, including severe illness. However, a sensitivity analysis restricted to cancer-free participants showed consistent findings. Weight loss can occur with aging, yet AF typically occurs in older ages.

Second, our statistical methods make several assumptions. Since we included seven time-dependent confounders measured at examinations, our analytic dataset was fairly high dimensional and prone to missingness, which motivated our decision to perform multiple imputation. We chose sequential multiple imputation in order to capture the longitudinal nature of the data and exclude participants after death.(Ning et al., 2013) Multiple imputation may not be appropriate if our data were missing not at random. Additionally, while time intervals should be approximately equal in a pooled logistic model, they must be equal in order to calculate the restricted mean survival time. However, FHS examinations are not equally spaced across participants. We chose linear interpolation to obtain covariate values at the yearly level, which assumes that covariate values change linearly between examinations.(Genolini et al., 2016) In addition, methods for causal inference including the parametric g-formula assume consistency, and in consequence

exchangeability (lack of confounding) and positivity (positive conditional probability of exposure), in order to make causal inference.(Robins et al., 2000; Hernán & Taubman, 2008) Consistency requires that interventions be well-defined, and does not typically hold for interventions based on BMI measurements. This is because we do not know the underlying reasons for BMI changes in FHS participants. We also do not account for the semi-competing risk of death, in which death may preclude incident AF.

Finally, the FHS participants are mostly of European ancestry, therefore our results may not be generalizable to all populations.

In conclusion, decreased BMI was associated with lower rate of new-onset AF after accounting for time-varying covariates which depend on previous BMI, which conventional approaches cannot do. The parametric g-formula is a flexible method to account for time-varying covariates which are also intermediate variables, which may commonly occur when assessing vascular risk factors. Furthermore, we illustrate how to derive and interpret absolute measures of association, including the difference in RMST. Absolute measures of association offer additional insight into the data and can improve our understanding of associations.

Directed acyclic graphs allow visualizing pathways between exposures, other covariates, and outcomes. Directed acyclic graphs can also help identifying time-varying confounding. Unlike Cox models, g-methods such as the parametric g-formula, marginal structural models with inverse probability weighting, and g-estimation, can adjust for time-dependent confounding (if any) when analyzing data from longitudinal cardiovascular studies. Marginal structural models can be used to assess fixed interventions, while the g-formula can also be used to assess dynamic interventions.(Bodnar et al., 2004) SAS macros and R functions are readily

available to implement these methods. Furthermore, we recommend the reporting of absolute measures of association, such as the difference in absolute risk and RMST, to give additional context to relative measures of association.

CHAPTER 4

Adjusted restricted mean times lost with competing risks

4.1 BACKGROUND

Competing risks data arise commonly in biomedical research when time to death from a specific disease may be of interest, but individuals can die from other causes. Similarly, semi-competing risks occur when the time to a non-terminal event is of interest but its observation is subject to a terminal event. Incidence of a competing event precludes observation of the event of interest, making it impossible to occur. In contrast, in the presence of censoring, an event can still occur but the follow-up time is not completely observed. In this context, between-group comparisons for a specific event have been studied using the cumulative incidence, cause-specific hazard, or subdistribution hazard.(Gray, 1988; Fine & Gray, 1999; Kulathinal & Gasbarra, 2002)

In the presence of competing risks, the difference in RMST is inadequate because it does not distinguish between event types, and the competing event is considered censored. Andersen (2013) and Zhao et al. (2018) proposed to measure the difference in restricted mean times lost (RMTL). With a single cause of failure, RMTL is equal to τ minus RMST. With competing risks, the RMTL is the area under the cause-specific cumulative incidence function. As a consequence, the difference in RMTL can be interpreted as the increase or decrease in lost life expectancy due to a specific cause of death or, in the case of a non-fatal outcome, the mean difference in disease-free time, up to time τ.

Methods for inference on the difference in RMTL with competing risks data and adjustment for covariates are scarce. Andersen (2013) described a generalized

linear model for the cause-specific RMTL using the pseudo-observation technique. In this chapter, we propose an IPW-adjusted estimator of the difference in cause-specific RMTL. We also present a regression model for the RMTL conditional on covariates estimated with inverse probability of censoring weighting (IPCW).

In Section 4.2, we discuss estimation of the difference in cause-specific RMTL, without adjustment for covariates. In Section 4.3, we present estimation of the marginal RMTL adjusted for covariates with IPW and its variance. In Sections 4.4, we describe an IPCW regression model for RMTL conditional on covariates. In Section 4.5, we present two simulation studies to assess the finite sample performance of the IPW estimator and the IPCW regression model. In Section 4.6, we apply the proposed methods to incident AF and cause-specific mortality after AF with data from the Framingham Heart Study.

4.2 DIFFERENCE IN CAUSE-SPECIFIC RESTRICTED MEAN TIMES LOST

Without loss of generality, we consider the case of a terminal or non-terminal event of interest $\epsilon = 1$, whose occurrence is subject to a competing event $\epsilon = 2$. Our notations extend to scenarios with k events. Let T be the time to event, C be the censoring time, $X = T \wedge C$ be the observed time to event, and $\delta = I(T \leq C)$ be an indicator of non-censoring. We assume that C is independent of T. We observe $(X_i, \delta_i \epsilon_i)$ for individuals $i = 1, \ldots, n$.

The cumulative incidence function for the event of interest is defined as $F_1(t) = P(T \leq t, \epsilon = 1) = \int_0^t S(u)\alpha_1(u)du$ where $S(t) = \exp(-\int_0^t A(u)du)$ is the overall survival function, $\alpha_k(u)$ is the cause-specific hazard function for event k, and $A(u) = \int_0^u \{\alpha_1(t) + \alpha_2(t)\}dt$ is the cumulative all-cause hazard function. Let $R = T \wedge \tau$ be the restricted time to event. We define the cause-specific RMTL,

$\nu_1(\tau)$, as the area under the cause-specific cumulative incidence function up to timepoint τ.

Here, we show that $\nu_1(\tau) = E\left[(\tau - R) \cdot I(\epsilon = 1)\right]$ is the area under the cause-specific cumulative incidence function, or cause-specific restricted mean time lost (RMTL). While we cannot take the expectation of the subdistribution failure time, $T \cdot I(\epsilon = 1)$, we can find the expectation of the random variable $R \cdot I(\epsilon = 1)$,

$$E\left[R \cdot I(\epsilon = 1)\right] = \int_0^\tau 1 - F_1(t)dt = \tau - \int_0^\tau F_1(t)dt,$$

which is the time alive and free of cause 1, but still includes the expected time lost to cause 2. To isolate the time lost due to cause 1, $\nu_1(\tau)$, we take the expectation of $(\tau - R) \cdot I(\epsilon = 1)$,

$$\begin{aligned} E\left[(\tau - R) \cdot I(\epsilon = 1)\right] &= E\left[(\tau - R) \cdot I(\epsilon = 1) | T \leq \tau\right] \cdot P(T \leq \tau) \\ &\quad + E\left[(\tau - R) \cdot I(\epsilon = 1) | T > \tau\right] \cdot P(T > \tau) \\ &= E\left[(\tau - T) \cdot I(\epsilon = 1) | T \leq \tau\right] \cdot P(T \leq \tau) \end{aligned}$$

Since $P(\tau - T \leq t, \epsilon = 1) = 1 - F_1(\tau - t)$ and $f_{1,\tau-t}(t) = \frac{d}{dt}(1 - F_1(\tau - t)) = f_1(\tau - t)$,

$$\begin{aligned} E\left[(\tau - R) \cdot I(\epsilon = 1)\right] &= \int_0^\tau \frac{(\tau - T)}{P(T \leq \tau)} f_1(\tau - t)dt \cdot P(T \leq \tau) \\ &= \int_0^\tau \tau f_1(\tau - t)dt - \int_0^\tau T f_1(\tau - t)dt. \end{aligned}$$

By integration by parts, it follows that

$$\begin{aligned} E\left[(\tau - R) \cdot I(\epsilon = 1)\right] &= \tau\left(F_1(0) - F_1(\tau)\right) - \left(\tau F_1(0) - \tau F_1(\tau) - \int_0^\tau F_1(t)dt\right) \\ &= \int_0^\tau F_1(t)dt \\ &= \nu_1(\tau). \end{aligned}$$

This relationship can be extended to the cause-specific RMTL conditional on covariates $\boldsymbol{Z}^*$.

We estimate the RMTL by integrating the Aalen-Johansen estimator of the cumulative incidence function

$$\hat{\nu}_1(\tau) = \int_0^\tau \hat{F}_1(t)dt. \tag{4.1}$$

We approximate Equation (4.1) with Riemann sums, $\hat{\nu}_1(\tau) = \sum_{t_j<\tau}(t_{j+1} - t_j)\hat{F}_1(t_j)$ where $t_j : t_1, \ldots, t_D$ are the ordered, unique event times prior to τ and $t_{D+1} = \tau$.

Let $N_k(t) = \sum_{i=1}^n I(X_i \leq t)\delta_i I(\epsilon_i = k)$ be the number of events of cause k by time t, $N(t) = N_1(t) + N_2(t)$ the total number of events by time t, $\Delta N_k(t)$ the increment $N_k(t) - N_k(t-)$, and $Y(t) = \sum_{i=1}^n I(X_i \geq t)$ the number at risk just before t. Let $\hat{S}(t)$ be the Kaplan Meier estimator of being alive and event-free at time t, $\hat{S}(t) = \prod_{t_j \leq t}\left(1 - \frac{\Delta N(t_j)}{Y(t_j)}\right)$, and $\hat{\theta}(t_j) = \hat{S}(t_{j-1})\frac{\Delta N_1(t_j)}{Y(t_j)}$. The cumulative incidence function is then estimated by the Aalen-Johansen estimator $\hat{F}_1(t) = \sum_{t_j \leq t}\hat{\theta}(t_j)$, where t_j are the unique event times. (Aalen & Johansen, 1978) The variance and covariance for the cumulative incidence function are estimated using Greenwood's

formula (Dinse & Larson, 1986; Gaynor et al., 1993),

$$\widehat{\text{var}}\left(\hat{F}_1(t)\right) = \sum_{t_j \le t} \widehat{\text{var}}\left(\hat{\theta}(t_j)\right) + 2\sum_{t_j < t}\sum_{t_j < t_k \le t} \widehat{\text{cov}}\left(\hat{\theta}(t_j), \hat{\theta}(t_k)\right)$$

and

$$\widehat{\text{cov}}\left(\hat{F}_1(t), \hat{F}_1(u)\right) = \widehat{\text{var}}\left(\hat{F}_1(t)\right) + \sum_{t_j \le t}\sum_{t < t_k \le u} \widehat{\text{cov}}\left(\hat{\theta}(t_j), \hat{\theta}(t_k)\right) \quad \text{for } t < u,$$

where

$$\widehat{\text{var}}\left(\hat{\theta}(t_j)\right) = \left(\hat{\theta}(t_j)\right)^2 \left(\frac{Y(t_j) - \Delta N_1(t_j)}{\Delta N_1(t_j) Y(t_j)} + \sum_{t_l < t_j} \frac{\Delta N(t_l)}{Y(t_l)(Y(t_l) - \Delta N(t_l))} \right)$$

and

$$\widehat{\text{cov}}\left(\hat{\theta}(t_j), \hat{\theta}(t_k)\right) = \hat{\theta}(t_j)\hat{\theta}(t_k) \left(-\frac{1}{Y(t_j)} + \sum_{t_l < t_j} \frac{\Delta N(t_l)}{Y(t_l)(Y(t_l) - \Delta N(t_l))} \right)$$

for $t_j < t_k$. Then, the variance of the cause-specific RMTL is

$$\begin{aligned}\widehat{\text{var}}\left(\hat{v}_1(\tau)\right) = &\sum_{t_j < \tau} (t_{j+1} - t_j)^2 \widehat{\text{var}}\left(\hat{F}_1(t_j)\right) \\ &+ 2\sum_{t_j < \tau}\sum_{t_k < t_j} (t_{j+1} - t_j)(t_{k+1} - t_k) \widehat{\text{cov}}\left(\hat{F}_1(t_j), \hat{F}_1(t_k)\right).\end{aligned}$$

With two groups defined by $A = 1$ and $A = 0$, the difference in cause-specific RMTL is given by $\delta = \hat{v}_1(\tau|A = 1) - \hat{v}_1(\tau|A = 0)$, with variance $\hat{\text{var}}(\delta) = \hat{\text{var}}\{\hat{v}_1(\tau|A = 1)\} + \hat{\text{var}}\{\hat{v}_1(\tau|A = 0)\}$.

With a single group, the timepoint τ should be less than or equal to the greatest observed time, $\tau \le \max(X_i)$, and may be chosen clinically or empirically, e.g. the

95th percentile of observed times or the maximum observed time, $\max(X_i)$, which approximates $\inf_\tau\{P(C \geq \tau) = 0\}$. (Tian et al., 2020) When taking the difference in RMTL between groups, τ should be less than or equal to the minimum of the largest event time across groups, $\tau \leq \min\{\max(X_i|A_i = 1), \max(X_i|A_i = 0)\}$. Tian et al. (2020) demonstrated that $\max(X_i)$ may be an inappropriate choice for τ if nearby observed times are sparse, as evidenced by a flat tail in the cumulative incidence curve, and should be chosen with caution.

4.3 ADJUSTED DIFFERENCE IN CAUSE-SPECIFIC RESTRICTED MEAN TIMES LOST

In observational studies, we must adjust for covariates. Here we demonstrate how to estimate the between-group difference in cause-specific RMTL adjusted for covariates. We extend Equation (4.1) by considering the area under the cumulative incidence function adjusted for covariates with IPW.

We again consider two groups defined by $A = 1$ and $A = 0$. Let $\boldsymbol{Z}$ denote a $n \times p$ matrix of covariates and $p_{ia} = P(A_i = a|\boldsymbol{Z_i})$. So p_{i1} is the propensity score for exposure. The inverse probability weight is defined as $w_i = \frac{I(A_i=1)}{p_{i1}} + \frac{I(A_i=0)}{1-p_{i1}}$. Then, the IPW Aalen-Johansen estimator of the cumulative incidence function is given by

$$\hat{F}_{1a}^W(t) = \sum_{t_j \leq t} \hat{\theta}_a^W(t_j)$$

where $\hat{\theta_a^W}(t_j) = \hat{S}_a^W(t_{j-1})\frac{\Delta N_{1a}^W(t_j)}{Y_a^W(t_j)}$, $N_{ka}^W(t) = \sum_{i:A=a} w_i \cdot I(X_i \leq t)\delta_i I(\epsilon_i = k)$ is the weighted number of events of cause k in exposure group $A = a$, $\Delta N_{ka}^W(t) = N_{ka}^W(t) - N_{ka}^W(t-)$ is the increment in weighted events in group a, $Y_a^W(t) = \sum_{i:A=a} w_i \cdot I(X_i \geq t)$ is the weighted number at risk in exposure group a, and

$\hat{S}_a^W(t) = \prod_{t_j \leq t} \left(1 - \frac{\Delta N_{1a}^W(t_j) + \Delta N_{2a}^W(t_j)}{Y_a^W(t_j)}\right)$ is the weighted overall survival function for being alive and event-free in group a. (Cole et al., 2015; Neumann & Billionnet, 2016) We define the estimated variance and covariance of the IPW Aalen-Johansen estimator as

$$\hat{\text{var}}\{\hat{F}_{1a}^W(t|A=a)\} = \sum_{t_j \leq t} \hat{\text{var}}\{\hat{\theta}_a^W(t_j)\} + 2 \sum_{t_j < t} \sum_{t_j < t_k \leq t} \hat{\text{cov}}\{\hat{\theta}_a^W(t_j), \hat{\theta}_a^W(t_k)\}$$

and

$$\hat{\text{cov}}\{\hat{F}_{1a}^W(t), \hat{F}_{1a}^W(u)\} = \hat{\text{var}}\{\hat{F}_1^W(t)\} + \sum_{t_j \leq t} \sum_{t < t_k \leq u} \hat{\text{cov}}\{\hat{\theta}_a^W(t_j), \hat{\theta}_a^W(t_k)\} \quad \text{for } t < u$$

where

$$\hat{\text{var}}\{\hat{\theta}_a^W(t_j)\} = \{\hat{\theta}_a^W(t_j)\}^2 \left\{ \frac{Y_a^W(t_j) - \Delta N_1^W(t_j)}{M_a(t_j) \Delta N_1^W(t_j)} + \sum_{t_l < t_j} \frac{\Delta N_a^W(t_l)}{M_a(t_l)(Y_a^W(t_l) - \Delta N_a^W(t_l))} \right\}$$

and

$$\hat{\text{cov}}\left\{\hat{\theta}_a^W(t_j), \hat{\theta}_a^W(t_k)\right\} = \hat{\theta}_a^W(t_j)\hat{\theta}_a^W(t_k) \left\{ -\frac{1}{M_a(t_j)} + \sum_{t_l < t_j} \frac{\Delta N_a^W(t_l)}{M_a(t_l)(Y_a^W(t_l) - \Delta N_a^W(t_l))} \right\}$$

for $t_j < t_k$, $M_a(t_j) = \frac{Y_a^W(t_j)^2}{\sum_{i:T_i > t_j} w_i^2}$, and assuming $1/M_a(t_j) \xrightarrow{p} 0$.(Choi et al., 2019)

We estimate the RMTL in exposure group a by integrating the inverse probability weighted cumulative incidence function

$$\hat{v}_{1a}^W(\tau) = \int_0^\tau \hat{F}_{1a}^W(t)dt. \tag{4.2}$$

We approximate Equation (4.2) as $\hat{v}_{1a}^W(\tau) = \sum_{t_j < \tau} (t_{j+1} - t_j)\hat{F}_{1a}^W(t_j)$. We estimate

the variance of the IPW estimator of the RMTL in exposure group a by

$$\hat{\text{var}}\{\hat{\nu}_{1a}^{W}(\tau)\} = \sum_{t_j<\tau}(t_{j+1}-t_j)^2\hat{\text{var}}\left\{\hat{F}_{1a}^{W}(t_j)\right\} + 2\sum_{t_j<\tau}\sum_{t_k<t_j}(t_{j+1}-t_j)(t_{k+1}-t_k)\hat{\text{cov}}\left\{\hat{F}_{1a}^{W}(t_j),\hat{F}_{1a}^{W}(t_k)\right\}. \quad (4.3)$$

Then, the adjusted difference in RMTL between exposure groups $A = 1$ and $A = 0$ is

$$\delta^{W} = \hat{\nu}_{1,1}^{W}(\tau) - \hat{\nu}_{1,0}^{W}(\tau)$$

with variance

$$\hat{\text{var}}\left(\delta^{W}\right) = \hat{\text{var}}\left\{\hat{\nu}_{1,1}^{W}(\tau)\right\} + \hat{\text{var}}\left\{\hat{\nu}_{1,0}^{W}(\tau)\right\}$$

assuming the weights are known. In practice, we estimate the propensity score p_{ia} and plug $\hat{p}_{ia}$ into the quantities introduced above.

4.4 INVERSE PROBABILITY OF CENSORING WEIGHTED GENERALIZED LINEAR MODEL FOR THE RESTRICTED MEAN TIME LOST

We describe an IPCW regression model to assess associations between covariates and the cause-specific RMTL.(Tian et al., 2014) Let $\widetilde{\delta} = I(R \leq C)$ and $\boldsymbol{Z}^* = (1, A, \boldsymbol{Z})$ denote the $n \times (p+2)$ matrix of predictors allowing an intercept term. We can model the RMTL for event $\epsilon = 1$ with a generalized linear model, $g\{\nu_1(\tau)|\boldsymbol{Z}_i^*\} = g[E\{(\tau - R_i) * I(\epsilon_i = 1)|\boldsymbol{Z}_i^*\}] = \boldsymbol{Z}_i^*\boldsymbol{\beta}$ with link function $g(\cdot)$. Model coefficients $\boldsymbol{\beta}$ represent differences in cause-specific RMTL when using the identity link, and log ratios of cause-specific RMTL when using the log link. Tian et al. (2014) also suggested a logit-like link function, $g\{\nu_1(\tau)\} = \log[\nu_1(\tau)/\{\tau - \nu_1(\tau)\}]$, which restricts the predicted RMTL to $(0, \tau)$.

Coefficients are estimated with the IPCW technique by solving the weighted estimating equation,

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \boldsymbol{Z}_i^* \{(\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta})\} = 0 \tag{4.4}$$

where $\hat{G}(t)$ is the Kaplan-Meier estimator of the non-censoring distribution $G(t)$. The outcome for each uncensored individual is either $\tau - R_i$ or 0 due to the event type indicator, $I(\epsilon_i = 1)$; essentially, individuals with a competing event do not lose any time to cause 1. The non-censoring indicator $\tilde{\delta}_i$ equals 0 for censored observations and 1 for uncensored observations; thus, censored observations have a weight of 0 and do not contribute to the estimating equation. However, the censoring distribution $\hat{G}(t)$ is estimated from the sample of all individuals.

We show that the IPCW estimating equation is unbiased. In the generalized linear model, we assume $E\left[(\tau - R) \cdot I(\epsilon = 1) | \boldsymbol{Z}^*\right] = g^{-1}(\boldsymbol{\beta Z}^*)$. Following the conditional expectation arguments of Bang & Tsiatis (2002) and Scheike et al. (2008), we have

$$\begin{aligned}
&E\left(\sum_{i=1}^{n} \frac{\widetilde{\delta}_i}{G(R_i)} \boldsymbol{Z}_i^* \{(\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{\beta Z}_i^*)\} \middle| \boldsymbol{Z}_i^*\right) \\
&= n\boldsymbol{Z}^* E\left(E\left[\frac{\widetilde{\delta}}{G(R)} \{(\tau - R) \cdot I(\epsilon = 1) - g^{-1}(\boldsymbol{\beta Z}^*)\} \middle| R, \boldsymbol{Z}^*, \epsilon\right] \middle| \boldsymbol{Z}^*\right) \\
&= n\boldsymbol{Z}^* E\left(\frac{E\left[\widetilde{\delta} \middle| R, \boldsymbol{Z}^*, \epsilon\right]}{G(R)} \{(\tau - R) \cdot I(\epsilon = 1) - g^{-1}(\boldsymbol{\beta Z}^*)\} \middle| \boldsymbol{Z}^*\right) \\
&= n\boldsymbol{Z}^* E\left((\tau - R) \cdot I(\epsilon = 1) - g^{-1}(\boldsymbol{\beta Z}^*) | \boldsymbol{Z}^*\right) \\
&= 0.
\end{aligned}$$

It follows that

$$E\left(E\left[\sum_{i=1}^{n}\frac{\widetilde{\delta}_i}{G(R_i)}\boldsymbol{Z}_{\boldsymbol{i}}^{*}\{(\tau-R_i)\cdot I(\epsilon_i=1)-g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\boldsymbol{\beta})\}\middle|\boldsymbol{Z}_{\boldsymbol{i}}^{*}\right]\right)=0.$$

Since $G(t)$ is unknown, we propose estimating $G(t)$ by the Kaplan-Meier estimator $\hat{G}(t)$ and solving the following equation to estimate $\boldsymbol{\beta}$,

$$\sum_{i=1}^{n}\frac{\widetilde{\delta}_i}{\hat{G}(R_i)}\boldsymbol{Z}_{\boldsymbol{i}}^{*}\{(\tau-R_i)\cdot I(\epsilon_i=1)-g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\boldsymbol{\beta})\}=0.$$

In Appendix C, we show the estimated coefficients $\hat{\boldsymbol{\beta}}$ are asymptotically normally distributed with covariance matrix estimated by the sandwich estimator $\hat{\text{var}}(\hat{\boldsymbol{\beta}})=\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ where

$$\hat{A}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_{\boldsymbol{i}}^{*\otimes\mathbf{2}}\frac{d}{d\hat{\boldsymbol{\beta}}}g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\hat{\boldsymbol{\beta}}),$$

$$\begin{aligned}\hat{B}=\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{\delta}_i}{\hat{G}(R_i)}\Bigg\{&\left[\boldsymbol{Z}_{\boldsymbol{i}}^{*\otimes\mathbf{2}}\left\{(\tau-R_i)\cdot I(\epsilon_i=1)-g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\hat{\boldsymbol{\beta}})\right\}^{2}\right]\\&+\int_0^{\tau}\left(\left[\boldsymbol{Z}_{\boldsymbol{i}}^{*}\{(\tau-R_i)\cdot I(\epsilon_i=1)-g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\hat{\boldsymbol{\beta}})-\hat{K}(u)\}\right]^{\otimes 2}I(R_i\geq u)\right)\frac{d\hat{\Lambda}_c(u)}{\hat{G}(u)}\Bigg\},\end{aligned}$$

and

$$\hat{K}(u)=\frac{1}{\hat{S}(u)n}\sum_{i=1}^{n}\left[\boldsymbol{Z}_{\boldsymbol{i}}^{*}\{(\tau-R_i)\cdot I(\epsilon_i=1)-g^{-1}(\boldsymbol{Z}_{\boldsymbol{i}}^{*}\hat{\boldsymbol{\beta}})\}I(R_i\geq u)\right].$$

4.5 SIMULATION STUDY

We conducted Monte Carlo simulation studies to assess the performance of the IPW method in estimating the marginal difference in RMTL between exposure groups in the presence of covariates, and the IPCW regression method in estimating the difference in RMTL conditional on covariates. We generated data with proportional and non-proportional subdistribution hazards (crossing cumulative incidence functions). (Fine & Gray, 1999; Li et al., 2015; Zhou et al., 2013) We estimated the mean bias, mean relative bias, root mean squared error, relative standard error (SE), and empirical coverage rate of both methods. We define the relative standard error as the ratio of the mean estimated SE and the Monte Carlo empirical SE.

4.5.1 Inverse probability weighted adjusted difference in restricted mean time lost

4.5.1.1 Data generation

We first generated 5 independent covariates $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4, Z_5)$, each $\mathcal{N}(0,1)$. Then we generated the binary exposure, A, from a Bernoulli distribution with $p = \text{expit}(\boldsymbol{Z'\alpha})$, where $\boldsymbol{\alpha} = (\alpha_0, \ln(0.7), \ln(0.6), \ln(0.5), \ln(0.9), \ln(0.8))^T$ and $\boldsymbol{Z'} = (1, \boldsymbol{Z})$. We set the intercept α_0 to allow 25% and 50% exposure.

To generate event times, we let the baseline subdistribution hazard function $\lambda_{10}(t) = \gamma \exp(\rho t)$ follow a Gompertz distribution. We then defined the subdistribution hazard function as $\lambda_1(t; \tilde{\boldsymbol{Z}}) = \lambda_{10}(t) \exp\{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1} + \tilde{\boldsymbol{Z}}\boldsymbol{\psi_2} t\}$, where $\tilde{\boldsymbol{Z}} = (A, \boldsymbol{Z})$, $\boldsymbol{\psi_1} = \{\psi_{11}, \ln(0.9), \ln(0.8), \ln(0.7), \ln(0.6), \ln(0.5)\}^T$, and $\boldsymbol{\psi_2} = (\psi_{21}, 0, 0, 0, 0, 0)^T$, in which ψ_{11} is the main effect of A and ψ_{21} is the time-varying effect of A. We then defined the cumulative incidence function for event

$\epsilon = 1$ as $F_1(t; \tilde{\boldsymbol{Z}}) = P(T \leq t, \epsilon = 1 | \tilde{\boldsymbol{Z}}) = 1 - \exp[-\int_0^t \gamma \exp\{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1} + \tilde{\boldsymbol{Z}}\boldsymbol{s}\boldsymbol{\psi_2} + \rho s\} ds]$, and the cumulative incidence function for event $\epsilon = 2$ is $F_2(t; \tilde{\boldsymbol{Z}}) = P(T \leq t, \epsilon = 2 | \tilde{\boldsymbol{Z}}) = \exp\left\{\frac{\gamma \exp(\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1})}{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_2} + \rho}\right\} [1 - \exp\{-t \exp(\tilde{\boldsymbol{Z}}\boldsymbol{\psi_3})\}]$, where $\boldsymbol{\psi_3} = \{\ln(0.67), \ln(0.5), \ln(0.6), \ln(0.7), \ln(0.8), \ln(0.9)\}^T$ is the vector of effects for $\tilde{\boldsymbol{Z}}$ on the subdistribution hazard of event 2.

We assigned the event type by generating a Bernoulli variable with probability $P(\epsilon = 1 | \tilde{\boldsymbol{Z}}) = 1 - \exp\left\{\frac{\gamma \exp(\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1})}{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_2} + \rho}\right\}$ where $\tilde{\boldsymbol{Z}}\boldsymbol{\psi_2} + \rho < 0$, which is the limit of $F_1(t; \tilde{\boldsymbol{Z}})$ with respect to t. We generate the event times by simulating a random variable $U \sim \text{Uniform}(0, 1)$ and determining the inverse of the conditional cumulative incidence function with respect to t. We generate event times from the respective conditional cumulative incidence functions: for event $\epsilon = 1$, $P(T \leq t | \epsilon = 1, \tilde{\boldsymbol{Z}}) = \left(1 - \exp[-\int_0^t \gamma \exp\{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1} + \tilde{\boldsymbol{Z}}\boldsymbol{s}\boldsymbol{\psi_2} + \rho s\} ds]\right) / \left[1 - \exp\left\{\frac{\gamma \exp(\tilde{\boldsymbol{Z}}\boldsymbol{\psi_1})}{\tilde{\boldsymbol{Z}}\boldsymbol{\psi_2} + \rho}\right\}\right]$; and for event $\epsilon = 2$, $P(T \leq t | \epsilon = 2, \tilde{\boldsymbol{Z}}) = 1 - \exp\{-t \exp(\tilde{\boldsymbol{Z}}\boldsymbol{\psi_3})\}$.

We applied independent right-censoring by generating censoring times $C \sim \text{Uniform}(0, c)$, where the parameter c was chosen to yield mild and heavy right censoring (10% and 25% proportion of individuals). Then, we determined if the event time was censored, $\delta = I(T \leq C)$, and the final observed event type, $\delta \cdot \epsilon$.

4.5.1.2 Scenarios and true values

We considered both proportional, $(\gamma, \rho, \psi_{11}, \psi_{21}) = (1.2, -1.5, -0.3, 0)$, and non-proportional, $(\gamma, \rho, \psi_{11}, \psi_{21}) = (1.5, -1.5, 0.5, -2)$ subdistribution hazards settings. In total, there were 16 unique scenarios, with varying censoring (10%, 25%), proportion of exposure (25%, 50%), pattern of cumulative incidence functions (proportional vs. non-proportional subdistribution hazards), and sample size (500, 1,000). We generated 6,000 datasets for each scenario.

We determined the true marginal difference in RMTL at $\tau = 365$ days by generating $\boldsymbol{Z}$ for a sample of size $n = 1,000,000$. (Austin & Fine, 2019) We first set everyone to be exposed ($A = 1$), and generated the event type and corresponding event time as described in 4.5.1.1. Then, we repeated the process where everyone was unexposed ($A = 0$), yielding two event sets of outcomes per person (event type and time). The true marginal difference in RMTL for event $\epsilon = 1$ at $\tau = 1$ year is the unadjusted difference in RMTL for event $\epsilon = 1$ when $A = 1$ and $A = 0$; it was equal to -22.63 days for the proportional subdistribution hazards setting (small, protective effect) and 8.03 days for the non-proportional subdistribution hazards setting (very small, harmful effect). The true cumulative incidence functions are depicted in Figure 4.1.

