Mediation Analysis with a Survival Mediator:

A Simulation Study of Different Indirect Effect Testing Methods

by

Han Joe Kim

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2017 by the
Graduate Supervisory Committee:

David P. MacKinnon, Co-chair
Jenn-Yun Tein, Co-chair
Stephen G. West
Kevin J. Grimm

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

Time-to-event analysis or equivalently, survival analysis deals with two variables simultaneously: *when* (time information) an event occurs and *whether* an event occurrence is observed or not during the observation period (censoring information). In behavioral and social sciences, the event of interest usually does not lead to a terminal state such as death. Other outcomes after the event can be collected and thus, the survival variable can be considered as a predictor as well as an outcome in a study. One example of a case where the survival variable serves as a predictor as well as an outcome is a survival-mediator model. In a single survival-mediator model an independent variable, X predicts a survival variable, M which in turn, predicts a continuous outcome, Y. The survival-mediator model consists of two regression equations: X predicting M (M-regression), and M and X simultaneously predicting Y (Y-regression). To estimate the regression coefficients of the survival-mediator model, Cox regression is used for the M-regression. Ordinary least squares regression is used for the Y-regression using complete case analysis assuming censored data in M are missing completely at random so that the Y-regression is unbiased. In this dissertation research, different measures for the indirect effect were proposed and a simulation study was conducted to compare performance of different indirect effect test methods. Bias-corrected bootstrapping produced high Type I error rates as well as low parameter coverage rates in some conditions. In contrast, the Sobel test produced low Type I error rates as well as high parameter coverage rates in some conditions. The bootstrap of the natural indirect effect produced low Type I error and low statistical power when the censoring proportion was non-zero. Percentile

bootstrapping, distribution of the product and the joint-significance test showed best performance. Statistical analysis of the survival-mediator model is discussed. Two indirect effect measures, the ab-product and the natural indirect effect are compared and discussed. Limitations and future directions of the simulation study are discussed. Last, interpretation of the survival-mediator model for a made-up empirical data set is provided to clarify the meaning of the quantities in the survival-mediator model.

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor and co-chair, Dr. David MacKinnon and my other co-chair, Dr. Jenn-Yun Tein, and the rest of my committee: Dr. Stephen West and Dr. Kevin Grimm for their endless support, patience and immense knowledge. Their guidance helped me with research and writing of this dissertation. I could not have imagined having bettor mentors for my Ph.D study.

Other than my committee, I sincerely thank all the rest of the current and past faculty members of the quantitative program at Arizona State University: Dr. Leona Aiken, Dr. Roger Millsap, Dr. Craig Enders, Dr. Hye Won Suk, and Dr. Michael Edwards. All of them provided me with the best quantitative education I could have in the world and the knowledge that I have learned from them has enabled me to finish this dissertation. I express my special thanks to my former advisor, Dr. Roger Millsap for accepting me as a student six years ago and for providing me numerous insights.

My sincere thanks also goes to the professors, researchers and staff at the prevention research center (now called as the "REACH" institute) at Arizona State University: Dr. Thomas Dishion, Dr. Irwin Sandler, Dr. Nancy Gonzalez, Dr. Sharlene Wolchik and many others. I especially thank Dr. Dishion for providing me graduate support for the first three years of my graduate education. Thanks to also Dr. Jan Hughes at Texas A&M University and Dr. Steve West for their current graduate support.

Thanks to all the graduate student friends and colleagues for the endless support, for the stimulating discussions and advices, and for all the fun we had together during the last six years.

I would like to thank my family: my parents and my sister for supporting me both mentally and financially throughout my graduate study. Last but not the least, thanks to my beloved and proud wife, Dr. Ahyeon Yun who endured and survived the experience of graduate school together and provided me with endless encouragement and support.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

INTRODUCTION

Survival analysis is a statistical method that focuses on the timing of an event. Survival analysis was first developed in biomedical and engineering research and traditionally focused on events such as death or failure of a system or a machine. Naturally, the timing variable is a dependent variable and researchers have developed models to predict the event timing with a set of variables. The most widely used survival model is the Cox proportional hazards model (Cox, 1972) where the log of the hazard rate is modeled as a linear combination of a set of predictors. More recently, the application of survival models has broadened to social science (*e.g.*, economics, sociology and psychology) research where the event of interest is not necessary a terminal state, such as, reemployment (Kiefer, 1998; Lancaster & Nickell, 1979; Lancaster, 1980), recidivism (Sherman & Berk, 1984) and gazing behavior in couples (Gardner & Griffin, 1989; Gardner, 1993). However, in spite of the various types of events, survival analyses are still associated with research where the timing variable is an outcome. This paper investigates a situation where the timing variable can both be a predictor as well as an outcome in a mediation model.

Mediation analysis is a crucial research methodology in psychology and many other social science areas. Mediation represents a hypothetical relationship where one variable (independent variable) affects a second variable (mediator variable) and in turn, affects a third variable (dependent variable). Statistical methods for mediation analysis have been actively developed and widely used in substantive research throughout the last three decades. Various combinations of a predictor-mediator-outcome single mediator model have been studied where the three variables in the model can be a combination of

categorical variables or continuous variables (Lacobucci, 2012; MacKinnon, 2008). Single mediator models including a survival outcome has also been studied including recent statistical developments on causal mediation analysis (Lange & Hansen, 2011; VanderWeele, 2011). In other applications, the survival variable can be used as a mediator in the mediation model. However, less investigation has been made on mediation models with a survival variable as the mediator. This dissertation investigates different methods to evaluate mediated effects for a single survival mediator model, through a simulation study.

The first chapter is a literature review chapter and consists of three sections, 1) survival analysis, 2) mediation analysis, and 3) survival mediation analysis. The first section (survival analysis) focuses on the review of traditional survival analysis literature. The basic concepts in survival analysis are explained and different statistical models that include or do not include covariates for survival data are reviewed. The Cox model is discussed in depth since not only is the Cox model the most widely used model in the survival analysis literature but it is also the model used in this study for the survival mediator model. The second section (mediation analysis) gives a brief review of mediation analysis. This section focuses on a single mediator model and explains different methods to test the mediated effect. The third section (survival mediation analysis) discusses single mediator models with survival outcomes and with survival mediators. The focus of this section is to discuss the survival mediator model and identify the challenges of estimating and testing the model parameters and the mediated effect. In the second chapter, a simulation study to investigate the performance of different methods to test the mediated effect is introduced. The third chapter presents the

results of the simulation study. The last chapter discusses about the statistical model of the survival mediator model, the different measures of the indirect effect, the simulation study results and provide general guidelines for analysis of survival mediator models.

I. LITERATURE REVIEW

1. Survival Analysis

Survival analysis is a branch of statistical methods that deal with time-to-event variables (Allison, 2010). Traditionally, survival analysis emerged from biomedical and engineering research. The wording "survival" comes from biomedical research where the event of interest is the death of biological organisms (*e.g.*, Cox, 1972; Klein & Moeschberger, 2003). Survival analysis has been applied to other fields with different terms such as failure time analysis (or 'reliability analysis' to match with the more positive term 'survival analysis' rather than 'death analysis') in engineering, duration analysis in economics, and event history analysis in sociology research. For example, engineers are interested in when a machine (also could be a product or controller) starts to malfunction in order to predict and prevent any accidents and minimize the risk of loss (*e.g.*, Beason, & Morgan, 1984; Kalbfleisch & Prentice, 2002; Zhai & Lin, 2004). One of the problems that economists are interested in is the duration of unemployment; such that duration analysis was used to identify explanatory variables related to unemployment interval (*e.g.*, Kiefer, 1998; Lancaster & Nickell, 1979; Lancaster, 1980).

The history of survival analysis in psychology has been relatively shorter (it began in the early 1970s) than other fields but it is rapidly growing and recently innovative methods have been proposed. Singer and Willett (1991, 1993, 2003) provide a good overview of survival analysis in psychology with various examples. Some of the examples in psychology are related to investigating the effectiveness of a smoking relapse prevention program (Stevens, & Hollis, 1989), studying the relapse time of affective disorders (Lavori, Keller, & Klerman, 1984), examining the age at first onset of

an affective illness (Rice et al., 1987), modeling the instantaneous rate of absence at work (Finchman, 1989), and studying factors that influence the age at first placement of non-parental child care (Singer, Fuller, Keiley & Wolf, 1998). Another rapidly growing area of application of survival models in psychology is with observational data. Early studies of parallel streams of observational data by Gardner and Griffin had used continuous-time sequential analysis to study gazing behavior between married couples (Gardner & Griffin, 1989). In another study, a survival regression model was used to study the interaction between mother and child (Griffin & Gardner, 1989). More recent applications use observed parent-child interaction data to model the duration time in a particular behavioral or emotional state (Dagne & Snyder, 2011; Stoolmiller & Snyder, 2006; Stoolmiller & Snyder, 2013).

Although there are many applications and related methods across different literatures, a common fact in survival analysis is interest in a well-defined discrete event. That is, the event may occur or not at a certain time point. The event can be death, failure of a system, unemployment, or observation of a specific behavior. If we were just interested in the occurrence of these events, we can quantify these events into "0" (non-occurrence) or "1" (occurrence) and use statistical methods to analyze binary data such as logistic regression (Hosmer, Lemeshow, & Sturdivant, 2013). However, the *event occurrence* itself is not the only thing that is considered in survival analysis. Survival analysis also focuses on the *timing* of an event. For example, the time until the moment of death or failure of a system can be studied. In other studies, the duration of an unemployment spell or the time interval of showing a particular behavior can be studied.

As suggested above, there are two variables that are of interest in survival analysis: *event occurrence* and *timing*. The next two sections explain the two variables.

A. Event Occurrence and Censoring

In survival analysis, an event of interest should be a discrete event that is clearly defined by the researcher. For example, if the event of interest is death, the organism must be dead or alive. There cannot be a state such as "mostly dead" or "barely alive". One of the two main variables in survival analysis is an event occurrence indicator. An interesting feature of the event occurrence indicator in survival analysis is that an event can either have occurred or not observed. That is, rather than establishing that an event has occurred or not, the event is coded as occurred if the event clearly has happened during the observation period and coded as "censored" if the event has not been observed during the observation period.

There are two ways of categorizing censoring types. First, the censoring type can be categorized by considering whether the investigator *can control or cannot control* the determinant of censoring, and *what* is controlled by the investigator (Allison, 2010). Accordingly, there are three types of censoring: Type I, Type II and random censoring. *Type I censoring* is caused by the investigator's pre-specified research period. Every study has a finite timeframe and usually the investigator determines when the study ends. The end time naturally draws a line for censoring. If the event of interest did not occur before the end of the research period, the data are censored. *Type II censoring* is another situation where the investigator plays a role in determining the criterion for censoring. Type II censoring occurs when the investigator controls the total number of events rather

than controlling the research period. For example, if there were 100 patients and the investigator ended the research when 10% of the patients (*i.e.,* 10 patients) died, the rest of the patient's data will be censored. Last, *random censoring* occurs when the investigator cannot control the censoring factor. For example, participants can suddenly drop out of the study because of personal reasons and thus, their data will be censored. In other cases, random censoring occurs because of other competing events than the event of interest. For example, say that we are interested in death caused by lung cancer. If a participant died because of an illness (*e.g.*, ebola virus) other than lung cancer, the survival time should be treated as censored in this study. Allison (2010) describes that while statistical methods (*i.e.*, maximum likelihood and partial likelihood) can handle the Type I and Type II censoring to produce unbiased survival estimates, they cannot handle random censoring. Random censoring can cause severe biases. It is difficult to know the magnitude or direction of the bias will be. Therefore, the best solution to minimize the bias is to reduce random censoring as much as possible with careful study design and by including all the covariates that can possibly account for censoring to lessen the bias.

Another way to categorize censoring types is *when* the censoring occurs. There are three types of censoring that are considered in the conventional survival analysis context (Klein & Moeschberger, 2003). *Left censoring* occurs when the event of interest has already happened before the research has started. For example, in unemployment spell studies (Kiefer, 1998; Lancaster & Nickell, 1979; Lancaster, 1980), the event of interest is observing the end of unemployment (*i.e.*, reemployment). To be reemployed, one must be unemployed. There might be a person who already has been reemployed before the study. This person's reemployment data is censored since it occurred before

the study. Left censoring is conceptually different from *truncation* where for truncation we do not have data on subjects because the research is restricted to a fixed observational time window. In contrast to censoring, data truncation is deliberate and caused by study design. For example, in the reemployment study, the researchers might only want to consider participant's data that are during an economic depression period. Before the depression, people's employment status is not considered and thus the data is left truncated for these people.

Right censoring occurs when the event happens after a certain time point but we do not observe the event. Usually, right censoring occurs because of the end of the study and therefore the event is not observed during the study. In the unemployment example, this can be the people that remained unemployed during the study but they might have been reemployed sometime after the study. Right censoring is the most usual and most frequent type of censoring encountered in research. Another type of censoring is *interval censoring* which concerns censoring during a study. Sometimes the occurrence of an event may not be observed since the study participant has been lost from the study for a period. In the unemployment example, some people may refuse to answer whether they were reemployed or not for a particular period. In this case, we do not know whether they have or have not been reemployed and perhaps unemployed again for that particular period.

More recently, another type of censoring has been proposed named *competing risk censoring* (Stoolmiller, 2014). In some studies, there can be competing events that are present at the same time. Say that if someone wins the lottery, he/she will not be looking for a job again and thus winning the lottery can be a competing event to

reemployment. In this case, winning the lottery suppresses the occurrence of reemployment and thus the reemployment event is censored. Figure I-1 depicts the four different types of censoring available. The filled dot represents an observed event and the hollow dot represents a censored event.

The importance of recognizing a censored event instead of ignoring it allows the censoring information to be used to correct the bias of the estimates in a survival model. If we say that the event did not occur and ignore it, the probability of the event occurring is zero. However, if we say that the event is censored which means that the event did not happen until a certain point (usually the end of observation), the probability of the event occurring is not zero but in fact the probability would (usually) get higher as time goes by. More statistical details about the consideration of censored data in survival model estimations will be presented later in this chapter.

B. Timing, Survival Function, Hazard Rate and Mean Residual Life

The other important variable in survival analysis is the timing variable. The "timing" here indicates the time to the occurrence of the event (time-to-event). The time-to-event can indicate different timings based on the research question. The most common example would be the time to death. Since the state before death is being alive or surviving, here the time-to-event is literally the survival time. In the unemployment spell studies, the event of interest is the end of unemployment or reemployment. Therefore, the time-to-event captures the duration of unemployment. In recidivism studies the event of interest is a person's relapse into criminal behavior and the time-to-event is the interval

between the unlawful behaviors. In observational studies of parent-child interaction, the time-to-event can be the duration in a particular dyadic state.

The timing variable can either be a discrete or a continuous measure (Masyn, 2014). Discrete timing variables can be either 1) chunks of continuous time: 1-10 years, 11-20 years, etc. or 2) naturally discrete time-points like measurement occasion (*e.g.*, baseline, immediate follow-up, second follow-up, etc.) or grade (*e.g.*, first grade, second grade, etc.). Continuous timing variables measure the exact physical time in years, months, days, hours, minutes or seconds. As the following section will show, different models can be used depending on the metric that is used for the timing variable. That said, if the exact time information is available, it is preferable to use continuous time models rather than discrete time models (Allison, 2010). Arbitrary discrete intervals may cause loss in information that could be acquired from a continuous variable and further lead to spurious conclusions. In this paper, only continuous time models will be discussed.

In survival analysis, the times at which an event happens is a stochastic process. That is, T is a random variable that has a nonnegative probability distribution. Here, T denotes the timing of an event to occur. There are four different functions to characterize the distribution of T (Klein & Moeschberger, 2003). Namely, they are 1) the *probability density (or probability mass) function*, 2) *survival function*, 3) *hazard rate (function)*, and 4) the *mean residual life function*. The four different functions for continuous time analyses are introduced below.

To understand the survival function, we first take a look at the probability distribution of T. Suppose that the event of interest was death. Then T is the time-to-

death or the survival time. The random variable, *T* can have a cumulative distribution function (c.d.f.), *F (t)* which is defined as

$$F\ (t) = \ Pr\ (T\ \leq\ t) \tag{I.1.1}$$

As Equation I.1.1 shows, *F (t)* is the probability that the event will occur (*i.e.*, death) at a time equal or less than *t*. If *T* is a continuous variable, the probability distribution function (p.d.f.) of *T*, *f (t)* is defined as the derivative of *F (t)*,

$$f\ (t) = \lim_{\Delta t \to 0} \frac{\Pr(t\ \leq\ T\ \leq\ t + \Delta t)}{\Delta t} = \frac{dF\ (t)}{dt} \tag{I.1.2}$$

The c.d.f. is the integral of the p.d.f. over the range of $0\ \leq t\ \leq\ \infty$. That is,

$$F\ (t) = \int_0^t f\ (x)dx \tag{I.1.3}$$

The survival function is the probability of a study unit (usually individual participants in psychological research) to have experienced the event of interest (*e.g.*, death) beyond time *t*. A formal statistical definition can be expressed as

$$S\ (t) = \ Pr\ (T\ >\ t) = 1 - Pr\ \ (T\ \leq t) \tag{I.1.4}$$

Equation I.1.4 implies that the survival function is the complement of the c.d.f. of *T*. That is,

$$S\ (t) = 1 - F(t) = \ 1 - \int_0^t f\ (x)\ dx = \int_t^\infty f\ (x)\ dx \tag{I.1.5}$$

Equation I.1.5 implies that the survival function is a nonnegative (greater or equal to zero), monotonically decreasing (non-increasing) function of *t* since the c.d.f.s or probabilities (by the definition) are bound to [0, 1]. Other than this restriction, the survival function can take any shape but there are some distribution shapes that are more

often used than others. Some of the frequently used distributions are shown in the next section.

Another basic function that is fundamental in survival analysis is the hazard rate or the hazard function. Klein and Moeschberger (2003) list the different terms for hazard rate in different areas of study, "The hazard rate is also known as the *conditional failure rate* in failure-time analysis, the *force of mortality* in demography, the *intensity function* in stochastic processes, the *age-specific failure rate* in epidemiology, and the *inverse of the Mills ratio*[1] in economics." (p.27).

The hazard rate is the event rate at time *t* conditional that the event did not happen until time *t* (Allison, 2010). If *T* is continuous, the hazard rate, *h (t)* can be defined as

$$h\,(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t \,|\, T \ge t)}{\Delta t} \qquad \text{(I.1.6)}$$

There are two things to note in Equation I.1.6. First, the limit of Δt approaching to zero implies that the hazard rate is an instantaneous rate of an event to occur at time *t*. Equation I.1.6 is very similar to the p.d.f. Equation 1.2. The p.d.f. is the derivative (instantaneous change) of the probability of observing the event to occur at time *t*. However, the expression of conditional probability when $T \ge t$, differentiates Equation I.1.6 from Equation I.1.2. The hazard rate is conditional that the event of interest has not occurred before time *t*. The conditional probability makes sense since if the event has already occurred there would be no more risk of the event happening. For example, if someone is already dead from lung cancer, we cannot talk about the risk of dying from

[1] The Mills ratio is defined as the ratio of the complementary cumulative distribution function (1-*F*(*x*); also known as the survival function) to the probability density function (*f*(*x*)). The inverse of the Mills function is known to be the hazard function.

lung cancer for that person. Note that even though the numerator of Equation I.1.6 is a probability, the hazard rate itself is not a probability. It can have a value larger than 1.0. The only restriction of the hazard rate is that it is a nonnegative function and there is no upper bound, $0 \leq h\,(t) \leq \infty$.

The survival function, *S (t*), the p.d.f., *f (t*), and the hazard function *h (t*) are all closely related. If we know one of them, the other two can also be calculated. In the continuous *T* case, we can derive the following equation from Equation I.1.5,

$$F\,(t) = 1 - S\,(t) \tag{I.1.7}$$

If we take the derivatives of both sides of Equation I.1.7,

$$f\,(t) = -s\,(t), \text{ where } s\,(t) \text{ is the derivative of } S\,(t) \tag{I.1.8}$$

Therefore, the negative derivative of the survival function equals the p.d.f. of *T*. Also, by definition, an alternative expression of the hazard function is

$$h\,(t) = \frac{f\,(t)}{S\,(t)} \tag{I.1.9}$$

After some algebra (Allison, 2010; Klein & Moeschberger, 2003), it can be shown that

$$h\,(t) = \frac{-\,d \ln S\,(t)}{dt} \tag{I.1.10}$$

$$S\,(\mathrm{t}) = exp\left(-\int_0^t h\,(x)dx\right) \tag{I.1.11}$$

$$f\,(t) = h\,(t)exp\left(-\int_0^t h\,(x)dx\right) \tag{I.1.12}$$

Klein and Moeschberger (2003) describe another function that characterizes the distribution of *T*. The *mean residual life* function at time *t* is the expected remaining life

given that the individual survived until time *t* (Hall & Wellner, 1981). A statistical definition of the mean residual life at time *t*, *mrl (t)* is given as

$$mrl\ (t) = \ E\ (T - t \mid T > t),\ \ t > 0 \qquad \text{(I.1.13)}$$

Given by the definition it follows that the *mrl (t)* of a continuous variable *T* is

$$mrl\ (t) = \frac{\int_t^{\infty} S\ (x)\ dx}{S\ (t)} \qquad \text{(I.1.14)}$$

That is, the *mrl (t)* is the area under the survival curve after time *t* divided by *S (t)*. One of the useful derivations of the *mrl (t)* is that the *mean life*, $\mu = mlr\ (0)$. That is, the mean life is the integral (or summation) of the survival function over the whole range of *T*.

Among the four different (but closely related) expressions of the distribution of *T*, the most useful (in terms of interpretation) and thus widely utilized expression is the hazard rate or function (Allison, 2010; Klein & Moeschberger, 2003). The hazard function in particular provides the simplest expression in models with covariates. Therefore, the hazard function will be discussed more frequently throughout the rest of the paper, although the other functions will also be mentioned. To understand more about the functions of the random variable *T*, some of the frequently used distributions for *T* and simple parametric models will be reviewed.

C. Survival Models without Covariates

This section will focus on models that specify the hazard or survival function without considering any covariates in the model. Mainly, there are two types of models: parametric survival models and nonparametric survival models. In the parametric

survival models, the functional form of a hazard rate or survival function is assumed. That is, the survival variables are assumed to have some particular distribution. In contrast, nonparametric survival models do not have any distributional assumptions on the survival variables and instead, use the data to estimate the survival function. We first take a look at the parametric survival models.

C.1. Parametric Survival Models

In survival analysis, many different distributions can be used to describe the distribution of *T*. Survival models with specific underlying distributions are sometimes classified as *parametric models* (Allison, 1984). Klein and Moeschberger (2003) show the hazard rate, survival function, p.d.f. and mean life function for 11 different commonly used distributions. These include: the exponential, Weibull, gamma, log-normal, log-logistic, normal, exponential power, Gompertz, inverse Gaussian, Pareto, and generalized gamma distribution. Among these, the exponential, Gompertz, and Weibull distributions are more widely used and will be reviewed in depth. The reason why these three distributions are widely used is that the functions are relatively easily to express in a mathematical form and provide a deeper understanding of the hazard, survival, and p.d.f. function (Klein & Moeschberger, 2003).

C.1.a. Exponential Distribution

We can start from thinking about the simplest hazard function. That is, the hazard function being a constant, $h\,(t) = \lambda$, where $\lambda > 0$. From Equations I.1.11 and I.1.12, the survival function and p.d.f. become

$$S\,(t) = \exp(-\lambda t) \qquad \text{(I.1.15)}$$

$$f\ (t) = \lambda \exp(-\lambda t) \qquad \text{(I.1.16)}$$

Figure I-2 shows an example of the survival function of an exponential model. Note that as lambda (λ) increases, the survival function decreases more quickly. In other words, the higher the hazard function is, lower the survival function is at time *t*.

Although the exponential model is often too strict to fit to data and unrealistic to apply to real data applications, it does have an interesting statistical property called the "lack of memory property". The lack of memory property is saying that *T* is independent from past history. Because of this property, the mean residual life function can be derived to be the reciprocal of the hazard function which is a constant. That is, $mrl\ (t) = E\ (T - t\ |T > t) = E\ (T) = 1/\lambda$.

C.1.b. Gompertz Distribution

From the exponential model (where the hazard function is a constant), the next complexity is to consider a linear function of time to model the (natural) logarithm of the hazard function. That is, $\log[h\ (t)] = b_0 + b_1 t$. If we take the exponential value at both sides of the equation, we get $h\ (t) = e^{b0} e^{b_1 t}$. By Equations I.1.11 and I.1.12, the survival function and p.d.f. become

$$S\ (t) = \exp\left(\frac{e^{b_0}}{b_1}\right) \exp(-\frac{e^{b_0}}{b_1} e^{b_1 t}) \qquad \text{(I.1.17)}$$

$$f\ (t) = \mathrm{e}^{b_0} e^{b_1 t} \exp\left(\frac{e^{b_0}}{b_1}\right) \exp(-\frac{e^{b_0}}{b_1} e^{b_1 t}) \qquad \text{(I.1.18)}$$

The parameters in the Gompertz function p.d.f. are often denoted as θ (shape parameter) and α (scale parameter), where $\alpha = b_1$ and $\theta = e^{b0}$. The parameters are restricted to $\alpha > 0$ and $\theta > 0$. Figure I-3 shows the effect of each of the parameters on the survival

function after fixing the other parameter to a specific value. The upper panel shows the survival function as b_0 = -2, 0 and 2 while b_1 = 0.5 and the lower panel shows the survival function as b_1 = 0.1, 0.5 and 1 while b_0 = 0. The shape of the curve changes as b_0 changes. As b_0 becomes more negative, the survival function becomes more like a linear line parallel to the horizontal axis. The b_1 parameter represents the rapidness of the declining survival function. As b_1 increases, the slope gets steeper and the survival function reaches the lower asymptote (in this case, 0) quicker. Note that the exponential distribution is a special case of a Gompertz function when b_1 = 0.

C.1.c. Weibull Distribution

Another widely used distribution is the Weibull distribution. Again, we start with the logarithm of the hazard function modeled to be a linear (linear in the parameters) function of *t*. In particular, $\log h\,(t) = b_0 + b_1 \log t$. After imposing the exponential function on both sides of the equation it becomes $h\,(t) = e^{b_0} e^{b_1 \log t} = e^{b_0} e^{\log t^{b1}} = e^{b_0} t^{b1}$. The hazard function is still a non-negative function, but now the slope parameter, b_1 can be a negative value. If b_1 is negative, the hazard function will show a decreasing curve and if b_1 is positive, the hazard function will show an increasing curve. Figure I-4 shows how the b_0 and b_1 parameters affect the hazard function. The upper panel shows how different values of b_0 affect the hazard function while fixing b_1 at 0.5. As b_0 increases, the upper asymptote gets higher. The lower panel shows how different values of b_1 affect the hazard function while fixing b_0 at 0. When b_1 = 0, the hazard function is a straight line parallel to the horizontal axis. Thus, it reduces to the exponential distribution. When $b_1 > 0$, the hazard function has a positive relationship

with time. However, when $b_1 < 0$, the hazard function has a negative relationship (think of an inverse square root function) with time.

One feature of the Weibull models that is distinguished from the Gompertz models is the intercept of the hazard function at $t = 0$. The intercept is always zero for a Gompertz model (except the special exponential case). However, the intercept can be zero or positive infinite depending on the sign of the b_1 parameter for a Weibull model. If b_1 is positive, the intercept is always zero, but when b_1 is negative, the intercept approaches a positive infinite number.

Another way to express the Weibull hazard function is to use the Weibull distribution parameters, λ (shape parameter) and α (scale parameter). Then, the hazard function, $h\,(t) = \lambda\alpha t^{\alpha-1}$, where $\lambda > 0$ and $\alpha > 0$. Also, from Equations I.1.11 and I.1.12,

$$S\,(t) = \exp(-\lambda t^{\alpha}) \tag{I.1.19}$$

$$f\,(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha}) \tag{I.1.20}$$

Figure I-5 depicts how the λ and α affect the survival function at different values. The lower asymptote gets lower as λ gets larger and the slope gets steeper (or in other words, the survival function reaches zero quicker) as α increases.

Hougaard (2000) and Mills (2011) both mention some advantages and disadvantages of the parametric survival models. The advantage of the parametric survival models is that the hazard function and survival function can be estimated by a predetermined smooth curve function of *t*. This is a major advantage over the nonparametric methods (*e.g.*, Kaplan-Meier estimator) since the nonparametric methods only focuses on the survival function (and not the hazard function) and also the survival

function is not a smooth curve but a discrete step function as it will be shown in next section. However, the disadvantage of parametric models also comes from fitting a particular smooth curve to the data. We do not necessarily know the true model that shapes the phenomenon that we observe and we are fitting a hypothetical model to the observed data. Even though we knew the true model to fit, we obtain less good model fit relative to the fit of a nonparametric method which does not require a hypothetical model.

In the next section, a simple nonparametric model will be reviewed. The model is called "nonparametric" because it does not assume any form of the distribution of *T*.

C.2. Nonparametric Survival Model

Nonparametric survival models do not assume any functional form for the survival or hazard function but use the data to estimate the survival curves. In order to estimate the survival curves, *life tables* are constructed and examined. A life table (also called a mortality table or actuarial table) includes the most basic descriptive statistics in survival analysis (Klein & Moeschberger, 2003). The life table discretizes time into certain intervals and can show the number of survivors, the number of deaths, and the probability of (cumulative) survival in a particular time interval. Another version of the life table uses the actual time of death. Whenever someone dies, the number of survivors, the number of deaths, and the probability of cumulative survival probability can be computed.

The Kaplan-Meier (K-M) estimator (Kaplan & Meier, 1958) is one of the most widely utilized non-parametric methods that is used for estimating the cumulative survival probability in a life table. The K-M estimator is the nonparametric maximum

likelihood estimate of a cumulative survival probability. The (cumulative) survival function using the K-M estimator can be expressed as,

$$\hat{S}(t) = \prod_{t_i \le t} \frac{N_{t_i} - E_{t_i}}{N_{t_i}} \qquad \text{(I.1.21)}$$

where N_{t_i} is the number of people at risk of an event (*e.g.*, death) at t_i, E_{t_i} is the number of people who experienced the event at time t_i. The number of people at risk, $N_{t_{i+1}} = N_{t_i} - E_{t_i} - C_{t_i}$, where C_{t_i} is the number of censored observations between t_i and t_{i+1}. For example, consider the numbers in Table I-1. Based on Equation I.1.21, the K-M estimate at time 2, 10 and 15 are respectively, $\hat{S}(2) = \frac{100-1}{100} = 0.99$, $\hat{S}(10) = \frac{100-1}{100} \times \frac{99-8}{99} = 0.91$, and $\hat{S}(15) = \frac{100-1}{100} \times \frac{99-8}{99} \times \frac{90-30}{90} = 0.61$. The K-M estimator takes advantage of all of the data including the censored data. To understand how the K-M estimator is accounting for the censored data, consider the observations at t_2 and t_3. There was 1 censored observation between t_2 and t_3. If we were to omit the one censored observation in calculating the survival function, we are throwing away the valuable information that the person had lived until time t_2. To avoid the loss of information, the K-M estimator computes the survival function by multiplying the probabilities including the censored observation up to time t_2. Then, only at time t_3 the probability is computed without the censored observation and then multiplied with probability up to time t_2. This requires the assumptions that the event observations are independent of each other and also that the censoring information is independent of the event observations. Another important feature of the K-M estimator is that at time t_0 the survival function is 1.0 and it assumes that the event occurs at specified exact times. In

between the specified times, the survival function is constant. Therefore, the K-M plot shows a decreasing step function. See Figure I-6 for a K-M plot of Table 1.

Hougaard (2000) and Mill (2011) mention that while nonparametric survival analysis may fit the data well, sometimes it is hard to interpret or even present the results. A disadvantage of nonparametric methods is that it does not have a single (or perhaps a few) statistic (*e.g.*, shape parameter) that summarizes the study findings. A table or figure must be shown to represent the results of a nonparametric analysis. This can sometimes be overwhelming information that prevents readers or researchers to arrive at a clear and concrete conclusion.

Until now, this section has only concentrated in reviewing basic survival models that do not include any covariates or predictors in the model. The terms covariates and predictors will be used interchangeably in this paper to indicate basically the regressors in a regression model. Often the focus of a research is to study predictors of a hazard function (or survival function). Survival models with covariates have been studied a lot in both applied and quantitative studies and an introduction to some of the most utilized models will be given in the next section. In particular, the *accelerated failure time model* which is representative of the parametric models and the *Cox proportional hazards model* which is often classified as a semiparametric model will be reviewed.