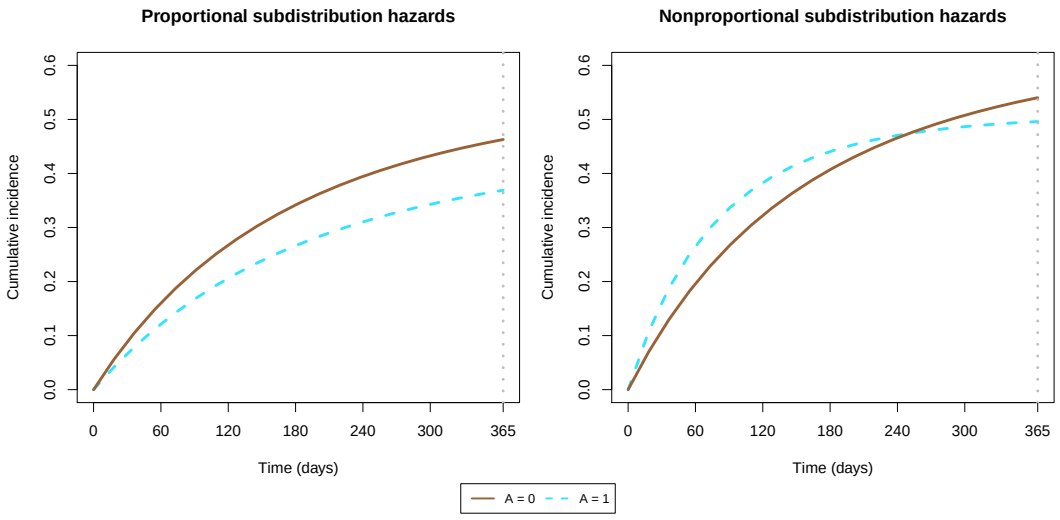


Figure 4.1: True cumulative incidence functions in simulation study of IPW estimator.

To apply the IPW method, we first derived the weights by estimating the probability of exposure with a logistic regression model, $\text{logit}\{P(A = 1|\boldsymbol{Z'})\} = \boldsymbol{Z'\alpha}$. We then estimated the IPW-adjusted RMTL in each exposure group, $A = 1$ and

$A = 0$, and the adjusted difference between groups.

4.5.1.3 Results

Overall, the IPW-estimator demonstrated little bias in all settings (Table 4.1). The bias was slightly greater in the presence of nonproportional subdistribution hazards in scenarios with smaller sample size ($n = 500$). The mean squared error was smaller in scenarios with larger sample size ($n = 1,000$) and equal exposure groups (50% exposed). The coverage proportion of 95% confidence intervals was greater than the nominal value, indicating that the confidence intervals tend to be conservative. The coverage proportion was sometimes slightly better in scenarios with balanced exposure groups (50% exposed) compared to settings with unbalanced exposure groups (25% exposed). The relative SE was greater than 1 in all scenarios, which indicates that the estimated SE was greater than the Monte Carlo empirical SE and likely explains the observed overcoverage.

4.5.2 Inverse probability of censoring weighted model for the restricted mean time lost

4.5.2.1 Data generation

To assess the IPCW model, we generated data according 2 binary covariates, $\boldsymbol{Z} = (Z_1, Z_2)$. We generated each binary covariate from independent Bernoulli distributions. We let the subdistribution hazard function follow a Gompertz distribution, $\lambda_1(t|\boldsymbol{Z}) = \gamma_z \exp(\rho_z t)$, where (γ_z, ρ_z) were set according to the 4 strata of (Z_1, Z_2).

We defined the cumulative incidence function for event 1 as $F_1(t|\boldsymbol{Z}) = P(T \leq t, \epsilon = 1|\boldsymbol{Z}) = 1 - \exp\{-\int_0^t \lambda_1(s|\boldsymbol{Z})ds\}$, and the cumulative incidence function

Table 4.1: Performance of IPW estimator of the difference in RMTL in simulation studies, $\tau = 365$ days

Proportional subdistribution hazards, δ = -22.63 days at $\tau = 365$ days							
Sample size	Exposed	Censoring	Bias, days	Rel bias	RMSE, days	Rel SE	Coverage
500	25%	10%	-0.026	0.001	14.269	1.262	0.984
		25%	0.231	-0.010	15.046	1.224	0.982
	50%	10%	0.118	-0.005	12.568	1.173	0.976
		25%	0.197	-0.009	13.179	1.146	0.974
1000	25%	10%	0.134	-0.006	10.153	1.248	0.984
		25%	0.424	-0.019	10.455	1.237	0.981
	50%	10%	-0.319	0.014	8.949	1.158	0.977
		25%	-0.182	0.008	9.267	1.146	0.974
Non-proportional subdistribution hazards, δ = 8.03 days at $\tau = 365$ days							
Sample size	Exposed	Censoring	Bias, days	Rel bias	RMSE, days	Rel SE	Coverage
500	25%	10%	0.721	0.090	16.702	1.215	0.980
		25%	0.759	0.095	17.433	1.196	0.978
	50%	10%	0.716	0.089	13.291	1.206	0.981
		25%	0.616	0.077	14.163	1.163	0.980
1000	25%	10%	0.344	0.043	11.679	1.219	0.982
		25%	0.264	0.033	12.070	1.213	0.980
	50%	10%	0.359	0.045	9.495	1.186	0.979
		25%	0.379	0.047	9.872	1.174	0.977

IPW: inverse probability weighted, RMTL: restricted mean time lost, Rel bias: mean bias relative to true parameter, RMSE: root of mean squared error, Rel SE: mean estimated standard error / Monte Carlo empirical error, Coverage: proportion of 95% confidence intervals which contain the true δ. To assess the IPW method, we generated competing risks data dependent on a binary covariate A and covariates $\boldsymbol{Z}$. We obtained δ, the true adjusted difference in RMTL between $A = 1$ and $A = 0$ at $\tau = 365$ days with a counterfactual approach in a sample of $n = 1000000$.

for event 2 as $F_2(t|\boldsymbol{Z}) = P(T \leq t, \epsilon = 2|\boldsymbol{Z}) = \exp(\gamma_z/\rho_z)\{1 - \exp(-t)\}$. We assigned the event type by generating a Bernoulli variable with probability $P(\epsilon = 1|\boldsymbol{Z}) = 1 - \exp(\gamma_z/\rho_z)$ where $\rho_z < 0$. We again generated times using the inversion method, with conditional cumulative incidence functions $P(T \leq t|\epsilon = 1, \boldsymbol{Z}) = [1 - \exp\{-\int_0^t \lambda_1(s|\boldsymbol{Z})ds\}] / [1 - \exp\{\gamma_z \exp(\gamma_z/\rho_z)\}]$ and $P(T \leq t|\epsilon = 2, \boldsymbol{Z}) = 1 - \exp(-t)$. We generated right censoring and determined the final observed event type as described in Section 4.5.1.1.

4.5.2.2 Scenarios and true values

Similar to 4.5.1.2, we included 16 unique scenarios with varying censoring (10%, 25%), proportion of exposure (both $P(Z_1 = 1)$ and $P(Z_2 = 1)$ equal to 25% or 50%), pattern of cumulative incidence functions (proportional vs. non-proportional subdistribution hazards), and sample size (500, 1,000). We generated 6,000 datasets for each scenario.

We define (β_1, β_2) as the true adjusted differences in RMTL at $\tau = 1$ year for covariates $\boldsymbol{Z} = (Z_1, Z_2)$. First we let $\nu_{1,z_1,z_2}(\tau) = \int_0^\tau F_1(t|\boldsymbol{Z})dt$. We set Gompertz parameters (ρ, γ) that yielded RMTLs for cause 1 of $\{\nu_{1,11}(\tau), \nu_{1,10}(\tau), \nu_{1,01}(\tau), \nu_{1,00}(\tau)\} = (0.3, 0.4, 0.45, 0.55)$ years. Then, the effect of Z_1 and Z_2 on the RMTL for cause 1 is additive such that $(\beta_1, \beta_2) = (-0.15, -0.1)$ years, or $(-54.75, -36.5)$ days. The Gompertz parameters are provided in Table C.1, and the true cumulative incidence functions are depicted in Figure 4.2.

We estimated IPCW regression models $g\{\nu_1(\tau)\} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ with an identity link. We then evaluated the performance of $(\hat{\beta}_1, \hat{\beta}_2)$ to estimate (β_1, β_2).

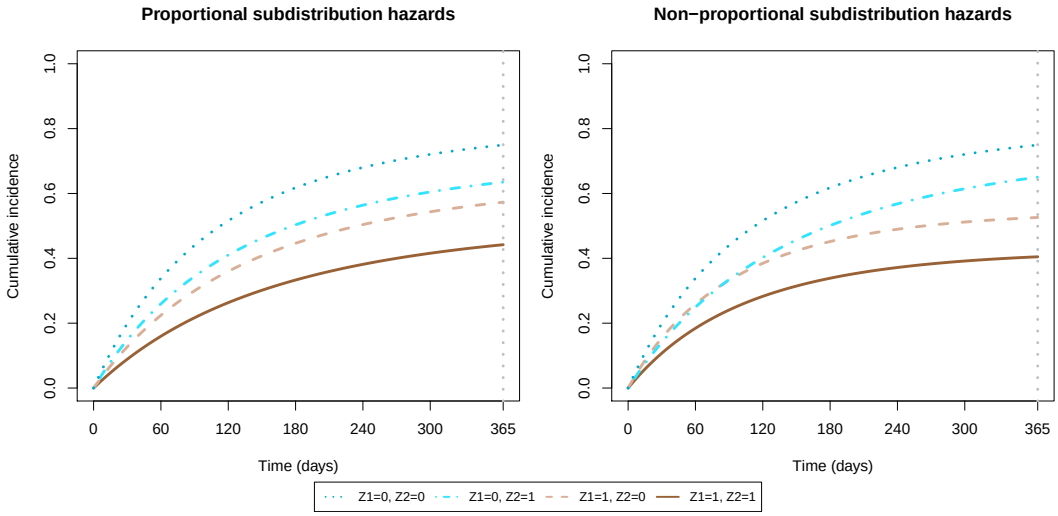


Figure 4.2: True cumulative incidence functions in simulation study of IPCW regression model.

4.5.2.3 Results

The IPCW model demonstrated very little bias in all settings (Table 4.2). The mean squared error decreased as sample size increased, censoring decreased, and when the proportion exposed was balanced compared to imbalanced. The coverage proportion of 95% confidence intervals ranged from 93% to 95%, which indicates slight undercoverage. The relative SE was less than 1 in all scenarios, which indicates that the estimated SE was smaller than the Monte Carlo empirical SE and likely explains the observed undercoverage.

Table 4.2: Performance of IPCW model of the difference in RMTL in simulation studies, $\tau = 365$ days

				β_1 = -54.75 days					β_2 = -36.5 days				
PSH	Sample size	Exposed	Censoring	Bias, days	Rel Bias	RMSE, days	Rel SE	Cov	Bias, days	Rel Bias	RMSE, days	Rel SE	Cov
Yes	500	50%	10%	0.076	-0.001	13.241	0.956	0.938	0.049	-0.001	13.251	0.956	0.938
			25%	0.135	-0.002	14.340	0.955	0.939	0.088	-0.002	14.282	0.955	0.942
		25%	10%	0.307	-0.006	15.096	0.977	0.942	0.116	-0.003	15.332	0.977	0.936
			25%	0.035	-0.001	16.661	0.957	0.935	0.481	-0.013	17.037	0.957	0.930
	1000	50%	10%	-0.039	0.001	9.251	0.968	0.943	-0.055	0.002	9.106	0.968	0.945
			25%	-0.149	0.003	10.132	0.957	0.937	0.097	-0.003	10.095	0.957	0.940
		25%	10%	-0.254	0.005	10.772	0.969	0.942	0.068	-0.002	10.852	0.969	0.933
			25%	0.006	0.000	11.811	0.954	0.937	-0.086	0.002	11.781	0.954	0.938
No	500	50%	10%	-0.204	0.004	13.411	0.963	0.942	-0.133	0.004	13.327	0.963	0.943
			25%	0.097	-0.002	14.458	0.967	0.943	-0.341	0.009	14.375	0.967	0.943
		25%	10%	0.077	-0.001	15.833	0.963	0.941	-0.276	0.008	15.311	0.963	0.941
			25%	-0.152	0.003	17.372	0.958	0.938	0.175	-0.005	17.405	0.958	0.938
	1000	50%	10%	0.189	-0.003	9.304	0.982	0.942	-0.087	0.002	9.386	0.982	0.944
			25%	-0.163	0.003	10.366	0.954	0.938	-0.134	0.004	10.363	0.954	0.937
		25%	10%	-0.056	0.001	11.058	0.974	0.944	0.016	0.000	10.818	0.974	0.941
			25%	0.034	-0.001	12.289	0.957	0.937	-0.190	0.005	12.066	0.957	0.937

IPCW: inverse probability of censoring weighted, RMTL: restricted mean time lost, PSH: proportional subdistribution hazards, RMSE: root of mean squared error, Rel SE: mean estimated standard error / Monte Carlo empirical error, Cov: coverage, or proportion of 95% confidence intervals which contain the true δ.. To assess the IPCW model, we generated competing risks data dependent on 2 binary covariates (Z_1, Z_2). We set the true unadjusted difference in RMTL at $\tau = 365$ days, $(\beta_1, \beta_2) = \{\nu_{1,1z_2}(\tau) - \nu_{1,0z_2}(\tau), (\nu_{1,z_11}(\tau) - \nu_{1,z_10}(\tau)\}$.

4.6 ILLUSTRATIVE EXAMPLES

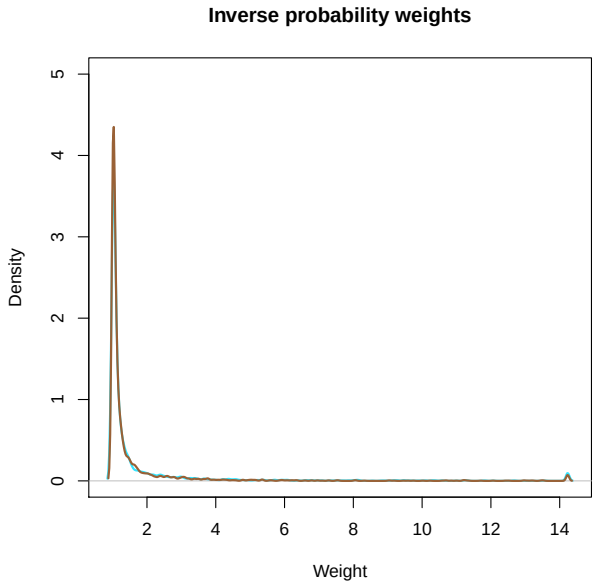
We present illustrative examples on AF in the Framingham Heart Study. (Staerk et al., 2018, 2020; Vinter et al., 2020) We first evaluated sex differences in incident AF. We also evaluated sex differences in cause-specific mortality after AF. In both examples, we apply the IPW and IPCW methods. We compared results with and without accounting for competing risks. The IPW method estimates the marginal RMTL in each sex and the difference in RMTL between sexes, adjusted for covariates, whereas the IPCW model estimates differences in RMTLs conditional on covariates. We also estimated subdistribution hazard ratios with Fine-Gray models. To protect the confidentiality of the Framingham Heart Study participants, the data from our illustrative examples are not publicly available. Participant level data from the Framingham Heart Study are available at the database of Genotypes and Phenotypes (https://www.ncbi.nlm.nih.gov/gap/) and BioLINCC (https://biolincc.nhlbi.nih.gov/home/).

4.6.1 Sex differences in incident atrial fibrillation

When assessing incident AF, individuals may die before developing AF; thus, death is considered a semi-competing risk. In the semi-competing risk setting and up to timepoint τ, the RMTL for AF is the mean AF-free time lost, or the mean time diagnosed with AF.

We included 6,016 participants who were free of AF at index age 55 years and attended an examination between ages 50 and 60 years. Risk factors were measured at the examination closest to age 55 years. Participants were followed until AF or atrial flutter, death, loss to follow-up, the last visit or medical contact. We set $\tau = 40$ years. At 40 years, there were 878 AF events and 1,284 deaths without

Figure 4.3: Examination of inverse probability weights in individuals at risk of atrial fibrillation in the Framingham Heart Study



AF. We adjusted for age, height, weight, current smoking status, systolic and diastolic blood pressure, hypertensive medication, current diabetes status, prior heart failure, and prior myocardial infarction.(Staerk et al., 2020) To address highly influential observations, leading to the instability of the IPW estimator, we replaced weights above the 99^{th} percentile of the distribution by the 99^{th} percentile.(Cole & Hernán, 2008)

First, we examine the distribution of the IP weights in Figure 4.3. We do not observe any extreme weights, and the distributions display reasonable overlap. In Figure 4.4, we show the IPW cumulative incidence curves and the complement of the IPW Kaplan-Meier curves (1-KM). The complement of the Kaplan-Meier curves overestimate cumulative incidence. Furthermore, the sum of cumulative incidence of AF and death without AF exceeds 1 at 30 years and later. In Table C.2, we show the adjusted RMTL in males and females, separately.

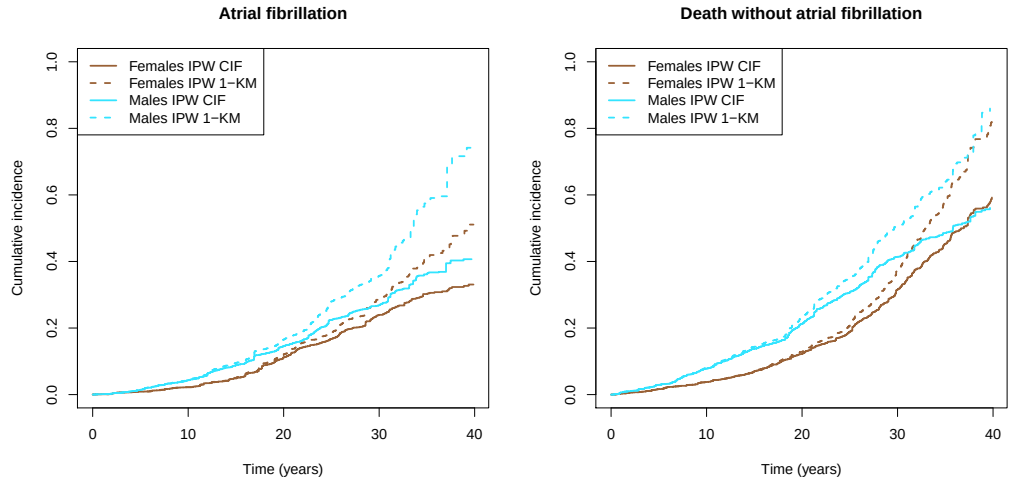


Figure 4.4: Inverse probability weighted cumulative incidence curves for incident atrial fibrillation in the Framingham Heart Study, with and without accounting for competing risks

IPW: inverse probability weighted, CIF: cumulative incidence function, KM: Kaplan-Meier.

With the IPW method, we found evidence of a marginal difference between sexes in RMTL for AF over 40 years (Table 4.3). Men lost an additional 1.4 (95% confidence interval [CI]: 0.5, 2.4) years of AF-free time as compared to women. With the IPCW model, there was no evidence of difference in the conditional difference in RMTL. On average females lost 0.4 (-0.92, 1.78) additional years of AF-free time as compared to males. When interpreting the results, it is important to note the two methods estimate different quantities. The IPW method estimates the population difference in RMTL between sexes marginalized over covariates, or the difference of the population being male or female.(Austin, 2013) In contrast, the IPCW regression model estimates the individual difference in RMTL between sexes conditional on covariates, or the difference in an individual being male or female.(Austin, 2013) The adjusted subdistribution hazard ratio for AF was 1.2 (1.0, 1.5), also indicating males have an increased incidence of AF compared to females. As for the competing event, death without AF, males lost more AF-free

life expectancy than females (IPW: 2.3 [1.2, 3.3], IPCW: 1.2 [0, 2.5]). The adjusted subdistribution hazard ratio from a Fine-Gray model for death without AF was 1.5 (1.3, 1.8).

Table 4.3: Adjusted differences in RMTL between males and females in 40-year incident atrial fibrillation

Estimators accounting for competing risks	**AF**	**Death without AF**
IPW	2.50 (1.61, 3.38)	1.44 (0.41, 2.46)
IPCW regression model	2.54 (1.65, 3.44)	1.79 (0.93, 2.64)
Estimators not accounting for competing risks	**AF**	**Death without AF**
IPW	3.89 (2.90, 4.88)	2.58 (1.67, 3.50)
IPCW regression model	2.78 (1.05, 4.52)	4.85 (3.74, 5.95)

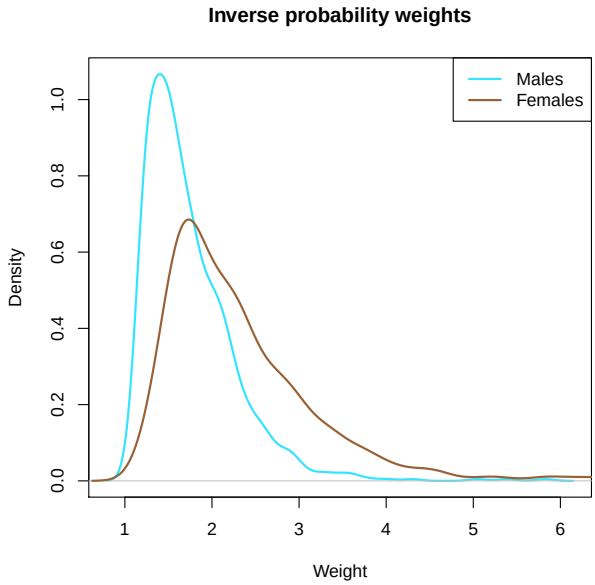
Data are differences in cause-specific RMTL (95% confidence intervals) for men vs. women. AF: atrial fibrillation; IPW: inverse probability weighted; IPCW: inverse probability of censoring weighted. The IPW method directly yields marginal differences in RMTL adjusting for covariates, while the IPCW regression models yield differences in RMTL conditional on covariates.

When ignoring competing risks, both the IPW and IPCW methods provided estimates larger in magnitude for both events (Table 4.3) as compared to results accounting for competing risks. In addition, both methods provided larger estimates of the RMTL in each sex, particularly the IPCW method (Appendix C.3, Table C.2).

4.6.2 Sex differences in cause-specific mortality after atrial fibrillation

Now, we illustrate our methods in a classic competing risk setting where all events are mortality, as opposed to disease incidence. We categorized death causes into cardiovascular disease (CVD), non-CVD, and unknown causes of death. We assessed sex differences in CVD and non-CVD mortality, while unknown cause of death was considered an additional competing risk. We included 1,483 individuals with incident AF who attended an exam within 10 years of diagnosis with non-missing covariates. We assessed sex differences in CVD and non-CVD mor-

Figure 4.5: Examination of inverse probability weights in individuals with atrial fibrillation in the Framingham Heart Study



tality, while unknown cause of death was considered an additional competing risk. Individuals entered the analysis at AF diagnosis, and were followed until death or loss to follow-up. We set $\tau = 10$ years. At 10 years of follow up, there were 839 total deaths: 329 due to CVD, 420 due to non-CVD causes, and 90 due to unknown causes. We estimated differences in cause-specific RMTL between men and women, adjusted for age at AF and risk factors measured at the closest examination prior to AF: body mass index, smoking status, systolic and diastolic blood pressure, hypertensive medication, and diabetes status. We also adjusted for prior heart failure and prior myocardial infarction, assessed at time of AF onset.

Again, we examine the distribution of the IP weights in Figure 4.5. We do not observe any extreme weights, and the distributions display reasonable overlap. In Figure 4.6, we show the IPW cumulative incidence curves and the complement of the IPW Kaplan-Meier curves (1-KM). In Table C.3 (Appendix D), we show the

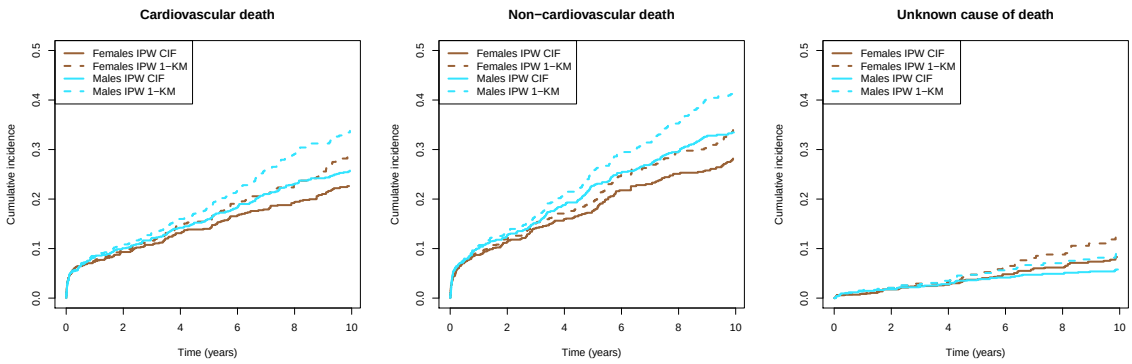


Figure 4.6: Inverse probability weighted cumulative incidence curves for cause-specific mortality among individuals with AF in the Framingham Heart Study, with and without accounting for competing risks

IPW: inverse probability weighted, CIF: cumulative incidence function, KM: Kaplan-Meier.

adjusted RMTLs for CVD death and CVD death in males and females, separately.

We did not find evidence of differences between sexes in 10-year cause-specific mortality (Table 4.4). When accounting for competing risks, both methods gave similar estimates for the difference in RMTL. The IPCW method provided slightly larger confidence intervals. On average, males had slightly greater lost life expectancy to CVD (IPW: 2.2 [-1.9, 6.4] months) and non-CVD causes (IPW: 4.0 [-0.5, 8.5] months) as compared to females over 10 years. The adjusted subdistribution hazard ratios from Fine-Gray models were consistent in direction with the differences in RMTLs (CVD-death: 1.2 [1.0, 1.5], non-CVD death: 1.3 [1.1, 1.5]).

When ignoring competing risks, both methods provided slightly larger estimates of the difference in RMTL for both CVD and non-CVD causes as compared to results accounting for competing risks. However, the change in estimates when accounting or not accounting for competing risks was much smaller than the previous example. Again, both methods also provided larger estimates of the RMTLs in each sex (Appendix C.3, Table C.3).

Table 4.4: Adjusted differences in RMTL between males and females in 10-year cause-specific mortality after atrial fibrillation

Estimators accounting for competing risks	**CVD death**	**Non-CVD death**
IPW	0.31 (-0.07, 0.68)	0.28 (-0.16, 0.71)
IPCW regression model	0.27 (-0.21, 0.76)	0.31 (-0.13, 0.76)
Estimators not accounting for competing risks	**CVD death**	**Non-CVD death**
IPW	0.40 (0.02, 0.79)	0.38 (-0.04, 0.80)
IPCW regression model	0.38 (0.03, 0.73)	0.48 (0.09, 0.87)

Data are differences in cause-specific RMTL (95% confidence intervals) for men vs. women. CVD: cardiovascular disease; IPW: inverse probability weighted; IPCW: inverse probability of censoring weighted. The IPW method directly yields marginal differences in RMTL adjusting for covariates, while the IPCW regression models yield differences in RMTL conditional on covariates.

4.7 DISCUSSION

In summary, we introduced methods to estimate and model the difference in RMTL in the presence of competing risks. We derived an IPW adjusted estimator of the difference in RMTL between exposure groups and its variance, and a generalized linear model for the RMTL conditional on covariates estimated with IPCW. In our illustrative examples, we demonstrate how ignoring competing risks leads to exaggerated estimates.

Previous works have demonstrated how ignoring competing events will lead to biased inference when estimating incidence.(Gaynor et al., 1993; Beiser et al., 2000) In this chapter, we demonstrate how ignoring competing events will lead to biased inference when reporting associations based on the RMST or RMTL. The complement of the Kaplan-Meier curve overestimates the cumulative incidence of an event and similarly, estimators of the RMTL must also account for the competing risk to avoid bias. In our illustrative examples, the estimated RMTLs were considerably larger when ignoring competing risks, which affected the between-

group differences in RMTLs. The IPCW method was particularly prone to giving larger estimates of the RMTL when the competing risk was ignored. This may be due to the two-fold bias of the IPCW model: if competing events are considered censored, the outcome is estimated only based on the event of interest rather than taking a value of 0 for competing events, and the censoring distribution $G(t)$ used to weight individuals is actually the distribution of a composite endpoint of censoring and the competing event.

In analyses of competing risks data, subdistribution and cause-specific hazard ratios are frequently used to measure associations.(Latouche et al., 2013) However, these measures can be difficult to interpret. First, they rely on proportionality over time, which often does not hold. Second, the subdistribution hazard for a given cause is a function of all cause-specific hazards, which allows the possibility that an exposure can indirectly influence the subdistribution hazard ratio through a competing event.(Latouche et al., 2013) Moreover, the cause-specific hazard ratio does not have a one-to-one relationship with the cumulative incidence function. This point also applies to RMTL-based measures of association, as a covariate can affect the cumulative incidence curve through the competing event. This can present difficulty in interpretation if the competing events are not also analyzed. To address this, we recommend reporting RMTL-based measures for both the event of interest and the competing event.

In a randomized experiment, Weir et al. (2019) demonstrated how clinicians may reach different conclusions when hazard ratios or differences in RMST are reported. Similarly, reporting the difference in RMTL as a complement to subdistribution and cause-specific hazard ratios may give additional insight to the data and offer a clinically meaningful interpretation. Unlike either hazard ratio, the un-

adjusted and IPW adjusted differences in RMTL directly give absolute measures in each group, which provides crucial background information.(Trinquart et al., 2017) Furthermore, the difference in RMTL is not subject to any proportionality assumptions. The assumption of proportional hazards is complex in the competing risk setting, as subdistribution hazards and cause-specific hazards cannot be simultaneously proportional and both must be explicitly assessed.(Beyersmann & Schumacher, 2007)

Other methods for the RMTL exist. In previous work, Calkins et al. (2018) estimated differences in RMST adjusted with IPW, and the bootstrap for confidence intervals. Andersen (2013) developed pseudo-observation model for cause-specific life years lost. Kipourou et al. (2020) also used the pseudo-observation technique to model life-years lost to various causes when cause of death is unreliable or missing. Sabathé et al. (2020) used the pseudo-observation approach to account for interval censoring in an illness-death model framework. Pan & Gastwirth (2013) modeled the restricted mean job duration, or 'lifetime', using Cox models for cause-specific event processes.

There are several limitations to our methods. First, if the IPW estimator weights are in fact estimated by estimating p_{ia}, the proposed variance of the RMTL does not account for variability in the estimated weights. In our simulation study, the coverage proportion of the IPW method was greater than 95%. Estimation of the weights can be accounted for using the bootstrap method.(Calkins et al., 2018) Second, when predicting the RMTL for a covariate profile using the IPCW model, one must be careful not to extrapolate outside the window of $(0, \tau)$, which may occur when using the identity or log link functions.(Andersen, 2013) In contrast, the logit-like link function, suggested by Tian et al. (2014), restricts the predicted

RMTL to $(0, \tau)$. Third, both of our proposed methods assume the event time and censoring are independent, which may not always hold. The IPCW framework has been extended to accommodate dependent censoring in other settings.(Wang & Schaubel, 2018) The IPCW model also exhibits a trade-off between estimation of the censoring distribution $G(t)$ and the number of uncensored observations. Ideally, one would have a large sample with sufficient censoring. One could improve estimation of the censoring distribution and overall model by including an augmentation term in the estimating equation.(Steingrimsson et al., 2016) Finally, we only considered covariates at baseline but it is possible that covariates are measured repeatedly over time. One could use g-computation and g-estimation to estimate the difference in RMTL while accounting for time-varying confounding.(Conner et al., 2019a; Hagiwara et al., 2019)

CHAPTER 5

Regression model for the lifetime risk using pseudo-observations

5.1 BACKGROUND

Lifetime risk is a measure of the absolute risk for developing a disease over the lifecourse. Estimation of the lifetime risk often requires accounting for left truncation (or delayed entry, because age is the time scale and individuals are observed at different starting ages), the competing risk of death, and inference at a fixed time point.(Gaynor et al., 1993; Beiser et al., 2000; Brookmeyer & Abdalla, 2019) Risk measures over longer time frames have been shown to motivate lifestyle behavior, particularly among younger individuals or those who may not be at short-term risk of disease.(Karmali & Lloyd-Jones, 2013) As a consequence, it is essential to predict the lifetime risk of conditions accurately. In this paper, we focus on modeling the residual lifetime risk, the risk among individuals who attained a given index age for developing a disease during their remaining lifespan. Because we use cohort data, we define the lifetime risk as the cumulative risk at an observable advanced age, e.g. 95 years. Moreover, participants are observed at periodic examinations and predictors are not collected exactly at the index age. Thus, we allow a window of time for participants to enter our analyses, but we restrict the window so risk factors are still comparable by age.(Staerk et al., 2018) Using age as the time scale with this entry time window induces left truncation.