D. Survival Models with Covariates

This section provides a review of statistical methods to incorporate covariate effects in the survival models. Before illustrating relatively more complicated regression type models, a simple method to compare survival curves for different groups is briefly

introduced. An option to statistically test whether the non-parametric survival curves are different for the different values of a categorical variable (*e.g.*, treatment status) is using the log-rank test (Peto & Peto, 1972). The log-rank test is similar to the Mantel-Haenszel test (Mantel & Haenszel, 1959) where the stratum variable is the k time-points of when an event occurs, and the qx2 table between a categorical variable with q categories (*e.g.*, q=2 for a treatment status variable) and an event occurrence variable is examined. The null hypothesis is the survival function for all categories are equivalent to each other for all k time points. This test is equivalent to testing whether the qx2 tables for k time-points show similar frequency distribution or not. However, this method is restricted to categorical variables and does not provide any estimates representing the magnitude of the categorical variable effect on the survival function. The more widely used survival models with covariates are the accelerated failure time (AFT) model and the Cox proportional hazards model (Cox model). The AFT and Cox models are in a regression format. Therefore, as for regression models, continuous or categorical variables can be used as covariates and the actual covariate effects are estimated as regression parameters in a model.

Before going into the details of the AFT model and the Cox model, Table I-2 compares some of the characteristics of the two models. First of all, the AFT model models the survival function while assuming a distributional form (parametric model), whereas, the Cox model models the hazard function without assuming any distributional form (semi-parametric model). The estimator used in the AFT model is the full maximum likelihood but the Cox model uses a special form of maximum likelihood estimation called the "partial maximum likelihood" (see below for details). The

advantage of the AFT model is that the interpretation is easy and straightforward. The most appealing feature of the Cox model is that there is no need to specify a baseline hazard function which leads to the proportional hazards assumption (to be explained below). One disadvantage of the AFT model is that the model is sensitive to the specification of the survival distribution. The semi-parametric Cox model circumvents this problem by not having to specify a distribution form for the hazard function, but the proportional hazards assumption can be violated and cause bias in the parameter estimates.

D.1. Accelerated failure time model

The accelerated failure time (AFT) model assumes that a set of covariates affect the event time scale by accelerating or decelerating it by a constant (to be estimated). The AFT model is a parametric model where the baseline function is specified by the researcher. In AFT models, the relationship between two individual's survival function is usually expressed as

$$S\,(t|\boldsymbol{X}) = S_0\{\exp[-(\boldsymbol{\beta}'\boldsymbol{X})]\,t\} = S_0\{\theta t\} \qquad \text{(I.1.22)}$$

where $\boldsymbol{X}$ is a set of covariates, $\boldsymbol{\beta}$ is the corresponding set of coefficients and $S_0[\,]$ denotes the baseline survival function where all $\boldsymbol{X}$ values are zero. Equation I.1.22 is saying that an individual's survival function is accelerated or decelerated by a constant (θ) which is a function of the covariates ($\theta = \exp[-(\boldsymbol{\beta}'\boldsymbol{X})]$). Therefore, if we say that an individual with some specific covariate value, $X = x$ has an acceleration factor of $\theta = 2$, it means that the individual with $X = x$ has a two-times-faster time scale than an individual at baseline (or $X = 0$). Allison (2010) gives an intuitive example comparing the survival probability of a dog and a human (*p*.72). Conventionally, a dog's aging is known to be 7

times faster than a human aging. That is, $\theta = 7$. If we think of *X* as an identifier of being a dog or a human (0 = dog and 1 = human), then a dog's survival probability is $S_0(t)$ whereas, by theory a human's survival probability is $S_0(7t)$. While a dog ages 1 year, a human is aging 7 years. Therefore, we can say that the human's time scale is 7 times slower than a dog's time scale or that a dog's time scale is 7 times faster than a human's time scale.

From equation I.1.22 and I.1.10 the hazard function can be derived to be

$$h\,(t|\boldsymbol{X}) = \frac{-dlnS_0[\theta t]}{dt} = \theta h_0[\theta t] \qquad \text{(I.1.23)}$$

where $\theta = \exp[-(\boldsymbol{\beta}'\boldsymbol{X})]$. Another important relationship expressed in the AFT model is the relationship between the natural logarithm (ln) of *T* and a set of covariates, $\boldsymbol{X}$:

$$\ln(T_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \sigma\varepsilon_i \qquad \text{(I.1.24)}$$

where ε_i is a random residual term which has a distribution with a variance of 1 and σ is a parameter to estimate which determines the variance of the ε_i distribution. Note that Equation I.1.24 is an expression for the natural logarithm of *T*. Therefore, the distribution of *T* is different from the distribution of ε_i. When ε_i follows an extreme value distribution (1 parameter), extreme value distribution (2 parameter), or a logistic distribution, *T* follows an exponential, Weibull or a log-logistic distribution, respectively. Other kinds of possible distributions for T are well documented in Table 6.1 in Mills (2011). Namely, they are the Gompertz, log-normal, log-logistic and generalized gamma (three different versions depending on the parameter value) distributions. The purpose of this paper does not cover parametric regression models in depth, however, descriptions of

the AFT models with different distributions are given in Mills (2011, *pp.* 117-125), Allison (2010, *pp.* 77-86) and Klein and Moeschberger (2003, *pp.* 375 – 388).

The most common estimation method of these AFT models is maximum likelihood (ML) accounting for censored data. The ML estimator has good statistical properties such as being consistent, asymptotically efficient and asymptotically normal when the sample size is large (Allison, 2010). The ML estimator is especially useful in survival analysis because the handling of censored data is straightforward. Not considering covariates and assuming that the data are uncensored, the likelihood function of the data, *L* is

$$L = \prod_{i=1}^{n} f_i\,(t_i) \tag{I.1.25}$$

where *i* denotes the individuals, and for each individual, $f_i\,(t_i)$ denotes the p.d.f. for time of event, t_i. Thus, the likelihood function, *L*, is telling us the joint probability of an event across all the individuals. If we further think of censored data, the likelihood function, *L* can be divided into two parts: the uncensored observations and the censored observations. Then, the *L* can be expressed as

$$L = \prod_{i=1}^{n} [f_i\,(t_i)]^{\delta_i} [S_i\,(t_i)]^{1-\delta_i} \tag{I.1.26}$$

where δ_i is a censoring indicator where 0 = censored and 1 = uncensored. Note that for the censored data part in Equation I.1.26, $[S_i\,(t_i)]^{1-\delta_i}$ includes the survival function which is the cumulative survival probability at time t_i. That is, when an individual is censored at time t_i, we can estimate the probability of an event to occur after time t_i by its survival function at time t_i. Once we choose a model (*e.g.*, an exponential AFT

model), the p.d.f. and survival function can be substituted in Equation I.1.26. Then, using the maximization process of the L function includes iterative methods like the Newton-Raphson algorithm (see Allison, 2010, pp. 92-93 for more details) and as a product of the maximization process we can get estimates of $\boldsymbol{\beta}$.

An advantage of the AFT model is that the model parameters are easily interpreted. The regression coefficients can be interpreted in their absolute terms (*e.g.*, number of years). From Equation I.1.24., the regression parameters directly increase (or decrease) the total ln(T) for a given X value. This is not true for the proportional hazard model where the regression parameters are interpreted as hazard ratios (more to be explained in next section). Also, the acceleration (or deceleration) factor, θ can be seen as a 'stretching' or 'contracting' the coefficient of survival time when comparing two groups with different covariate values. However, a limitation of the AFT model is that we often do not know what the true model that underlines T is. If the underlying T distribution is incorrectly specified, the researchers are at risk of reaching incorrect substantive conclusions (Mills, 2011). Also, it is difficult to include time-varying covariates in an AFT model, although it is not impossible (see Hougaard, 2000).

Next, we look at the Cox proportional hazards model which has properties that can overcome some of the problems that the parametric models have.

D.2. The Cox proportional hazards model

One of the most widely used survival models is the Cox proportional hazards model (from now on abbreviated as the Cox model; Cox, 1972). One of the reasons for the popularity of the Cox model is that the Cox model does not assume any of the complicated distributions we have looked at in the parametric models. Furthermore, the

regression coefficients are interpretable and including time-varying or time-invariant covariates in the model is relatively easy compared to the AFT model (Allison, 2010). For simplicity, a Cox model only including time-invariant covariates is presented first and extensions to include time-variant covariates are presented later in this chapter.

In a Cox model, the hazard rate is regressed on a set of predictors through a nonlinear function.

$$h_i\,(t) = \; h_0\,(t) \exp(\beta_1 x_{i1} + \cdots + \; \beta_p x_{ip}) \qquad \text{(I.1.27)}$$

where $h_0\,(t)$ denotes the baseline hazard function (when all covariate values are zero) at time *t,* and x_{ip} denotes the *p*th time-invariant covariate for individual *i.* Taking the natural logarithm on both sides of Equation I.1.27 becomes

$$\ln[\,h_i\,(t)] = \ln[h_0\,(t)] + \beta_1 x_{i1} + \cdots + \; \beta_p x_{ip} \qquad \text{(I.1.28)}$$

Equation I.1.28 shows that the logarithm of the hazard rate can be expressed as a linear combination of the covariates. If the logarithm of the baseline hazard function, $\ln[h_0\,(t)]$ is chosen to have a particular functional form (*e.g.*, the hazard function for a Weibull model), Equation I.1.28 becomes a representation for the parametric PH model. Unlike the parametric PH model, the Cox model does not need to assume any distribution of $\ln[\,h_i\,(t)]$ which is a function of the event time. That is, the baseline hazard function does not need to have a particular form and there are no parameters to estimate for the baseline hazard function. Therefore, the Cox model is often termed a semi-parametric model (Klein & Moeschberger, 2003). The only parameters estimated are the regression coefficients, the βs. Before explaining how the Cox model can estimate the regression coefficients without having to specify the baseline hazard, the reason why the Cox model is called as a "proportional hazards" model is illustrated.

An interesting fact about Equation I.1.27 is that the hazard rate for any given individual divided by another individual's hazard rate is a fixed proportion. This can be shown by calculating the hazard ratio for persons *i* and *j*:

$$\frac{h_i\ (t)}{h_j\ (t)} = \frac{h_0\ (t)\exp(\beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{h_0\ (t)\exp(\beta_1 x_{j1} + \cdots + \beta_p x_{jp})}$$

$$= \exp[\beta_1(x_{i1} - x_{j1}) + \cdots + \beta_p(x_{ip} - x_{jp})] \qquad \text{(I.1.29)}$$

Note that the baseline hazard, $h_o\ (t)$ cancels out because it is in the numerator and denominator of the second equality of Equation I.1.29. Thus, the ratio of two individuals becomes a constant value over time *t*. This ratio is called the *proportional hazards* (PH) property. The PH property is particularly important because we can say that the hazard rate only differs by the covariate values. This allows the Cox model to produce clear interpretations when the PH assumption is met. If the hazard rates for two people are actually different as a function of time as well as the covariate values, then the Cox model can produce severe biased results (Allison, 2010; Hosmer, Lemeshow, & May, 2008).

D.2.a. Estimation of the Cox model

In this section, the partial likelihood estimation method and a Bayesian method for estimating the parameters of the Cox model are discussed. In particular, the widely used partial likelihood estimation method is illustrated in detail. In contrast to the full likelihood method, the partial likelihood method allows estimation of the regression parameters without having to specify or estimate the baseline hazard function. Suppose we have observed *J* events in total during the observation period and each event occurred

to individual *i* at time t_i. Here we assume that there were no *ties*[2]. That is, for event *j*, only one individual experiences the event. The data are arranged in the order of time t_i. Table I-3 shows made-up data to illustrate how the partial likelihood works. The first column shows a numbering of the events that occurs during the observation and a dot represents that the event did not happen to that particular individual. The second column shows the individual IDs in the study, the third column contains the timing information of when the event happens and finally, the last right column shows the censoring information where, 0 = censored and 1 = event occurred.

The likelihood for event *j* can be computed by the hazard rate of when the event happened at t_i over the sum of this individual's hazard rate and all the individual's hazard rates who did not experience the event before t_i. That is, the likelihood for event 1, L_1 is

$$L_1 = \frac{h_{79}(5)}{h_{79}(5) + h_{25}(5) + h_{67}(5) + h_{43}(5) + h_8(5) + h_{17}(5) + \cdots} \quad \text{(I.1.30)}$$

where $h_i\ (t)$ denotes the hazard rate at time *t* for individual *i*. Note that in the denominator of L_1 everyone's hazard rate is included because it is the first event (we do not consider events before the observation started). The likelihood for event 2, L_2, is

$$L_2 = \frac{h_{67}(7)}{h_{67}(7) + h_{43}(7) + h_8(7) + h_{17}(7) + \cdots} \quad \text{(I.1.31)}$$

Note that the hazard rates for the deceased and censored prior to this event time are excluded in the denominator of L_2. That is, $h_{79}(7)$ is not included because individual 79

[2] This may or not be a valid assumption and there are existing methods to deal with tied data. Allison (2010) provides a good overview of the different estimation methods for tied data. Namely, there are two exact methods (one for continuous-time model and one for discrete-time model) and two approximate methods (Breslow, 1974 and Efron, 1977). The exact methods calculate the likelihood based on all possible combination of orderings for tied data. The two approximate methods are used to reduce the computational burden of exact methods and both of them work fairly well when there are not too many ties existing at a time point.

had already died at year 5, and $h_{25}(7)$ is not included because individual 25 dropped out of the study at year 6 making it impossible to calculate the hazard rate at year 7. The hazard rate of individual 43 at year 7 is included in the denominator, although individual 43's data was censored, we know that individual lived until year 7. Similarly, we can compute the likelihood for event 3, L_3 as

$$L_3 = \frac{h_8(11)}{h_8(11) + h_{17}(11) + \cdots} \tag{I.1.32}$$

Then the partial likelihood estimator can be expressed as multiplying all the likelihoods for event j. If we substitute the hazard rate, $h_i\ (t)$ using Equation 1.27 and then rewrite the likelihood in terms of multiplying all the individual likelihoods (i = 1 to n), a general expression of the partial likelihood estimator, *PL* becomes

$$PL = \prod_{i=1}^{n} \left[\frac{h_0(t)\exp(\boldsymbol{\beta x_i})}{\sum_{j=1}^{n} Y_{ij} h_0(t)\exp(\boldsymbol{\beta x_j})}\right]^{\delta_i} \text{ where } \begin{cases} Y_{ij} = 1\ if\ t_j \geq t_i \\ Y_{ij} = 0, else \end{cases} \tag{I.1.33}$$

Y_{ij} is an indicator to exclude individuals who already had experienced the event from the denominator and δ_i is the censoring indicator which has a value of 0 when censored and a value of 1 when data are present. The censoring indicator, δ_i allows consideration of only the likelihoods when an event had actually occurred. The baseline hazard, $h_0(t)$ in the numerator and denominator can be canceled out of Equation I.1.33 which leaves the PL to be independent of any specification of the baseline hazard, $h_0(t)$. Once the PL function is determined, the procedure remaining is the same as the ML estimation method we have looked at above. We usually take the logarithm of the *PL* (take the logarithm of the *PL* in Equation I.1.33) and then find the parameter estimates that maximize the log (*PL*) function. An iterative procedure (*e.g.*, Newton-Raphson method) can be used to maximize the function to acquire parameter estimates, $\boldsymbol{\beta}$.

Another estimation method employs the Bayesian framework. Bayesian methods are more flexible than the frequentist method (*i.e.*, MLE), especially when many unknown parameters are in the model (*e.g.*, Dagne & Snyder, 2009). In Bayesian estimation, the parameters are treated as random variables and thus can be expressed as having a distribution (*i.e.*, posterior distribution) rather than characterizing the parameter as a fixed population value. The posterior distribution can be approximated by a multiplicative function of a given prior distribution and the likelihood function. A prior distribution incorporates any previous beliefs or knowledge about each of the parameters. The prior distribution can be derived from an earlier study or pilot study, or a non-informative prior (*e.g.*, uniform distribution) can be given if the researcher is unsure which prior to use (this will give more weight to the likelihood function based on the data). Although the basic idea of Bayesian estimation is simple and appealing, the realization of the method was not so easy until the development of Markov chain Monte Carlo (MCMC) methods conducted with computers to do large amount of computations. The core of MCMC is to update parameter estimates based on their previous set of estimates (using the memoryless property of the Markov chain) until satisfactory convergence is reached. Note that convergence of the posterior distribution is evaluated rather than a point estimate. Further discussions of the Bayesian method are omitted due to the scope of this dissertation. Gelman et al. (2014) provide a good introduction of general Bayesian analysis and Ibrahim, Chen and Sinha (2001) provide a general discussion of Bayesian survival analysis.