Statistical methods are available to estimate the lifetime risk. Beiser et al. (2000) described a modified Kaplan-Meier estimator to allow delayed entries, based on estimators introduced by Gaynor et al. (1993) and Dinse & Larson (1986). Brookmeyer & Abdalla (2019) developed estimating equations for the lifetime risk using

a Markov multi-state model with age as the time scale. Carone et al. (2014) developed an estimator of the lifetime risk using cross-sectional survival data. To our knowledge, there are no methods for multivariable regression modeling of the lifetime risk. Adequate methods should accommodate covariates potentially associated with the cumulative incidence function, but not associated with the lifetime risk. For example, sex is associated with the cumulative incidence function for atrial fibrillation, as the cumulative incidence increases markedly after age 50 years in men but only after 60 years in women. However, the lifetime risk is similar between men and women. (Magnussen et al., 2017)

One can model the cumulative incidence function and predict the lifetime risk at a fixed age, τ. Fine & Gray (1999) extended the Cox proportional hazards model to the subdistribution hazard function of the cumulative incidence function, assuming proportional subdistribution hazards. Including time-covariate interactions allows modeling time-varying effects. Jeong & Fine (2006) proposed a cure model with Gompertz distribution. Scheike et al. (2008) introduced direct binomial regression allowing time-varying covariate effects, in the absence of left truncation. Royston-Parmar flexible parametric models allow a more complex baseline subdistribution hazard by using restricted cubic splines. (Lambert et al., 2017; Royston & Parmar, 2002)

Klein & Andersen (2005) and Graw et al. (2009) used pseudo-observations of the Aalen-Johansen estimator to model the cumulative incidence function over a grid of timepoints, although in the absence of left truncation. Moreno-Betancur & Latouche (2013) extended the approach to account for missing causes of failure. In this paper, we present a method for multivariable prediction of the lifetime risk using the pseudo-observation technique. We derive the pseudo-observations at a

fixed age, τ, which avoids assumptions about the functional form of risk factors over time. We also incorporate left truncation in the Aalen-Johansen estimator. (Grand et al., 2019) Throughout this chapter, we focus on the residual lifetime risk from an index age, τ_0, but refer to this quantity as the lifetime risk.

In Section 5.2, we discuss the non-parametric estimator of the lifetime risk and its connection to the Aalen-Johansen estimator. In Section 5.3, we propose the regression model for the lifetime risk using pseudo-observations. In Section 5.4, we describe how to use regression models for the cumulative incidence function to predict the lifetime risk at a fixed age, τ. In Section 5.5, we present an extensive simulation study to compare the methods' finite sample performance under a wide range of scenarios. In Section 5.6, we use the proposed method to predict the lifetime risk of atrial fibrillation with data from the Framingham Heart Study.

5.2 NON-PARAMETRIC ESTIMATION OF THE LIFETIME RISK

We consider two causes of failure, $\epsilon^* \in \{1, 2\}$ where $\epsilon^* = 1$ is the disease of interest (for example, atrial fibrillation) and $\epsilon^* = 2$ is the competing event (for example, death without atrial fibrillation). Age is the time scale. Let L^* be the entry age, T^* be the age at failure, C^* be the age at censoring, and $\delta^* = I(T \leq C)$ be an indicator of non-censoring. We assume C^* and L^* are independent of T^* and ϵ^*.

We consider an index age, τ_0, so that $L^* \geq \tau_0$. Moreover, individuals are observed conditional on $L^* < T^*$ and $L^* < C^*$. We denote $L, T, C, \delta, \epsilon$ the variables conditional on $L^* < T^*$ and $L^* < C^*$. We observe $X = T \wedge C$ the age at failure or censoring. We denote $\boldsymbol{Z}$ a $p \times 1$ vector of covariates. Then, we observe $(L_i, X_i, \delta_i \cdot \epsilon_i, \boldsymbol{Z_i})$ for individuals $i = 1, \ldots, n$.

The lifetime risk is defined as the cumulative risk for developing the disease

from index age τ_0 (e.g. 55 years) until an advanced age, τ (e.g. 95 years),

$$\begin{aligned} F_1(\tau|T^* > \tau_0) &= E\{1(T^* \leq \tau, \epsilon^* = 1|T^* > \tau_0)\} \\ &= P(T^* \leq \tau, \epsilon^* = 1|T^* > \tau_0) \\ &= \int_{\tau_0}^{\tau} S(t|t > \tau_0)\alpha_1(t)dt, \end{aligned}$$

in which $\alpha_1(t)$ is the cause-specific hazard function for event 1, $A(t)$ is the cumulative all-cause hazard function, and $S(t|t > \tau_0) = exp(-\int_{\tau_0}^{t} A(u)du)$ is the overall survival function (here, of being alive and event-free).(Jeong & Fine, 2009)

Let $N_{k,i}(t) = I(L_i < X_i \leq t)\delta_i I(\epsilon_i = k)$, $N_k(t) = \sum_{i=1}^{n} N_{k,i}(t)$ be the number of events of cause k by age t, $N(t) = N_1(t) + N_2(t)$ be the number of total events by age t, $\Delta N_k(t)$ be the increment $N_k(t) - N_k(t-)$, and $Y(t) = \sum_{i=1}^{n} I(L_i < t \leq X_i)$ be the number at risk just before age t. The lifetime risk is estimated by

$$\hat{F}_1(\tau|T > \tau_0) = \sum_{\tau_0 < t_j \leq \tau} \hat{S}(t_{j-1}) \frac{\Delta N_1(t_j)}{Y(t_j)}$$

where $\tau_0 < t_j \leq \tau$ are the unique ages at failure and $\hat{S}(t)$ is the Kaplan Meier estimator of being alive and event-free at age t with a modified risk set $Y(t)$ to allow delayed entries, $\hat{S}(t) = \prod_{\tau_0 < t_j \leq t} \left(1 - \frac{\Delta N(t_j)}{Y(t_j)}\right)$ (Beiser et al., 2000; Gaynor et al., 1993; Du, 2010; Andersen et al., 1993). The variance is estimated using Greenwood's formula. (Gaynor et al., 1993; Allignol et al., 2010)

$$\widehat{\text{var}}\left\{\hat{F}_1(\tau|T > \tau_0)\right\} = \sum_{\tau_0 < t_j \leq \tau} \left\{ \hat{S}(t_{j-1}) \frac{\Delta N_1(t_j)}{Y(t_j)} \right\}^2$$

$$\left\{ \frac{Y(t_j) - \Delta N_1(t_j)}{Y(t_j)\Delta N_1(t_j)} + \sum_{t_l < t_{t_j}} \frac{\Delta N(t_l)}{Y(t_l)(Y(t_l) - \Delta N(t_l))} \right\}$$
$$+ 2 \sum_{\tau_0 < t_j < \tau} \sum_{t_k \in (t_j, \tau]} \hat{S}(t_{j-1}) \frac{\Delta N_1(t_j)}{Y(t_j)} \hat{S}(t_{k-1}) \frac{\Delta N_1(t_k)}{Y(t_k)}$$
$$\left\{ -\frac{1}{Y(t_j)} + \sum_{t_l < t_j} \frac{\Delta N(t_l)}{Y(t_l)(Y(t_l) - \Delta N(t_l))} \right\}.$$

The estimator of the lifetime risk, $\hat{F}_1(\tau|T > \tau_0)$, is an element of the Aalen-Johansen estimator with three states (alive and disease-free, disease, and death) and allowing left truncation (i.e., age as the time scale). (Andersen et al., 1993; Aalen & Johansen, 1978) Connecting the lifetime risk to the Aalen-Johansen estimator allows us to explore statistical methodologies for competing risks data, such as pseudo-observations, which have not yet been used for modeling the lifetime risk. (Klein & Andersen, 2005; Graw et al., 2009) For instance, we can use an alternative representation of the cumulative incidence with inverse probability of censoring and left truncation weights. Note that the conditional distribution $dF_{L,T,C,\delta,\epsilon}(l,t,c,u,k)$ is equal to $dF_{L^*,T^*,C^*,\delta^*,\epsilon^*}(l,t,c,u,k)/\zeta$, where $\zeta = P(L^* \leq T|T > \tau_0)$, $G(t) = P(C^* > t|T^* > \tau_0)$, and $H(t) = P(L^* \leq t|T^* > \tau_0)$. Then using conditional expectations,

$$E\left\{ \frac{\zeta}{n} \sum_{i=1}^{n} \frac{N_{1,i}(\tau)}{G(T_i)H(T_i)} \right\} = E\left[E\left\{ \frac{I(L_i^* < T_i^* \leq \tau, \delta_i^* = 1, \epsilon_i^* = 1|T_i^* > \tau_0)}{G(T_i^*)H(T_i^*)} \middle| L_i^*, T_i^*, C_i^* \right\} \right]$$
$$= E\{I(T_i^* \leq \tau, \epsilon_i^* = 1|T_i^* > \tau_0)\}$$
$$= F_1(\tau|T^* > \tau_0).$$

This motivates an alternative, weighted estimator introduced by Geskus (Geskus,

2011),

$$\hat{F}_1(\tau|T > \tau_0) = \frac{1}{\hat{n}} \sum_{i=1}^{n} \frac{N_i(\tau)}{\hat{G}(X_i)\hat{H}(X_i)}$$

where $\hat{n}^{-1} = \{\sum_{i=1}^{n} \hat{H}(X_i)\}^{-1}$ estimates ζ/n, and $\hat{G}(t)$ and $\hat{H}(t)$ are Kaplan-Meier estimators for $G(t)$ and $H(t)$.

5.3 PSEUDO-OBSERVATION MODEL FOR THE LIFETIME RISK

We propose a regression model for the lifetime risk using the pseudo-observation technique. (Klein & Andersen, 2005; Andersen et al., 2003; Andersen & Pohar Perme, 2010) Let a generalized linear model for the lifetime risk be

$$g[E\{1(T_i^* \leq \tau, \epsilon_i^* = 1|T_i^* > \tau_0)|\boldsymbol{Z_i'}\}] = g\{F_1(\tau|T_i^* > \tau_0, \boldsymbol{Z_i'})\} = \boldsymbol{Z_i'}\boldsymbol{\beta}$$

where $g(\cdot)$ is a link function and $\boldsymbol{Z_i'} = (1, \boldsymbol{Z_i})$ is the covariate matrix allowing an intercept term. Here we let $g(\theta)$ denote a logit link function, $g(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$. Alternatively, the complementary log-log link function, $g(\theta) = \ln\{-\ln(1-\theta)\}$, or the identity link function, $g(\theta) = \theta$, can be used. We replace $1(T_i^* \leq \tau, \epsilon_i^* = 1|T_i^* > \tau_0)$ with pseudo-observations for the ith individual at age τ. In the presence of left truncation, the pseudo-observation for individual i at age t is given by

$$\hat{\theta}_i(t) = \left\{\sum_{i=1}^{n} 1(L_i \leq t)\right\} \hat{F}_1(t|T_i > \tau_0) - \left\{\sum_{i=1}^{n} 1(L_i \leq t) - 1\right\} \hat{F}_1{}^{-(i)}(t|T_i > \tau_0)$$

for individuals who have entered the study by age t ($L_i \leq t$), where $\hat{F}_1{}^{-(i)}(t|T_i > \tau_0)$ is the estimated cumulative incidence function at time t without the ith individual. (Grand et al., 2019) The lifetime risk is evaluated at a fixed age, τ, so we estimate the pseudo-observations at a single age τ. (Klein & Andersen, 2005; Klein

et al., 2007; Chen et al., 2018) Since all individuals have entered the study before τ, the pseudo-observation is

$$\hat{\theta}_i(\tau) = n\hat{F}_1(\tau|T_i > \tau_0) - (n-1)\hat{F}_1^{-(i)}(\tau|T_i > \tau_0),$$

where $\hat{F}_1^{-(i)}(\tau|T_i > \tau_0)$ is the estimated lifetime risk without the ith individual. We then model

$$g[E\{\hat{\theta}_i(\tau)|\boldsymbol{Z_i'}\}] = \boldsymbol{Z_i'}\boldsymbol{\beta}.$$

In the presence of independent right-censoring, it has been previously shown that the pseudo-observations are asymptotically unbiased for the cumulative incidence function.(Graw et al., 2009; Logan et al., 2011; Overgaard et al., 2017)

The estimated coefficients $\hat{\boldsymbol{\beta}}$ can be obtained by solving the estimating equation

$$U(\boldsymbol{\beta}) = \sum_i \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} F_1(\tau|T_i > \tau_0, \boldsymbol{Z_i'}) \right\}^T V_i^{(-1)} \{\hat{\theta}_i - F_1(\tau|T_i > \tau_0)\} = 0,$$

where V_i denotes the variance function. When using the logit link, the coefficients $\boldsymbol{\beta}$ are interpreted as log lifetime odds ratios. Note that if data are correlated, for example when pooling data from multiple cohorts, we can estimate the coefficients by using generalized estimating equations and we estimate the working correlation matrix V_i with a sandwich estimator. (Logan et al., 2011; Liang & Zeger, 1986)

We derive the estimated lifetime risk with the inverse link function. With the logit link, the estimated lifetime risk is

$$\hat{F}_1^{PO}(\tau|T_i > \tau_0, \boldsymbol{Z_i'}) = \frac{1}{1 + \exp(-\boldsymbol{Z_i'}\hat{\boldsymbol{\beta}})}.$$

The estimated difference in lifetime risk between individuals with covariate profiles i and j is $\hat{\mu} = \hat{F}_1^{PO}(\tau|T_i > \tau_0, \boldsymbol{Z_i'}) - \hat{F}_1^{PO}(\tau|T_i > \tau_0, \boldsymbol{Z_j'})$. We estimate the variance with the multivariate delta method, $\nabla(\hat{\mu})^T \Sigma \nabla(\hat{\mu})$, where $\nabla(\hat{\mu})$ is the gradient and Σ is the covariance matrix of the estimated model coefficients.

5.4 MULTIVARIABLE MODELS FOR THE CUMULATIVE INCIDENCE FUNCTION WITH LEFT TRUNCATION

We describe how to fit regression models for the cumulative incidence function, and how to use the fitted model coefficients to predict the lifetime risk. We first discuss the Fine-Gray model, which is the default model choice for competing risks data.(Fine & Gray, 1999) However, this model assumes proportional subdistribution hazards and does not include the baseline subdistribution hazard in the likelihood. We also describe including time-dependent effects in the Fine-Gray model. Finally, we describe the flexible parametric model, which directly models the baseline subdistribution hazard function using restricted cubic splines and also permits time-dependent effects.(Lambert et al., 2017; Royston & Parmar, 2002) The fully parametric form allows more complexity, straightforward predictions, and confidence intervals with the delta method.

5.4.1 Fine-Gray model of the subdistribution hazard

The Fine-Gray model is given by

$$\lambda_1(t|T_i^* > \tau_0, \boldsymbol{Z_i}) = \lambda_{1,0}(t|T_i^* > \tau_0) \exp(\boldsymbol{Z_i}\boldsymbol{\pi})$$

where $\lambda_{1,0}(t|T_i^* > \tau_0)$ is the baseline subdistribution hazard and $\lambda_1(t|T_i^* > \tau_0, \boldsymbol{Z}) = -\frac{d}{dt}\log\{1 - F_1(t|T_i^* > \tau_0, \boldsymbol{Z})\}$ is the subdistribution hazard.(Fine & Gray, 1999) We use the extension of the Fine-Gray model by Geskus to incorporate left truncation with a variation of inverse probability of censoring weighting (IPCW), in which individuals receive time-dependent weights and parameters $\boldsymbol{\pi}$ are estimated by solving a weighted likelihood.(Geskus, 2011) An individual i receives weight $w_i(t) = 1$ at all ages t before their age at failure or censoring, X_i. After age X_i, the individual receives weight $w_i(t) = \hat{G}(t)\hat{H}(t)/\hat{G}(X_i)\hat{H}(X_i)$ if they experienced a competing event, where $G(t)$ and $H(t)$ are the probabilities of not being censored and entering the study with delay at age t, each estimated by Kaplan-Meier estimators. Otherwise, $w_i(t) = 0$.

The lifetime risk can be obtained by predicting the cumulative incidence function at τ,

$$\begin{aligned}\hat{F}_1^{FG1}(\tau|T_i > \tau_0, \boldsymbol{Z_i}) &= 1 - \exp\left\{-\int_{\tau_0}^{\tau} \hat{\lambda}_{1,0}(t|T_i > \tau_0)\exp(\boldsymbol{Z_i}\hat{\boldsymbol{\pi}})dt\right\} \\ &= 1 - \exp\left\{-\hat{\Lambda}_{1,0}(\tau|T_i > \tau_0)\exp(\boldsymbol{Z_i}\hat{\boldsymbol{\pi}})\right\}\end{aligned}$$

where $\hat{\Lambda}_{1,0}(t|T_i > \tau_0)$ is a modified Breslow estimator of the baseline cumulative subdistribution hazard and $\hat{\Lambda}_{1,0}(\tau_0|T_i > \tau_0) = 0$. The difference in lifetime risk between individuals with covariate profiles i and j is $\hat{F}_1^{FG1}(\tau|T_i > \tau_0, \boldsymbol{Z_i}) - \hat{F}_1^{FG1}(\tau|T_i > \tau_0, \boldsymbol{Z_j})$. We construct 95% confidence intervals for the difference in lifetime risk between covariate profiles using the bootstrap method.

The Fine-Gray model assumes proportional subdistribution hazards, which may not hold in real-life applications. If the Fine-Gray model is misspecified, i.e. the proportional subdistribution hazards assumption is violated, subsequent pre-

diction of the cumulative incidence function will be affected. If a covariate has a proportional effect on the cause-specific hazards, it will not have a proportional effect on the sub-distribution hazards.(Beyersmann et al., 2011; Latouche et al., 2007; Beyersmann & Schumacher, 2007; Muñoz et al., 2013) Therefore, one cannot assume proportional cause-specific hazards and proportional subdistribution hazards simultaneously. Furthermore, hypothesis tests of the model parameters do not align with hypothesis tests of the lifetime risk. In Appendix D Section D.1, we illustrate this by assuming the cause-specific hazards have a Weibull form.

5.4.2 Fine-Gray model of the subdistribution hazard with time-dependent effects

We include time-dependent effects in the Fine-Gray model by adding time-covariate interactions

$$\lambda_1(t|T_i^* > \tau_0, \boldsymbol{Z_i}) = \lambda_{1,0}(t|T_i^* > \tau_0) \exp\{\boldsymbol{Z_i}\boldsymbol{\pi_1} + \boldsymbol{Z_i}\log(t)\boldsymbol{\pi_2}\}$$

where $\boldsymbol{\pi_1}$ denotes the time-fixed covariate effects and $\boldsymbol{\pi_2}$ denotes the effects for the interactions between the covariates and $\log(t)$. To predict the lifetime risk, we calculate the numerical value of the integral accounting for the time-covariate interaction term

$$\hat{F}_1^{FG2}(\tau|T_i > \tau_0, \boldsymbol{Z_i}) = 1 - \exp\left[-\int_{\tau_0}^{\tau} \hat{\lambda}_{1,0}(t|T_i > \tau_0) \exp\{\boldsymbol{Z_i}\boldsymbol{\pi_1} + \boldsymbol{Z_i}\log(t)\boldsymbol{\pi_2}\}\, dt\right].$$

Similar to Section 4.1, the difference in lifetime risk between individuals with covariate profiles i and j is $\hat{F}_1^{FG2}(\tau|T_i > \tau_0, \boldsymbol{Z_i}) - \hat{F}_1^{FG2}(\tau|T_i > \tau_0, \boldsymbol{Z_j})$ and 95% confidence intervals are obtained using the bootstrap.

5.4.3 Flexible parametric model of the cumulative subdistribution hazard

Finally, we describe how to model the cumulative incidence function by using the Royston-Parmar flexible parametric model estimated with IPCW to allow competing risks and left truncation. (Lambert et al., 2017; Royston & Parmar, 2002; Geskus, 2011) In contrast to the Fine-Gray model, the flexible parametric approach allows a more complex baseline subdistribution hazard via restricted cubic splines. We model the log cumulative subdistribution hazard as

$$\begin{aligned}
&\log[-\log\{1 - F_1(t|T_i^* > \tau_0, \boldsymbol{Z_i})\}] \\
&= \log\{\Lambda_1(t|T_i^* > \tau_0, \boldsymbol{Z_i})\} \\
&= \log\{\Lambda_{1,0}(t|T_i^* > \tau_0,)\} + \boldsymbol{Z_i\mu_1} + \boldsymbol{Z_i} \cdot s\{\log(t)|\boldsymbol{\gamma}, \boldsymbol{k_0}\} \cdot \boldsymbol{\mu_2} \\
&= s\{\log(t)|\boldsymbol{\gamma}, \boldsymbol{k_0}\} + \boldsymbol{Z_i\mu_1} + \boldsymbol{Z_i} \cdot s\{\log(t)|\boldsymbol{\gamma}, \boldsymbol{k_0}\} \cdot \boldsymbol{\mu_2},
\end{aligned}$$

where the log baseline subdistribution hazard, $\log\{\Lambda_{1,0}(t|T_i^* > \tau_0,)\}$, is modeled by $s\{\log(t)|\boldsymbol{\gamma}, \boldsymbol{k_0}\}$, a restricted cubic spline function with a vector of knots, $\boldsymbol{k_0} \in \{k_1, \ldots, k_K\}$, basis functions $\boldsymbol{\eta}$, and associated parameters $\boldsymbol{\gamma}$. More specifically, the restricted cubic spline function $s\{\log(t)|\boldsymbol{\gamma}, \boldsymbol{k_0}\} = \gamma_0 + \gamma_1\eta_1 + \cdots + \gamma_{K-1}\eta_{K-1}$ where $\eta_1 = log(t)$, $\eta_j = (log(t) - k_j)^3 - \nu_j (log(t) - k_1)^3 - (1 - \nu_j) (log(t) - k_K)^3$ for $j = 2, \ldots, K-1$, and $\nu_j = \frac{k_K - k_j}{k_K - k_1}$. Time-dependent effects, $\boldsymbol{\mu_2}$, are modeled by including interaction terms between covariates and the restricted cubic spline for $\log(t)$. Estimation of the model incorporates IPCW, as described in Section 4.1. (Lambert et al., 2017)

The lifetime risk is estimated by

$$\hat{F}_1^{FP}(\tau|T_i^* > \tau_0, \boldsymbol{Z_i})$$

$$= 1 - \exp(-\exp[s\{\log(\tau)|\hat{\gamma}, \boldsymbol{n_0}\} + \boldsymbol{Z_i}\hat{\boldsymbol{\mu}} + \boldsymbol{Z_i} \cdot s\{\log(\tau)|\gamma, \boldsymbol{k_0}\} \cdot \boldsymbol{\mu_2}]),$$

and the difference in lifetime risk between individuals with covariate profiles i and j is $\hat{F}_1^{FP}(\tau|T_i^* > \tau_0, \boldsymbol{Z_i}) - \hat{F}_1^{FP}(\tau|T_i^* > \tau_0, \boldsymbol{Z_j})$. We obtain standard errors with the multivariate delta method.

5.5 SIMULATION STUDIES

We conducted Monte Carlo simulation studies to assess and compare the performance of the pseudo-observation, Fine-Gray, and flexible parametric models in estimating the lifetime risk and its difference between groups. We generated data under both cause-specific hazards (Beyersmann et al., 2009, 2011) and subdistribution hazards frameworks (Zhou et al., 2013; Li et al., 2015) to assess if our findings depend on the data generation mechanism. Our data generation scenarios cover a variety of patterns, including non-proportional subdistribution hazards where the cumulative incidence functions cross or converge. We compared the bias, relative bias, mean squared error, and empirical coverage rate for all methods. We also assessed the Type I error and power of our method in detecting a difference in lifetime risk.

5.5.1 Data generation

5.5.1.1 Cause-specific hazard framework

We first generated one binary covariate, Z^*, from a Bernoulli distribution. To generate event times, we set the cause-specific hazard $\alpha_\epsilon(t; Z^*)$, $\epsilon \in \{1, 2\}$ to follow a Weibull form, $\alpha_\epsilon(t; Z^*) = \frac{a_{\epsilon,Z^*}}{(b_{\epsilon,Z^*})^{a_{\epsilon,Z^*}}} t^{a_{\epsilon,Z^*}-1}$ with shape parameter a_{ϵ,Z^*} and scale parameter b_{ϵ,Z^*}. We then determined the cumulative all-cause haz-

ard, $A(t;Z^*) = \int_0^t \alpha_1(u;Z^*) + \alpha_2(u;Z^*)du$ and its inverse function, $A^{-1}(t;Z^*)$. We obtain the generated times using the inversion method, where we generate a random variable $U \sim \text{Uniform}(0,1)$ and the simulated time to event is then $T^* = A^{-1}(-\ln(1-U))$. To determine the inverse function A^{-1}, we used the uniroot function in R. Then, we assigned the event type, ϵ^*, by a Bernoulli experiment, with probability $p = \frac{\alpha_1(T^*;Z^*)}{\alpha_1(T^*;Z^*)+\alpha_2(T^*;Z^*)}$ for cause $\epsilon = 1$.

To use age as the time scale, we generated entry times $L^* \sim \text{Uniform}(0,5)$ for 50% and 80% of individuals and 0 otherwise. We shifted times by adding 55 to L^* and T^*, to be consistent with our motivating example in which we examine residual lifetime risk from age 55. We included independent right-censoring by generating censoring times $C^* \sim \text{Uniform}(c,d)$, where the parameters (c,d) were chosen to yield approximately 25% and 75% probability of censoring at $\tau = 95$. Observed event ages were given by $X^* = T^* \wedge C^*$. Then, we determined if the event time was censored, $\delta^* = I(T^* \leq C^*)$, and the final observed event type, $\delta^* \cdot \epsilon^*$. Participants entered the study at entry age L^* if $L^* < X^*$; otherwise, they were not observed. This yields $(L, X, \delta \cdot \epsilon, Z)$.

5.5.1.2 Subdistribution hazard framework

We first generated one binary covariate, Z^*, from a Bernoulli distribution. Let $\lambda_{10}(t) = \gamma \exp(\rho t)$ denote the baseline subdistribution hazard function which follows a Gompertz distribution, and $\lambda_1(t;Z^*) = \lambda_{10}(t)\exp\{(\psi_1 + \psi_2 \cdot t)Z^*\}$ denote the subdistribution hazard function, where (ψ_1, ψ_2) are the covariate and time-dependent covariate effects on the subdistribution hazard of event 1. We assume the cumulative incidence function for event 1 with non-proportional subdistribution hazards is given by $F_1(t;Z^*) = P(T^* \leq t, \epsilon = 1|Z^*) =$

$1 - \exp[-\int_0^t \gamma \exp\{(\rho + \psi_2 Z^*)s + \psi_1 Z^*\} ds]$, and the cumulative incidence function for event 2 is $F_2(t; Z^*) = P(T^* \leq t, \epsilon = 2 | Z^*) = \exp\left\{\frac{\gamma \exp(\psi_1 Z)}{\rho + \psi_2 Z^*}\right\}[1 - \exp\{-t \exp(\psi_3 Z^*)\}]$, where ψ_3 is the covariate effect on the subdistribution hazard of event 2. (Zhou et al., 2013)

We assigned the event type by generating a Bernoulli variable with probability $P(\epsilon = 1 | Z^*) = 1 - \exp\left\{\frac{\gamma \exp(\psi_1 Z)}{\rho + \psi_2 Z}\right\}$ where $\rho + \psi_2 Z^* < 0$, which is the limit of $F_1(t | Z^*)$ with respect to t. We generate times for the event of interest from the conditional cumulative incidence functions for event 1, $P(T^* \leq t | \epsilon = 1, Z) = \left(1 - \exp[-\int_0^t \gamma \exp\{(\rho + \psi_2 Z^*)s + \psi_1 Z^*\} ds]\right) / [1 - \exp\{\gamma \exp(\psi_1 Z^*)/\rho + \psi_2 Z^*\}]$, and event 2 $P(T^* \leq t | \epsilon = 2, Z^*) = 1 - \exp\{-t \exp(\psi_3 Z^*)\}$. We generate the event times by simulating a random variable $U \sim \text{Uniform}(0, 1)$ and determining the inverse of the conditional cumulative incidence function with respect to t. We rescaled all times by multiplying by 40. (Zhou et al., 2013) We included independent right-censoring and delayed entry times as described in Section 4.1.1. We shifted times by adding $\tau_0 = 55$ to L^* and X^*, in order to be consistent with our motivating example in which we examine residual lifetime risk from age 55 years. Participants entered the study at entry age L^* if $L^* < X^*$; otherwise, they were not observed. This yields $(L, X, \delta \cdot \epsilon, Z)$.

5.5.2 Scenarios

The true cumulative incidence functions under cause-specific and subdistribution hazard data generation are available in Figure D.1. A description of each setting, including the parameters of the Weibull scenarios for cause-specific hazard settings 1-4 and the parameters for the Gompertz scenarios for subdistribution hazard settings 5-8, are provided in Tables D.1 and D.2.

Parameters were selected to obtain a variety of shapes in the cumulative incidence function (converging, diverging, and crossing curves), and a lifetime risk of approximately 40%, similar to the actual lifetime risk of atrial fibrillation in our motivating example (Staerk et al., 2018). The cumulative incidence curves in settings 1 and 5 converge, which shows an association with the cumulative incidence function but not the lifetime risk. The cumulative incidence curves in settings 2 and 6 cross. In the cause-specific hazard setup, we also varied if Z was associated with one or both events. In the subdistribution hazard setup, we varied how the association changes with time; whereas the associations in settings 5-7 include a main effect ψ_1 and time-varying term ψ_2, whereas the association in setting 8 is completely dependent on time ($\psi_1 = 0$).

We assessed 64 unique scenarios, with varying censoring (25%, 75%), left truncation (50%, 80%), and sample size (500, 1,000). We generated 2,000 datasets of size $n = 500$ and $n = 1,000$ for each scenario. The number of replicates was determined to obtain precision of 0.01 in detecting an empirical size of 0.05. Due to left truncation, the effective sample sizes are smaller; we report the average sample size for each scenario.

5.5.3 Statistical analysis

We estimated the lifetime risk at $\tau = 95$ from an index age of $\tau_0 = 55$ in each group and the difference in lifetime risk between groups using our pseudo-observation method, the Fine-Gray model with and without an interaction with log(time), and the flexible parametric model. For our pseudo-observation method and the Fine-Gray model, we used R 3.6.0. We used the `finegray()` function in R, and fit models with and without interactions between Z and log(time). For the flexible parametric

method, we used the `stpm2` procedure in Stata. We fit flexible parametric models with a time-varying effect by including an interaction between the covariate Z and log(t), $\log[-\log\{1 - F_1(t|T_i^* > \tau_0, Z_i)\}] = \gamma_0 + \gamma_1 \log(t) + \gamma_2\eta_2(t) + \gamma_3\eta_3(t) + \beta_1 Z_i + \beta_2 Z_i \log(t)$. All models included 4 knots for the baseline subdistribution hazard and 2 knots for the time-varying term. We chose the number of knots by comparing the Akaike's Information Criterion (AIC) when fitting models with 2 to 6 knots for the baseline subdistribution hazard in a random sample of simulated datasets. There was little change in AIC between 4, 5, and 6 knots, which justified our choice to use 4 knots.

5.5.4 Assessment of performance

For all methods, we compared estimation of the lifetime risk in each group and the difference in lifetime risk between groups according to mean bias, mean relative bias, and mean squared error. We compared the empirical coverage rate for the pseudo-observation and flexible parametric models only, since the Fine-Gray models require bootstrapping the confidence intervals. We also assessed the type I error and power of the estimated coefficient (log odds ratio) for the pseudo-observation method. We present results using nested loop plots.(Rücker & Schwarzer, 2014) The true lifetime risk at τ is given by $F_1(\tau|T^* > \tau_0, Z) = \int_{\tau_0}^{\tau} e^{-A(u;Z)} \cdot \alpha_1(u;Z)du = 1 - \exp[-\int_{\tau_0}^{t} \gamma \exp\{(\rho + \psi_2 Z)s + \psi_1 Z\}ds]$, where the parameters are given in Tables D.1 and D.2. Then, the true difference in lifetime risk, δ, between two groups is $\mu = F_1(\tau|T^* > \tau_0, Z = 1) - F_1(\tau|T^* > \tau_0, Z = 0)$.