D.2.b. Statistical testing of model parameters

After fitting the model of interest, a researcher may be interested in performing statistical tests to evaluate parameter estimate(s) or to do model comparisons between nested models: the Wald test, the likelihood ratio test (LRT) and the score test. The three tests are asymptotically equivalent (when the sample size is large). The tests use different statistics; all three statistics approximately follow the chi-square distribution when the null hypothesis is true. All three tests can be used to test a single parameter (univariate test) and also can be extended to test multiple parameters (multivariate test). The Wald statistic is computed by taking the squared difference between the ML estimate and the value under the null hypothesis (usually H_0: $\boldsymbol{\beta} = \mathbf{0}$), and dividing it by the estimated variance of the ML estimate. The Wald statistic approximately follows a chi-square distribution with *p* (number of parameters tested) degrees of freedom. The LRT can be used in comparing nested models by subtracting the -2LL from the larger model (with more parameters including the parameters of interest) from the -2LL from the more restricted model (with fewer parameters, often, the model that satisfies the null hypothesis). Then the -2LL difference between the larger and smaller model approximately follows a chi-square distribution with *p* (the parameter number difference between the two models) degrees of freedom. Lastly, the score test (or the "Lagrange multiplier test") is computed by dividing the squared score function term by the variance of the ML estimate evaluated at the parameter under the null hypothesis. The score function is defined by taking the derivative of the log likelihood function in respect to the parameter(s). The resulting statistic approximately follows a chi-square distribution with *p* (number of constraints the null hypothesis imposes) degrees of freedom.

D.2.c. Proportional hazards assumption checking

The Cox PH model works only when the PH assumption has met. The PH assumption is saying that the effect of each covariate in the model is the same at all time-points. When the PH assumption is clearly violated, there are two methods to extend the Cox model to allow nonproportional hazards. One is to explicitly include the interaction between the covariate and time in the model. For example, we can have the following Cox model including an interaction effect between *x* and t.

$$\ln h\,(t) = \ln[h_0\,(t)] + \,(\beta_1 + \,\beta_2 t)x \qquad \text{(I.1.34)}$$

In Equation I.1.34, β_1 is the effect of covariate *X* when *t* = 0 and β_2 is the linearly additive (or subtractive) effect of *X* as a function of *t*, thus the effect of *X* increases linearly with time if β_2 is positive and decreases linearly with time if β_2 is negative.

The other approach is to use a stratified Cox model. The stratification method is most useful when there is a categorical variable that interacts with time but is not of main interest. The categorical variable serves as a stratification variable in the model. The stratified Cox model can be expressed as

$$\ln h\,(t) = \ln[h_S\,(t)] + \,\beta x \qquad \text{(I.1.35)}$$

where $h_S\,(t)$ denotes the arbitrary baseline hazard for a stratification variable, *S* (*e.g.*, school ID, gender or age). Note that while the effect of *X* is fixed (β) to be the same across *s* (levels of *S*), the baseline hazard function is allowed to be different across *s*. The β parameter is estimated by first constructing separate PLs for each of the *s* levels, then multiplying the PLs together and at last choosing the β estimate that maximizes the multiplied PL function. The stratified method has disadvantages such that we cannot specify a main effect for the stratification variable in the model as a parameter nor we can

specify an interaction effect between the stratification variable and time in the model as a parameter. However, the stratified method can be useful when different levels of the stratification variable are expected or known to have different functions of time. The effect of a variable can only increase or decrease linearly with time in the interaction model. However, the stratification method allows that the stratification effect can be reversed (once or multiple times) in relation to time. It is recommended that the stratification variable is not a variable of interest but a nuisance variable (*e.g.*, a clustering variable such as school ID) whose effect should be controlled (Allison, 2010).

Allison (2010) argues that some crucial questions about the Cox model might be ignored while researchers focus on testing the PH assumption. Researchers are sometimes overly sensitive to nonproportional hazards only because the model is called the "proportional hazards" model. However, this assumption is just like any other assumption that we make in a statistical model. For example, in many cases, the 2-way (or even higher order) interaction effects are ignored in an ordinary least squares multiple regression model. The interaction effect is included in a regression model when only the researcher has an interest in it. This is the same situation as for a PH model. If a researcher believes that the hazard ratio for a covariate will not change with time, they are relying that PH assumption holds and estimates the model. If the researcher believes that the hazard ratio is a function of time, then the researchers can use one of the two methods (the stratification method or the interaction model) that are introduced above. Allison argues that there are more important considerations that a researcher must not neglect are: 1) are there any other confounding variable(s)?, 2) is the censoring mechanism noninformative?, and 3) is the measurement error in the covariates acceptably

low?. If these basic assumptions are not met, the parameter estimates can be severely biased.

D.2.d. Time-dependent covariates

A researcher may incorporate time-dependent covariates expecting that a covariate effect would change with time (*i.e.*, the PH assumption is violated). By including time-dependent covariates, the basic Cox PH model in Equation I.1.27 becomes

$$h_i\,(t) = \; h_0\,(t) \exp(\beta_1 x_{i1}(t) + \cdots + \; \beta_p x_{ip}(t)) \qquad \text{(I.1.36)}$$

The only difference in Equation I.1.36 from Equation I.1.27 is that now x_{ip} is replaced with $x_{ip}(t)$ with implies that the covariate values are allowed to vary over time, *t*. Here we might assume that the covariate values are known at an instant just before time *t*, so that we can say that the hazard rate is predictable based on the covariate values. However, as Allison (2010) and Mills (2010) have recommended, it is often useful to create lagged variables to ensure that changes in the time-dependent variable precede the event at time *t* thus allowing for a causal interpretation.

Conclusion of the Survival Analysis Section

Survival analysis involves statistical analysis of "whether" an event occurs and "when" the event occurs. Survival and hazard functions are used to accommodate both whether and when information together. Researchers can be interested in different type of questions: 1) what does the survival/hazard function look like as a function of time?; 2) what are the relationships between the survival/hazard function and covariates? For the former question, parametric or non-parametric survival/hazard functions can be drawn from a pre-existing known distribution (parametric) or from purely from the data

(non-parametric). For the latter question, the most popular choice of model is the Cox proportional hazards model where the hazard rate is expressed as a function of the baseline hazard rate and a non-linear (exponential) function of the covariates. The strength of the Cox model is that a function of the baseline hazard rate (which can be complicated sometimes) needs not to be estimated while estimating the regression coefficients of the covariates. Because of this property, the Cox model is used for statistical analysis of a mediation model with a survival mediator in this dissertation. Before introducing details of the survival-mediator model, in the next section, statistical mediation analysis with continuous variables is discussed.

2. Mediation Analysis

Mediation analysis is a crucial research methodology in psychology and many other social science areas. Mediation represents a hypothetical relationship where one variable (independent variable) affects a second variable (mediator variable) and in turn, affects a third variable (dependent variable). The mediator variable explains "how" an independent variable affects a dependent variable. For example, in an intervention study, a randomized intervention trial is given to each family to reduce a child's behavior problems. Hopefully, the intervention trial compared to a control trial reduces child's behavior problems. To investigate further "how" the intervention reduces child's behavior problems, the investigators might measure parenting practices during the trials and use this variable as a mediator and conduct a mediation analysis. The goal of this chapter is to provide a brief introduction to mediation analysis. The single mediator model with a continuous mediator and continuous outcome is introduced and different methods testing the mediated effect are reviewed.

A. The Single Mediator Model

The single mediator model represents the inter-relationships among three variables, the independent variable, *X*, the mediator, *M*, and the dependent variable, *Y*. The single mediator model with continuous *M* and *Y* variables (*X* can be either a continuous or a discrete variable) can be expressed as three regression equations, one regression equation for the mediator, *M*, and two regression equations for the outcome *Y*,

$$M = i_1 + aX + e_1 \qquad \text{(I.2.1)}$$

$$Y = i_2 + c'X + bM + e_2 \quad \text{(I.2.2)}$$

$$Y = i_3 + cX + e_3 \quad \text{(I.2.3)}$$

where i_1, i_2 and i_3 are intercepts for each regression equation, *a* is the coefficient for *M* regressed on *X*, *c'* is the coefficient for *Y* regressed on *X* after controlling for *M*, *b* is the coefficient for *Y* regressed on *M* controlling for *X*, *c* is the coefficient for *Y* regressed only on *X*, and e_1, e_2 and e_3 are residuals following independent normal distributions, $e_1 \sim N(0, \sigma_1^2)$, $e_2 \sim N(0, \sigma_2^2)$, and $e_3 \sim N(0, \sigma_3^2)$. Figure I-7 depicts the relationships represented in the above equations.

Different methods can be used to estimate the parameters in the single mediator model. The first method is to use the ordinary least squares (OLS) estimator. The goal of the OLS estimator is to find the regression coefficient estimates that minimize the sum of squares of the residuals. Mathematically the OLS estimator provides a closed form solution to estimate the regression parameters. The OLS estimator can be used when *M* and *Y* are both continuous variables (*X* can be a continuous or a categorical variable). However, the OLS estimator cannot be used in situations where the *M* and/or *Y* variable is a categorical variable or other type of variables (*e.g.*, count). If either *M* or *Y*, or both *M* and *Y* are categorical variables, logistic regression or probit regression is used to estimate the parameters in the mediation model (MacKinnon, 2008). In this case, the maximum likelihood (ML) estimator is used instead of the OLS estimator (the ML estimator can also be used for continuous mediators and outcomes). The likelihood function is a function that describes how well a certain set of parameter estimates would fit the given data. The goal of the ML estimator is to find regression coefficient estimates that maximizes the likelihood function. Usually, the ML is found using an iterative

process such as, the expectation-maximization (E-M) algorithm. The E-step calculates the expectation of the likelihood function given the current estimates of the regression parameters. Then, the M-step updates the regression estimates which maximizes the expected likelihood value in the E-step. The updated regression estimates are used in the next E-M step until a certain criterion (e.g., the estimate difference between two iterations is smaller than some criterion value, δ) is met.

B. Tests of the Indirect Effect

The mediated effect is the effect of *X* on *Y* contributed by the mediator, *M* (the effect of *X* on *Y* through *M*). Since, *X* indirectly affects *Y* through *M* (the *a* and *b* path in Figure I-7), the mediated effect is also called the indirect effect. There is also an effect in which *X* directly affects *Y* in the model (the *c'* path in Figure I-7) and this is called the direct effect. In the case in which both *M* and *Y* are continuous variables, the sum of the indirect effect and the direct effect is the total effect (the *c* path in Figure I-7). There are two different ways the indirect effect can be quantified, 1) the product of coefficients estimate: $a \times b$ (or ab), and 2) the difference in coefficients estimate: $c - c'$. The two quantities equal ($ab = c - c'$) only when *M* and *Y* are both continuous variables and the sample size for estimating the parameters in regression equations for *M* and *Y* are the same (MacKinnon, 2008). If the mediator and/or the outcome are non-continuous variables (*e.g.*, a categorical variable, a survival variable), $ab \neq c - c'$. In this paper, the product of coefficients (ab) estimate is chosen over the difference in coefficient ($c - c'$) estimate because of two reasons: 1) For mediation models with binary outcomes, simulation studies (MacKinnon & Dwyer, 1993; MacKinnon, et al., 2007) have shown

that the $c - c'$ estimate did not change (or even decreased) as the true mediated effect increased, whereas the *ab* estimate properly reflected increasing of the true mediated effect, and 2) In general, the statistical tests of the *ab* estimate performed better than the tests of the $c - c'$ estimate in terms of Type I error rates (MacKinnon et al., 2002). In the following Section B.2., several different statistical tests for the *ab* estimate is explained.

There are three different ways to test the indirect effect. The first group of methods evaluates separate parameters in the mediation model and arrives at a conclusion of whether the mediated effect is significant or not. Namely, the tests in this group are the causal steps approach (also known as the "Baron and Kenny" approach) and the "joint significance test". The second group evaluates the indirect effect estimate (e.g., *ab*) whether it is significant or not. As mentioned above, the focus will be testing the *ab* estimate instead of the $c - c'$ estimate. Several different methods would be considered in the second group including the Sobel test, bootstap methods, and the distribution of product method. The third group is a relatively new area in mediation analysis termed "causal mediation". Using the counterfactual approach of causal effects, the natural indirect effects are derived using conditional probabilities for the potential outcomes. The next sections explain these different methods to evaluate the indirect effect of a single mediator model.

B.1. Separate Evaluation of the Mediation Parameters

In this first group of methods, each relationship between the variables in a mediation model is tested and the mediated effect is evaluated as a collection of the separate significant results.

B.1.a. The Baron and Kenny Approach

The causal steps approach or "Baron and Kenny" approach is one of the most widely used methods to evaluate a mediated effect. In the initial statement of this approach, Judd and Kenny (1981) proposed the following tests were necessary to conclude that mediation exists: 1) The effect of *X* on *Y* (*c*-path in Figure I-7) is significant, 2) The effect of *X* on *M* (*a*-path in Figure I-7) is significant, 3) The effect of *M* on *Y* controlling for *X* (*b*-path in Figure I-7) is significant, and 4) the effect of *X* on *Y* controlling for *M* (*c'*-path in Figure I-7) is zero. Baron and Kenny (1986) argued for a set of similar tests with the major difference from Judd and Kenny being in the fourth condition. If all the other conditions are true and the fourth condition, $c' = 0$ is also true, then, the effect of *X* on *Y* is fully mediated through *M*. Although Judd and Kenny argue for a complete mediation scenario, Baron and Kenny relaxes the fourth condition so that the *c'*-path can have a value different than zero meaning that there can be partial mediation.

The Baron and Kenny approach should be credited for providing statistical tests to demonstrate that there is mediation by laying out the logical relationships among the *X*, *M*, and *Y*. However, there are some limitations with the Baron and Kenny approach. First of all, the Baron and Kenny approach cannot explain some particular cases of mediation such as an "inconsistent" mediation model where the indirect effect (*ab*) cancels out the direct effect (*c'*), $ab + c' = c = 0$. This is a violation of the first condition saying that the *c* coefficient must be significant. Another limitation is that the Baron and Kenny approach does not provide a statistic that represents the indirect effect and therefore, it is also difficult to quantify the size of the indirect effect.

B.1.b. The Joint-Significance Test

The joint-significance test can also be seen as a variant of the causal steps approach focusing on just the *X* to *M* and *M* to *Y* relationship. The joint-significance test consists of two independent statistical tests: 1) A test of the effect of *X* on *M* (*a*-path in Figure I-7), and 2) A test of the effect of *M* on *Y* (*b*-path in Figure I-7). The mediated effect is claimed to be significant only if the two tests are jointly significant. The joint-significance test also has a limitation that there is no direct estimate of the indirect effect. Nonetheless, the joint-significance test is simple to conduct and also evidence from simulation studies has revealed that the joint-significance test performs better than the Baron and Kenny approach in terms of Type I error and statistical power (MacKinnon, et al., 2002).

B.2. Evaluation of the *ab* Product Estimate

The indirect effect is often estimated by the product term of the *a* and *b* coefficients, *ab*. The following methods describe different statistical tests used to evaluate the *ab* estimate.

B.2.a. The Sobel Test

Sobel (1982) derives the standard error for *ab* using the multivariate delta method. Given that a vector of random variables is normally distributed, any function (such as *ab* indirect effect) of the random variables is asymptotically (as sample size approaches infinity) normally distributed by the multivariate delta theorem. The variance of the normal distribution of the new function (e.g., the *ab* coefficient) of random variables are given as $\nabla h(\cdot)^T \boldsymbol{\Sigma} \nabla h(\cdot)$ where, $\nabla h(\cdot)$ is the gradient (first derivative) of a given function, $h(\cdot)$, $\nabla h(\cdot)^T$ is the transpose of $\nabla h(\cdot)$, and $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the

original random variables. In the mediation model, we are interested in the function *ab*, which denotes the mediated effect thus, $h(\cdot) = ab$. Applying the multivariate delta theorem the following standard errors of *ab* are derived as,

$$SE(ab) = \sqrt{a^2\sigma_b^2 + b^2\sigma_a^2} \quad \text{(I.2.4)}$$

where σ_a^2 and σ_b^2 are the standard errors for the *a* and *b* estimates, respectively. Once *SE(ab)* is estimated, a *z*-statistic can be computed as $z_{ab} = \frac{ab}{SE(ab)}$. Then, a critical value (e.g., $|z_{crit.}|$=1.96 for a two-tailed test) given a Type-I error rate (e.g., $\alpha = .05$) is compared to the z_{ab} statistic to conclude whether the mediated effect, *ab*, is significant (e.g., significant if $z_{ab} > |z_{crit.}|$). Since the Sobel test is based on an approximation (the Taylor series approximation is used to prove the multivariate delta theorem) and is asymptotically true with large sample size, the test may not be accurate especially at small sample sizes. Also, as explained further in the next section, the distribution of the product (*ab*) is not normally distributed even though the sample size is large. In the simulations of MacKinnon et al. (2002) comparing different statistical tests for *ab*, the Sobel test had low Type-I errors overall and very low statistical power especially when the sample size was small as 50 and the mediated effect was small-medium.