5.5.5 Results

The nested loop plots in Figures 5.1 and 5.2 present the mean bias and mean squared error using the pseudo-observation, Fine-Gray, and flexible parametric methods. Each panel corresponds to a different scenario indicated by the legend. The scenarios were organized to best illustrate any patterns in the simulation results. Figures for the mean relative bias, coverage, Type I error, and power are available in Appendix D (Figures D.2-D.4). We assessed relative bias and power for scenarios where the true difference in lifetime risk was not 0, and type I error in scenarios where the true difference in lifetime risk was 0; all other criteria were assessed for all scenarios. The flexible parametric model did not reach convergence in all simulated datasets; the presented results include only datasets where the model converged. The convergence rate for this method ranged from 98% to 100%.

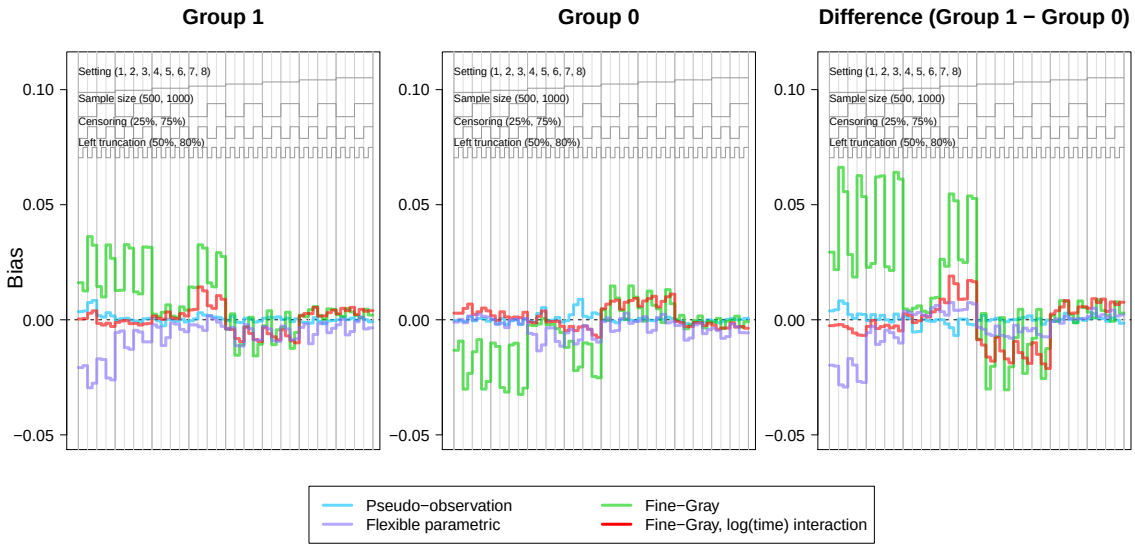


Figure 5.1: Nested loop plot showing the simulation study results: bias.

The legend at the top describes the organization of results by setting, sample size, censoring, and left truncation. Each vertical panel is a unique scenario, with a total of 64 scenarios.

The bias and relative bias were lower for the pseudo-observation approach than the Fine-Gray and flexible parametric approaches for most scenarios (Figures 5.1 and D.2). The Fine-Gray model without log(time) interaction demonstrated visibly greater bias for the difference in lifetime risk than the other methods across scenarios, particularly in settings with heavy censoring and where the cumulative incidence functions cross or converge (settings 1, 2, 4, 5 and 6). The Fine-Gray model with log(time) interaction showed a similar pattern, performing more poorly with heavy censoring but with smaller magnitude. Smaller sample size slightly increased the bias of the pseudo-observation method, but did not appear to heavily impact the bias of the other methods.

The mean squared error followed a consistent pattern across all methods, except the Fine-Gray without log(time) interaction which displayed notably greater mean squared error (Figure 5.2). Overall, mean squared error increased with smaller sample size, heavy right censoring, or heavy left truncation. The Fine-Gray model's fluctuation in mean squared error was greater in settings generated from cause-specific hazards (settings 1-4); but fairly equal among all methods in settings generated from subdistribution hazards (settings 5-8).

The coverage of the pseudo-observation approach hovered around 95% for all scenarios (Figure D.3). The coverage of the flexible parametric model was close to 95% for many scenarios, but slightly deviated in some scenarios, particularly setting 1 where there was no difference in lifetime risk. Finally, the Type I error of the pseudo-observation method hovered around 5% (Figure D.4). The power of the pseudo-observation method was high for most scenarios, but low when the difference in lifetime risk was small (Figure D.4).

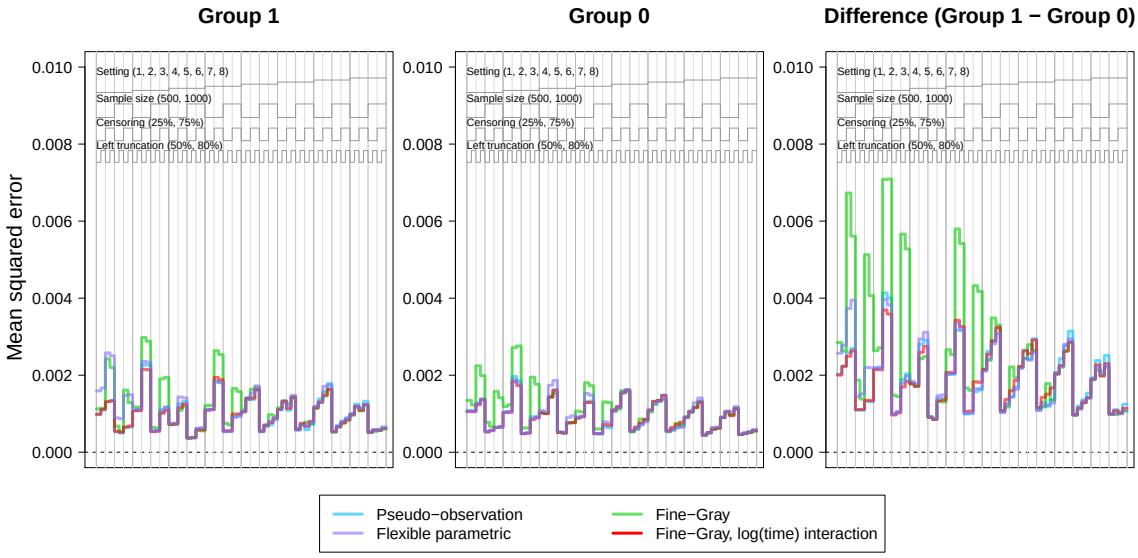


Figure 5.2: Nested loop plot showing the simulation study results: mean squared error.

The legend at the top describes the organization of results by setting, sample size, censoring, and left truncation. Each vertical panel is a unique scenario, with a total of 64 scenarios.

5.6 ILLUSTRATIVE EXAMPLE: LIFETIME RISK OF ATRIAL FIBRILLATION

(Staerk et al., 2018) previously estimated the lifetime risk of atrial fibrillation by calculating the Aalen-Johansen estimator in various risk factor strata. We apply the pseudo-observation, flexible parametric, and Fine-Gray methods to model the lifetime risk of atrial fibrillation in the Framingham Heart Study.

We included 5,613 Framingham Heart Study participants who were free of atrial fibrillation at an index age of 55 years and attended an examination between ages 50 and 60 years. Risk factors were measured at the examination closest to age 55 years. Participants were followed until atrial fibrillation or flutter, death, age 95 years, loss to follow-up, or last visit or medical contact in which the participant was free of atrial fibrillation. In all, 789 participants developed new-onset atrial fibrillation over a median follow-up of 14.0 years. We included the follow-

ing risk factors: age at entry, biological sex (male vs. female), body mass index, current smoking status, elevated vs. optimal alcohol use (elevated defined as 14 or more drinks per week for males and seven or more drinks per week for females), systolic and diastolic blood pressure, hypertensive medication use, prior diabetes, prior heart failure, and prior myocardial infarction. (Staerk et al., 2018) Participant characteristics are available in Appendix D (Table D.3).

5.6.1 Model fitting and lifetime risk prediction

We fit the pseudo-observation model with a logit link, the Fine-Gray model, and the flexible parametric model for the lifetime risk of atrial fibrillation. All models employed age as the time scale.

When fitting the Fine-Gray model, we considered whether time-varying effects improved the model fit according to AIC. We fit additional models, each with an interaction between log(time) and a single covariate. In a forward fashion, we selected the model with the lowest AIC, and repeated this process for the remaining covariates. (Sauerbrei et al., 2007) The final model contained interactions between log(time) and sex, smoking status, body mass index, prior heart failure, and prior myocardial infarction. For each binary covariate, we predicted the difference in lifetime risk between groups with a least squares mean approach, where all other covariates were set to the sample average or proportion. For continuous covariates, we predicted the difference in lifetime risk for one standard deviation increase from the mean.

When fitting the flexible parametric model, we considered models with 2 to 6 knots for the baseline subdistribution hazard and chose the value that yielded the lowest AIC, which was 5 knots. Then we fit additional models, each with an inter-

action between log(time) and a single covariate (which corresponds to 2 knots for covariate interactions). In a forward fashion, we selected the model with the lowest AIC, and repeated this process for the remaining covariates. (Sauerbrei et al., 2007) The final model contained interactions between log(time) and sex, smoking status, and hypertensive medication.

We also fit models for the lifetime risk of death without atrial fibrillation to better understand the competing risk. Fine-Gray models including interactions between log(time) and entry age, sex, smoking status, hypertensive medication, prior heart failure, and prior myocardial infarction. Flexible parametric models for death without atrial fibrillation included 4 knots for the baseline subdistribution hazard, determined by comparing the AIC for models with varying numbers of knots. The final model contained interactions between log(time) and sex, systolic blood pressure, hypertensive medication, and prior diabetes.

For all models, we predicted the lifetime risks by appropriately back-transforming the estimated model coefficients using a least squares means approach. For each binary covariate, we predicted the difference in lifetime risk between groups, where all other covariates were set to the sample average or proportion. For continuous covariates, we predicted the difference in lifetime risk for one standard deviation increase from the mean. We obtained 95% confidence intervals using the delta method for the pseudo-observation model with logit link and flexible parametric models, and the 0.025 and 0.975 percentiles of bootstrap resamples for the Fine-Gray model.

We generated calibration plots for each method, available in Figure D.5 of Appendix D. The pseudo-observation method demonstrates the highest calibration, followed by the flexible parametric model and then the Fine-Gray model.

The pseudo-observation underestimated the lifetime risk in the seventh and tenth deciles of predicted risk. The Fine-Gray model over-estimated the lifetime risk in the second, fourth, fifth, and tenth deciles, and under-estimated in the seventh and ninth deciles. Finally, the flexible parametric model over-estimated the lifetime risk in the third, fifth, sixth, and tenth deciles. The mean squared error between predicted and observed lifetime risk was 0.0007 for the pseudo-observation method, 0.0017 for the Fine-Gray model, and 0.0016 for the flexible parametric model.

5.6.2 Results

In Table 5.1, we present the predicted differences in lifetime risk of atrial fibrillation at age 95 from an index age of 55. With the pseudo-observation model, there was evidence of increase in lifetime risk of atrial fibrillation associated with male sex (difference in lifetime risk [95% CI]: 10.6% [5.2, 16.0], higher systolic blood pressure (5.7% [1.3, 10.1], lower diastolic blood pressure (-4.3% [-7.9, -0.7]), and use of hypertension medication (7.2% [1.4, 12.9]). There also was evidence of decrease in lifetime risk of atrial fibrillation associated with smoking (-10.2% [-17.3, -3.2]) and prior diabetes (-11.7% [-18.8, -4.6]). The Fine-Gray and flexible parametric model found similar results for sex, systolic blood pressure, diastolic blood pressure, use of hypertension medication (Table 5.1). The Fine-Gray and flexible parametric models also predicted decreases in lifetime risk with smoking and prior diabetes, though smaller in magnitude. Results for prior myocardial infarction and heart failure varied by method, which may be explained by the small number of participants with these risk factors (Table D.3). The model coefficients, which are lifetime odds ratios for the pseudo-observation model and subdistribution hazards ratios for the Fine-Gray and flexible parametric models, are available in the

Appendix (Table D.5).

Table 5.1: Predicted percentage differences in lifetime risk of atrial fibrillation at age 95, from index age 55

		Cumulative incidence models	
Risk factor	**Pseudo-observation model**	**Fine-Gray**	**Flexible parametric**
Age at entry, (SDU)	-2.38 (-5.30, 0.55)	-1.61 (-2.89, 0.04)	-1.61 (-3.65, 0.43)
Male vs. female	10.61 (5.19, 16.03)	10.23 (0.13, 13.48)	8.52 (2.79, 14.25)
Systolic blood pressure, (SDU)	5.69 (1.32, 10.05)	5.44 (0.02, 7.17)	5.88 (2.58, 9.17)
Diastolic blood pressure, (SDU)	-4.28 (-7.88, -0.68)	-2.91 (-4.26, -0.00)	-3.26 (-6.22, -0.30)
Hypertension medication, yes vs. no	7.15 (1.39, 12.91)	9.56 (0.03, 12.74)	4.32 (-2.42, 11.06)
Smoker vs. non-smoker	-10.24 (-17.27, -3.22)	-5.38 (-8.21, 3.48)	-5.84 (-11.20, -0.48)
Alcohol use, elevated vs. optimal)	4.36 (-6.76, 15.48)	7.71 (-0.10, 13.14)	8.68 (-0.22, 17.58)
Body mass index, (SDU)	2.09 (-0.17, 4.36)	1.94 (0.02, 4.04)	5.68 (3.13, 8.23)
Prior diabetes, yes vs. no	-11.69 (-18.75, -4.63)	-3.92 (-8.44, 0.81)	-4.11 (-11.58, 3.36)
Prior heart failure, yes vs. no	3.75 (-29.00, 36.51)	-10.92 (-24.59, 24.59)	15.49 (-16.74, 47.72)
Prior myocardial infarction, yes vs. no	-9.80 (-21.74, 2.13)	-12.65 (-19.36, 6.60)	3.06 (-8.68, 14.80)

SDU: standard deviation unit. Results are % differences in lifetime risk and 95% confidence intervals. Differences in lifetime risk for continuous risk factors are the differences for a 1 SDU increase in the covariate from the mean value of the covariate. For the pseudo-observation model and flexible parametric models, differences in lifetime risk were obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are obtained with the delta method. For the Fine-Gray model, differences in lifetime risk were also obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are the .025 and .975 percentile of bootstrap resamples.

In Table D.4, we present predicted differences in lifetime risk of death without atrial fibrillation at age 95 from an index age of 55. With all methods, we observe positive associations for smoking status (pseudo-observation 19.7 [11.3, 28.1], Fine-Gray 21.3 [15.6, 26.3], flexible parametric 25.9 [21.7, 30.0]) and prior diabetes (pseudo-observation 9.0 [-1.9, 19.9], Fine-Gray 21.6 [14.6, 28.9], flexible parametric 23.3 [15.9, 30.8]). The model coefficients are available in the Appendix (Table D.6).

The inverse associations of smoking and the lifetime risk of atrial fibrillation may seem counterintuitive, as smoking and diabetes are established risk factors for atrial fibrillation. However, smoking is strongly associated with the competing risk, death without atrial fibrillation. If individuals who smoke have a greater hazard of death, they may die before they can develop atrial fibrillation. It would be misleading to report the association between smoking status and atrial fibrillation in isolation, as the association with the competing risk is also important here. However, the inverse association is still an important result from a public health perspective. In quantifying the burden of atrial fibrillation at advanced ages, individuals who smoke or with diabetes may not develop atrial fibrillation because they did not live long enough to develop atrial fibrillation. The individuals we observe developing atrial fibrillation later in life may not exhibit these risk factors from the very fact that they have survived to a later age due to the absence of these risk factors.

5.7 DISCUSSION

In summary, we demonstrate how the pseudo-observation technique can be used to fit a multivariable model for the lifetime risk. Unlike models for the cumulative

incidence function, our approach avoids the assumption of proportional subdistribution hazards and complex time-varying effects by focusing on a single point in time. Models for the cumulative incidence function are designed to estimate the subdistribution hazard ratio over all time points, whereas we leverage the pseudo-observation technique to focus on our timepoint of interest. In simulation studies, our approach had lower bias and similar mean squared error to the Fine-Gray and flexible parametric models. In the Framingham Heart Study, we identified mid-life risk factors associated with the residual lifetime risk of atrial fibrillation. We focus on the pseudo-observation method with logit link in this paper. Alternatively, one could directly model differences in lifetime risk with the pseudo-observation method by using the identity link with caution of estimating lifetime risk outside of the $(0, 1)$ range.

The flexible parametric model and Fine-Gray model with log(time) interaction also performed well in predicting the lifetime risk in many scenarios. However, we observed considerable bias and error when the Fine-Gray model did not include an interaction term between the covariate and log(time), thus the model did not try to account for a time-varying association. This bias was particularly large in settings where the cumulative incidence functions crossed or in the presence of heavy censoring. Both the flexible parametric and Fine-Gray models sometimes showed improved performance in the presence of heavier left truncation as compared to lighter left truncation; this may be explained by the weights, which require estimation of the entry distribution, $H(t)$. If the amount of left truncation is small, $\hat{H}(t)$ may not be estimated correctly which in turn affects the weights.

The regression coefficients of the Fine-Gray and flexible parametric models are log subdistribution hazard ratios for the cumulative incidence function. In our il-

lustrative example with age as the time scale, covariates with log(time) interactions often showed large or small coefficients due to the range of time (55 to 95 years). This could be alleviated by rescaling the age variables or the log(time) interaction, i.e. substracting τ_0. With the flexible parametric model, the lifetime risk at τ can be predicted and confidence intervals or hypothesis tests can be easily obtained with the delta method. In contrast, confidence intervals for the Fine-Gray predicted lifetime risk require the bootstrap method. When including time-dependent effects in the Fine-Gray model, prediction of the lifetime risk is not always straightforward.(Austin et al., 2020) On GitHub, we share our R code for predicting the cumulative incidence function at a fixed time point from the Fine-Gray model with left truncation and time-varying effects using a long dataset approach. The long dataset divides follow-up time into smaller increments and can be considerably large, which makes bootstrapping the confidence intervals computationally intensive.

Interpretations of associations in the competing risk setting require careful consideration. As explored in our simulation study, is possible for a variable to appear to have an association with an event's cumulative incidence function, when in fact it is associated with the competing event's cumulative incidence function. Analyzing the cumulative incidence functions of both events, rather than only focus on one event, enables capturing the entire picture. (Beyersmann et al., 2011) In our illustrative example, we found smoking status and prior diabetes had inverse associations with the lifetime risk of atrial fibrillation using the pseudo-observation approach. The Fine-Gray and flexible parametric model included a significant interaction between smoking status and log(time), thus illustrating how the association with the cumulative incidence function is changing with time. In these

models, we see smoking status is highly, positively associated with death. As individuals die, they are no longer at risk of developing atrial fibrillation, hence the inverse associations in the atrial fibrillation models.

Our simulation study compared three methods, but was not exhaustive of all methods for modeling the cumulative incidence function. In Section 1, we described several alternative models for modeling the cumulative incidence function, which could be fit to obtain predictions of the lifetime risk at τ. Furthermore, the cumulative incidence function can also be estimated with other approaches to the flexible parametric model. For example, one can fit a flexible parametric model for each log cumulative cause-specific hazard or for the cause-specific cumulative incidence function using a direct likelihood approach. (Hinchliffe & Lambert, 2013; Mozumder et al., 2018)

Grand et al. (2019) studied the performance of pseudo-observations for the survival function and restricted mean survival time in the presence of left truncation, but not the cumulative incidence function. Grand et al. (2019) considered two methods of calculating the pseudo-observations, using either the total sample size or a 'stopped' approach with the number of entries at each timepoint. Our results confirm their finding that using the number of entries, rather than the total sample size, performs better. When predicting the lifetime risk, we avoid issues that may usually arise with left truncation, such as small numbers at risk at early event times. However, such issues may arise in other applications.

Jacobsen & Martinussen (2016) and Overgaard et al. (2017) recently examined the asymptotic properties of pseudo-observations and found that the sandwich estimator of the variance is biased, although the bias is small in many applications. Overgaard et al. (2018) proposed an alternative variance estimator, which

was found to perform well in large samples but poorly in smaller samples. Here, we only consider the sandwich estimator variance, so our results may be conservative. However, the pseudo-observation model demonstrated adequate coverage in our simulation study.

There are several limitations to our approach. First, it may take a long time to compute the pseudo-observations in larger datasets. However, once the pseudo-observations are computed, any number of models can be fit. Second, while jackknifing one timepoint is advantageous for the lifetime risk, we are not modeling earlier timepoints. Third, here we consider risk factors in mid-life, but have not extended the models to account for risk factors which may change over time. Finally, we assume that left truncation and right censoring are independent of the event time and covariates. Dependent truncation or censoring could be accommodated by modeling pseudo-observations of a modified Aalen-Johansen estimator or adjusting for the entry time as a covariate.(Lok et al., 2018; Stegherr et al., 2020; Pencina et al., 2007)

In conclusion, our approach using pseudo-observations produces unbiased predictions of the lifetime risk and allows hypothesis testing directly on the lifetime risk, rather than the cumulative incidence function. Furthermore, by focusing on one age τ, we avoid specifying complex patterns over time, such as the baseline subdistribution hazard function or time-varying associations.

CHAPTER 6
Discussion

6.1 SUMMARY

In summary, we provide new methods to estimate the RMST, RMTL, and lifetime risk. We address statistical challenges for the RMST common in observational studies or non-randomized clinical trials, including an IPW estimator to adjust for covariates, the parametric g-formula with multiple imputation and linear interpolation for time-varying confounding, and an IPW estimator and IPCW regression model to account for competing risks. Our IPW estimator for the RMST is unbiased, and well as our IPW estimator and IPCW model for the RMTL with competing risks. We also demonstrated how to estimate the RMST while accounting for time-varying confounding using the parametric g-formula, and address missing data challenges which arise with multiple risk factor measurements. We also developed a regression model for the lifetime risk that is unbiased and provides direct inference for the lifetime risk. Our method outperforms models for the cumulative incidence function to predict lifetime risk, such as the Fine-Gray and flexible parametric model, in settings with crossing or converging cumulative incidence functions.

We apply our methods to AF and coronary heart disease in observational data from the Framingham Heart Study. However, our methods are broadly applicable to other diseases or applications outside of cardiovascular research, and may be used for covariate adjustment to gain precision in randomized controlled trials. Moreover, our illustrative examples focused on exposure-disease relationships, but the proposed methods are applicable to non-randomized studies of treatment ef-

fects. Our methods also improve overall interpretation of absolute effect measures, which improves our understanding of AF epidemiology.

6.2 COMPARISON OF OUR METHODS WITH EXISTING METHODS

As discussed in Chapter 2, there are several existing regression models which can be used to estimate the adjusted RMST conditional on covariates. These models include generalized linear models for the RMST estimated with pseudo-observations or IPCW, as well as generalized linear models for the log cumulative hazard function with restricted cubic splines for time.(Andersen & Pohar Perme, 2010; Tian et al., 2014; Royston & Parmar, 2013) As demonstrated in our simulation study, the pseudo-observation and IPCW approaches are identical in the absence of right-censoring. Direct models for the RMST allow parameters with direct interpretations for the RMST, such as the difference or ratio in RMST. These methods assume the event and censoring times are independent, but have been extended to address dependent censoring.(Schaubel & Wei, 2011) For instance, one can estimate the IPC weights with a Cox model conditional on covariates or derive pseudo-observations of a modified Kaplan-Meier estimator. Similarly, as discussed in Chapter 4, one can model the adjusted RMTL conditional on covariates in the presence of competing risks. Andersen (2013) introduced a generalized linear model for the cause-specific RMTL estimated with pseudo-observations. In Chapter 4, we introduced the IPCW model for the cause-specific RMTL, which reduces to Tian et al. (2014)'s model in the absence of competing risks.

In contrast, one can model the hazard and transform the estimated hazard to predict the RMST for a covariate profile.(Chen & Tsiatis, 2001; Karrison, 1987) Flexible parametric models are an appealing alternative to the Cox model framework

due to their use of restricted cubic splines, which can readily incorporate more complex baseline hazards or non-proportional covariate effects. While parameters of various hazard models do not have a direct interpretation on the RMST, the flexible parametric model allows inference on the RMST via the delta method. In theory, one could estimate the RMTL in the presence of competing risks by modeling all cause-specific hazards or the cause-specific subdistribution hazard. To our knowledge this has yet to be explored using Cox and Fine-Gray models, but Mozumder et al. (2020) describes estimating the RMTL using flexible parametric models in work under review.

Unlike the regression methods, our IPW approach in Chapter 2 estimates the marginal RMST and is congruent with the original concept non-parametric RMST estimator as the area under the Kaplan-Meier curve. Our IPW approach in Chapter 4 estimates the marginal, cause-specific RMTL and mirrors this concept by taking the area under an adjusted Aalen-Johansen cumulative incidence curve. For both the RMST and RMTL, the IPW approach is better suited for binary or categorical exposures of interest, while the regression models can accommodate continuous exposures or examining several variables simultaneously.

If the marginal RMST or RMTL difference is solely of interest and one has an exposure of interest in mind, we recommend our IPW approach. Our IPW approach is unbiased for the marginal difference in RMST and RMTL, and corresponds with IPW-adjusted Kaplan-Meier or cumulative incidence curves.(Xie & Liu, 2005; Choi et al., 2019) An alternative approach to estimate the marginal RMST difference would be to fit a hazards model, obtain subject-specific conditional survival predictions at each timepoint when exposed and unexposed, average the predictions to obtain marginal survival curves, and take the area under these marginal sur-

vival curves.(Hernán, 2010) With large sample sizes and numerous event times, this approach may be computationally intensive. In contrast, our IPW approach involves fewer steps to estimate the marginal RMST or RMTL difference. We also provide a formula for the RMST variance.

If the conditional RMST or RMTL difference is solely of interest, we recommend either the pseudo-observation or IPCW models due to their advantage of estimating model parameters directly interpretable for the RMST without transformation. The two methods perform similarly, but in small sample sizes or in the presence of little right-censoring, the IPCW model may give unstable estimates due to instability of the estimated IPC weights.

If one is interested in obtaining various measures of association with a unified statistical approach (such as the HR, RMST difference, and cumulative risk difference), we recommend the flexible parametric cumulative hazard model.(Royston & Parmar, 2013) The flexible parametric model allows estimating multiple measures of associations by including a parameter for the baseline hazard. Inference can be performed with the delta method. Some model selection is advised to ensure the splines fit the data well, which may become complex when considering covariate-time interactions.(Rutherford et al., 2015) As of now, predicted RMST using flexible parametric models has only been explored in the absence of competing risks but we look forward to new developments which address competing risks.(Mozumder et al., 2020)

Chen & Tsiatis (2001) demonstrated how to estimate the difference in RMST by fitting stratified Cox models, thus avoiding the proportional hazards assumption for the exposure of interest. This approach may be appealing to researchers familiar with the Cox model, and could also be used to calculate hazard ratios

or cumulative risk differences. However, this approach does not easily extend to continuous exposures.

In Chapter 3, we discussed estimation of the marginal RMST difference with time-varying confounders using the parametric g-formula. Software exists to report RMST using the parametric g-formula, but to our knowledge, the estimation of RMST had not been described in detail.(Logan et al., 2018) The parametric g-formula assumes equally spaced intervals, which may be unrealistic for some data types. While this assumption may have minimal consequence when estimating risk, the RMST is on the time scale and requires logical time increments. We described linear interpolation of covariates at yearly increments in order to apply the parametric g-formula to FHS data with unequally spaced examinations.

Other methods exist to account for time-varying confounding. Recently, Hagiwara et al. (2019) introduced g-estimation of a structural nested RMTL model. Robins et al. (2000) and Naimi et al. (2017) introduced marginal structural models for the survival function, estimated by Cox models with time-varying IP weights; in theory, one could take the area under the marginal survival curves to estimate the RMST difference but this has not been done to our knowledge. It would be interesting to compare these approaches in estimating the RMST or RMTL while accounting for time-varying confounding.

In Chapter 5, we introduced a direct regression model for the lifetime risk using pseudo-observations. Beiser et al. (2000) previously described a non-parametric estimator of the lifetime risk which allows delayed entries, based on estimators introduced by Gaynor et al. (1993) and Dinse & Larson (1986). However, to our knowledge there were no methods for multivariable regression modeling of the lifetime risk. Instead, one could model the cumulative incidence function and pre-

dict the lifetime risk at a fixed age, τ.(Fine & Gray, 1999; Lambert et al., 2017) Klein & Andersen (2005) and Graw et al. (2009) previously used pseudo-observations of the Aalen-Johansen estimator to model the cumulative incidence function over a grid of timepoints, instead of the residual lifetime risk at age τ. Our approach using pseudo-observations produces unbiased predictions of the lifetime risk and allows hypothesis testing directly on the lifetime risk, rather than the cumulative incidence function. By directly modeling the lifetime risk at the age of interest, τ, we avoid specifying complex patterns over time, such as the baseline subdistribution hazard function or time-varying associations.

6.3 LIMITATIONS AND FUTURE WORK

In Sections 2.6 and 4.7, we discussed limitations of the IPW-adjusted RMST and RMTL difference. One of these limitations is the possibility of extreme IP weights, which may lead to unstable estimators of the RMST and RMTL difference. Extreme weights can arise when the distributions of IP weights in each exposure group demonstrate poor overlap.(Li et al., 2018, 2019) For instance, in Chapter 2 the IP weights for total cholesterol demonstrate poor overlap in both men and women. Li et al. (2018) proposed overlap weighting as an alternative approach to adjust for confounders. Overlap weighting is used to estimate a different estimand, the average treatment effect of the overlap population. In contrast, IPW is used to estimate the marginal or average treatment effect.

If the exposure groups of interest demonstrate poor overlap, one might consider using overlap weights to instead estimate the average treatment effect of the overlap population. One could then estimate adjusted RMST or RMTL differences by using the overlap weights instead of IP weights using our proposed estima-

tors. To our knowledge, overlap weights have not been previously applied to the difference in RMST or RMTL.

Furthermore, it is possible to improve both the IP weights and IPC weights in Chapters 2 and 4 by using augmented weights.(Steingrimsson et al., 2016; Zhang & Schaubel, 2012b) Augmented weights incorporate two different estimates: an estimate of the propensity score (for example with a logistic model), and an estimate of the outcome conditional on covariates. Therefore, the resulting estimator is less prone to bias, more efficient, and doubly robust.

In Chapters 2 and 3, we did not consider the competing risk of death. From our results in Chapter 4, it is possible that ignoring the competing risk leads to overestimating the risk of AF and RMST (time alive and AF-free). Thus, Chapter 4 is an evolution of the methods proposed in Chapter 2 for time-to-event outcomes other than overall mortality. It would be interesting to further investigate the parametric g-formula for the cumulative incidence function, calculation of the cause-specific RMTL difference, and any challenges which may arise. For instance, Austin et al. (2020) described how time-varying covariates present challenges with Fine-Gray models for competing risks data as the covariates are not observed after death.

In Section 5.7, we described limitations to our regression model for the lifetime risk. Briefly, calculation the pseudo-observations may be computationally intensive in larger datasets and assume the left truncation and right-censoring times are independent of the event time. Furthermore, there is little guidance on performing model selection or assess model fit when fitting the pseudo-observation model. Pavlič et al. (2019) described an approach by assessing the influence function of the pseudo-observations. Sachs et al. (2019) applied the SuperLearner algorithm, an ensemble of machine learning algorithms with cross-validation, to pseudo-

observations of the cumulative incidence function. Since the pseudo-observations are observed for each individual, they can be used in machine learning algorithms for continuous data instead of algorithms for right-censored survival data. It would be interesting to apply the SuperLearner algorithm or other machine learning methods to our pseudo-observation model for the lifetime risk, and perform variable selection for the lifetime risk of AF.