B.2.b. Distribution of the Products Test

MacKinnon, Lockwood, and Hoffman (1998) introduced a method to test the indirect effect based on the actual distribution of the product of two normally distributed random variables. The distribution of the products test assumes that the *a* and *b* estimates are normally distributed, however the product term, *ab*, is not normally distributed. The distribution of *ab* is usually not normal such that the kurtosis and skewness indices depart

from those of a normal distribution (MacKinnon, Lockwood, & Williams, 2007). Since the distribution of the products is usually skewed, the confidence limits for *ab* are asymmetric.

The procedure for the distribution of the products method is as the following (Tofighi & MacKinnon, 2011):

1) Compute the standardized values for *a* and *b*, $z_a = \frac{a}{\sigma_a}$ and $z_b = \frac{b}{\sigma_b}$.
2) Compute the cumulative distribution function of $w_{ab} = z_a z_b$ using the method provided by Meeker and Escobar (1994)
3) Compute asymmetric confidence limits using the distribution from step 2.

The complicated evaluation of the w_{ab} distribution and the confidence limits can be obtained using statistical programs such as "PRODCLIN" (MacKinnon et al., 2007) and "RMediation" (Tofighi & MacKinnon, 2011).

MacKinnon et al. (2002) show that the distribution of the products test performed better (accurate Type I error rates and high statistical power) than most of the other methods (e.g., Baron and Kenny test, Sobel test) proposed in the study.

B.2.c. Bootstrap Methods

The statistical tests of the *ab* effect discussed above all assume some kind of distribution (usually a normal distribution except for the distribution of the products method) and evaluate the indirect effect from the presumed distribution. However, in some cases, the presumed distribution might not represent the population of the *ab* effect accurately. Another field of inferential statistics evaluates the population distribution from an empirical basis using resampling methods. The basic idea of bootstrapping a statistic (e.g., the *ab* effect) involves resampling a sample dataset (with replacement) *P*

times and compute the *ab* effect for each resampled dataset. Then, an empirical distribution can be drawn based of the *P* number of *ab* effects. The estimate, *ab* is evaluated using confidence limits from the empirical distribution. If the confidence limit does not contain zero, the *ab* estimate is significant, otherwise, the *ab* estimate is not significant.

The percentile bootstrap confidence limits are obtained using the method introduced by Efron and Tibshirani (1993). With a given Type I error rate, α, the (1- α) confidence limit is evaluated at $^{\alpha}/_{2}$ (lower limit) and $1 - \alpha/2$ (upper limit) percentile from the bootstrap distribution. For example, for a 95% confidence limit, the 2.5 percentile and 97.5 percentile values from the bootstrap distribution are used.

Another widely used bootstrap method in the mediation analysis literature is the bias-corrected bootstrap method. The bias-corrected bootstrap method corrects for the bias that is produced by estimating *ab* (usually the average value from the bootstrap distribution) instead of using the true population value of *ab*. In other words, the bias comes from using a sample to estimate the *ab* parameter instead of using the population values which we cannot always obtain. If the average bootstrap estimate is $\widehat{ab}_{bootstrap}$ and the true population value is $\widehat{ab}$, the bias is $bias[\widehat{ab}] = \widehat{ab}_{bootstrap} - \widehat{ab}$ (Efron, 1987). MacKinnon et al. (2004) incorporate a correction for the bias expressed by $\widehat{z_0}$, which is the *z* percentile score of the observed sample indirect effect. Then, the confidence limit is computed by $[2\widehat{z_0} + z_{\alpha/2},\ 2\widehat{z_0} + z_{1-\alpha/2}]$. MacKinnon et al. (2004) argue that the bias corrected bootstrap method performed better than other bootstrap methods. However, Fritz, Taylor, and MacKinnon (2012) argue that while the bias

corrected bootstrap method has high statistical power, it might be an effect of the increased-than-normal Type I error rate.

B.3. Causal Mediation Analysis using the Potential Outcomes Framework

Causal mediation analysis using the potential outcomes approach is an emerging area of research in mediation analysis. The potential outcomes approach (Rubin, 1974) can be explained by first focusing on an individual's observation. Suppose that a randomized trial (X) was conducted to a person where X=1 denotes that the person received the treatment and X=0 denotes the person received a control condition. Then, a measure of an outcome, Y can be expressed as $Y(X=1)$ if Y were measured after the person received the treatment condition and $Y(X=0)$ if Y were measured after the person received the control condition. The individual's causal effect is computed by taking the difference between the two potential outcomes, $Y(X=1) - Y(X=0)$. However, this comparison is not possible because a person cannot receive the treatment and control condition at the same time. Instead, the average causal effect is computed averaging across the individuals, $E[Y(X=1) - Y(X=0)]$, assuming that the treatment and control condition were randomly assigned to people. The potential outcomes framework applied to mediation analysis is important in two ways: 1) It clarifies the underlying assumptions of traditional mediated effects and 2) It provides a framework to estimate mediated effects in complex models including non-linear relationships between variables and also models with confounding variables.

Causal mediation analysis involves estimation of several different effects (Pearl, 2001; VanderWeele & Vansteelandt, 2009): 1) Controlled direct effects, 2) Natural direct

effects, and 3) Natural indirect effects. The average controlled direct effect (CDE) is the effect of *X* on *Y* controlling for *M* at a specific value.

$$CDE = E[Y(X = 1, M = m)] - E[Y(X = 0, M = m)] \qquad \text{(I.2.5)}$$

The average natural direct effect (NDE) is different from the CDE such that instead of *Y* evaluated at a specific value of *M*, *M* can have different values conditional on *X*.

$$NDE = E[Y\big(X = 1, M(X = x)\big)] - E[Y(X = 0, M(X = x))] \qquad \text{(I.2.6)}$$

Based on values of X, the NDE is further distinguished into the average pure natural direct effect (PNDE) when *X*=0 and the average total natural direct effect (TNDE) when *X*=1.

$$PNDE = E[Y\big(X = 1, M(X = 0)\big)] - E[Y(X = 0, M(X = 0))] \qquad \text{(I.2.7)}$$

$$TNDE = E[Y\big(X = 1, M(X = 1)\big)] - E[Y(X = 0, M(X = 1))] \qquad \text{(I.2.8)}$$

To understand Equations I.2.5 – I.2.8, note that the *X* value changes but the *M* value is fixed in the first and second expectations. Qualitatively, this is the definition of a "direct" effect where we are looking at the effect of *X* on *Y* without any influence of the *M* value since we are controlling *M* at a constant. Then, the difference between the controlled and natural direct effect is whether the *M* value is controlled at a fixed value not considering *X* (CDE) or *M* is a value conditional on *X* (NDE). Finally, the distinction between the PNDE and TNDE is that the PNDE evaluates the *M* value at *X*=0 (absence of treatment) and the TNDE evaluates the *M* value at *X*=1 (existence of treatment).

The natural indirect effect (NIE) is the effect of *X* on *Y* only through *M*. The following equation shows how the average NIE is quantified.

$$NIE = E[Y\big(X = x, M(X = 1)\big)] - E[Y(X = x, M(X = 0))] \qquad \text{(I.2.9)}$$

Since *X* is equal at a fixed value of *x* in both expectations in Equation I.2.9, the only difference in the expectations of *Y* are determined by the *M* values that are conditional on the *X* values (*X*=1 or *X*=0). Equation I.2.9 follows the qualitative definition of the indirect effect. That is, the indirect effect is the effect of *X* on *Y* that only operates through the conditional values of *M*. With similar logic shown in the PNDE and TNDE, the average pure natural indirect effect (PNIE) and the average total natural indirect effect (TNIE) can be quantified as the following.

$$PNIE = E[Y\big(X = 0, M(X = 1)\big)] - E[Y(X = 0, M(X = 0))] \qquad \text{(I.2.10)}$$

$$TNIE = E[Y\big(X = 1, M(X = 1)\big)] - E[Y(X = 1, M(X = 0))] \qquad \text{(I.2.11)}$$

The only thing that changes in Equation I.2.10 and Equation I.2.11 is the first argument in the *Y* expectations. The PNIE is the NIE when *X*=0 (absence of treatment) and the TNIE is the NIE when *X*=1 (existence of treatment). In other words, the PNIE expresses how much on average *Y* changes as *M* changes from *M* (*X*=1) to *M* (*X*=0) when *X* is controlled at *X*=0 and the TNIE is the average change of *Y* as *M* changes from *M* (*X*=1) to *M* (*X*=0) when *X*=1. As other studies do, the TNIE will be the focus of this study (Muthén & Asparouhov, 2014).

It is important to think about the assumptions related to the causal mediation with potential outcomes framework in order to obtain the estimated effects described above. The first basic assumption that applies to any potential outcomes causal model is the stable unit treatment value assumption (SUTVA). The SUTVA assumes that a person's outcome relies only on the treatment that the person was assigned and must not be affected by any other's treatment assignment. Other than the SUTVA, there are four

additional assumptions made for causal mediation (Pearl, 2001; VanderWeele & Vansteelandt, 2009).

1) There are no unmeasured variables that confound the relationship between *X* and *Y*.
2) There are no unmeasured variables that confound the relationship between *M* and *Y*.
3) There are no unmeasured variables that confound the relationship between *X* and *M*.
4) There are no unmeasured *M* to *Y* confounders that are affected by *X*.

Assumptions 1 and 2 are essential to estimate the CDE. Also, additional to the first two assumptions, assumptions 3 and 4 are essential to estimate the NDEs and NIEs. Assumptions 1 and 3 is satisfied if *X* is a treatment indicator and the treatment is randomly assigned. However, assumptions 2 and 4 are more difficult to meet since, although *X* is randomized, *M* is usually not randomized but rather, the values of *M* are presumably reflections of the individual's characteristics. Therefore, it is inevitable to have some degree of confounding effect between *M* and *Y*. In this paper, sequential ignorability (Imai, Keele, Tingley, & Yamamoto, 2011) is assumed. Sequential ignorability has two parts: 1) There are no unmeasured confounders that affect the *X* to *M* and *X* to *Y* relationship, and 2) There are no unmeasured confounders that affect the *M* to *Y* relationship after controlling for treatment and baseline (before treatment) measures of the mediator. This assumption is a strong assumption that is in fact, used in the single mediator model of Figure I-7. In other words, this paper assumes that there are no

confounders of any part of the single mediator model and the model in Figure I-7 is the true model.

The evaluation of the TNIE can be done by applying bootstrapping methods as described in the previous Section B.2.c. The bootstrapping methods are useful for the causal mediation models since it could be applied to nonlinear models where the mediator and/or outcome is non-normal (*e.g.*, categorical, survival). Trying to use the normal distribution assumptions for the TNIE can be inappropriate because the distributional assumptions will not hold for models with non-normal mediators and/or outcomes.

C. Comparison of the Methods

MacKinnon and his colleagues (2002, 2004, 2007) have compared the performance (in terms of Type I error, statistical power, and required sample size) of different statistical methods through simulation studies. For mediation models with a continuous mediator and a continuous outcome, among the various methods, the joint significant test, distribution of the products test, and bias-corrected bootstrap had superior performance to the other proposed methods (Fritz and MacKinnon, 2007). The Baron and Kenny approach resulted in low Type I error rate and low statistical power (MacKinnon et al., 2002), requiring a large sample size to achieve .8 power with small effect sizes (Fritz & MacKinnon, 2007). The Sobel test produced lower than the stated normal Type I error rate and low statistical power for small sample size and small to medium effect sizes (MacKinnon et al., 2002); the required sample size was larger than the recommended methods (Fritz & MacKinnon, 2007). In general, bootstrap methods

performed better than methods based on distributional assumptions (i.e., the distribution of the products method) and the bias corrected bootstrap performed a little bit better than the percentile bootstrap method (MacKinnon et al., 2004) in terms of statistical power. However, it is noted that the bias corrected bootstrap may produce spuriously high statistical power because of the increased Type I error rate (Fritz et al., 2012). There is not yet a study comparing the performance of testing the natural indirect effects with the other traditional statistical methods to test the indirect effect. This is probably due to the fact that the causal mediation analysis is a relatively new area of research and papers have been more focused on methods of estimating the natural indirect effects in various different conditions (e.g., introducing different types of confounders, non-linear relationship between variables).

Conclusion of the Mediation Analysis Section

Statistical mediation analysis can answer questions about "how" (the process or mechanism) an independent variable affects a dependent variable via a mediator variable. The mediated effect is the effect of an independent variable, *X*, on a dependent variable, *Y*, attributable to the mediator variable, *M*. Among the various ways to evaluate the mediated effect, six of them will be considered in this dissertation for the methods to evaluate the mediated effect for the survival-mediator model: 1) The Sobel test, 2) The distribution of the products test, 3) Percentile bootstrap, 4) Bias-corrected bootstrap, 5) The joint-significance test, and 6) Causal mediation using the potential outcomes framework (using bootstrapping). In this section, the different mediated effect evaluation methods were discussed and performance of these methods were illustrated in the case of

a single mediator model with a continuous mediator and a continuous outcome. In the next section, mediation models with survival variables as a dependent variable or as a mediator will be discussed.

3. Survival Mediation Analysis

In this chapter, mediation analysis with survival variables is reviewed. Survival variables can be an outcome, a mediator or both in the mediation model. The survival events that are studied in biomedical and engineering research usually lead the subjects (participants or other physical objects) to a terminal state (i.e., death, failure of a system or machine). If the event of interest brings an end to the research subject, then the survival variable (the time to event and whether that event occurred to a subject) is often used as a dependent variable (outcome) and a set of covariates is used to predict the survival variable. However, the event of interest in psychology can be a non-terminal state, such as smoking relapse. If the event of interest does not mean eternal termination of a subject (a participant or a part of a machine), there is a chance that the survival variable can be used as a predictor of another outcome (e.g., lung cancer) as well as being used as an outcome to regress on a predictor (e.g., participation or not in a prevention program to cease smoking).

First, research on mediation models where the outcome is a survival variable is reviewed. Then, a mediation model where the mediator is a survival variable is illustrated. When a survival variable is used as a predictor, the censored values cause bias in the estimation of some parameters in the mediation model. Methods of treating the censored data will be discussed. Finally, the causal mediation framework reviewed in the previous chapter (see Chapter 2 Section B.3.) will be applied to the case when the mediator is a survival variable. The derivation of the natural effects will be shown in depth for the survival-mediator case.

A. Mediation with a Survival Outcome

If the event of interest is a terminal state such as death, the survival variable is usually the outcome of a mediation model. There might be exceptions in a prospective study where time-to-death is a predictor of a future outcome (e.g., time-to-death since debut as a singer predicting post-death album sales). Nonetheless, the common case is that a survival variable concerning a terminal state is used as an outcome. Mediation analysis of a survival outcome has been an active area of research especially in epidemiology. Epidemiologists investigate the effect of predictors on the survival rate (or hazard rate) and are interested in identifying mediators of the predictor-survival outcome relationship. For example, Jung et al. (2012) studied the effect of age on survival after metastatic breast cancer being mediated by a comorbidity measure (the Charlson comorbidity score (CCS) plus hypertension). In their study, a Cox regression was used to study the relationship between age and hazard rate of death controlling for comorbidity. The Baron and Kenny method and the hazard rates with and without the mediator were examined to evaluate mediation. They concluded that age was no longer a significant predictor of the hazard rate after controlling for comorbidity, especially including hypertension.