APPENDIX A

Appendix to Chapter 2: adjusted restricted mean survival times

Figure A.1: Weibull hazard functions defining simulation scenarios

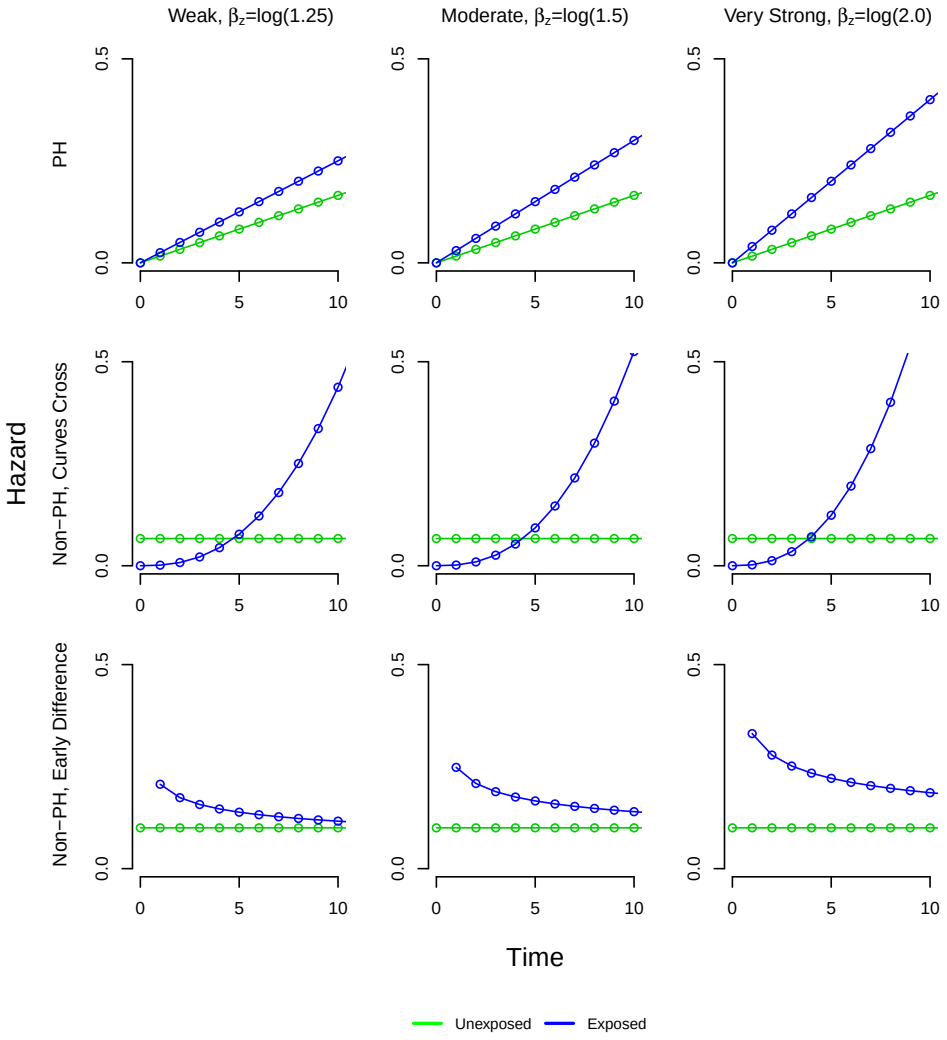


Figure A.2: Weibull survival functions defining simulation scenarios

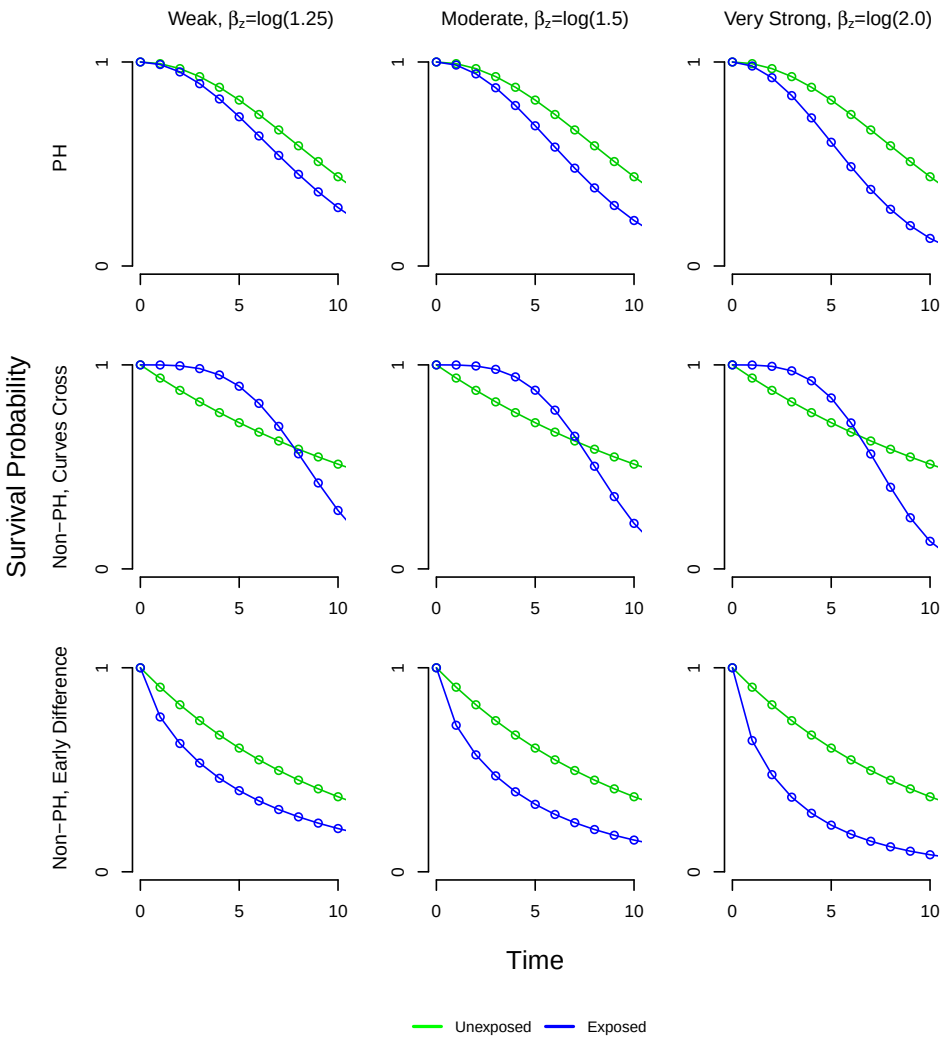
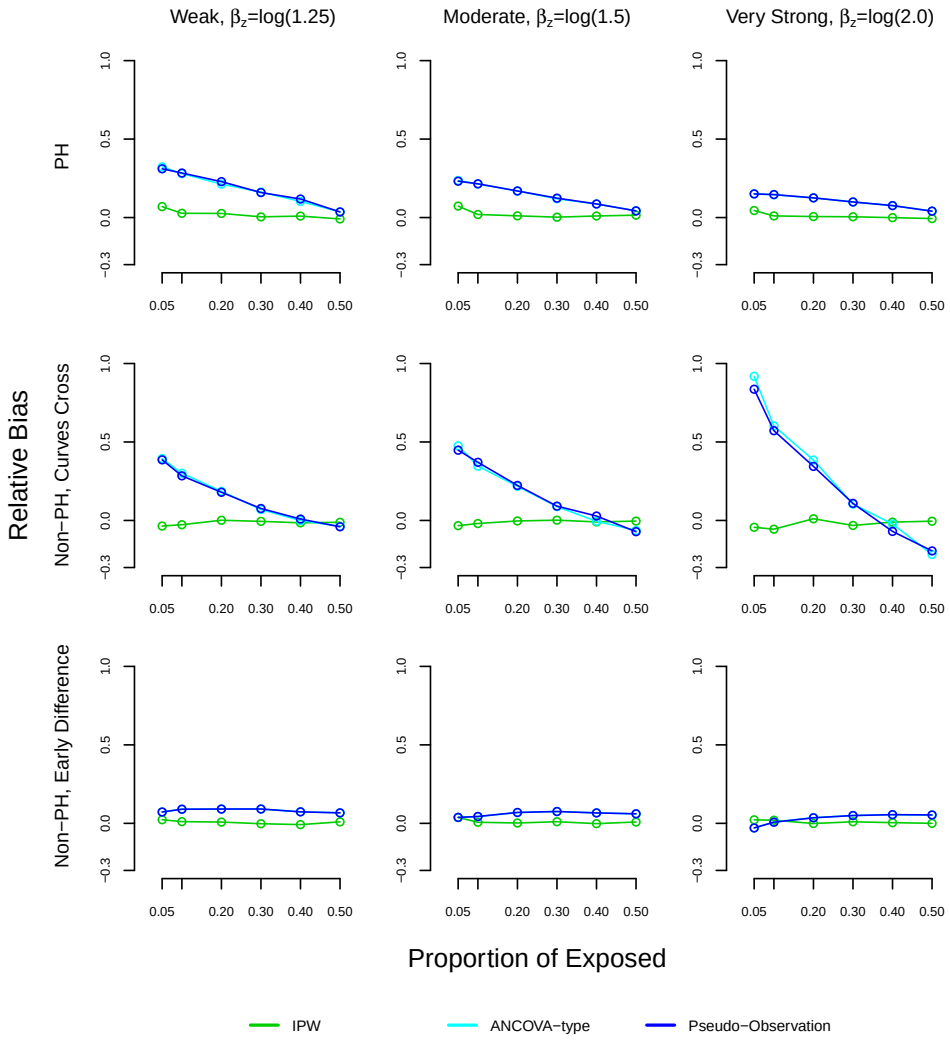
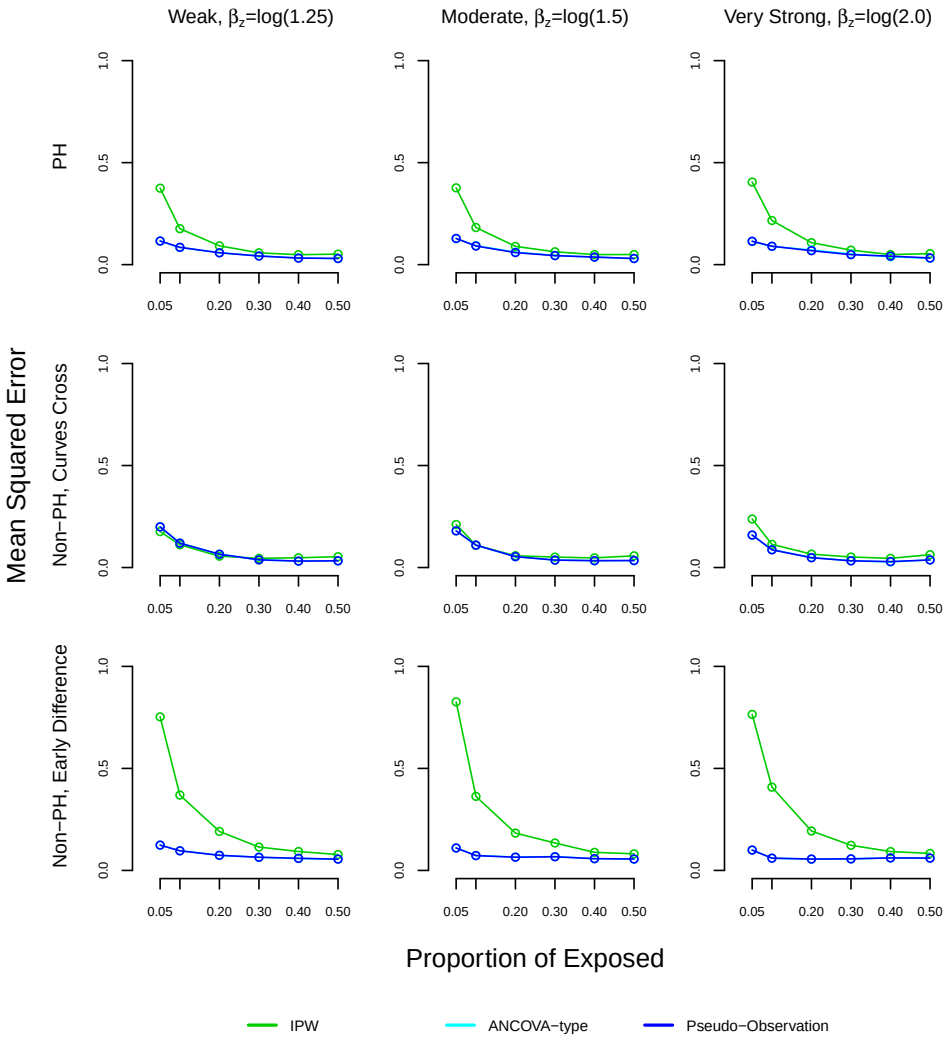


Figure A.3: Relative bias in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=1000



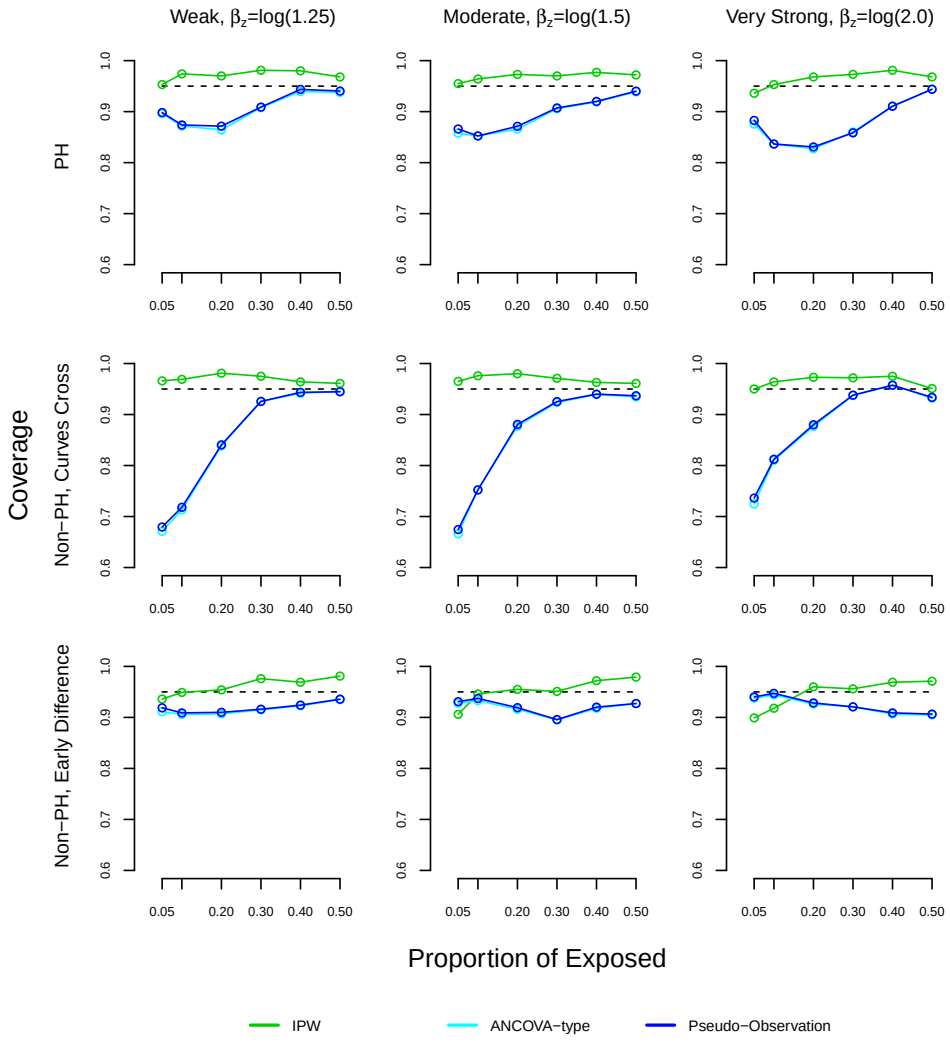
Note: The ANCOVA and pseudo-observation approaches perform similarly, which is why the lines appear overlapping.

Figure A.4: Mean squared error in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=1000



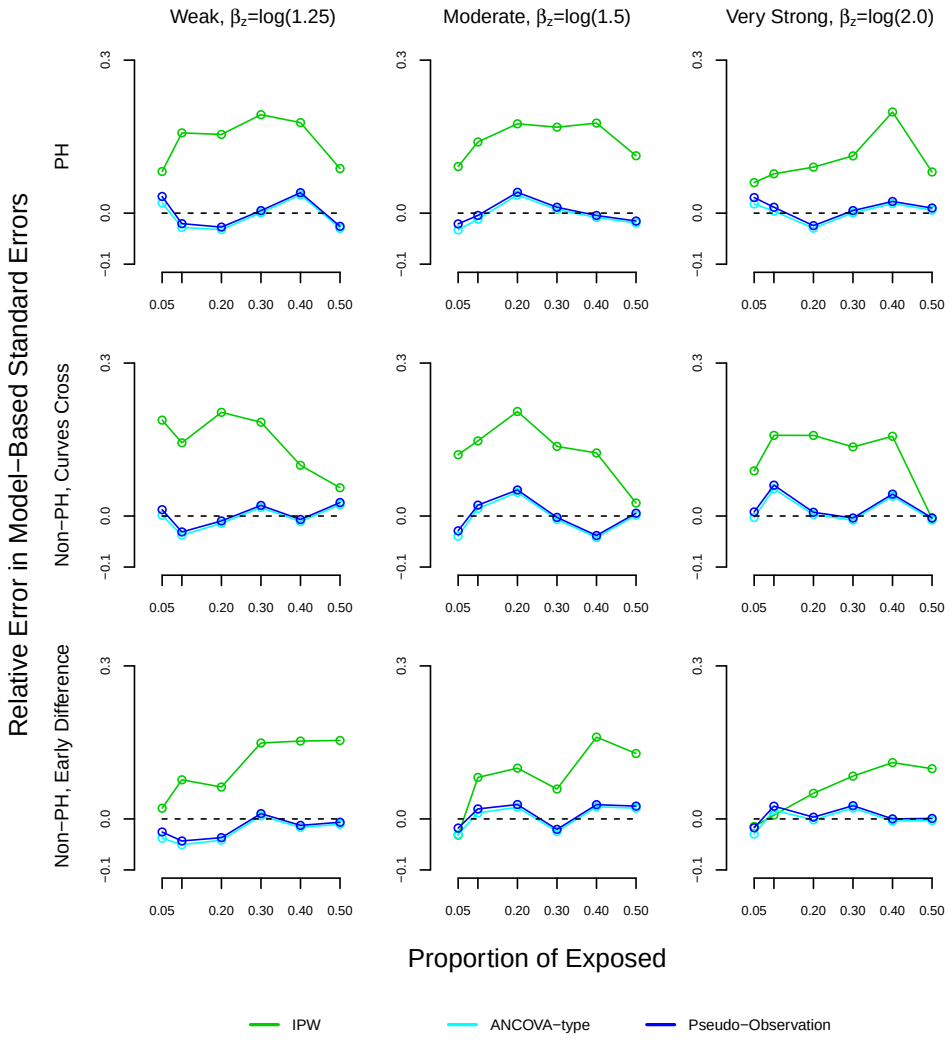
Note: The ANCOVA and pseudo-observation approaches perform similarly, which is why the lines appear overlapping.

Figure A.5: Coverage in 95% confidence intervals in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=1000



Note: The ANCOVA and pseudo-observation approaches perform similarly, which is why the lines appear overlapping.

Figure A.6: Relative error in model-based standard errors in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=1000



Note: The ANCOVA and pseudo-observation approaches perform similarly, which is why the lines appear overlapping.

The relative error is defined as $\frac{ModSE - EmpSE}{EmpSE}$, where $ModSE = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{V}(\hat{\theta}_i)}$ and $EmpSE = \sqrt{\frac{1}{n_{sim}-1} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})}$.

Figure A.7: Relative bias in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=500

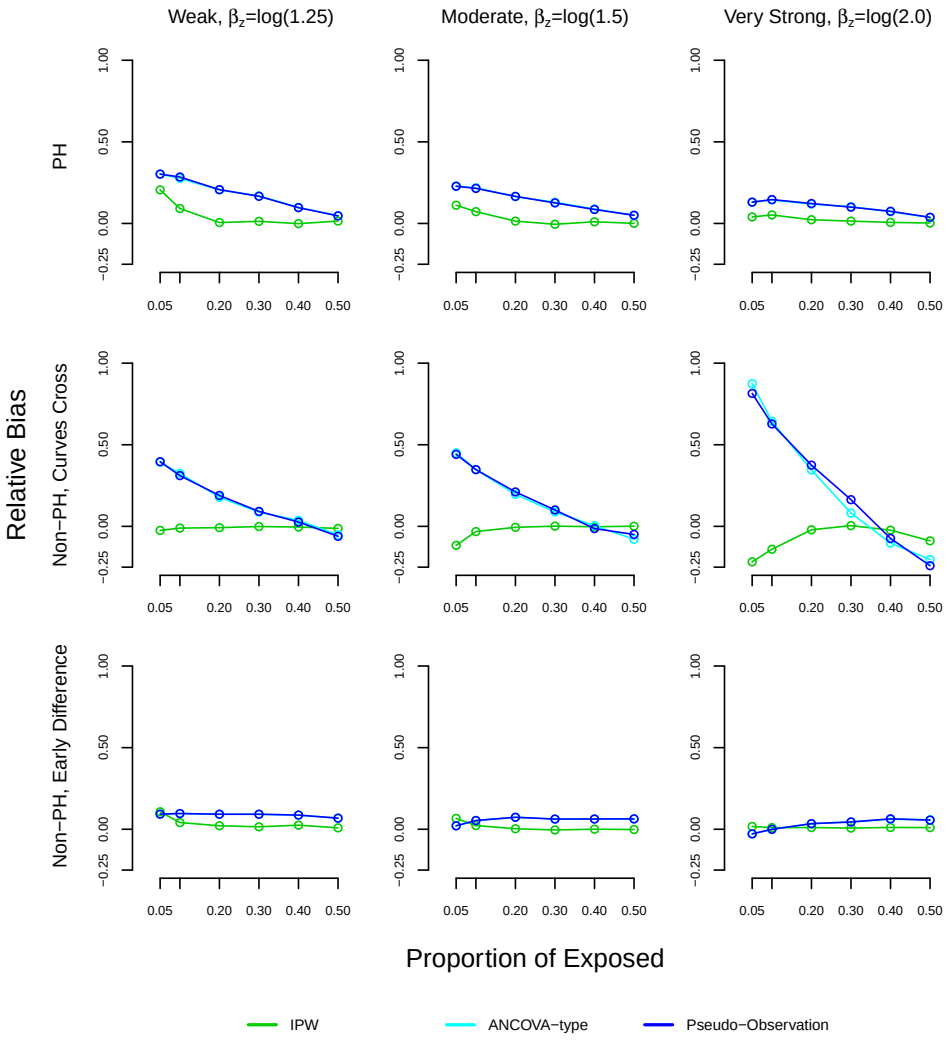
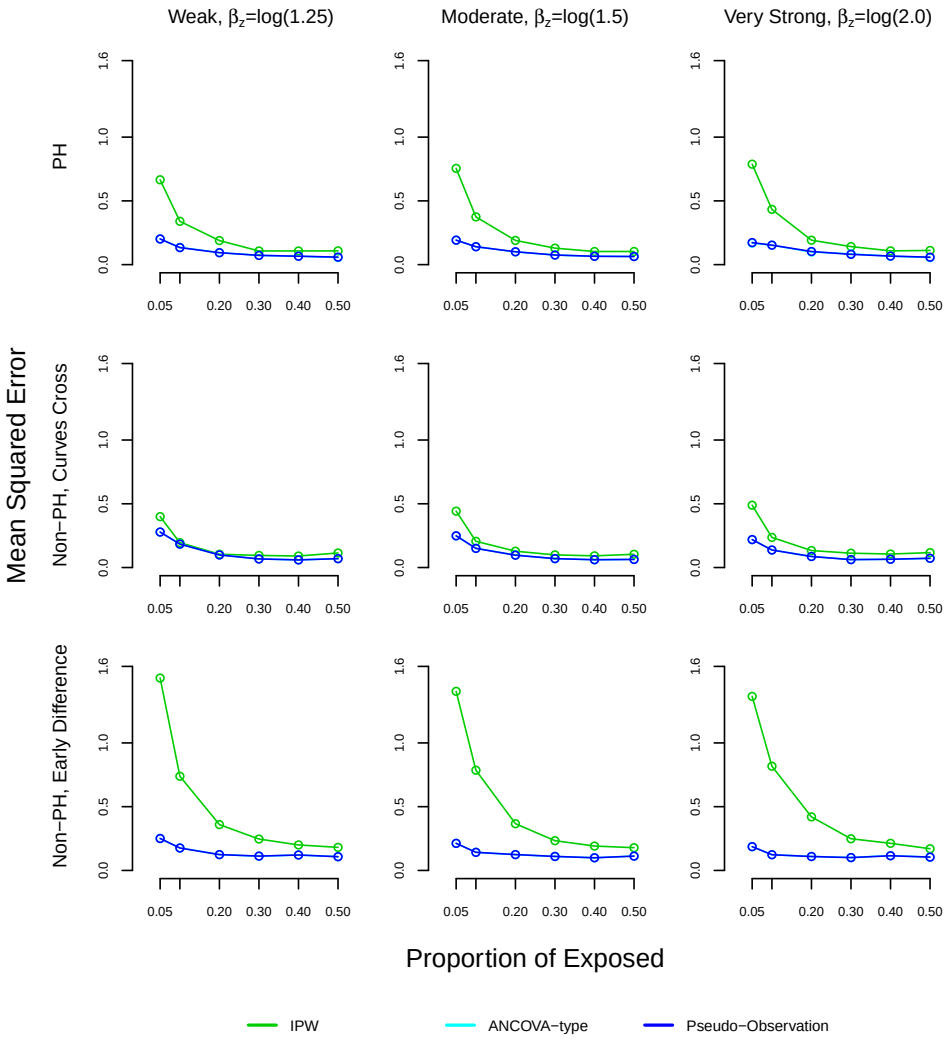


Figure A.8: Mean squared error in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=500



Note: The ANCOVA and pseudo-observation approaches perform similarly, which is why the lines appear overlapping.

Figure A.9: Coverage in 95% confidence intervals in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=500

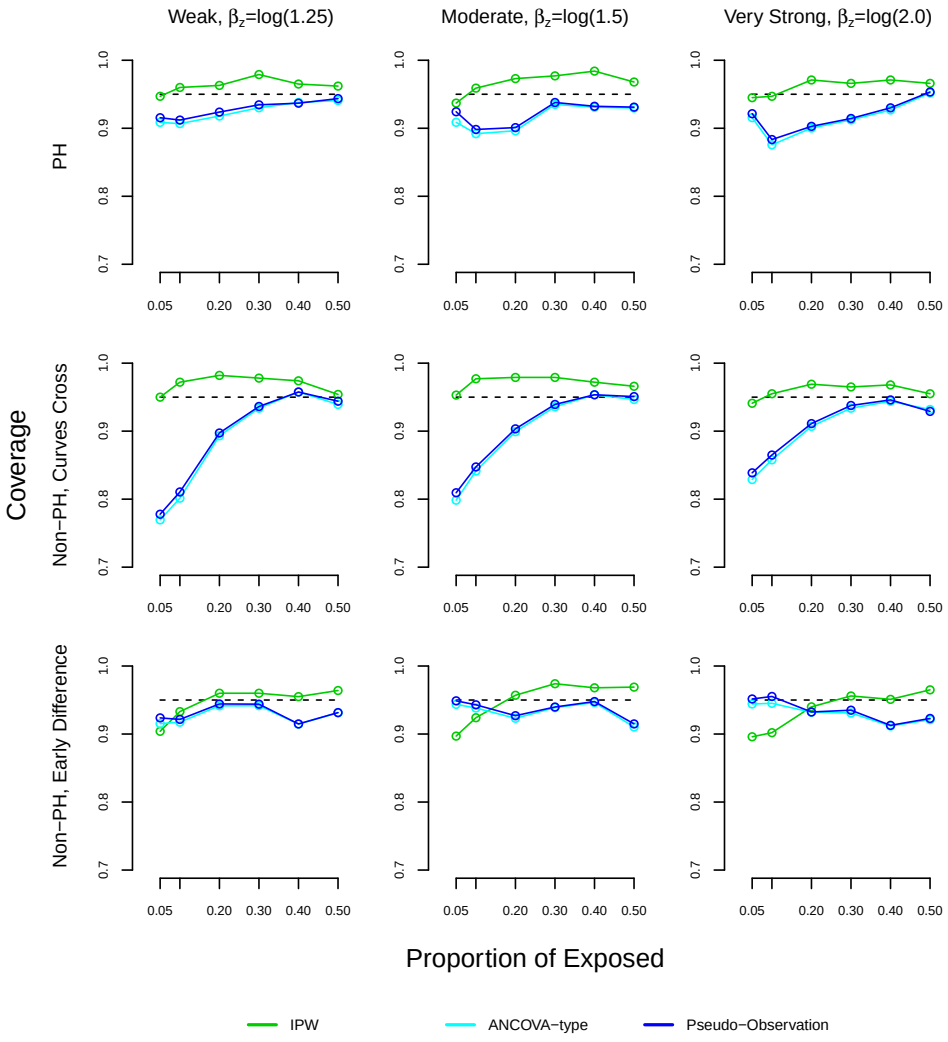
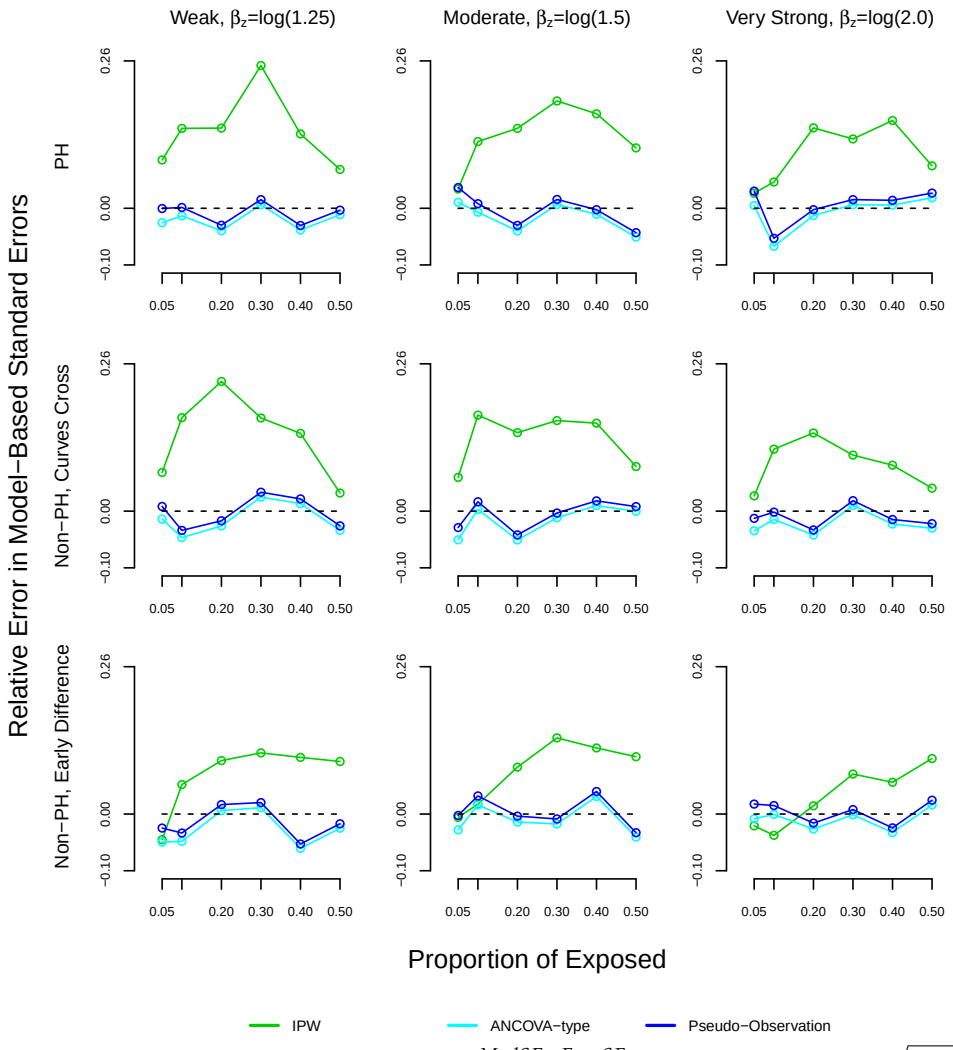


Figure A.10: Relative error in model-based standard errors in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=500



Note: The relative error is defined as $\frac{ModSE - EmpSE}{EmpSE}$, where $ModSE = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{V}(\hat{\theta}_i)}$ and $EmpSE = \sqrt{\frac{1}{n_{sim}-1} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})}$.