Another active area of research in mediation with survival outcomes is causal mediation using the counterfactual approach. Lange and Hansen (2011) argue that the traditional method in epidemiology comparing the hazard rates between a Cox model with a mediator and without a mediator cannot be given a causal interpretation and is mathematically inaccurate to compare the two models. Instead, Lange and Hansen use the counterfactual approach and formulate the potential outcomes to estimate the natural

direct and natural indirect effect in a model that has a survival outcome. They also provide standard errors of the estimates in order to conduct statistical tests. As an illustration of their proposed method, Lange and Hansen reanalyzed data from Christensen, Labriola, Lund, and Kivimäki (2008). The effect of socioeconomic position (SEP-5 groups: executive manager and/or academics, middle managers and/or 3-4 years of further education, other white-collar workers, skilled blue-collar workers, and semi-skilled or unskilled workers) on onset of a long term sickness absence in workplace was mediated by work environment (physical and psychosocial factors). VanderWeele (2011) derived natural direct and indirect effects for survival models that are widely used in the survival literature, that is the Cox model and the accelerated failure time (AFT) model. VanderWeele gave a proof in the online Appendix that the test of the product method (*ab*) provides a valid test so that the mediated effect can be evaluated although the value itself does not provide an accurate measure of the indirect effect. Fulcher, Tchetgen and Williams (2016) examined the AFT model and provided empirical evidence that the difference in coefficients (*c*-*c'*) method would fail even when the product of coefficients (*ab*) method is unbiased, especially when there is censoring and an other-than-normal (*e.g.*, Weilbull) distribution is specified in the AFT model. Gelfand, MacKinnon, DeRubeis, and Baraldi (2016) compared the performance of the AFT model with the Cox model. In general, the AFT model outperformed the Cox model in terms of Type I error rate and statistical power of the natural indirect effects.

A.1. The Single Mediator Model with a Survival Outcome

Formally, the single mediator model with a survival outcome can be expressed with two regression equations: the *M*-regression and the *Y*-regression. Here, the mediator

is assumed to be a continuous variable and thus, the *M*-regression can be expressed as a normal regression equivalent to Equation I.2.1. The *Y*-regression is the regression analogous to Equation I.2.2 where *X* and *M* are both included as predictors of *Y*. However, since the outcome is a survival variable, the *Y*-regression requires the use of other than OLS regression to avoid biased estimates. One possible choice for the *Y*-regression is the Cox model since it does not require assumptions about the baseline hazard rate. In other words, Equation I.1.27. can be used as the following:

$$h(t) = h_0(t)\exp(c'X + bM) \qquad \text{(I.3.1)}$$

where *h(t)* is the hazard rate at time *t*, $h_0(t)$ is the baseline hazard rate at time *t*, *c'* and *b* are the corresponding hazard rates for *X* and *M*.

The *Y*-regression with only *X* as a predictor (analogous to Equation I.2.3) is omitted because the difference in coefficients method (*c*-*c'*) is less accurate than the product of coefficients method (*ab*) when a non-normal variable (e.g., categorical, survival) is included in the model. A non-normal variable used as an outcome causes issues that some of the parameters in the model are fixed at a particular value to identify the model. For example, the residual variance of a binomial logistic regression is fixed at $\frac{\pi^2}{3}$ in order to estimate the threshold and slope. This constraint causes problems that the scales are not the same across different models. The same kind of issue is pertinent with a survival outcome where the specified error hazard function is not normally distributed in an AFT model (Fulcher et al., 2016).

B. Mediation with a Survival Mediator

In social science, the survival event is not always a terminal event. Some examples of events in social science are reemployment (Kiefer, 1998; Lancaster & Nickell, 1979; Lancaster, 1980), recidivism (Sherman & Berk, 1984), smoking relapse (Stevens, & Hollis, 1989), relapse of affective disorders (Lavori, Keller, & Klermaet al., 1984), first onset of an affective illness (Rice et al., 1987), and absence at work (Finchman, 1989). The duration of time before the occurrence of these events can be an outcome that may be regressed on a set of predictors (e.g., treatment); however, the duration of time before occurrence of the events can be a used as a predictor of future outcomes. For example, the duration of being unemployed can predict work competence upon reemployment, the time after quitting smoking until relapse can predict health outcomes, and the incubation period of an affective illness can predict future conflicts with family members.

If a survival variable is used simultaneously as an outcome and as a predictor in a model, the proposed model can be called as a survival mediator model. In the following sections, a single survival mediator model is formally introduced and some of the issues with a survival mediator model are discussed. Of importance, the censored predictor issue in the *Y*-regression is discussed and the natural direct and indirect effects are derived using the potential outcomes approach in causal mediation.

B.1. The Single Survival-Mediator Model

Similar to the mediation model with a survival outcome, the single survival mediator model consists of two regressions: the *M*-regression which regresses the *M* variable on *X* and the *Y*-regression which regresses the *Y* variable on both *X* and *M*

simultaneously. In case of a mediation model where *M* is a survival variable and *Y* is a continuous variable, a Cox model can be used for the *M*-regression. Using the notation used in Equation I.1.27 for the *M*-regression,

$$h(t) = h_0(t)\exp(aX) \qquad \text{(I.3.2)}$$

where $exp(a)$ denotes the hazard ratio between the treatment group (X=1) and the control group (X=0) when X is a treatment indicator. The Cox model for the *M*-regression correctly controls for the probability of the censored cases while estimating the hazard ratios (see Chapter I, 1. Survival Analysis for details).

The remaining part of the survival mediator model is the *Y*-regression. If *Y* is a continuous variable, the *Y*-regression is an OLS-regression (analogous to Equation I.2.2) with the complication that *M* is a censored predictor in the model. A problem that arises with a censored predictor in the *Y*-regression is explained in the next section.

B.2. Censored Predictor in the *Y*-regression

The censored values for predictor *M* in the *Y*-regression can cause bias in estimating the regression coefficients. When there is a right-censored predictor in the model, the regression coefficient can be inflated, either positively or negatively, compared to what the coefficient would be with known values of the predictor. Figure I-8 shows a hypothetical example of what may happen with a censored predictor in a model. In Figure I-8, the solid dots represent the data points with true *M* and *Y* values, the hollow dots are the *Y* values at censored *M* values (*e.g.*, time at the end of observation), line-A depicts the regression line when the known true values of *M* are used in the *Y*-regression, and line-B represents the regression line when the censored values are used for *M*. The slope of line-B is greater than line-A because of the high *Y*-values at

censoring time of *M*. Therefore, the regression is biased in line-B when the censored values of *M* is used in the *Y*-regression.

B.2.a. Treating Censored Values as a Missing Data Problem

In practice, we do not know the true value of *M* after a certain time point (*e.g.*, time at the end of study). One approach is to use the last known observed time (*e.g.*, time at the end of the study). As demonstrated in Figure I-8, the regression coefficient can be biased using the last known observed time. Another approach is to treat *M* as having a missing value instead of the last known observed time value. Then, the censored predictor issue can be viewed as a missing data problem. Schafer and Graham (2002) introduced three different mechanisms of missing data (also see Enders, 2010 for details): missing completely at random (MCAR); missing at random (MAR); and not missing at random (NMAR). The MCAR mechanism assumes that the missingness is totally independent of any other possible variables in a dataset. The MAR mechanism assumes that the missingness is random only after controlling for a set of variables. In other words, the missing values of a variable can be predicted by other variables in a dataset and after the prediction, the error would be random. The NMAR mechanism assumes that the missingness of a variable is attributable to the values of the variable itself. For example, a person who drinks a lot may be too drunk to answer a questionnaire about drinking problems.

Different missing data techniques can be used depending on the missing mechanism (see Enders, 2010 and Schafer & Graham, 2002 for more details). If the MCAR assumption is true, listwise deletion (a.k.a. complete case analysis) can be used. Listwise deletion uses only data with non-missing values for all the variables in the

model. By definition of the MCAR assumption, the distribution of the missing values is equivalent to the distribution of the non-missing values and therefore, the model estimates using only non-missing observations are consistent to have using the full data. If the MAR assumption is true, modern missing data techniques such as multiple imputation (MI) and (full information) maximum likelihood (FIML; or just ML) are recommended. Both methods can include auxiliary variables that help predict the missing values and model estimates are consistent after controlling for auxiliary variables and variables in the model. The FIML missing data method is widely used in research because its implementation in statistical software such as Mplus (Muthén & Muthén, 1998-2015). In Mplus, the FIML is the default estimator and therefore researchers do not need to make an extra effort to deal with missing data. MI is also widely used and implemented in statistical software (e.g., SPSS, SAS and Mplus) but multiple steps (1. generate multiple imputed datasets, 2. run statistical analysis for each of the datasets, and 3. merge the results from each dataset) are needed. At last, if the NMAR assumption is true, the missing data mechanism model needs to be included into the main analysis. However, since we do not know the true missing data model, a sensitivity analysis is recommended.

B.2.b. A Simulation Study Addressing the Censoring Predictor Issue

In this section, I report the results of a small simulation study conducted to examine the performance of different methods to deal with the censoring predictor issue in the *Y*-regression (Equation I.2.2) of a survival mediation model. The different methods that are compared are 1) OLS regression using true values of *M*, 2) OLS regression using censored values of *M*, 3) treating censored *M* values as missing and using FIML for the

Y-regression, and 4) treating censored *M* values as missing and OLS regression with non-missing observations only (Complete Case). Method 1 is expected to produce unbiased *Y*-regression estimates (*i.e.*, *b* and *c'*) since the exact values are known that generated the data. That is, the correct *M*-values that generated the *Y*-values based on the *Y*-regression model are used for model estimation and thus the model estimates are expected to resemble the true parameter values within statistical error. On the other hand, method 2 using the censored *M*-values would be expected to produce biased estimates because of the shift of *Y*-values at the censoring *M*-value (see Figure I-1). The most interesting part of this simulation study is the comparison of the two missing-data approaches. It would be helpful to see which of the two methods would produce statistically better (in terms of Type I error and statistical power) results and resemble the estimates derived from Method 1. Theoretically, method 3 using FIML would not lose any observations in the data analysis and thus is expected to display more statistical power than method 4 using only complete cases.

For the true population model, a single survival-mediator model with small effect size path coefficients for the *a*-path and *b*-path (both .14) and a zero *c'*-path was specified. Predictor *X* was assumed to be a continuous variable following a normal distribution with a mean of zero and a standard deviation of one. The true *M*-values were generated from an exponential hazard (λ = .1) baseline Cox model with *X* as a predictor (*a* = .14). Under the MCAR assumption, the censoring values were generated from an independent (random) censoring process following an exponential function (λ = .043). Then, the censored *M*-values were evaluated at the true timing value when the true timing value was smaller than the censoring value and evaluated at the censoring value when the

true timing value was equal or larger than the censoring value. The censored observations were adjusted to be approximately 30% of the sample. Finally, the *Y*-values were generated from the true regression model with *X* (*c'* = 0) and *M* (*b* = .14) as predictors. The true *M*-values were used in generating the *Y*-values. Therefore, the generated *Y*-values are the values assuming we knew all the true *M*-values (without any censoring) and the *X* values. One thousand replication datasets were generated with a sample size of 300 for each of the datasets. Generation of data and all analyses were done in SAS 9.4 (SAS Institute Inc, 2014). Using a Type I error rate of .05 for the analyses, the statistical power of the *b*-coefficient and the Type I error rate of the *c'*-coefficient was evaluated from the 1,000 replications as well as the average estimate. Since the true *b*-value was .14, *b* should be statistically significant (different from zero) in the replicated analyses. Statistical power was determined as the percentage of significant *b*s out of 1,000 replications. On the other hand, the true value of *c'* in the population model was zero. Therefore, Type-I error was determined as the percentage of significant *c'*s (false alarms) out of 1,000 replications.

Table I-4 summarizes the results from the simulation study using the four different methods. First of all, as expected, using the true *M* values (method 1) produced accurate estimates of *b* (mean = 0.14) and *c'* (mean = -0.001) with 100% statistical power and 5% Type I error, respectively. The censored *M* values (method 2) produced problems with the *c'* parameter; biased estimates of *c'* (mean = -0.11) were obtained and the Type I error was 24.5% compared to the nominal α = .05. However, the estimates for *b* were unbiased (mean = -.001) and the statistical power for *b* was 100%. Among the missing data techniques, the complete case method (method 4) worked better than the

FIML method (method 3). Method 4 worked as well as method 1; average $b = 0.14$ with statistical power of 100% and average $c' = -0.001$ with 5% Type I error. However, method 3 produced a relatively more biased estimate of b (mean = 0.17) and the statistical power drastically dropped to 12.6%. Also, there was relatively more bias in the c' estimate (mean = -0.02) and the Type I error was smaller than expected (should be 5%) at 0%.

In conclusion, given MCAR, the small simulation study suggested that the complete case method (or equivalently, the listwise deletion method) can produce unbiased estimates for b and c' for the Y-regression. This was expected since listwise deletion produces unbiased estimates when the MCAR assumption is satisfied (Enders, 2010). However, unexpectedly, FIML produced biased estimates in the small simulation study where MCAR was assumed. Normally, FIML is known to produce unbiased estimates when the missing data mechanism is MCAR or MAR (Enders, 2010). The unexpected result seems to be an artifact of the simulation where the only variables that help estimation of the missing M values are the X and Y variables. Although these variables should not have any relationship with the missingness of M under the MCAR assumption, they seem to have pulled the missing values of M towards their favor since they are the only variables that are in the model and thus causing bias in the Y-regression. The results also supports Paul Allison's argument that the listwise deletion method is robust to the missing data mechanism assumptions (MCAR or MAR) when there is only missing data in the independent variable (a.k.a. predictor) in a linear regression (2001). Therefore, in the main simulation proposed in the next chapter, complete case analysis will be used for the Y-regression assuming that the missing data mechanism is MCAR.

B.3. The Natural Indirect Effect of a Single Survival Mediator Model

In this section, causal mediation using the counterfactual approach (see Chapter I, 2. Mediation Analysis, Section B.3.) is applied to the situation where the mediator is a survival variable and the outcome is a continuous variable. In particular, derivation of the natural indirect effect (derivations of other effects can be done using similar logic) is shown using the "law of iterated expectations" (VanderWeele, 2011) and the "Darth Vader rule" (Muldowney, Ostaszewski, & Wojdowski, 2012). First, the law of iterated expectations is described in the following equation,

$$E[Y] = \int_x E(Y|X)P(X)\mathrm{d}x \qquad \text{(I.3.3)}$$

Intuitively, Equation I.3.3 integrates across all *X* values to compute the marginal expectation of *Y*. Equation I.3.3 applied to the single mediator model becomes,

$$E[Y|M, X = x] = \int_m E\big(Y\big|M(X = x)\big)P(M(X = x))dm \qquad \text{(I.3.4)}$$

In Equation I.3.4, the expectation of *Y* is computed by integrating through all possible values of *M* given a fixed value of *X*=*x*. Next, the Darth Vader rule is expressed as,

$$E[M] = \int_0^\infty S(M)\, dm \qquad \text{(I.3.5)}$$

where *S*(*M*) is the survival function of *M* (see Equation I.1.5 for definition of the survival function). In other words, the expected value of a survival variable is computed by integrating the survival function across all possible values of *M*.

The following derivations are based on a strong set of assumptions: 1) No confounders in the single mediator model (Figure I-7), and 2) Sequential ignorability (see Chapter I, 2. Mediation Analysis, Section B.3) is true. Starting from the definition of the

NIE given in Equation I.2.9, using Equation I.2.2 for the *Y*-regression along with the law of iterated expectations and the Darth Vader rule provides,

$$NIE = E[Y\big(X = x, M(X = 1)\big)] - E[Y(X = x, M(X = 0))] \quad \text{From Equation I.2.9}$$

$$= \int_m E[Y|X = x, M = m]P(M|X = 1)$$

$$- \int_m E[Y|X = x, M = m]P(M|X = 0)$$

$$= \int_m (bm + c'x)P(M|X = 1) - \int_m (bm + c'x)P(M|X = 0)$$

$$= \int_m bmP(M|X = 1) + c'x - \int_m bmP(M|X = 0) + c'x$$

$$= b\{E[M|X = 1] - E[M|X = 0]\}$$

$$= b\left\{\int_0^\infty S(M|X = 1)dm - \int_0^\infty S(M|X = 0)dm\right\} \quad \text{(I.3.6)}$$

The equation in the second step is the direct application of the law of iterated expectations. In the third step, Equation I.2.2 is applied for the *Y*-regression (the expected value for *Y*). The fourth step is rearrangement of the third step. From step four, note that only the first part in the integrals are relevant to the integral function. The latter part of the integrals, *c'x* is a constant value and thus is eliminated in the fifth step. Note that because of the elimination of *c'x*, the estimates of the TNIE and the PNIE do not depend on the *X* value and thus, the two estimates are equivalent. Also, using the definition of expectations, $E(X) = \int xf(x)dx$, the equation in step 4 reduces to the equation in step 5. Finally, the Darth Vader rule is applied and gives the result of Equation I.3.6. After substituting the survival functions in Equation I.3.6 with a given

model (*e.g.*, Cox model), further computation of the integrals is mathematically extremely difficult if not impossible (involves integration of double exponential functions). From Equation I.1.11, the survival functions are expressed as an exponential function of the negative integral of the hazard function. Substituting the hazard function (*i.e.*, Equation I.3.2) in the survival function results in a complicated function that includes integrals of an exponential function of another exponential function. Although, a simple form of the NIE is difficult to derive, the NIE can be empirically evaluated through Equation I.3.6. Equation I.3.6. involves two major steps, 1) Estimation of the *b* parameter from the *Y*-regression using the complete case method described in Chapter I, 3. Survival Mediation Analysis, Section B.2, and 2) Compute the difference of the summed survival functions for a given *X* value (*X*=1 and *X*=0). For a given *X* value, the second step involves: estimation of the survival functions at each time point and then calculating the area under the curve[3] across the time points. After completing the two major steps, the estimates from the two steps are multiplied to get an estimate of the NIE.

Since the NIE estimate itself is a complicated function already, the standard error of the NIE would be a complicated or else an intractable solution. Therefore, instead of trying to estimate a test statistic (*t* or *z* statistic) and evaluate the NIE using the test statistic under certain known distributions (*t* or *z* distribution), a bootstrap method is employed to evaluate the NIE. An empirical distribution is generated by bootstrapping the NIEs. Then, a (1-α) confidence limit is computed for the empirical distribution. If

[3] In this study, the area under the survival curves were estimated by a numerical integration approach using the trapezoidal rule. For a given small time interval (*a*, *b*), the area under the curve can be estimated by calculating the area of a small trapezoid. The time difference, (*b* - *a*) is multiplied by the average survival function evaluated at the two time points, (S(*a*)+S(*b*))/2. Then, the whole area under the curve is computed by adding up all the trapezoids over all possible time intervals in the data.

the confidence limit contains zero, the NIE is not significant, otherwise, the confidence limit does not contain zero, the NIE is significant.

Conclusion of the Survival Mediation Analysis Section

In a mediation model, the mediator or the outcome can be a survival variable. The focus of this dissertation is a mediation model with a survival mediator. A Cox model can be used for the *M*-regression which gives an unbiased estimate of the *X* on *M* effect without having to specify a baseline hazard function. However, biased estimates can be obtained in the *Y*-regression because of the censored *M* predictor in the *Y*-regression. One method to correct this problem is to treat the censored *M* values as missing and use modern missing data techniques. A small simulation study showed evidence that complete case analysis produced unbiased estimates for the *Y*-regression assuming the censoring data mechanism is MCAR. This section has also shown derivation of the NIE for the single survival-mediator model. In contrast to the *ab* estimate of the mediated effect, the NIE provides an accurate estimate of the indirect effect for the survival-mediator model. More about the *ab* mediated effect and the NIE will be discussed in the Discussion chapter.

In the next chapter, a simulation study is proposed to evaluate performance of different mediated effect testing methods for the single survival-mediator model.

II. METHOD

1. Purpose of the Simulation Study

The purpose of this Monte-Carlo simulation study was to compare the performance of different methods of evaluating a mediation effect (i.e., indirect effect) in a single survival mediator model. There are various methods to evaluate the indirect effect such as the joint-significance test (evaluating the *a* and *b* coefficients separately; MacKinnon et al., 2002), the Sobel test (Sobel, 1982), distribution of the products test (MacKinnon, Lockwood, & Hoffman, 1998), bootstrapping methods (MacKinnon, Lockwood, & Williams, 2004; MacKinnon, 2008), and causal mediation methods using the counterfactual approach (Muthén, 2011; Pearl, 2001; VanderWeele & Vansteelandt, 2009). Most of this work has focused on mediation models with continuous mediators and outcomes. However, less research has been done investigating a survival mediation model and comparing performance of different methods to evaluate the indirect effect for a single survival mediator model. The current study is important in two ways: 1) providing accurate statistical methods for estimating a single survival mediator model and 2) providing empirical comparison of different methods evaluating the indirect effect for a single survival mediator model.

Figure II-1 shows the single-mediator model where *X* is a binary variable (0=control and 1=treatment), *M* is a survival variable, and *Y* is a continuous variable. Fitting the single-mediator model in Figure II-1 consists of fitting two regressions: the *M*-regression and *Y*-regression. The *a*-path coefficient is estimated from the *M*-regression and the *b*-path and *c'*-path coefficients are estimated from the *Y*-regression. If the *M*-variable is a timing variable, the *M*-regression is a typical survival analysis model using

for example a Cox regression. In case of censored values for the *M* variable, it is crucial to use a Cox regression instead of an OLS regression because the Cox regression appropriately adjusts for bias in the regression estimates, whereas OLS regression would not be free from bias.

Since the *Y*-variable is a continuous variable in the model in Figure II-1, the *Y*-regression can be appropriately estimated with OLS regression using the *X* and *M* variables as predictors. One of the complications of a survival mediator is that the *b*-path or/and *c'*-path can be biased when using the original *M*-variable with censored values in the *Y*-regression. As a fix to this issue, using only the uncensored cases for the *Y*-regression is proposed. As discussed in the previous chapter (Chapter I, 3. Survival Mediation Analysis, Section B.2.b), using only the uncensored cases for the *Y*-regression works well assuming the MCAR assumption holds for censoring values of *M*.

After fitting the appropriate statistical models for the single survival mediator model, the indirect effect can be evaluated using different methods. Namely, the different methods of interest are the joint-significance test, Sobel test, distribution of the products test, percentile bootstrap of *ab*, bias-corrected bootstrap of *ab*, and percentile bootstrap of the natural indirect effect (NIE). Simulation studies comparing different indirect effect test methods for a mediation model with a continuous mediator and a continuous outcome has shown that: 1) the Sobel test produces very low statistical power with a small effect size and small sample size, 2) the bias-corrected bootstrap of *ab* produces higher than nominal level (α = .05) Type I errors and inflated statistical power in some cases, and 3) the distribution of the product test, joint-significance test, and percentile bootstrapping perform best among the methods. Performance of the causal

mediation approach is unclear because previous simulation studies did not include this approach. Although performance of the different indirect effect tests for a continuous mediator and outcome model is revealed from previous simulation studies, less is known about how these different methods work in a single survival mediator model. The following proposed simulation study provides empirical evidence of how well these approaches would perform under certain conditions.

2. Overview of the Simulation Procedure

The basic model structure used for the simulation study is a single-mediator model where the mediator is a survival variable and the outcome is a continuous variable. Figure II-1 shows the basic mediation model used for the simulation study. Five different factors were manipulated for data generation. The five factors were 1) the size of the *a*-path (0, 0.21, or 0.52), 2) the size of the *b*-path (0, 0.14, or 0.39), 3) the size of the *c'*-path (0 or 0.39), 4) the censoring proportion of the survival variable (0% or 30%), and 5) the sample size (150 or 300). Details of these factors are given in the next section. Manipulation of the first five factors yield 72 (3x3x2x2x2) different combinations of true population values of the proposed model. One thousand datasets were created under each of these different combinations. Then, for each of the datasets, the model in Figure II-1 was fitted to each generated dataset. The single-mediator model consists of fitting two regression models: the *M*-regression and the *Y*-regression. The Cox model was used for the *M*-regression and OLS regression was used for the *Y*-regression. To obtain unbiased estimates for the *Y*-regression, the censored values of *M* were treated as missing. Then, assuming the missing mechanism is MCAR, only the complete cases (uncensored

observations) were used in the OLS *Y*-regression. After fitting the mediation model, the performance of six different indirect test methods (Sobel test, the distribution of the product test, the percentile bootstrap, the bias-corrected bootstrap, the joint-significance test, and the bootstrap test of the total natural indirect effect) were compared. The performance was evaluated with 1) empirical Type I error rate, 2) empirical statistical power, 3) parameter coverage rate, 4) average raw bias, 5) average relative bias, 6) sign of the bias (whether it is a negative or positive bias) and 7) mean squared error (MSE). There were 6 factors in total, 3 (*a* path size) x 3 (*b* path size) x 2 (*c'* path size) x 2 (censoring proportion) x 2 (sample size) x 6 (indirect effect test method) = 432 conditions. Based on the metric of the outcomes, logistic regression or analysis of variance (ANOVA) was employed to examine the effects of the factors. Interaction terms up to 3-way interactions were included in the logistic regression or ANOVA model (discussed later). A flow chart of the simulation study procedure is given in Figure II-2.

3. Factors of the Simulation Study

In this section, the six factors in the simulation study are discussed in detail. The first five factors determine the true parameter values of the data-generating model and the sample size of the datasets. Namely they are, 1) the size of the *a*-path, 2) the size of the *b*-path, 3) the size of the *c'*-path, 4) the censoring proportion of the *M*-variable, and 5) the sample size. These five factors were included in the simulation study to control for their effects and to examine whether there were conditions where one indirect test method performs better than another method.

The sixth factor and the main interest in this simulation study is the method used to evaluate the indirect effect. Once each dataset was created and the proposed survival-mediator model was fitted to each of the datasets, six different methods were used to evaluate the indirect effect. Namely they are, 1) the Sobel test, 2) the distribution of the products test, 3) percentile bootstrap, 4) bias-corrected bootstrap, 5) joint-significance test, and 6) percentile bootstrap of the NIE.

A. Size of the Regression Coefficients (*a*, *b*, and *c'*)

The magnitude of the regression coefficients *a*, *b* and *c'* was manipulated in the simulation study to see whether the results changed as the magnitude of the indirect effect (*ab*) changed.

The *Y*-regression is an OLS regression with *X* and *M* as predictors in the model. Coefficient *b* is the regression of *Y* on *M* controlling for *X* and coefficient *c'* is the regression of *Y* on *X* controlling for *M*. The magnitude of the *b*-path was manipulated to have values of 0, 0.14 or 0.39 in the population to approximately match Cohen's recommendations of zero, small (2% of explained variance in the outcome) and medium (14% of explained variance in the outcome) effect sizes, respectively (MacKinnon et al., 2002). For the *c'*-parameter, two levels, 0 and 0.39 were selected to each represent a situation when there was zero direct effect (fully mediated model) and when there was a direct effect with a medium effect size.

The *a*-parameter is the regression coefficient in the *M*-regression. In this study, the true model for *M*-regression was a Cox regression with *X* as a predictor. The *a*-coefficient in a Cox regression is interpreted as a log hazard ratio and the exponential value of *a* is the hazard ratio. The *a*-values used for the current simulation study were 0,

0.21, or 0.52 (corresponding hazard ratios, exp(a) = 1, 1.23, or 1.69) which correspond to Cohen's guidelines of zero, small, and medium effect sizes. The following procedure was used to determine the effect sizes for *a*. These effects were estimated based on Hsieh and Lavori (2000) formula of calculating sample size required for a Cox regression model.

From Hsiesh and Lavori (2000), the number of event occurrences (i.e., number of deaths in their example), *D*, for a Cox regression model is:

$$D = (Z_{1-\alpha} + Z_{1-\beta})^2[\sigma_x^2 a^2]^{-1} \qquad \text{(II.1)}$$

where, $Z_{1-\alpha}$ is the standard normal deviate at the desired one-sided significance level α, $Z_{1-\beta}$ is the standard normal deviate at the desired power $1 - \beta$, σ_x^2 is the variance of a predictor variable *X*, and a is the regression coefficient related to a unit increase in *X*. Once *D* is computed for a given α-level, $1 - \beta$ level and effect size, a, then, the required total sample size is *D*/*P*, where *P* is the overall death rate. The overall death rate, *P* is the number of deaths in proportion to the total sample.

From Equation II.1, the equation can be rearranged relative to a,

$$a = \sqrt{\frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{D\sigma_x^2}} \qquad \text{(II.2)}$$

In Equation II.2, we can use $Z_{1-\alpha/2}$ instead of $Z_{1-\alpha}$ to evaluate *a* based on a two-tailed test instead of a one-tailed test.

Meanwhile, for a two-sided test to detect a regression estimate with an effect size of Cohen's *d* = 0.2 and 0.5 (small and medium) with α = .05, $1 - \beta$ = .80, the required sample size is respectively, *n* = 788 and 128 using a sample size software, GPower 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). Taking these sample sizes into a survival data

and assuming 10%[4] of data are right-censored, the number of event occurrences is respectively, D = 709.2 and 115.2.

X is a binary variable in which 50% of the cases have a value of 0 (control group) and 50% of the cases have a value of 1 (treatment group). The variance of the binary predictor, σ_x^2 = 0.5(1-0.5) = 0.25. The required effect size for α = .05, $1 - \beta$ = .80 and respective D values (709.2 and 115.2) can be calculated by plugging in these numbers in Equation II.2. This results in a = 0.2104 and 0.5221, corresponding to the D values of 709.2 and 115.2, respectively. These numbers translate into respective hazard ratios of $exp(a)$ = 1.2342 and 1.6856.

B. Censoring Proportion of the *M* Variable

In this study, the mediator variable is a survival variable. One important aspect of a survival variable is that censoring may exist in the data. Right-censoring is the most frequently observed type of censoring in real data and therefore, only right-censoring is considered in this study. Right-censoring typically occurs when the observation period has reached its end but also it can occur for other reasons such as attrition. Therefore, rather than using a fixed time point that indicates the end of observation, the censoring time is assumed to follow a specific distribution. In this study, the censoring time was assumed to follow an exponential distribution with a parameter of λ_C. That is, the hazard rate of censoring in this study is a constant value of λ_C over the entire time interval.

The censoring proportion which is the ratio of the number of censored observations to the total observations is manipulated to have about 0% or 30% censoring

[4] Ten percentage of censoring was arbitrary decided based on the author's experience with survival data as a mediator. Higher percentage of censoring will require larger effect size of *a* to achieve a specific value of statistical power (e.g., *power*=.80) with a given Type I error rate (e.g., α=.05).

in this study. To illustrate how the censoring proportion is manipulated in this study, the distribution of the *M*-variable's time-to-event needs to be explained first. For this study, the time-to-event (*T*) for the *M*-variable was assumed to follow an exponential distribution with a parameter of λ_T = 0.1. That is, the hazard rate is a constant (= 0.1) throughout the time interval and the average time-to-event is $1/\lambda_T$ = 10. As seen in the parametric survival models in Chapter I, the time-to-event can follow various kinds of distributions such as the exponential distribution, the Gompertz distribution and the Weibull distribution. Among the many different distributions, the exponential distribution was chosen because of mathematical simplicity to calculate the censoring proportion.

Consider two independent exponential random variables *T* (time-to-event) and *C* (censoring) with parameters λ_T and λ_C, respectively. By independence, the joint probability density function of *T* and *C* is

$$f_{T,C}(\mathrm{t},\mathrm{c}) = \lambda_T\lambda_C \exp(-\lambda_T t)\exp(-\lambda_C c) \qquad t > 0,\ c > 0 \qquad \text{(II.3)}$$

Also, consider a new random variable, $Y = C - T$. The censoring proportion can be calculated by deriving the cumulative function $F_Y(y)$ of *Y* for the range of $y \leq 0$, which follows to be

$$\begin{aligned} F_Y(\mathrm{y}) &= \mathrm{P}(\mathrm{Y} \leq y) \\ &= P(C - T \leq y) \\ &= P(T \geq \mathrm{C} - y) \\ &= \int_0^\infty \int_{c-y}^\infty f_{T,C}(\mathrm{t},\mathrm{c})dtdc \\ &= \int_0^\infty \int_{c-y}^\infty \lambda_T\lambda_C \exp(-\lambda_T t)\exp(-\lambda_C c)dtdc \end{aligned}$$

$$= \frac{\lambda_C}{\lambda_T + \lambda_C} \exp(\lambda_C y) \qquad y \leq 0 \qquad \text{(II.4)}$$

If $y = 0$, this implies that $c \leq t$, which indicates that the time-to-event is greater than the censoring time meaning that the timing value would be censored at the censoring time in this case. From Equation II.4, the censoring proportion is $\mathrm{P}(\mathrm{Y} \leq 0) = F_Y(0) = \frac{\lambda_C}{\lambda_T + \lambda_C}$. Therefore, λ_C = 0 for 0% censoring and λ_C = 0.0430 for 30% censoring.