Figure A.11: Relative bias in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=250

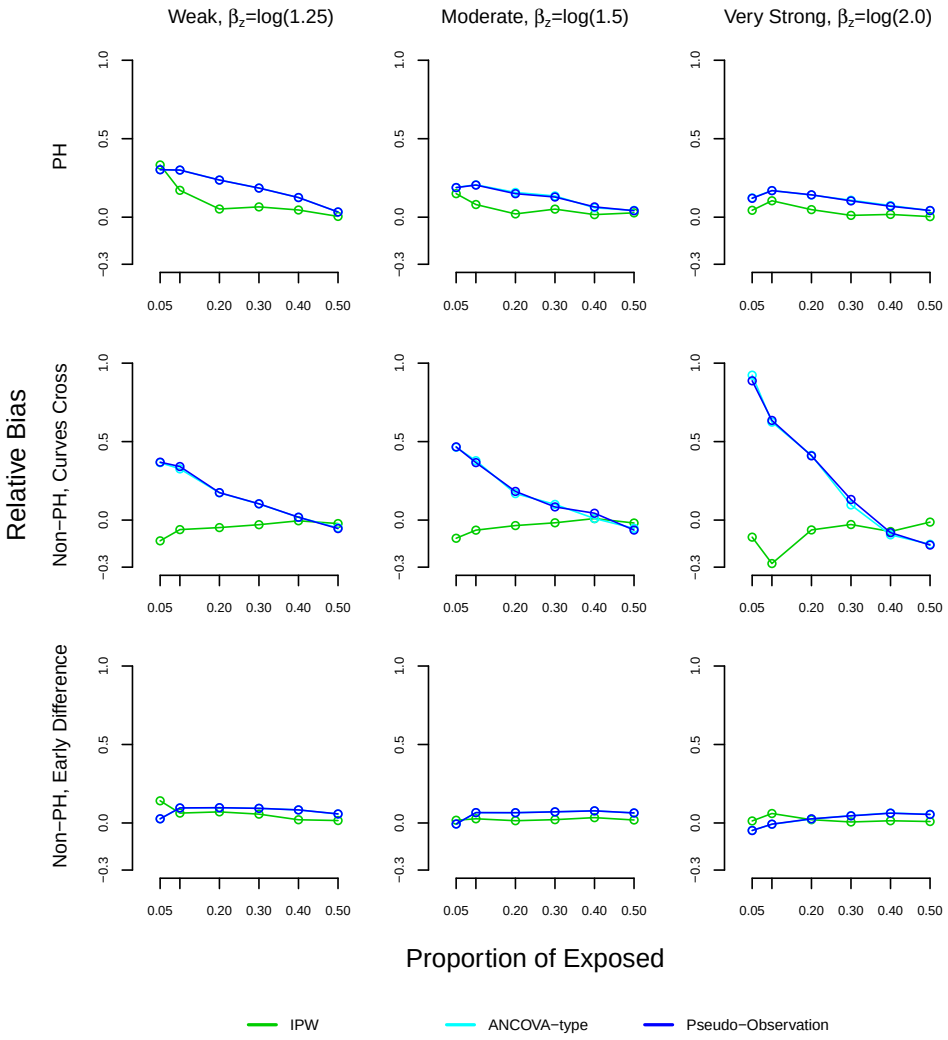


Figure A.12: Mean squared error in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=250

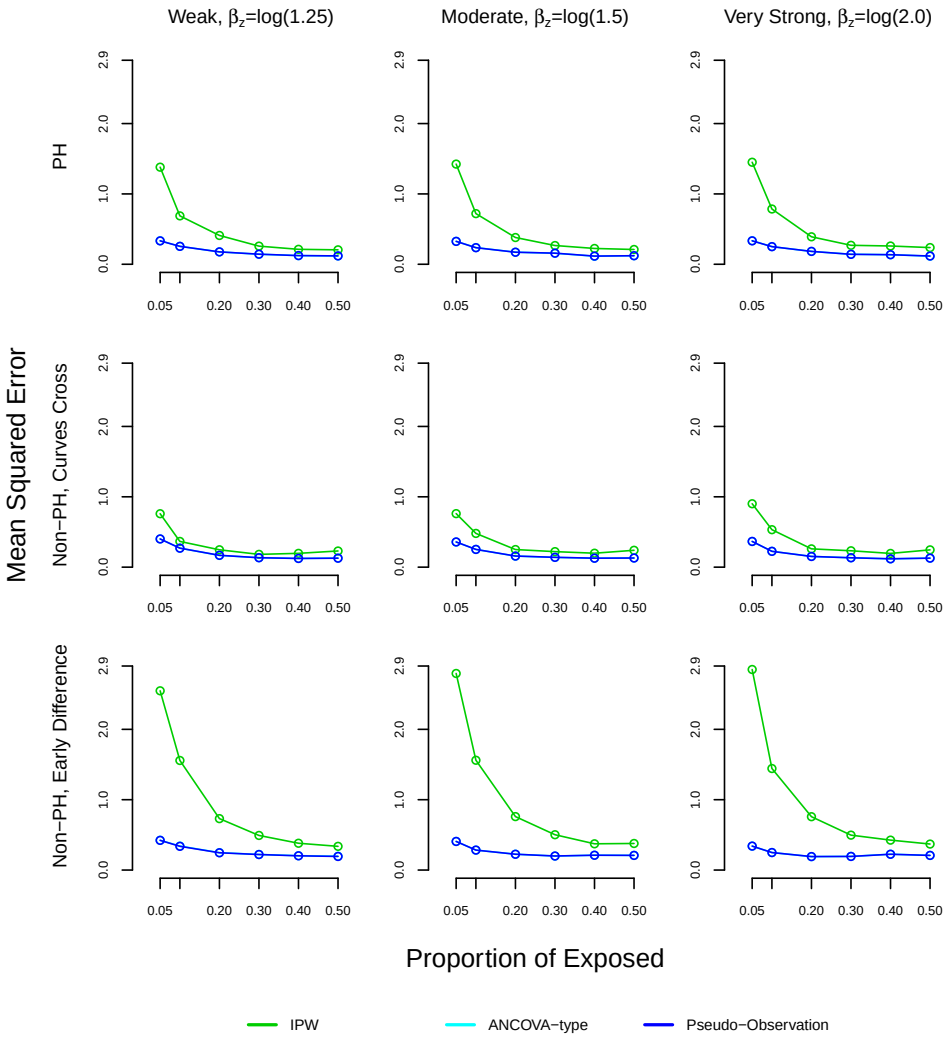


Figure A.13: Coverage in 95% confidence intervals in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=250

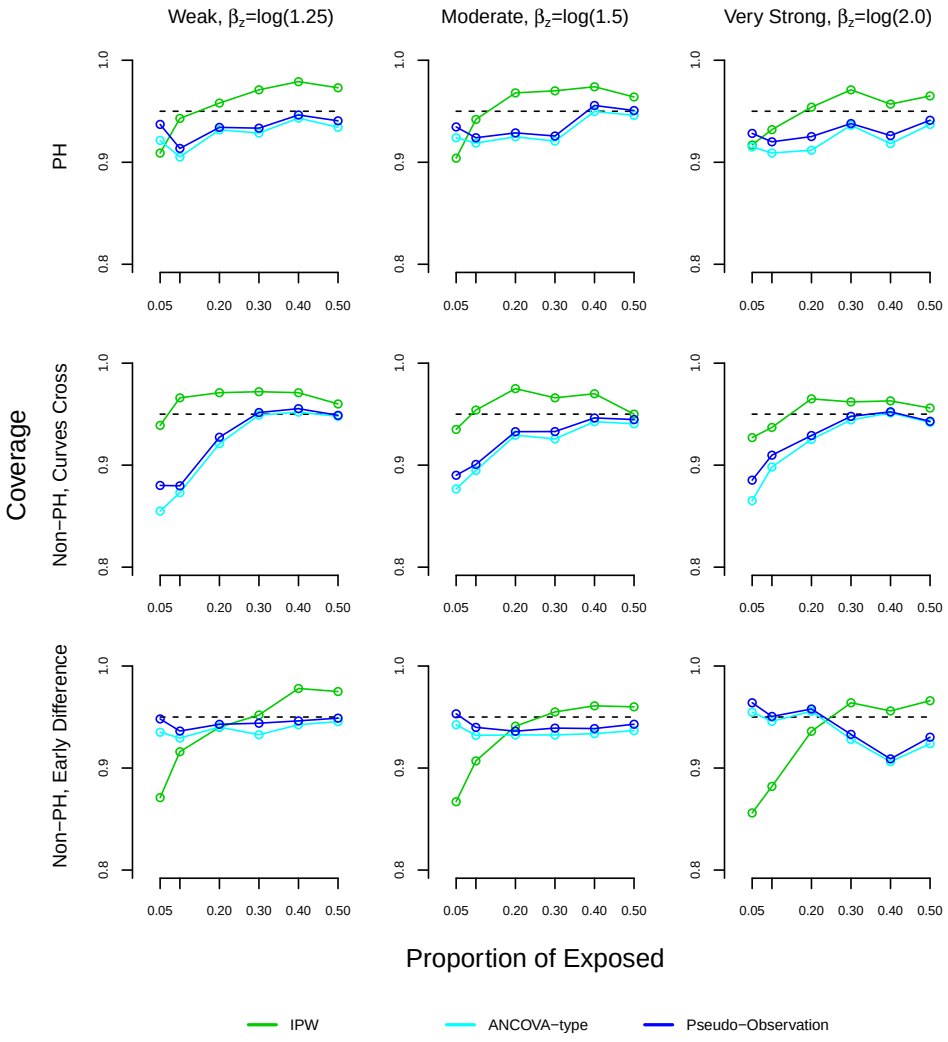
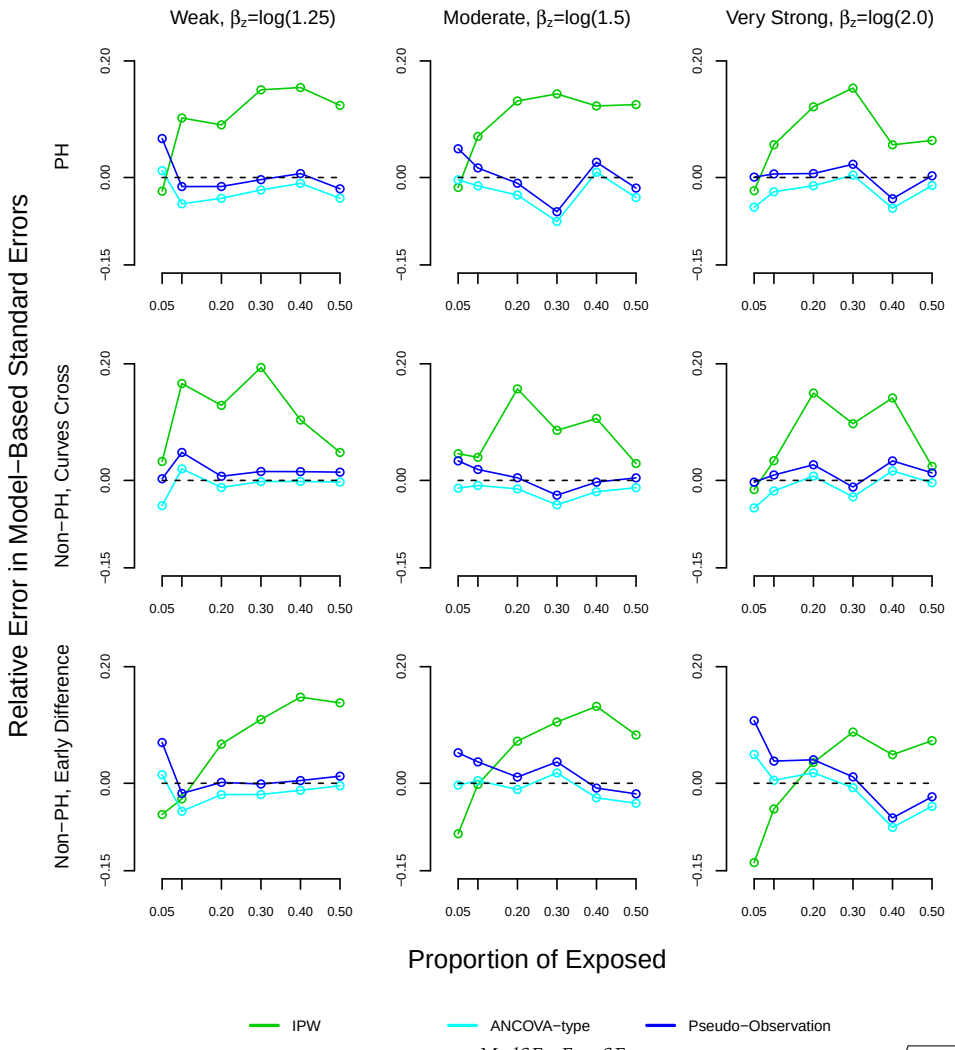


Figure A.14: Relative error in model-based standard errors in simulation study for inverse probability weighting, ANCOVA-type, and pseudo-observation methods, sample size n=250



Note: The relative error is defined as $\frac{ModSE - EmpSE}{EmpSE}$, where $ModSE = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{V}(\hat{\theta}_i)}$ and $EmpSE = \sqrt{\frac{1}{n_{sim}-1} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})}$.

APPENDIX B

Appendix to Chapter 3: adjusted restricted mean survival times with time-varying confounding

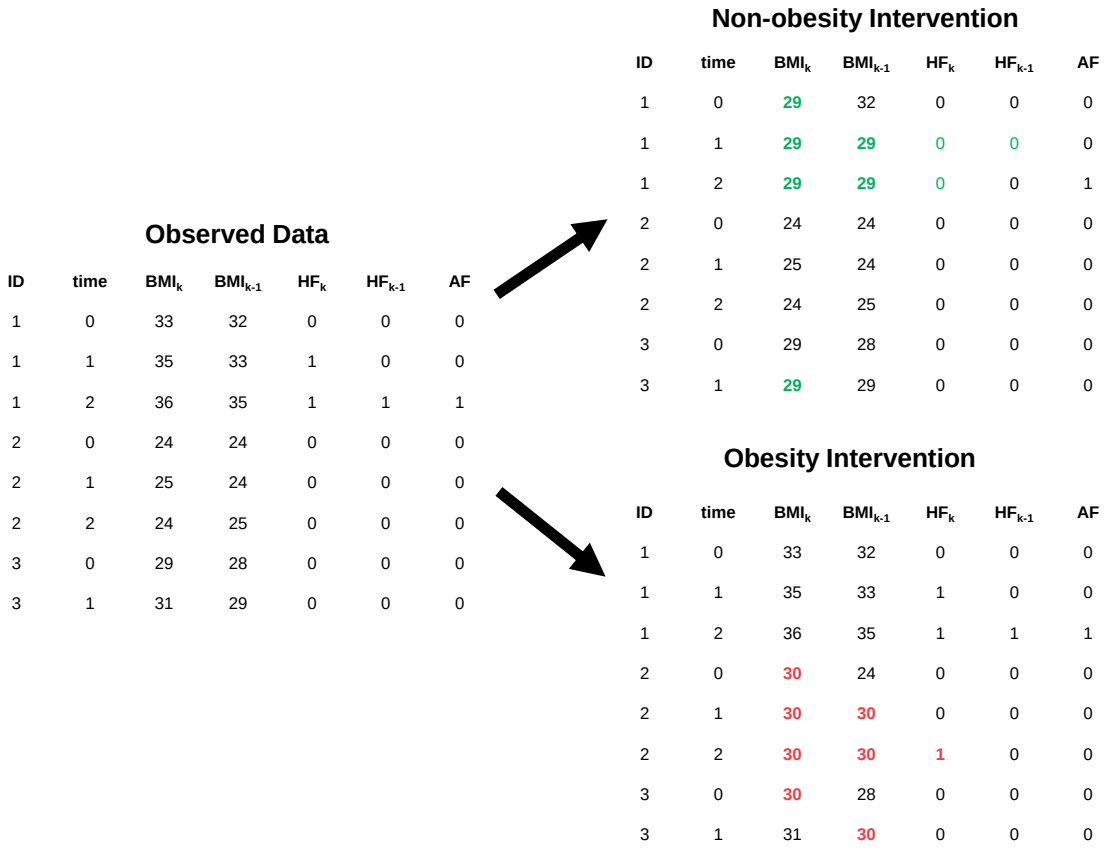


Figure B.1: Diagram of g-formula Monte Carlo simulation

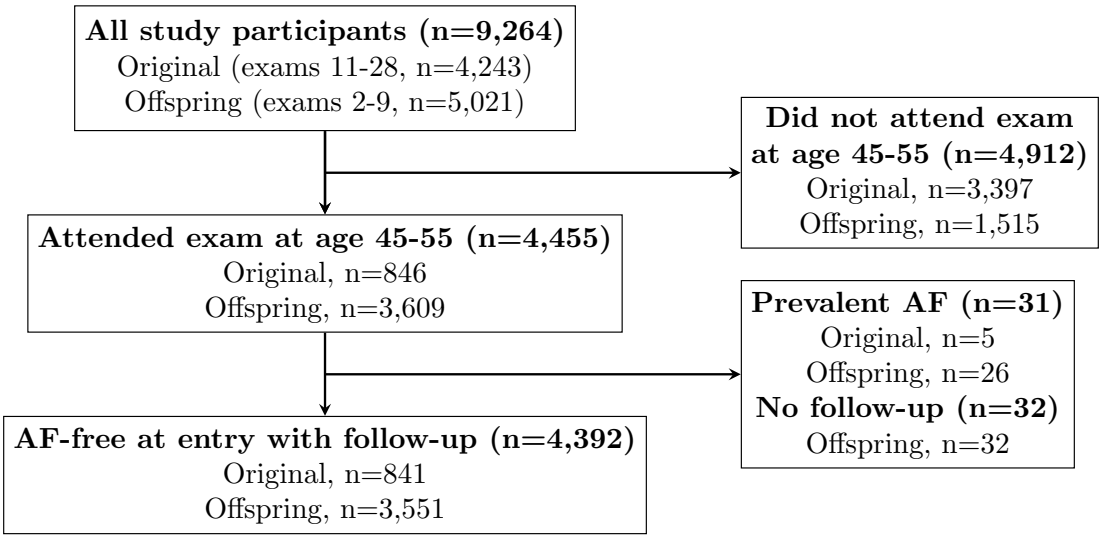


Figure B.2: Flow diagram of study participants

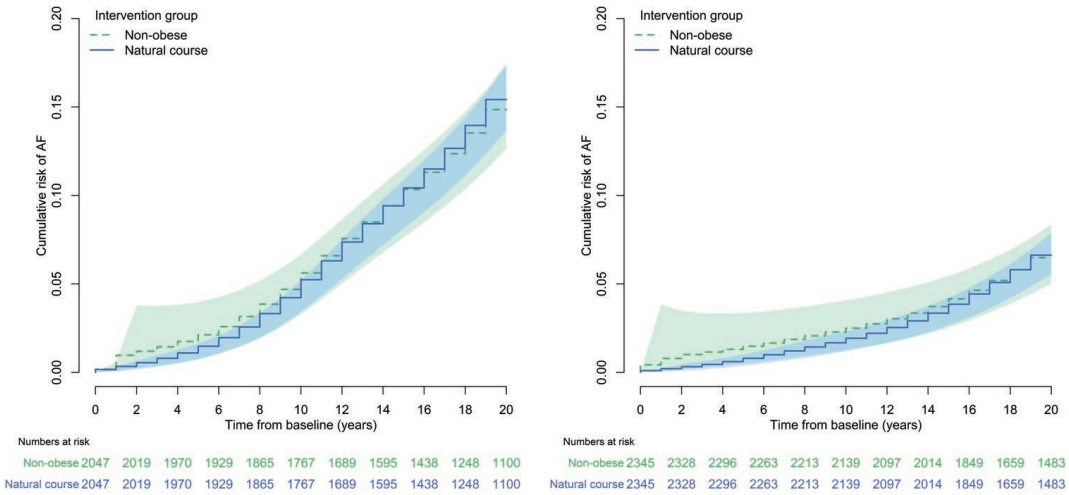


Figure B.3: G-formula survival probabilities for cumulative risk of atrial fibrillation comparing simulated obese and non-obese populations in males (left) and females (right)

Table B.1: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up, men.

Men (n=2,047)	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.81 (0.63, 1.04)
Risk, %	14.81 (12.33, 17.29)	18.11 (14.79, 21.42)	Risk ratio	0.82 (0.66, 1.01)
			Risk difference	-3.30 (-6.89, 0.30)
RMST, years	18.78 (18.41, 19.14)	18.62 (18.25, 18.98)	Difference in RMSTs, mos.	1.89 (-3.67, 7.44)
Non-obese vs. natural course				
			Hazard ratio	0.96 (0.82, 1.14)
Risk, %	14.87 (12.37, 17.37)	15.51 (13.63, 17.39)	Risk ratio	0.96 (0.85, 1.08)
			Risk difference	-0.64 (-2.48, 1.20)
RMST, years	18.77 (18.40, 19.14)	18.82 (18.66, 18.98)	Difference in RMSTs, mos.	-0.61 (-4.70, 3.48)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.97 (0.87, 1.07)
Risk, %	14.97 (13.02, 16.93)	15.49 (13.64, 17.34)	Risk ratio	0.97 (0.92, 1.02)
			Risk difference	-0.52 (-1.30, 0.26)
RMST, years	18.84 (18.68, 19.01)	18.82 (18.66, 18.98)	Difference in RMSTs, mos.	0.28 (-0.37, 0.93)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

Table B.2: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up, women.

Women (n=2,345)	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.69 (0.48, 1.01)
Risk, %	6.24 (4.52, 7.97)	8.94 (6.45, 11.42)	Risk ratio	0.70 (0.50, 0.98)
			Risk difference	-2.69 (-5.37, -0.02)
RMST, years	19.48 (19.21, 19.76)	19.38 (19.18, 19.59)	Difference in RMSTs, mos.	1.19 (-2.34, 4.71)
Non-obese vs. natural course				
			Hazard ratio	0.93 (0.72, 1.21)
Risk, %	6.24 (4.50, 7.98)	6.63 (5.46, 7.81)	Risk ratio	0.94 (0.76, 1.16)
			Risk difference	-0.39 (-1.80, 1.01)
RMST, years	19.48 (19.20, 19.76)	19.53 (19.44, 19.63)	Difference in RMSTs, mos.	-0.58 (-3.74, 2.58)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.97 (0.82, 1.14)
Risk, %	6.37 (5.20, 7.53)	6.63 (5.46, 7.81)	Risk ratio	0.96 (0.91, 1.01)
			Risk difference	-0.27 (-0.62, 0.08)
RMST, years	19.54 (19.45, 19.64)	19.53 (19.44, 19.63)	Difference in RMSTs, mos.	0.11 (-0.13, 0.35)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

Table B.3: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up in cancer-free participants. (n=3,870)

	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.73 (0.55, 0.97)
Risk, %	10.10 (8.01, 12.19)	13.60 (10.73, 16.46)	Risk ratio	0.74 (0.58, 0.95)
			Risk difference	-3.49 (-6.45, -0.54)
RMST, years	19.22 (18.94, 19.50)	19.05 (18.84, 19.27)	Difference in RMSTs, mos.	1.95 (-2.03, 5.92)
Non-obese vs. natural course				
			Hazard ratio	0.92 (0.76, 1.12)
Risk, %	10.09 (8.00, 12.17)	10.93 (9.24, 12.62)	Risk ratio	0.92 (0.80, 1.07)
			Risk difference	-0.85 (-2.34, 0.65)
RMST, years	19.22 (18.94, 19.49)	19.23 (19.11, 19.34)	Difference in RMSTs, mos.	-0.11 (-3.23, 3.01)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.95 (0.85, 1.07)
Risk, %	10.41 (8.72, 12.11)	10.93 (9.25, 12.61)	Risk ratio	0.95 (0.90, 1.00)
			Risk difference	-0.51 (-1.07, 0.04)
RMST, years	19.25 (19.14, 19.36)	19.23 (19.11, 19.34)	Difference in RMSTs, mos.	0.30 (-0.12, 0.71)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

Table B.4: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up, with complete case data at entry. (n=3,102)

	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.71 (0.51, 0.98)
Risk, %	9.76 (7.64, 11.87)	13.58 (9.81, 17.34)	Risk ratio	0.72 (0.54, 0.96)
			Risk difference	-3.82 (-7.69, 0.05)
RMST, years	19.21 (18.84, 19.57)	19.00 (18.44, 19.55)	Difference in RMSTs, mos.	2.52 (-4.49, 9.54)
Non-obese vs. natural course				
			Hazard ratio	0.97 (0.78, 1.20)
Risk, %	9.77 (7.65, 11.90)	10.15 (8.89, 11.42)	Risk ratio	0.96 (0.81, 1.14)
			Risk difference	-0.38 (-2.15, 1.39)
RMST, years	19.20 (18.84, 19.56)	19.25 (19.15, 19.36)	Difference in RMSTs, mos.	-0.62 (-4.74, 3.49)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.96 (0.84, 1.10)
Risk, %	9.64 (8.25, 11.03)	10.14 (8.86, 11.43)	Risk ratio	0.95 (0.88, 1.02)
			Risk difference	-0.51 (-1.19, 0.17)
RMST, years	19.27 (19.16, 19.39)	19.26 (19.15, 19.36)	Difference in RMSTs, mos.	0.23 (-0.32, 0.78)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

Table B.5: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up, entry at age 60 plus or minus five years. (n=6,149)

	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.87 (0.77, 0.99)
Risk, %	19.90 (18.35, 21.45)	22.66 (20.33, 25.00)	Risk ratio	0.88 (0.80, 0.97)
			Risk difference	-2.76 (-4.89, -0.63)
RMST, years	18.35 (18.18, 18.52)	18.22 (18.02, 18.42)	Difference in RMSTs, mos.	1.56 (-0.88, 4.01)
Non-obese vs. natural course				
			Hazard ratio	0.96 (0.87, 1.06)
Risk, %	19.92 (18.36, 21.47)	20.65 (19.21, 22.09)	Risk ratio	0.96 (0.93, 1.00)
			Risk difference	-0.73 (-1.50, 0.03)
RMST, years	18.35 (18.18, 18.51)	18.33 (18.21, 18.46)	Difference in RMSTs, mos.	0.17 (-1.17, 1.52)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.99 (0.90, 1.08)
Risk, %	20.44 (18.98, 21.89)	20.64 (19.18, 22.09)	Risk ratio	0.99 (0.97, 1.01)
			Risk difference	-0.20 (-0.62, 0.23)
RMST, years	18.34 (18.21, 18.46)	18.33 (18.21, 18.46)	Difference in RMSTs, mos.	0.01 (-0.35, 0.37)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

Table B.6: G-formula associations per body mass index intervention group and contrasts between intervention groups at 20 years of follow-up, negative binomial model for atrial fibrillation. (n=4,392)

	Intervention	Comparator	Measure of Association	
Non-obese vs. obese				
			Hazard ratio	0.72 (0.58, 0.89)
Risk, %	9.78 (8.40, 11.16)	13.18 (11.22, 15.14)	Risk ratio	0.74 (0.63, 0.88)
			Risk difference	-3.40 (-5.45, -1.34)
RMST, years	19.24 (19.05, 19.43)	19.04 (18.87, 19.20)	Difference in RMSTs, mos.	2.39 (-0.28, 5.06)
Non-obese vs. natural course				
			Hazard ratio	0.91 (0.78, 1.07)
Risk, %	9.79 (8.41, 11.17)	10.74 (9.64, 11.84)	Risk ratio	0.91 (0.83, 1.00)
			Risk difference	-0.95 (-1.93, 0.04)
RMST, years	19.24 (19.05, 19.43)	19.20 (19.11, 19.29)	Difference in RMSTs, mos.	0.44 (-1.58, 2.46)
10% decrease in BMI per year vs. natural course				
			Hazard ratio	0.97 (0.86, 1.09)
Risk, %	10.35 (9.15, 11.54)	10.77 (9.59, 11.95)	Risk ratio	0.96 (0.92, 1.00)
			Risk difference	-0.42 (-0.84, -0.01)
RMST, years	19.22 (19.12, 19.32)	19.20 (19.10, 19.30)	Difference in RMSTs, mos.	0.23 (-0.10, 0.56)

Numbers are estimates and 95% confidence intervals, obtained with 500 bootstrap samples.

APPENDIX C

Appendix to Chapter 4: adjusted restricted mean time lost in the presence of competing risks

C.1 ASYMPTOTIC VARIANCE OF IPCW ESTIMATING EQUATION

Following Tian et al. (2014) and Bang & Tsiatis (2000, 2002), we leverage martingale identities

$$\frac{\delta_i}{G(X_i)} = 1 - \int_0^\infty \frac{dM_i^c(u)}{G(u)},$$

$$\frac{\hat{G}(t)}{G(t)} = 1 - \int_0^t \frac{\hat{G}(u-)}{G(u)} \frac{dM^c(u)}{Y(u)},$$

and

$$Y(u)/n = \hat{G}(u-)\hat{S}(u-).$$

The estimating equation evaluated at the truth $\boldsymbol{\beta_0}$ can be written as asymptotically linear,

$$\begin{aligned}
&n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta_0}) \\
&= n^{-1/2}\sum_{i=1}^n \left\{ \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} + \frac{\widetilde{\delta}_i}{G(R_i)} - \frac{\widetilde{\delta}_i}{G(R_i)} \right\} \boldsymbol{Z_i^*}\{(\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\boldsymbol{\beta})\} \\
&= n^{-1/2}\sum_{i=1}^n \left[\frac{\widetilde{\delta}_i}{G(R_i)} + \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \left\{ \frac{G(R_i) - \hat{G}(R_i)}{G(R_i)} \right\} \right] \boldsymbol{Z_i^*}\{(\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\boldsymbol{\beta})\} \\
&= n^{-1/2}\sum_{i=1}^n \left[1 - \int_0^\tau \frac{dM_i^c(u)}{G(u)} + \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \left\{ \int_0^{R_i} \frac{\hat{G}(u-)dM^c(u)}{G(u)Y(u)} \right\} \right] \\
&\quad \times \boldsymbol{Z_i^*}\{(\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\boldsymbol{\beta})\} \\
&= n^{-1/2}\sum_{i=1}^n \boldsymbol{Z_i^*}\{(\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\boldsymbol{\beta_0})\}
\end{aligned}$$

$$
\begin{aligned}
&- n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \} - \hat{K}(\boldsymbol{\beta_0}, u) \right] \frac{dM_i^C(u)}{G(u)} \\
&= n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \} \\
&\quad - n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \} - K(\boldsymbol{\beta_0}, u) \right] \frac{dM_i^C(u)}{G(u)} + o_p(1)
\end{aligned}
$$

where

$$
\hat{K}(\boldsymbol{\beta_0}, u) = \frac{1}{n\hat{S}(u)} \sum_{i=1}^{n} \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \left[\boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \} I(R_i \geq u) \right]
$$

and

$$
K(\boldsymbol{\beta_0}, u) = \frac{1}{S(u)} E \left[\boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \} I(R_i \geq u) \right].
$$

The first two terms are independent due to martingale properties, so

$$
\begin{aligned}
&\text{var}(n^{-1/2} \boldsymbol{U}(\boldsymbol{\beta_0})) \\
&= E \left[\boldsymbol{Z}_i^{*\otimes 2} \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{\beta_0} \boldsymbol{Z}_i^*) \}^2 \right] \\
&+ \int_0^{\tau} E \left[\left(\boldsymbol{Z}_i^* \{ (\tau - R_i) \cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{\beta_0} \boldsymbol{Z}_i^*) \} - K(\boldsymbol{\beta_0}, u) \right)^{\otimes 2} I(R_i \geq u) \right] \frac{d\Lambda_c(u)}{G(u)}
\end{aligned} \tag{C.1}
$$

Then, $\text{var}\{n^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})\} = A^{-1} B A^{-1}$ where B is the variance of the estimating equation (C.1) and

$$
A = E \left[\boldsymbol{Z}_i^{*\otimes 2} \frac{d}{d\boldsymbol{\beta_0}} g^{-1}(\boldsymbol{Z}_i^* \boldsymbol{\beta_0}) \right].
$$

This covariance matrix can be estimated by $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ in which

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z_i^*}^{\otimes 2} \frac{d}{d\hat{\beta}} g^{-1}(\boldsymbol{Z_i^*}\hat{\boldsymbol{\beta}})$$

and

$$\begin{aligned}\hat{B} = &\frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \left[\boldsymbol{Z_i^*}^{\otimes 2}\left((\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\hat{\boldsymbol{\beta}})\right)^2\right] \\ &+ \int_0^\tau \frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{\delta}_i}{\hat{G}(R_i)} \left[\boldsymbol{Z_i^*}\{(\tau - R_i)\cdot I(\epsilon_i = 1) - g^{-1}(\boldsymbol{Z_i^*}\hat{\boldsymbol{\beta}}) - \hat{K}(\hat{\boldsymbol{\beta}}, u)\}\right]^{\otimes 2} \\ &\times I(R_i \geq u)\frac{d\hat{\Lambda}_c(u)}{\hat{G}(u)},\end{aligned}$$

where $\hat{\Lambda}_c(t) = \int_0^t \frac{dN_c(u)}{\pi(u)} du$, $N_c(u) = \sum_i I(R_i \leq u, \widetilde{\delta}_i = 0)$, and $\pi(u) = \sum_i I(R_i \geq u)$.

C.2 PARAMETERS FOR DATA GENERATION IN SIMULATION STUDY OF IPCW REGRESSION MODEL

Table C.1: Gompertz parameters for data generation in simulation study of IPCW regression model

Proportional subdistribution hazards, $\tau = 365$ days					
ρ	γ	Z_1	Z_2	RMTL $\mu_1(\tau)$, years	RMTL $\mu_1(\tau)$, days
-2.8	1.546588	1	1	0.30	109.5
-2.9	2.291237	1	0	0.40	146.0
-1.4	1.952647	0	1	0.45	164.25
-1.7	2.879494	0	0	0.55	200.75
Non-proportional subdistribution hazards, $\tau = 365$ days					
ρ	γ	Z_1	Z_2	RMTL $\mu_1(\tau)$, years	RMTL $\mu_1(\tau)$, days
-1.7	1.213444	1	1	0.30	109.5
-1.7	1.770851	1	0	0.40	146.0
-1.7	2.096025	0	1	0.45	164.25
-1.7	2.879494	0	0	0.55	200.75

To assess the IPCW model, we generated competing risks data dependent on 2 binary covariates (Z_1, Z_2). We set the true unadjusted difference in RMTL at $\tau = 365$ days, $(\beta_1, \beta_2) = \{\mu_{1,1z_2}(\tau) - \mu_{1,0z_2}(\tau), (\mu_{1,z_11}(\tau) - \mu_{1,z_10}(\tau)\}$.

C.3 ADJUSTED RESTRICTED MEAN TIMES LOST BY SEX FROM ILLUSTRATIVE EXAMPLES

For the IPCW models, we predicted the RMTL for each sex with a least squares means approach by summing the estimated model coefficients at the mean value for continuous covariates and the sample proportion for binary covariates.

Table C.2: Adjusted RMTLs for 40-year incident atrial fibrillation by sex.

		RMTL, years	
Estimators accounting for competing risks		**AF**	**Death without AF**
Males	IPW	6.74 (5.98, 7.49)	9.74 (8.90, 10.57)
	IPCW regression model	6.64 (5.78, 7.51)	8.95 (8.17, 9.74)
Females	IPW	5.33 (4.73, 5.93)	7.48 (6.86, 8.10)
	IPCW regression model	7.07 (6.37, 7.78)	7.71 (7.06, 8.37)
		RMTL, years	
Estimators not accounting for competing risks		**AF**	**Death without AF**
Males	IPW	9.50 (8.55, 10.45)	12.07 (11.18, 12.97)
	IPCW regression model	10.85 (9.87, 11.82)	12.80 (12.00, 13.60)
Females	IPW	6.67 (5.96, 7.39)	9.15 (8.48, 9.81)
	IPCW regression model	12.08 (10.86, 13.29)	10.83 (10.03, 11.63)

AF: atrial fibrillation. IPW: inverse probability weighted. IPCW: inverse probability of censoring weighted.

Table C.3: Adjusted RMTLs for 10-year cause-specific death after atrial fibrillation by sex.

		RMTL, years	
Estimators accounting for competing risks		**CVD death**	**Non-CVD death**
Males	IPW	1.62 (1.40, 1.84)	2.13 (1.88, 2.37)
	IPCW regression model	1.58 (1.28, 1.89)	2.13 (1.85, 2.41)
Females	IPW	1.44 (1.17, 1.70)	1.79 (1.51, 2.08)
	IPCW regression model	1.39 (1.05, 1.73)	1.76 (1.44, 2.09)
		RMTL, years	
Estimators not accounting for competing risks		**CVD death**	**Non-CVD death**
Males	IPW	1.95 (1.70, 2.19)	2.47 (2.20, 2.73)
	IPCW regression model	2.50 (2.28, 2.72)	3.01 (2.76, 3.25)
Females	IPW	1.66 (1.37, 1.95)	2.04 (1.73, 2.35)
	IPCW regression model	2.16 (1.89, 2.43)	2.49 (2.20, 2.78)

CVD: cardiovascular disease. IPW: inverse probability weighted. IPCW: inverse probability of censoring

APPENDIX D

Appendix to Chapter 5: regression model for the lifetime risk using pseudo-observations

D.1 ILLUSTRATION OF NON-PROPORTIONAL SUBDISTRIBUTION HAZARDS WITH PROPORTIONAL CAUSE-SPECIFIC HAZARDS

Let the cause-specific hazards take Weibull form, $\alpha_1(t;Z) = \frac{a_1}{b_1^{a_1}} t^{a_1-1} e^{\gamma_1 Z}$ and $\alpha_2(t;Z) = \frac{a_2}{b_2^{a_2}} t^{a_2-1} e^{\gamma_2 Z}$, where Z is a binary covariate with a proportional effect on both cause-specific hazards. The all-cause cumulative hazard is $A(t;Z) = \frac{t}{b_1}^{a_1} e^{\gamma_1 Z} + \frac{t}{b_2}^{a_2} e^{\gamma_1 Z}$. The proportional effect of Z on the cause-specific hazard can be verified by deriving the cause-specific hazard ratio (CSHR),

$$\frac{\alpha_1(t;Z=1)}{\alpha_1(t;Z=0)} = \frac{\frac{a_1}{b_1^{a_1}} t^{a_1-1} e^{\gamma_1}}{\frac{a_1}{b_1^{a_1}} t^{a_1-1}} = e^{\gamma_1}.$$

The subdistribution hazard can be represented as a function of the cause-specific hazard, $\lambda_1(t) = \alpha_1(t)/\left(1 + \frac{F_2(t)}{S(t)}\right)$, where $F_k(t) = \int_0^t S(u)\alpha_k(u)du = 1 - \exp(-\int_0^t \lambda_k(u)du)$. The subdistribution hazard ratio (SDHR) is then

$$\frac{\lambda_1(t;Z=1)}{\lambda_1(t;Z=0)} = \frac{\alpha_1(t;Z=1)/\left(1 + \dfrac{F_2(t;Z=1)}{S(t;Z=1)}\right)}{\alpha_1(t;Z=0)/\left(1 + \dfrac{F_2(t;Z=0)}{S(t;Z=0)}\right)} = e^{\gamma_1} \cdot \left(\frac{1 + \dfrac{F_2(t;Z=0)}{S(t;Z=0)}}{1 + \dfrac{F_2(t;Z=1)}{S(t;Z=1)}}\right).$$

The SDHR is not constant, but equal to the CSHR multiplied by a function of time.

D.2 TABLES AND FIGURES

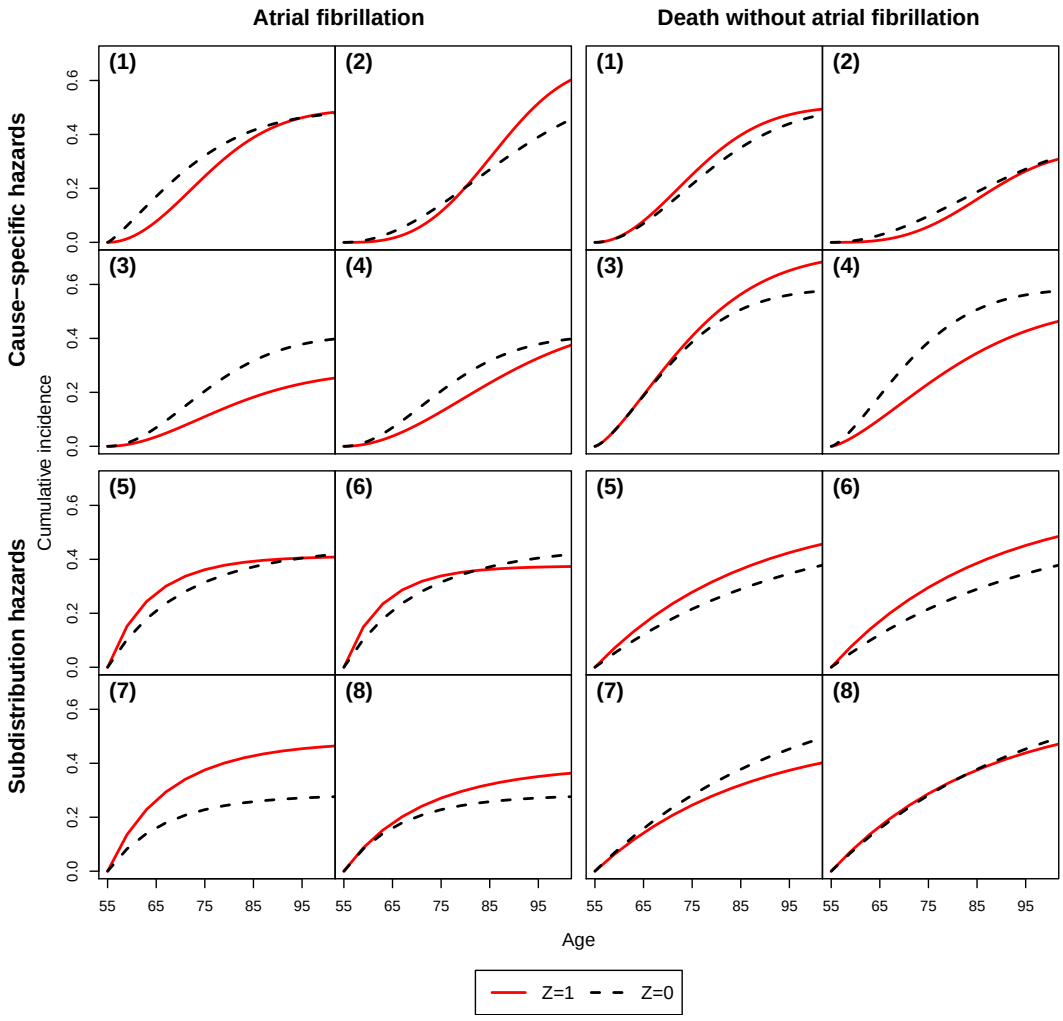


Figure D.1: True cumulative incidence functions in simulation study of lifetime risk

Table D.1: Settings for cause-specific hazards data generation

	Setting 1	**Setting 2**	**Setting 3**	**Setting 4**
Z associated with atrial fibrillation	Yes	Yes	Yes	No
Z associated with death	No	Yes	No	Yes
Proportional cause-specific hazards	No	No	Yes	Yes
CSHR for atrial fibrillation	$0.006*t^{1.9}$	.059*t	0.5	1.0
CSHR for death	1.0	.043*t	1.0	0.5
Weibull distribution of CSH				
CSH of atrial fibrillation, Z=0	Weibull(1.2, 39.5)	Weibull(2, 50)	Weibull(2, 35)	Weibull(2, 35)
CSH of atrial fibrillation, Z=1	Weibull(3.1, 29.5)	Weibull(3, 40)	Weibull(2, 49.5)	Weibull(2, 35)
CSH of death, Z=0	Weibull(1.4, 23.4)	Weibull(2, 60)	Weibull(1.4, 30)	Weibull(1.4, 30)
CSH of death, Z=1	Weibull(1.4, 23.4)	Weibull(3, 50)	Weibull(1.4, 30)	Weibull(1.4, 49.22)
Lifetime risk, Z=1	0.34	0.52	0.23	0.33
Lifetime risk, Z=0	0.34	0.39	0.38	0.38
Difference in lifetime risk	0.00	0.12	-0.15	-0.05

CSH: cause-specific hazard function. CSHR: cause-specific hazard ratio. Sample size before left truncation was set to n=500 and n=1,000. Right censoring was generated from a Uniform(c,d) distribution, where parameters c and d were set to obtain 25% and 75% probability of censoring at τ. Left truncation times were generated from a Uniform(0,5) distribution for 50% and 80% of individuals.

Table D.2: Settings for non-proportional subdistribution hazards data generation

	Setting 5	**Setting 6**	**Setting 7**	**Setting 8**
Main effect of Z on atrial fibrillation	Yes	Yes	Yes	No
Time-varying effect of Z on atrial fibrillation	Yes	Yes	Yes	Yes
Main effect of Z on death	Yes	Yes	Yes	Yes
Subdistribution HR for atrial fibrillation	$1.6487e^{-1.72t}$	$1.6487e^{-2.2t}$	$1.6487e^{0.5t}$	$e^{0.5t}$
Gompertz parameters				
γ	1.2	1.2	1	1
ρ	-2	-2	-3	-3
Effect of Z on atrial fibrillation, ψ_1	0.5	0.5	0.5	0
Time-varying effect of Z on atrial fibrillation, ψ_2	-1.72	-2.2	0.5	1
Effect of Z on death, ψ_3	0.5	0.5	0.5	0.5
Lifetime risk, Z=1	0.40	0.37	0.45	0.35
Lifetime risk, Z=0	0.40	0.40	0.27	0.27
Difference in lifetime risk	0.00	-0.03	0.18	0.08

HR: cause-specific hazard ratio. Sample size before left truncation was set to n=500 and n=1,000. Right censoring was generated from a Uniform(c,d) distribution, where parameters c and d were set to obtain 25% and 75% probability of censoring at τ. Left truncation times were generated from a Uniform(0,5) distribution for 50% and 80% of individuals.

Table D.3: Participant characteristics at index age 55 in the Framingham Heart Study

Risk factor	n=5,613
Age at entry, years	55.68 (1.03)
Male	2,660 (47.39)
Systolic blood pressure, mmHg	125.78 (17.09)
Diastolic blood pressure, mmHg	78.46 (9.85)
Hypertension medication use	1,227 (21.86)
Smokers	1,186 (21.13)
Alcohol use, elevated	474 (8.44)
Body mass index, kg/m^2	27.68 (5.29)
Prior diabetes	426 (7.59)
Prior heart failure	19 (0.34)
Prior myocardial infarction	135 (2.41)

Values are means (standard deviation) or frequency (%).

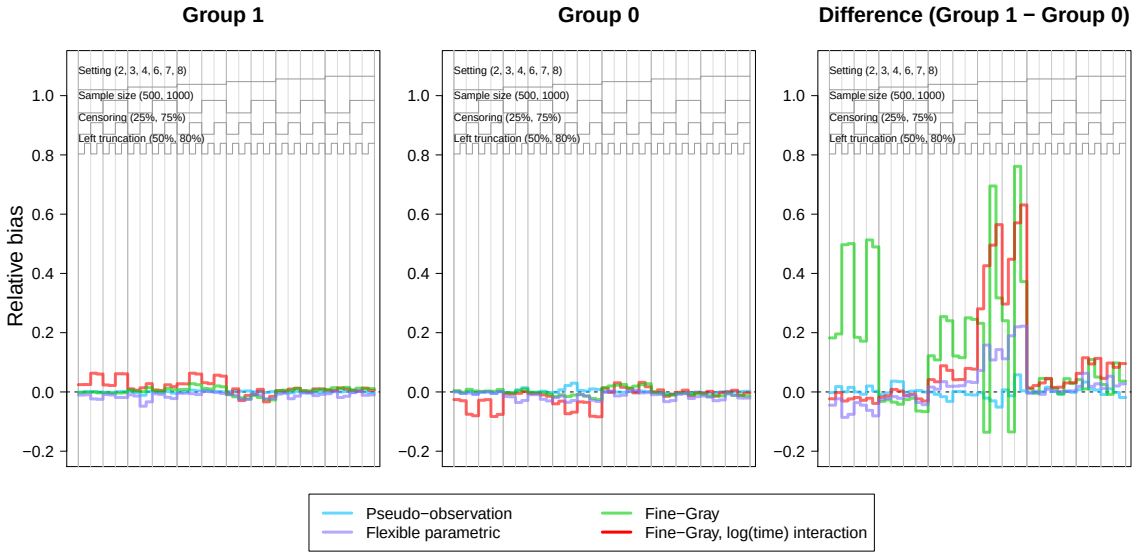


Figure D.2: Nested loop plot showing the simulation study results: relative bias

Table D.4: Predicted percentage differences in lifetime risk of death without atrial fibrillation at age 95, from index age 55

		Cumulative incidence models	
Risk factor	**Pseudo-observation model**	**Fine-Gray**	**Flexible parametric**
Age at entry, (SDU)	-1.44 (-5.21, 2.33)	1.50 (-0.36, 4.33)	0.48 (-1.56, 2.51)
Male vs. female	-5.40 (-12.02, 1.22)	2.87 (-4.33, 7.21)	4.01 (-1.44, 9.45)
Systolic blood pressure, (SDU)	-3.44 (-8.06, 1.19)	1.80 (-1.23, 5.23)	1.67 (-1.52, 4.86)
Diastolic blood pressure, (SDU)	6.34 (1.95, 10.73)	4.63 (1.10, 7.79)	4.61 (1.33, 7.89)
Hypertension medication, yes vs. no	-7.40 (-14.08, -0.72)	-8.03 (-15.82, -2.94)	-3.2 (-9.3, 2.91)
Smoker vs. non-smoker	19.69 (11.26, 28.12)	21.31 (15.60, 26.26)	25.85 (21.67, 30.03)
Alcohol use, elevated vs. optimal)	-5.69 (-18.96, 7.59)	2.34 (-5.60, 9.83)	2.48 (-5.37, 10.33)
Body mass index, (SDU)	0.18 (-2.43, 2.78)	-2.22 (-5.04, 0.43)	-2.93 (-5.5, -0.37)
Prior diabetes, yes vs. no	8.96 (-1.92, 19.85)	21.61 (14.57, 28.92)	23.32 (15.85, 30.79)
Prior heart failure, yes vs. no	4.15 (-25.61, 33.91)	-21.15 (-43.59, 11.06)	18.98 (-6.33, 44.3)
Prior myocardial infarction, yes vs. no	16.36 (3.09, 29.62)	13.28 (-4.12, 26.39)	22.21 (12.56, 31.85)

SDU: standard deviation unit. Results are % differences in lifetime risk and 95% confidence intervals. Differences in lifetime risk for continuous risk factors are the differences for a 1 SDU increase in the covariate from the mean value of the covariate. For the pseudo-observation model and flexible parametric models, differences in lifetime risk were obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are obtained with the delta method. For the Fine-Gray model, differences in lifetime risk were also obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are the .025 and .975 percentile of bootstrap resamples.

Table D.5: Multivariable models for incident atrial fibrillation in the Framingham Heart Study, from index age 55 to age 95

			Cumulative incidence models			
	Pseudo-observation		Fine-Gray		Royston-Parmar	
Risk factor at entry age	**OR (95% CI)**	**p-value**	**SDHR (95% CI)**	**p-value**	**SDHR (95% CI)**	**p-value**
Age at entry	0.90 (0.79, 1.03)	0.119	0.94 (0.88, 1.01)	0.102	0.95 (0.88, 1.02)	0.127
Male	1.59 (1.25, 2.01)	<.001	8.63e+04 (435.94, 1.71e+07)	<.001	1.82e+04 (379.46, 8.72e+05	<.001
Male x log(time)			0.08 (0.02, 0.28)	<.001	0.12 (0.05, 0.30)	<.001
Systolic blood pressure	1.27 (1.06, 1.52)	0.009	1.20 (1.09, 1.33)	<.001	1.21 (1.09, 1.33)	<.001
Diastolic blood pressure	0.83 (0.70, 0.98)	0.025	0.90 (0.81, 1.00)	0.045	0.89 (0.81, 0.99)	0.035
Hypertension medication	1.36 (1.06, 1.73)	0.014	1.38 (1.17, 1.64)	<.001	352.86 (5.54, 2.25e+04)	0.006
Hypertension medication x log(time)					0.28 (0.11, 0.73)	0.009
Current smoking	0.63 (0.45, 0.88)	0.008	2.46e+05 (1.11e+03, 5.45e+07)	<.001	4.86e+03 (112.14, 2.10e+05)	<.001
Current smoking x log(time)			0.06 (0.02, 0.20)	<.001	0.15 (0.06, 0.35)	<.001
Alcohol (elevated vs. optimal)	1.20 (0.76, 1.92)	0.434	1.30 (0.99, 1.70)	0.059	1.32 (1.01, 1.72)	0.045
Body mass index	1.09 (0.99, 1.21)	0.069	172.77 (14.39, 2.07e+03)	<.001	1.20 (1.11, 1.30)	<.001
Body mass index x log(time)			0.31 (0.17, 0.56)	<.001		
Prior diabetes	0.58 (0.40, 0.83)	0.003	0.87 (0.66, 1.13)	0.288	0.87 (0.67, 1.13)	0.296
Prior heart failure	1.17 (0.30, 4.61)	0.819	4.91e+12 (1.57e-04, 1.54e+29)	0.132	1.60 (0.65, 3.93)	0.310
Prior heart failure x log(time)			1.06e-03 (1.11e-07, 10.00)	0.142		
Prior myocardial infarction	0.63 (0.34, 1.16)	0.141	3.55e+11 (2.74e+05, 4.60e+17)	<.001	1.10 (0.76, 1.61)	0.601
Prior myocardial infarction x log(time)			1.92e-03 (6.72e-05, 0.05)	<.001		

SDU: standard deviation unit. Results are differences in lifetime risk and 95% confidence intervals. Differences in lifetime risk for continuous risk factors are the differences for a 1 SDU increase in the covariate from the mean value of the covariate. For the pseudo-observation model and flexible parametric models, differences in lifetime risk were obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are obtained with the delta method. For the Fine-Gray model, differences in lifetime risk were also obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are the .025 and .975 percentile of bootstrap resamples.

Table D.6: Multivariable models for incident death without atrial fibrillation in the Framingham Heart Study, from index age 55 to age 95

			Cumulative incidence function models			
	Pseudo-observation		Fine-Gray		Flexible parametric	
Risk factor at entry age	**OR (95% CI)**	**p**	**SDHR (95% CI)**	**p**	**SDHR (95% CI)**	**p**
Age at entry	0.94 (0.81, 1.10)	0.451	0.20 (0.03, 1.34)	0.098	1.01 (0.96, 1.07)	0.647
Age at entry x log(time)			1.43 (0.94, 2.18)	0.094		
Male	0.8 (0.61, 1.05)	0.112	1.12e+04 (225.23, 5.55e+05)	<.001	1.41e+05 (2484.2, 8.05e+06)	<.001
Male x log(time)			0.12 (0.05, 0.30)	<.001	0.08 (0.03, 0.19)	<.001
Systolic blood pressure	0.87 (0.72, 1.05)	0.142	1.05 (0.97, 1.15)	0.236	5.01 (0.71, 35.13)	0.105
Systolic blood pressure x log(time)					0.71 (0.46, 1.09)	0.117
Diastolic blood pressure	1.31 (1.08, 1.58)	0.007	1.14 (1.05, 1.25)	0.003	1.14 (1.04, 1.24)	0.006
Hypertension medication	0.74 (0.57, 0.97)	0.030	2.81e+03 (27.80, 2.85e+05)	0.001	5.17e+03 (20.06, 1.33e+06)	0.003
Hypertension medication x log(time)			0.15 (0.05, 0.46)	0.001	0.15 (0.04, 0.51)	0.003
Current smoking	2.36 (1.56, 3.57)	0.000	746.44 (16.82, 3.31e+04)	0.001	2.05 (1.82, 2.31)	<.001
Current smoking x log(time)			2.54e-01 (1.05e-01, 0.08)	0.002		
Alcohol (elevated vs. optimal)	0.79 (0.47, 1.35)	0.397	1.07 (0.86, 1.33)	0.532	1.07 (0.86, 1.33)	0.534
Body mass index	1.01 (0.91, 1.12)	0.895	0.94 (0.87, 1.01)	0.072	0.92 (0.86, 0.99)	0.027
Prior diabetes	1.46 (0.90, 2.36)	0.122	1.83 (1.50, 2.24)	<.001	6.20e+04 (718.66, 5.35e+06)	<.001
Prior diabetes x log(time)					0.10 (0.04, 0.27)	<.001
Prior heart failure	1.19 (0.34, 4.18)	0.788	3.12e+20 (1.41e+06, 6.92e+34)	0.005	1.69 (0.82, 3.46)	0.155
Prior heart failure x log(time)			1.39e-05 (4.46e-09, 0.04)	0.006		
Prior myocardial infarction	2.08 (1.06, 4.09)	0.034	5.39e+03 (0.66, 4.43e+07)	0.062	1.85 (1.39, 2.45)	<.001
Prior myocardial infarction x log(time)			0.15 (0.02, 1.29)	0.085		

SDU: standard deviation unit. Results are differences in lifetime risk and 95% confidence intervals. Differences in lifetime risk for continuous risk factors are the differences for a 1 SDU increase in the covariate from the mean value of the covariate. For the pseudo-observation model and flexible parametric models, differences in lifetime risk were obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are obtained with the delta method. For the Fine-Gray model, differences in lifetime risk were also obtained by transforming model coefficients at the mean value of other covariates (LSmeans approach), and 95% CIs are the .025 and .975 percentile of bootstrap resamples.

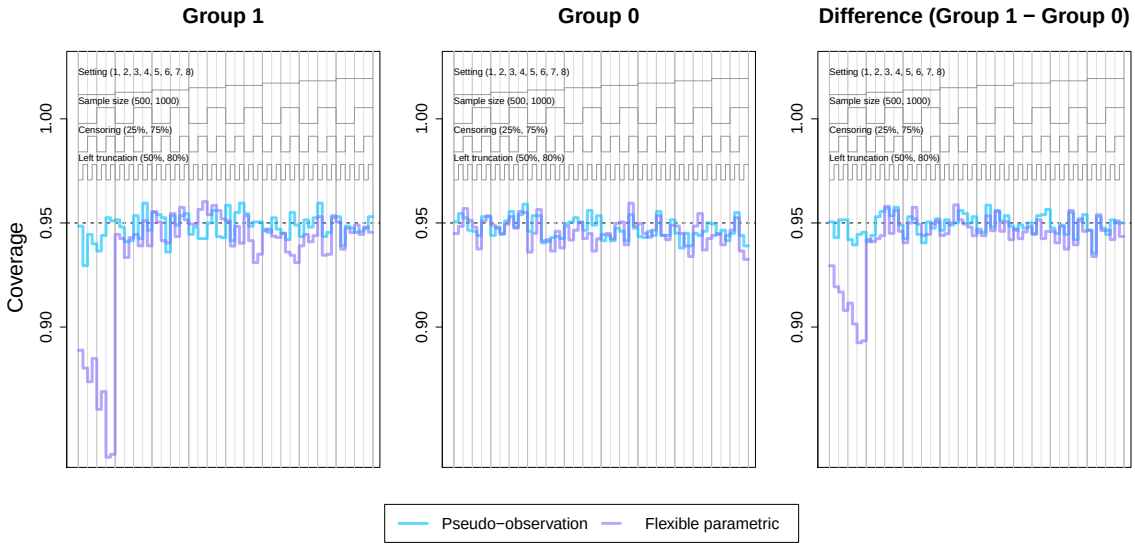


Figure D.3: Nested loop plot showing the simulation study results: coverage

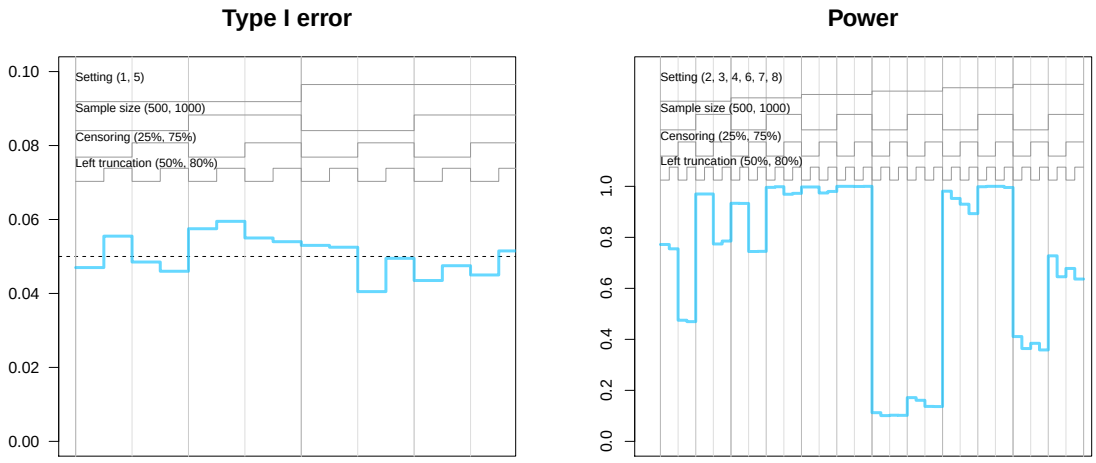


Figure D.4: Nested loop plot showing the simulation study results: type I error and power of pseudo-observation method

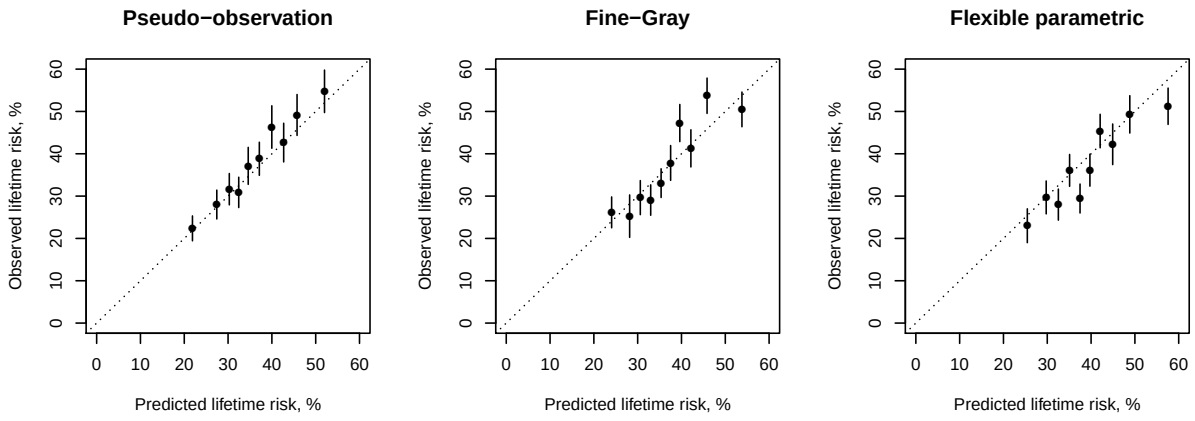


Figure D.5: Calibration of atrial fibrillation models in the Framingham Heart Study

BIBLIOGRAPHY

Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, (pp. 141–150).

Abed, H. S., Wittert, G. A., Leong, D. P., Shirazi, M. G., Bahrami, B., Middeldorp, M. E., Lorimer, M. F., Lau, D. H., Antic, N. A., Brooks, A. G., et al. (2013). Effect of weight reduction and cardiometabolic risk factor management on symptom burden and severity in patients with atrial fibrillation: a randomized clinical trial. *Jama*, *310*(19), 2050–2060.

Allignol, A., Schumacher, M., & Beyersmann, J. (2010). A note on variance estimation of the aalen–johansen estimator of the cumulative incidence function in competing risks, with a view towards left-truncated data. *Biometrical Journal*, *52*(1), 126–137.

Allignol, A., Schumacher, M., Beyersmann, J., et al. (2011). Empirical transition matrix of multi-state models: the etm package. *Journal of Statistical Software*, *38*(4), 1–15.

Alonso, A., Bahnson, J. L., Gaussoin, S. A., Bertoni, A. G., Johnson, K. C., Lewis, C. E., Vetter, M., Mantzoros, C. S., Jeffery, R. W., Soliman, E. Z., et al. (2015). Effect of an intensive lifestyle intervention on atrial fibrillation risk in individuals with type 2 diabetes: the look ahead randomized trial. *American heart journal*, *170*(4), 770–777.

Alonso, A., Krijthe, B. P., Aspelund, T., Stepas, K. A., Pencina, M. J., Moser, C. B., Sinner, M. F., Sotoodehnia, N., Fontes, J. D., Janssens, A. C. J., et al. (2013). Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the charge-af consortium. *Journal of the American Heart Association*, *2*(2), e000102.

Andersen, P. K. (2013). Decomposition of number of life years lost according to causes of death. *Stat Med*, *32*(30), 5278–5285.

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*, chap. IV. Nonparametric Estimation, (pp. 279–280). New York, NY, USA: Springer-Verlag.

Andersen, P. K., Hansen, M. G., & Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal*, *10*(4), 335–350.

Andersen, P. K., Klein, J. P., & Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, *90*(1), 15–27.

Andersen, P. K., & Perme, M. P. (2010). Pseudo-observations in survival analysis. *Stat Methods Med Res*, *19*, 71–99.

Andersen, P. K., & Pohar Perme, M. (2010). Pseudo-observations in survival analysis. *Statistical methods in medical research*, *19*(1), 71–99.

Aune, D., Sen, A., Schlesinger, S., Norat, T., Janszky, I., Romundstad, P., Tonstad, S., Riboli, E., & Vatten, L. J. (2017). Body mass index, abdominal fatness, fat mass and the risk of atrial fibrillation: a systematic review and doseresponse meta-analysis of prospective studies. *European Journal of Epidemiology*, *32*, 181–192.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*, *28*(25), 3083–3107.

Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in medicine*, *32*(16), 2837–2849.

Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine*, *33*(7), 1242–1258.

Austin, P. C. (2016). Variance estimation when using inverse probability of treatment weighting (iptw) with survival analysis. *Statistics in medicine*, *35*(30), 5642–5655.

Austin, P. C., & Fine, J. P. (2019). Propensity-score matching with competing risks in survival analysis. *Statistics in Medicine*, *38*(5), 751–777.

Austin, P. C., Latouche, A., & Fine, J. P. (2020). A review of the use of time-varying covariates in the fine-gray subdistribution hazard competing risk regression model. *Statistics in Medicine*, *39*(2), 103–113.

Bang, H., & Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika*, *87*(2), 329–343.

Bang, H., & Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics*, *58*(3), 643–649.

Beiser, A., D'Agostino Sr, R. B., Seshadri, S., Sullivan, L. M., & Wolf, P. A. (2000). Computing estimates of incidence, including lifetime risk: Alzheimer's disease in the framingham study. the practical incidence estimators (pie) macro. *Statistics in medicine*, *19*(11-12), 1495–1522.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Stat Med*, *24*(11), 1713–1723.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., Floyd, J., Fornage, M., Gillespie, C., Isasi, C., et al. (2017). Heart disease and stroke statistics-2017 update: a report from the american heart association. *circulation*, *135*(10), e146–e603.

Berkovitch, A., Kivity, S., Klempfner, R., Segev, S., Milwidsky, A., Erez, A., Sabbag, A., Goldenberg, I., Sidi, Y., & Maor, E. (2016). Body mass index and the risk of new-onset atrial fibrillation in middle-aged adults. *American Heart Journal*, *173*, 41–48.

Berry, J. D., Dyer, A., Cai, X., Garside, D. B., Ning, H., Thomas, A., Greenland, P., Van Horn, L., Tracy, R. P., & Lloyd-Jones, D. M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, *366*(4), 321–329.

Beyersmann, J., Allignol, A., & Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

Beyersmann, J., Latouche, A., Buchholz, A., & Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in medicine*, *28*(6), 956–971.

Beyersmann, J., & Schumacher, M. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*, *26*(7), 1649.

Blagoev, K. B., Wilkerson, J., & Fojo, T. (2012). Hazard ratios in cancer clinical trials - a primer. *Nat Rev Clin Oncol*, *9*(3), 178–183.

Bodnar, L. M., Davidian, M., Siega-Riz, A. M., & Tsiatis, A. A. (2004). Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *American Journal of Epidemiology*, *159*(10), 926–934.

Brookmeyer, R., & Abdalla, N. (2019). Multistate models and lifetime risk estimation: Application to alzheimer's disease. *Statistics in medicine*, *38*(9), 1558–1565.

Calkins, K. L., Canan, C. E., Moore, R. D., Lesko, C. R., & Lau, B. (2018). An application of restricted mean survival time in a competing risks setting: comparing time to art initiation by injection drug use. *BMC medical research methodology*, *18*(1), 27.

Carone, M., Asgharian, M., & Jewell, N. P. (2014). Estimating the lifetime risk of dementia in the canadian elderly population using cross-sectional cohort survival data. *Journal of the American Statistical Association*, *109*(505), 24–35.

Chen, J., Hou, Y., & Chen, Z. (2018). Statistical inference methods for cumulative incidence function curves at a fixed point in time. *Communications in Statistics-Simulation and Computation*, (pp. 1–16).

Chen, P.-Y., & Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*, *57*(4), 1030–1038.

Choi, S., Kim, C., Zhong, H., Ryu, E.-S., & Han, S. W. (2019). Adjusted-crude-incidence analysis of multiple treatments and unbalanced samples on competing risks. *Statistics and Its Interface*, *12*(3), 423–437.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, *168*(6), 656–664.

Cole, S. R., Lau, B., Eron, J. J., Brookhart, M. A., Kitahata, M. M., Martin, J. N., Mathews, W. C., & Mugavero, M. J. (2015). Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to aids after initiation of antiretroviral therapy. *American Journal of Epidemiology*, *181*(4), 238–245.

Collaboration, H.-C., et al. (2018). Effect of immediate initiation of antiretroviral treatment on the risk of acquired hiv drug resistance. *AIDS (London, England)*, *32*(3), 327.

Conner, S. C., Lodi, S., Lunetta, K. L., Casas, J. P., Lubitz, S. A., Ellinor, P. T., Anderson, C. D., Huang, Q., Coleman, J., White, W. B., et al. (2019a). Refining the association between body mass index and atrial fibrillation: G-formula and restricted mean survival times. *Journal of the American Heart Association*, *8*(16), e013011.

Conner, S. C., Sullivan, L. M., Benjamin, E. J., LaValley, M. P., Galea, S., & Trinquart, L. (2019b). Adjusted restricted mean survival times in observational studies. *Statistics in Medicine*.

Cupples, L. A., D'Agostino, R. B., Anderson, K., & Kannel, W. B. (1988). Comparison of baseline and repeated measure covariate techniques in the framingham heart study. *Statistics in medicine*, *7*(1-2), 205–218.

D'Agostino, R. B., Lee, M.-L., Belanger, A. J., Cupples, L. A., Anderson, K., & Kannel, W. B. (1990). Relation of pooled logistic regression to time dependent cox regression analysis: the framingham heart study. *Statistics in medicine*, *9*(12), 1501–1515.

Danaei, G., Pan, A., Hu, F. B., & Hernán, M. A. (2013). Hypothetical mid-life interventions in women and risk of type 2 diabetes. *Epidemiology (Cambridge, Mass.)*, *24*(1), 122.

Danaei, G., Robins, J. M., Young, J., Hu, F. B., Manson, J. E., & Hernán, M. A. (2016). Estimated effect of weight loss on risk of coronary heart disease and mortality in middle-aged or older women: sensitivity analysis for unmeasured confounding by undiagnosed disease. *Epidemiology (Cambridge, Mass.)*, *27*(2), 302.

Díaz, I., Colantuoni, E., Hanley, D., & Rosenblum, M. (2018). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. (pp. 1–30).

Dinse, G. E., & Larson, M. G. (1986). A note on semi-markov models for partially censored data. *Biometrika*, *73*(2), 379–386.

Du, Y. (2010). *Measuring Effects of Risk Factors on Cumulative Incidence and Remaining Lifetime Risk in the Presence of Competing Risks*. Ph.D. thesis, Boston University.

Edwards, J. K., Cole, S. R., Lesko, C. R., Mathews, W. C., Moore, R. D., Mugavero, M. J., & Westreich, D. (2016). An Illustration of Inverse Probability Weighting to Estimate Policy-Relevant Causal Effects. *American Journal of Epidemiology*, *184*(4), 336–344.

Feinleib, M., Kannel, W. B., Garrison, R. J., McNamara, P. M., & Castelli, W. P. (1975). The framingham offspring study. design and preliminary data. *Preventive medicine*, *4*(4), 518–525.

Feuer, E. J., Wun, L.-M., Boring, C. C., Flanders, W. D., Timmel, M. J., & Tong, T. (1993). The lifetime risk of developing breast cancer. *JNCI: Journal of the National Cancer Institute*, *85*(11), 892–897.

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

Finegold, J. A., Shun-Shin, M. J., Cole, G. D., Zaman, S., Maznyczka, A., Zaman, S., Al-Lamee, R., Ye, S., & Francis, D. P. (2016). Distribution of lifespan gain from primary prevention intervention. *Open Heart*, *3*(1).

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis*, vol. 998. John Wiley & Sons.

Frost, L., Benjamin, E. J., Fenger-Grøn, M., Pedersen, A., Tjønneland, A., & Overvad, K. (2014). Body fat, body fat distribution, lean body mass and atrial fibrillation and flutter. a danish cohort study. *Obesity*, *22*(6), 1546–1552.

Garcia-Aymerich, J., Varraso, R., Danaei, G., Camargo Jr, C. A., & Hernán, M. A. (2014). Incidence of adult-onset asthma after hypothetical interventions on body

mass index and physical activity: an application of the parametric g-formula. *American journal of epidemiology*, *179*(1), 20–26.

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Straus, D. J., Clarkson, B. D., & Brennan, M. F. (1993). On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association*, *88*(422), 400–409.

Genolini, C., Lacombe, A., Écochard, R., & Subtil, F. (2016). Copymean: a new method to predict monotone missing values in longitudinal studies. *Computer methods and programs in biomedicine*, *132*, 29–44.

Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, *67*(1), 39–49.

Gifford Jr, R. (1993). The fifth report of the joint national committee on detection, evaluation, and treatment of high blood pressure: insights and highlights from the chairman. *Cleveland Clinic journal of medicine*, *60*(4), 273–277.

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*(3), 515–526.

Grand, M. K., Putter, H., Allignol, A., & Andersen, P. K. (2019). A note on pseudo-observations and left-truncation. *Biometrical Journal*, *61*(2), 290–298.

Graw, F., Gerds, T. A., & Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis*, *15*(2), 241–255.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, (pp. 1141–1154).

Green, M. S., & Symons, M. J. (1983). A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *Journal of chronic diseases*, *36*(10), 715–723.

Hagiwara, Y., Shinozaki, T., & Matsuyama, Y. (2019). G-estimation of structural nested restricted mean time lost models to estimate effects of time-varying treatments on a failure time outcome. *Biometrics*.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology*, *21*(1), 13.

Hernán, M. A., & Taubman, S. L. (2008). Does obesity shorten life? the importance of well-defined interventions to answer causal questions. *International journal of obesity*, *32*(3), S8–S14.

Hinchliffe, S. R., & Lambert, P. C. (2013). Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC medical research methodology*, *13*(1), 13.

Hindricks, G., Potpara, T., Dagres, N., Arbelo, E., Bax, J. J., Blomström-Lundqvist, C., Boriani, G., Castella, M., Dan, G.-A., Dilaveris, P. E., et al. (2020). 2020 esc guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the european association of cardio-thoracic surgery (eacts) the task force for the diagnosis and management of atrial fibrillation of the european society of cardiology (esc) developed with the special contribution of the european heart rhythm association (ehra) of the esc. *European Heart Journal*.

Jackson, J. W., Schmid, I., & Stuart, E. A. (2017). Propensity scores in pharmacoepidemiology: Beyond the horizon. *Current Epidemiology Reports*, (pp. 1–10).

Jacobsen, M., & Martinussen, T. (2016). A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. *Scandinavian Journal of Statistics*, *43*(3), 845–862.

Jain, P., Danaei, G., Robins, J. M., Manson, J. E., & Hernán, M. A. (2016). Smoking cessation and long-term weight gain in the framingham heart study: an application of the parametric g-formula for a continuous outcome. *European journal of epidemiology*, *31*(12), 1223–1229.

Jamaly, S., Carlsson, L., Peltonen, M., Jacobson, P., Sjöström, L., & Karason, K. (2016). Bariatric surgery and the risk of new-onset atrial fibrillation in swedish obese subjects. *Journal of the American College of Cardiology*, *68*(23), 2497–2504.

Jeong, J.-H., & Fine, J. (2006). Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *55*(2), 187–200.

Jeong, J.-H., & Fine, J. (2009). A note on cause-specific residual life. *Biometrika*, *96*(1), 237–242.

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J. (1961). Factors of risk in the development of coronary heart diseasesix-year follow-up experience: the framingham study. *Annals of internal medicine*, *55*(1), 33–50.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc*, *53*(282), 457–481.

Karmali, K. N., & Lloyd-Jones, D. M. (2013). Adding a life-course perspective to cardiovascular-risk communication. *Nature Reviews Cardiology*, *10*(2), 111.

Karrison, T. (1987). Restricted mean life with adjustment for covariates. *J Am Stat Assoc*, *82*(400), 1169–1176.

Keil, A. P., Edwards, J. K., Richardson, D. R., Naimi, A. I., & Cole, S. R. (2014). The parametric g-formula for time-to-event data: towards intuition with a worked example. *Epidemiology (Cambridge, Mass.)*, *25*(6), 889.

Khan, S. S., Ning, H., Wilkins, J. T., Allen, N., Carnethon, M., Berry, J. D., Sweis, R. N., & Lloyd-Jones, D. M. (2018). Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA cardiology*, *3*(4), 280–287.

Kipourou, D.-K., Perme, M. P., Rachet, B., & Belot, A. (2020). Direct modeling of the crude probability of cancer death and the number of life years lost due to cancer without the need of cause of death: a pseudo-observation approach in the relative survival setting. *Biostatistics*.

Klein, J. P., & Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, *61*(1), 223–229.

Klein, J. P., Gerster, M., Andersen, P. K., Tarima, S., & Perme, M. P. (2008). Sas and r functions to compute pseudo-values for censored data regression. *Comput Methods Programs Biomed*, *89*(3), 289–300.

Klein, J. P., Logan, B., Harhoff, M., & Andersen, P. K. (2007). Analyzing survival curves at a fixed point in time. *Statistics in medicine*, *26*(24), 4505–4519.

Kulathinal, S., & Gasbarra, D. (2002). Testing equality of cause-specific hazard rates corresponding to m competing risks among k groups. *Lifetime Data Analysis*, *8*(2), 147–161.

Lajous, M., Willett, W. C., Robins, J., Young, J. G., Rimm, E., Mozaffarian, D., & Hernán, M. A. (2013). Changes in fish consumption in midlife and the risk of coronary heart disease in men and women. *American journal of epidemiology*, *178*(3), 382–391.

Lambert, P. C., Wilkes, S. R., & Crowther, M. J. (2017). Flexible parametric modelling of the cause-specific cumulative incidence function. *Statistics in medicine*, *36*(9), 1429–1446.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., & Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology*, *66*(6), 648–653.

Latouche, A., Boisson, V., Chevret, S., & Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in medicine*, *26*(5), 965–974.

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, *113*(521), 390–400.

Li, F., Thomas, L. E., & Li, F. (2019). Addressing extreme propensity scores via the overlap weights. *American journal of epidemiology*, *188*(1), 250–257.

Li, J., Scheike, T. H., & Zhang, M.-J. (2015). Checking fine and gray subdistribution hazards model with cumulative sums of residuals. *Lifetime Data Analysis*, *21*(2), 197–217.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Litwin, S. E. (2010). Cardiac remodeling in obesity: time for a new paradigm.

Lloyd-Jones, D. M., Leip, E. P., Larson, M. G., dAgostino, R. B., Beiser, A., Wilson, P., Wolf, P. A., & Levy, D. (2006). Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation*, *113*(6), 791–798.

Lloyd-Jones, D. M., Wang, T. J., Leip, E. P., Larson, M. G., Levy, D., Vasan, R. S., DAgostino, R. B., Massaro, J. M., Beiser, A., Wolf, P. A., et al. (2004). Lifetime risk for development of atrial fibrillation: the framingham heart study. *Circulation*, *110*(9), 1042–1046.

Lodi, S., Phillips, A., Logan, R., Olson, A., Costagliola, D., Abgrall, S., van Sighem, A., Reiss, P., Miró, J. M., Ferrer, E., et al. (2015). Comparative effectiveness of immediate antiretroviral therapy versus cd4-based initiation in hiv-positive individuals in high-income countries: observational cohort study. *The lancet HIV*, *2*(8), e335–e343.

Logan, B. R., Zhang, M.-J., & Klein, J. P. (2011). Marginal models for clustered time-to-event data with competing risks using pseudovalues. *Biometrics*, *67*(1), 1–7.

Logan, R., Young, J., Taubman, S., Lodi, S., Picciotto, S., Danaei, G., & Hernán, M. (2018). Gformula sas macro version 3.0. *Harvard University*.

Lok, J. J., Yang, S., Sharkey, B., & Hughes, M. D. (2018). Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms. *Lifetime Data Analysis*, *24*(2), 201–223.

Magnani, J. W., Hylek, E. M., & Apovian, C. M. (2013). Obesity begets atrial fibrillation: a contemporary summary. *Circulation*, *128*(4), 401–405.

Magnussen, C., Niiranen, T. J., Ojeda, F. M., Gianfagna, F., Blankenberg, S., Njølstad, I., Vartiainen, E., Sans, S., Pasterkamp, G., Hughes, M., et al. (2017). Sex differences and similarities in atrial fibrillation epidemiology, risk factors, and mortality in community cohorts: results from the biomarcare consortium (biomarker for cardiovascular risk assessment in europe). *Circulation*, *136*(17), 1588–1597.

Mahajan, R., Nelson, A., Pathak, R. K., Middeldorp, M. E., Wong, C. X., Twomey, D. J., Carbone, A., Teo, K., Agbaedeng, T., Linz, D., et al. (2018). Electroanatomical remodeling of the atria in obesity: impact of adjacent epicardial fat. *JACC: Clinical Electrophysiology*, *4*(12), 1529–1540.

Misayaka, Y., Barnes, M., Gersh, B., Cha, S., Bailey, K., Abhayaratna, W., et al. (2006). Secular trends in incidence of atrial fibrillation in olmsted county minnesota, 1980 to 2000 and implications on the projection for future prevalence. *Circulation*, *114*, 119–25.

Moreno-Betancur, M., & Latouche, A. (2013). Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-values. *Statistics in medicine*, *32*(18), 3206–3223.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in medicine*.

Mozumder, S. I., Lambert, P., & Rutherford, M. (2020). Estimating restricted mean survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models. https://www.researchsquare.com/article/rs-14402/v1.

Mozumder, S. I., Rutherford, M., & Lambert, P. (2018). Direct likelihood inference on the cause-specific cumulative incidence function: A flexible parametric regression modelling approach. *Statistics in Medicine*, *37*(1), 82–97.

Muñoz, A., Abraham, A. G., Matheson, M., & Wada, N. (2013). Non-proportionality of hazards in the competing risks framework. In *Risk Assessment and Evaluation of Predictions*, (pp. 3–22). Springer.

Naimi, A. I., Cole, S. R., & Kennedy, E. H. (2017). An introduction to g methods. *International journal of epidemiology*, *46*(2), 756–762.

Nascimento, G. G., Peres, M. A., Mittinty, M. N., Peres, K. G., Do, L. G., Horta, B. L., Gigante, D. P., Corrêa, M. B., & Demarco, F. F. (2017). Diet-induced overweight and obesity and periodontitis risk: an application of the parametric g-formula in the 1982 pelotas birth cohort. *American Journal of Epidemiology*, *185*(6), 442–451.

Neumann, A., & Billionnet, C. (2016). Covariate adjustment of cumulative incidence functions for competing risks data using inverse probability of treatment weighting. *Computer Methods and Programs in Biomedicine*, *129*, 63–70.

Nevalainen, J., Kenward, M. G., & Virtanen, S. M. (2009). Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Statistics in medicine*, *28*(29), 3657–3669.

Ngwa, J. S., Cabral, H. J., Cheng, D. M., Pencina, M. J., Gagnon, D. R., LaValley, M. P., & Cupples, L. A. (2016). A comparison of time dependent cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the framingham heart study. *BMC medical research methodology*, *16*(1), 148.

Ning, Y., McAvay, G., Chaudhry, S. I., Arnold, A. M., & Allore, H. G. (2013). Results differ by applying distinctive multiple imputation approaches on the longitudinal cardiovascular health study data. *Experimental aging research*, *39*(1), 27–43.

Norby, F. L., Soliman, E. Z., Chen, L. Y., Bengtson, L. G., Loehr, L. R., Agarwal, S. K., & Alonso, A. (2016). Trajectories of cardiovascular risk factors and incidence of atrial fibrillation over a 25-year follow-up: the aric study (atherosclerosis risk in communities). *Circulation*, *134*(8), 599–610.

Overgaard, M., Parner, E. T., & Pedersen, J. (2018). Estimating the variance in a pseudo-observation scheme with competing risks. *Scandinavian Journal of Statistics*, *45*(4), 923–940.

Overgaard, M., Parner, E. T., Pedersen, J., et al. (2017). Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations. *The Annals of Statistics*, *45*(5), 1988–2015.

Pan, Q., & Gastwirth, J. L. (2013). Estimating restricted mean job tenures in semi-competing risk data compensating victims of discrimination. *The Annals of Applied Statistics*, (pp. 1474–1496).

Pathak, R. K., Middeldorp, M. E., Lau, D. H., Mehta, A. B., Mahajan, R., Twomey, D., Alasady, M., Hanley, L., Antic, N. A., McEvoy, R. D., et al. (2014). Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the arrest-af cohort study. *Journal of the American College of Cardiology*, *64*(21), 2222–2231.

Pavlič, K., Martinussen, T., & Andersen, P. K. (2019). Goodness of fit tests for estimating equations based on pseudo-observations. *Lifetime Data Analysis*, *25*(2), 189–205.

Pencina, M. J., Larson, M. G., & D'Agostino, R. B. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine*, *26*(6), 1343–1359.

Rahman, F., Yin, X., Larson, M. G., Ellinor, P. T., Lubitz, S. A., Vasan, R. S., McManus, D. D., Magnani, J. W., & Benjamin, E. J. (2016). Trajectories of risk factors and risk of new-onset atrial fibrillation in the framingham heart study. *Hypertension*, *68*(3), 597–605.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Mathematical modelling*, *7*(9-12), 1393–1512.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

Rosengren, A., Hauptman, P. J., Lappas, G., Olsson, L., Wilhelmsen, L., & Swedberg, K. (2009). Big men and atrial fibrillation: effects of body size and weight gain on risk of atrial fibrillation in men. *European heart journal*, *30*(9), 1113–1120.

Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*, *21*(15), 2175–2197.

Royston, P., & Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med*, *30*(19), 2409–2421.

Royston, P., & Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*, *13*(1), 152.

Royston, P., et al. (2015). Estimating the treatment effect in a clinical trial using difference in restricted mean survival time. *Stata J*, *15*(4), 1098–1117.

Rücker, G., & Schwarzer, G. (2014). Presenting simulation results in a nested loop plot. *BMC medical research methodology*, *14*(1), 129.

Rutherford, M. J., Crowther, M. J., & Lambert, P. C. (2015). The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, *85*(4), 777–793.

Sabathé, C., Andersen, P. K., Helmer, C., Gerds, T. A., Jacqmin-Gadda, H., & Joly, P. (2020). Regression analysis in an illness-death model with interval-censored

data: A pseudo-value approach. *Statistical Methods in Medical Research*, *29*(3), 752–764.

Sachs, M. C., Discacciati, A., Everhov, Å. H., Olén, O., & Gabriel, E. E. (2019). Ensemble prediction of time-to-event outcomes with competing risks: a case-study of surgical complications in crohn's disease. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(5), 1431–1446.

Sauerbrei, W., Royston, P., & Look, M. (2007). A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal*, *49*(3), 453–473.

Schaubel, D. E., & Wei, G. (2011). Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics*, *67*(1), 29–38.

Scheike, T. H., Zhang, M.-J., & Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, *95*(1), 205–220.

Schnabel, R. B., Sullivan, L. M., Levy, D., Pencina, M. J., Massaro, J. M., D'Agostino Sr, R. B., Newton-Cheh, C., Yamamoto, J. F., Magnani, J. W., Tadros, T. M., et al. (2009). Development of a risk score for atrial fibrillation (framingham heart study): a community-based cohort study. *The Lancet*, *373*(9665), 739–745.

Schomaker, M., & Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in medicine*, *37*(14), 2252–2266.

Seshadri, S., Beiser, A., Kelly-Hayes, M., Kase, C. S., Au, R., Kannel, W. B., & Wolf, P. A. (2006). The lifetime risk of stroke: estimates from the framingham study. *Stroke*, *37*(2), 345–350.

Shakiba, M., Mansournia, M. A., Salari, A., Soori, H., Mansournia, N., & Kaufman, J. S. (2018). Accounting for time-varying confounding in the relationship between obesity and coronary heart disease: analysis with g-estimation: the aric study. *American journal of epidemiology*, *187*(6), 1319–1326.

Staerk, L., Preis, S. R., Lin, H., Casas, J. P., Lunetta, K., Weng, L.-C., Anderson, C. D., Ellinor, P. T., Lubitz, S. A., Benjamin, E. J., et al. (2020). Novel risk modeling approach of atrial fibrillation with restricted mean survival times: Application in the framingham heart study community-based cohort. *Circulation: Cardiovascular Quality and Outcomes*, *13*(4), e005918.

Staerk, L., Wang, B., Lunetta, K. L., Helm, R. H., Ko, D., Sherer, J. A., Ellinor, P. T., Lubitz, S. A., McManus, D. D., Vasan, R. S., et al. (2017). Association between leukocyte telomere length and the risk of incident atrial fibrillation: the framingham heart study. *Journal of the American Heart Association*, *6*(11), e006541.

Staerk, L., Wang, B., Preis, S. R., Larson, M. G., Lubitz, S. A., Ellinor, P. T., McManus, D. D., Ko, D., Weng, L.-C., Lunetta, K. L., et al. (2018). Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. *British Medical Journal*, *361*, k1453.

Stegherr, R., Allignol, A., Meister, R., Schaefer, C., & Beyersmann, J. (2020). Estimating cumulative incidence functions in competing risks data with dependent left-truncation. *Statistics in Medicine*, *39*(4), 481–493.

Steingrimsson, J. A., Diao, L., Molinaro, A. M., & Strawderman, R. L. (2016). Doubly robust survival trees. *Statistics in medicine*, *35*(20), 3595–3612.

Taubman, S. L., Robins, J. M., Mittleman, M. A., & Hernán, M. A. (2009). Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International journal of epidemiology*, *38*(6), 1599–1611.

Tian, L., Jin, H., Uno, H., Lu, Y., Huang, B., Anderson, K. M., & Wei, L. (2020). On the empirical choice of the time window for restricted mean survival time. *Biometrics*.

Tian, L., Zhao, L., & Wei, L. (2014). Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*, *15*(2), 222–233.

Trinquart, L., Jacot, J., Conner, S. C., & Porcher, R. (2016). Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol*, *34*(15), 1813–1819.

Trinquart, L., Jacot, J., Conner, S. C., & Porcher, R. (2017). Reply to KH Eng et al. *Journal of Clinical Oncology*, *35*(4), 466.

Ueda, P., Gulayin, P., & Danaei, G. (2018). Long-term moderately elevated ldl-cholesterol and blood pressure and risk of coronary heart disease. *PloS one*, *13*(7), e0200017.

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L., et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*, *32*(22), 2380–2385.

Vangen-Lønne, A. M., Ueda, P., Gulayin, P., Wilsgaard, T., Mathiesen, E. B., & Danaei, G. (2018). Hypothetical interventions to prevent stroke: an application of the parametric g-formula to a healthy middle-aged population. *European journal of epidemiology*, *33*(6), 557–566.

Vinter, N., Huang, Q., Fenger-Gron, M., Frost, L., Benjamin, E., & Trinquart, L. (2020). 45-year trend in excess mortality associated with atrial fibrillation in the community-based Framingham Heart Study cohorts. *BMJ*.

Wang, T. J., Parise, H., Levy, D., DAgostino, R. B., Wolf, P. A., Vasan, R. S., & Benjamin, E. J. (2004). Obesity and the risk of new-onset atrial fibrillation. *Jama*, *292*(20), 2471–2477.

Wang, X., & Schaubel, D. E. (2018). Modeling restricted mean survival time under general censoring mechanisms. *Lifetime data analysis*, *24*(1), 176–199.

Wei, G. (2008). *Semiparametric methods for estimating cumulative treatment effects in the presence of non-proportional hazards and dependent censoring*. Ph.D. thesis, University of Michigan.

Weir, I., Marshall, G., Schneider, J., Sherer, J., Lord, E., Gyawali, B., Paasche-Orlow, M., Benjamin, E., & Trinquart, L. (2019). Interpretation of time-to-event outcomes in randomized trials: an online randomized experiment. *Annals of Oncology*, *30*(1), 96–102.

Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, *97*(18), 1837–1847.

Witteman, J. C., D'Agostino, R. B., Stijnen, T., Kannel, W. B., Cobb, J. C., de Ridder, M. A., Hofman, A., & Robins, J. M. (1998). G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the framingham heart study. *American Journal of Epidemiology*, *148*(4), 390–401.

Wong, C. X., Sullivan, T., Sun, M. T., Mahajan, R., Pathak, R. K., Middeldorp, M., Twomey, D., Ganesan, A. N., Rangnekar, G., Roberts-Thomson, K. C., et al. (2015). Obesity and the risk of incident, post-operative, and post-ablation atrial fibrillation: a meta-analysis of 626,603 individuals in 51 studies. *JACC: clinical electrophysiology*, *1*(3), 139–152.

Xie, J., & Liu, C. (2005). Adjusted kaplan–meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*, *24*(20), 3089–3110.

Zhang, M., & Schaubel, D. E. (2011). Estimating differences in restricted mean lifetime using observational data subject to dependent censoring. *Biometrics*, *67*(3), 740–749.

Zhang, M., & Schaubel, D. E. (2012a). Contrasting treatment-specific survival using double-robust estimators. *Statistics in medicine*, *31*(30), 4255–4268.

Zhang, M., & Schaubel, D. E. (2012b). Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics*, *68*(4), 999–1009.

Zhao, L., Claggett, B., Tian, L., Uno, H., Pfeffer, M. A., Solomon, S. D., Trippa, L., & Wei, L. (2016). On the restricted mean survival time curve in survival analysis. *Biometrics*, *72*(1), 215–221.

Zhao, L., Tian, L., Claggett, B., Pfeffer, M., Kim, D. H., Solomon, S., & Wei, L.-J. (2018). Estimating treatment effect with clinical interpretation from a comparative clinical trial with an end point subject to competing risks. *JAMA Cardiology*, *3*(4), 357–358.

Zhou, B., Fine, J., & Laird, G. (2013). Goodness-of-fit test for proportional subdistribution hazards model. *Statistics in Medicine*, *32*(22), 3804–3811.

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. *J Am Stat Assoc*, *93*(442), 702–709.

CURRICULUM VITAE