An infinite number of λ_T values (with restriction, $\lambda_T \geq 0$) can be used and there can be different underlying distributions (e.g., Weibull distribution) for the time-to-event other than the exponential distribution. However, neither the type of distribution nor the shape or location of the distribution affects the regression parameter estimates in the mediation model because of the utilization of the Cox regression in this study. One of the greatest advantages of using Cox regression is that it does not need to specify an underlying function of the time-to-event while estimating the regression parameters.

C. Sample Size

Sample size is an important factor in simulation studies. With larger sample sizes, the parameter estimates have smaller standard errors and thus higher statistical power. In contrast, if the sample size is small, the parameter estimates have larger standard errors and thus lower statistical power. In some cases, the fitted model might fail to converge due to small sample sizes. In this study, sample sizes of 150 or 300 were used that are comparable to common sample sizes used in social science.

D. Different Methods to Evaluate the Indirect Effect

In this study, six different modern methods to evaluate the indirect effect were compared. The six methods can be classified into three groups. Four out of the six

methods are in the first group which is related to evaluation of the product term, *ab*. Namely, they are the Sobel test, percentile bootstrap, bias-corrected bootstrap, and distribution of the product method (see Chapter I, 2. Mediation Analysis, Section B for details). The joint-significance test alone is in the second group of testing indirect effects and is done by looking at the significance of the *a* and *b* coefficients separately and the indirect effect is only significant when both of the regression coefficients are significantly different from zero. The last method is bootstrapping the NIE derived from using the counterfactual framework in causal mediation (see Chapter I, 3. Survival Mediation Analysis, Section B.3 for details). In summary, the NIE for the survival-mediator model can be computed by *b* multiplied by the difference between the integral of a survival function given an *x*-value (e.g., $x = 1$) and the integral of a survival function given an *x'*-value (e.g., $x' = 0$). The true parameter NIE value is unknown because it is not just a function of the model parameters. The NIE can only be estimated by the data.

4. Data Generation and Estimation

The data generation process for the simulation study is given as a flow chart in Figure II-3. First of all, data for *X* was generated. The *X*-values for a continuous *X*, X_{con}, was randomly picked from a standard normal distribution. Then, the *X*-values for a binary *X*, X_{bin}, was determined as 0 when $X_{con} < 0$ and 1, otherwise. The other variables, *T*, *M* and *Y* all depend on the binary values of X_{bin} in the following data generation process.

In the second data generation step, values for variable *T* (time-to-event) were generated by the method proposed by Bender, Augustin, and Blettner (2005). Their

method uses a uniformly distributed random variable to generate the values of time-to-event assuming a Cox regression (inverse transform method). The method can be explained in four major steps. First, from the Cox model representation and the relationship between the hazard function and cumulative distribution function from Chapter I (see Equations I.1.7 and I.1.11), the distribution function of the Cox model is

$$F(T|x) = 1 - \exp[-H_0(T)\exp(ax)] \qquad \text{(II.5)}$$

where $H_0(t) = \int_0^t h_0(u)\,du$ denotes the cumulative baseline hazard function and a is the regression coefficient for the *X* on *M* effect. Then, from the fact that the cumulative function for a random variable always follow a uniform distribution on the interval [0, 1], U ~ Uni[0, 1], and also the fact that (1-U) ~ Uni[0, 1], Equation II.5 becomes

$$U = \exp[-H_0(T)\,exp(ax)] \sim Uni[0,1] \qquad \text{(II.6)}$$

In the third step, H_0 can be inverted and the time-to-event, *T*, can be expressed as

$$T = H_0^{-1}[-\log(U)\,exp(-ax)] \qquad \text{(II.7)}$$

In the last step, the inverse of the cumulative baseline hazard function, H_0^{-1} is substituted into Equation II.7. In the case of an exponential baseline hazard function, $H_0^{-1}(t) = \lambda_T^{-1}t$ and therefore inserting this into Equation II.7 results in

$$T = \lambda_T^{-1}[-\log(U)\,exp(-ax)] \qquad \text{(II.8)}$$

In summary, Equation II.8 is used to generate the time-to-event values. In this study, the λ value was fixed to .1 and the magnitude of *a* was a manipulated factor (*a* = 0, .21 or .52) of the study. Using Equation II.7, *T* can be generated using different cumulative baseline hazard functions, $H_0^{-1}(t)$, but again, the a-estimate is not affected by the different functions of the baseline hazard since Cox regression is used when estimating the regression parameters.

In the third data generation step, the censoring timing values (*C*) were randomly generated from an independent exponential distribution with λ_C parameter. Then, the *M*-variable values were determined as *T* if the time-to-event value was smaller than the censoring time and *C*, otherwise. Note that λ_C can have different values ($\lambda_C = 0$ and $\lambda_C = 0.043$) as a product of the censoring proportion manipulation (0% or 30%) as seen in details in Section 3.2 above. The MCAR assumption is satisfied in the current simulation study since the time-to-event (*T*) was generated independently of the censoring time (*C*).

In the last data generation step (Step 4), the *Y*-values were generated based on the *Y*-regression equation. Note that the time-to-event variable, *T* was used in this equation instead of *M*. The main difference between *T* and *M* is that *T* is an uncensored variable and *M* is a censored variable. That is, *T* assumes that all the time-to-event values are known and *M* is a censored version of *T* where all values above *C* are censored. Therefore, the *Y*-values in this simulation were based on the uncensored true values of *T*. In reality, we cannot observe *T* but only *M* because of the right censoring. The *b* (*b* = 0, 0.14, or 0.39) and *c'* (*c'* = 0 or 0.39) coefficients were manipulated factors in this study. Finally, a residual term, *r* was added to the *Y*-regression. The residual term was assumed to follow a standard normal distribution ($r \sim N\ (0, 1)$) for mathematical simplicity.

SAS 9.4 was used to generate all data for this simulation study. A macro was created to generate data based on specific values of the study factors. Inside the DATA step in SAS, the RANNOR and RANUNI command were used to randomly generate values from the standard normal distribution and the standard uniform distribution, respectively. The RAND("EXPONENTIAL") command was used multiplied by a quantity defined as the inverse of the λ_C parameter to generate values from an

exponential distribution with λ_C. All the other commands in the DATA step follow the specifications in Figure II-3.

Once 1,000 replication samples were generated based on the specific factor levels, the estimation procedure is straightforward. There are two regressions that were fitted for the survival-mediator model: the *M*-regression and the *Y*-regression. For the *M*-regression, the Cox model was fitted. For the *Y*-regression, the censored values in the *M*-variable were treated as missing values and only the complete cases were used in the *Y*-regression. Since the data generation process satisfies the MCAR assumption, using only the complete cases of *M* in the *Y*-regression would recover the parameter values better than using the original *M*-variable with censored values (as shown in the small simulation study in Chapter I, 3. Survival Mediation Analysis, Section B.2.b).

The entire estimation procedure was done in SAS 9.4. PROC PHREG was used for the Cox regression of the *M*-variable and PROC REG was used for the OLS regression of the *M*-variable and the *Y*-variable. For bootstrapping procedures, PROC SURVEYSELECT was used.

5. Dependent Variables (Outcomes)

Seven dependent variables were used to measure the performance of different test methods for the indirect effect. Namely, they are the empirical Type I error rate, empirical statistical power, parameter coverage rate, average raw bias, average relative bias, sign (negative or positive) of the bias, and mean squared error (MSE). Among the seven outcomes, the first three outcomes (Type I error rate, statistical power, and parameter coverage rate) were used to compare performance of the different indirect

effect tests. The latter four outcomes (average raw bias, average relative bias, sign of the bias, and MSE) were used to check whether the proposed statistical methods retrieved the true parameter values accurately under different conditions.

First, the empirical Type I error rate was measured by the number of error decisions out of the 1,000 replications saying that the indirect effect was significantly different from zero when the true indirect effect was actually zero, $a = 0$, $b = 0$, or $a = b = 0$. Out of the 9 interaction levels of a and b (3 levels of a and 3 levels of b), there were 5 conditions that satisfied $ab = 0$. In this study, the 5% level of significance was used. Therefore, the indirect effect was expected to be statistically significant in 50 (5%) of the 1,000 replication samples when the true indirect effect equals zero. If the Type I error rate is largely different from 5% using a particular indirect effect testing method, the performance of that method is not preferable.

The empirical statistical power was measured as the proportion of times out of the 1,000 replications that the indirect effect to be significantly different from zero given that the true indirect effect was actually not a zero value, $a \neq 0$ and $b \neq 0$. There were 4 out of 9 conditions that satisfied $a*b \neq 0$. Higher values of the statistical power indicate better performance of an evaluation method of the indirect effect.

For each of the replications, parameter coverage was marked by an indicator that has a value of 1 when the confidence interval (or confidence limit for the bootstrapping methods) of the indirect effect includes the true parameter and a value of 0 when the confidence interval does not include the true indirect effect. The parameter coverage rate was quantified by the number of replications with successful parameter coverage out of 1,000 replications. Ideally, the parameter coverage rate would be close to .95 (the

interval coverages the true parameter value 950 times out of 1,000 replications) given α = .05. The parameter coverage rate was only computed for the *ab*-product methods (Sobel test, distribution of the product test, percentile bootstrap and bias-corrected bootstrap). The parameter coverage rate could be different for the four different *ab*-product methods because they all estimated the confidence intervals using different methods. The joint-significance test does not estimate an indirect effect and confidence intervals and the true parameter value for the NIE is unknown. Therefore, the parameter coverage rate was limited to compare performance of the four *ab*-product test methods.

The raw bias was calculated using the difference between the true indirect effect (θ) and the estimated indirect effect ($\hat{\theta}$) in a replication sample using a particular method. The average of 1,000 replications was computed. A relatively small average raw bias indicates a good performance of a method. Since the joint-significance test does not evaluate an indirect value and the true NIE is unknown, the average raw bias is only relevant to the *ab*-product methods. Furthermore, the four *ab*-product methods (Sobel test, distribution of the product test, percentile bootstrap, and bias-corrected bootstrap) did not produce different raw biases because the same point estimate of the indirect effect, *ab*, was used for these methods. Therefore, the raw bias could not be used as an outcome to compare performance of the different indirect effect tests.

Since the raw bias depends on the metric of the coefficient, the relative bias was also reported for this study. The relative bias was computed by

$$Relative\ bias = \frac{\hat{\theta} - \theta}{\theta} \qquad \text{(II.9)}$$

Note that the relative bias cannot be calculated when $\theta = 0$. The relative bias was computed first for each of the replications and then averaged across the 1,000 replications. A relatively small average relative bias indicates a good performance of a method. Similar as the average raw mean bias, the relative bias is irrelevant to comparing performance of the different indirect effect tests because the *ab* estimates are identical for all tests.

The average raw and relative bias show how accurately the model parameters recover the true values for a given condition. However, it does not reveal whether the parameter estimate is lower or higher than the true parameter value. The sign of the raw bias was investigated to see if there was a systematic tendency of a specific condition of the study. If the estimate was higher than the true parameter value, the raw bias was a positive value and if the estimate was lower than the true parameter value, the raw bias was a negative value. The raw bias was zero when the estimate was equal to the parameter value. Again, the sign of the raw bias is not applicable to compare performance of the different indirect effect tests because the *ab* estimates are identical for all tests.

Finally, the MSE was examined. The MSE was computed as

$$MSE = \frac{\sum\left(\hat{\theta} - \theta\right)^2}{R} \qquad \text{(II.10)}$$

where *R* is the number of replications (*R* = 1,000 for this study). The MSE can also be expressed as the sum of the variance of $\hat{\theta}$ and the bias of $\hat{\theta}$ squared: $\text{MSE} = \text{Var}\left(\hat{\theta}\right) + \text{Bias}(\hat{\theta})^2$. Therefore, the MSE captures the variance as well as the bias of an estimate. A relatively smaller MSE indicates good performance of a method. Again, the MSE is

irrelevant to compare performance of the different indirect effect tests because the *ab* estimates are identical for all tests.

6. Logistic Regression and Analysis of Variance

Logistic regression or ANOVA was conducted for the relevant conditions based on the different dependent variables. Logistic regression analyses were conducted for Type I error rate, the statistical power, the parameter coverage rate, and the sign of the raw bias. ANOVAs were conducted for the average raw bias, the average relative bias, and the MSE.

The simulation design of the current study is a 3x3x2x2x2x6 factorial design with each of the cells having 1,000 observations of the outcomes. The six factors are in order of: the size of *a* (factor "A"), the size of *b* (factor "B"), the size of *c'* (factor "C"), the censoring proportion (factor "CP"), the sample size (factor "SS"), and the different testing methods of the indirect effect (factor "IE"). Although factor IE is technically a within-replication factor (all six of the indirect effect tests were conducted "within" a replication sample), it was treated as a between-replication factor in the analyses for ease of analysis[5].

Different cells were compared for different outcomes. For Type I error rate, (5)x2x2x2x6 = 240 cells were compared because only 5 (*a*=0, *b*=0.14; *a*=0, *b*=0.39; *a*=0.21, *b*=0; *a*=0.52, *b*=0; and *a*=*b*=0) out of the 9 conditions created by the (AxB)

[5] The standard error of the estimates will be larger without considering the dependencies that can arise by testing the indirect effects with different methods within a replication sample. Factor IE is treated as a between-replication factor although the standard errors might be larger than normal. A conservative approach (less Type I errors rather than high statistical power) is taken in this study and also, statistical significance of the effects is not of primary importance in this study.

interaction were related with the Type I error. Similarly, (4)x2x2x2x6 = 192 cells were compared for the Statistical Power because 4 (*a*=0.21, *b*=0.14; *a*=0.21, *b*=0.39; *a*=0.52, *b*=0.14; *a*=0.52, *b*=0.39) out of the 9 (AxB) conditions were relevant. For the parameter coverage rate, 3x3x2x2x2x4 = 288 cell comparisons were made since the IE factor only had four levels (*ab*-product term: Sobel test, distribution of the product test, percentile bootstrap, and bias-corrected bootstrap) that were relevant. For the average raw bias, sign of the raw bias, and the MSE, 3x3x2x2x2 = 72 cell comparisons were made using the *ab*-product method. For the average relative bias, the true *ab* parameter value cannot be zero. Therefore, only (4)x2x2x2 = 32 cell comparisons were made.

The focal variable of this study is the IE factor. The main effects of the other five factors were included in the logistic regression or ANOVA model to control for their effects. Also, interactions related to the IE factor were of interest and included in the logistic regression or ANOVA model. There can be up to a 6-way interaction between the 6 factors, but since it is practically too complicated to interpret the interactions that are higher than 3-way interactions, only the 2-way and 3-way interactions were included in the logistic regression and ANOVA models. All possible 2-way and 3-way interactions were included in the ANOVA or logistic regression model. However, only the 2-way and 3-way interactions that include the IE factor were interpreted. Other 2-way and 3-way interactions were included to control for their effects.

Statistical significance of the main and interaction effects was examined from the ANOVA and logistic regression analyses. Then, for a significant main or interaction effect, the marginal means were further examined both numerically and graphically. This

strategy was taken since there was no a priori hypothesis for the simple effects especially, for effects related to the six different indirect effect test methods.

Since there were a large number of observations (i.e., 1,000 replications) in each cell of the ANOVA, even diminutive effects could be statistically significant. Interpretation of the ANOVA results was focused on effects that have a partial omega square ($\widehat{\omega^2}$) over .01 which is a small effect size based on Cohen's recommendations (Cohen, 1988). For logistic regression, effect coding was used for the binary or categorical predictors in the model. Although, the interpretation of each regression coefficient can be less clear for effect coding rather than dummy coding[6], effect coding was used because the individual regression coefficients were not the focus of this study. Instead, the main and interaction effects for each categorical variable or combination of the categorical variables were of interest. In SAS PROC LOGISTIC, joint tests using the Wald chi-square statistic (χ^2) were conducted for each of the main and interaction effects. The joint test can be a multiple degrees of freedom test (more than 1 degree of freedom) that tests a collection of tests that add up to a main effect or interaction effect. For example, for the main effect of factor A, there can be two simple effects (setting a=0 as the reference): 1) the difference between the mean when a=0.21 and the mean of a=0.21 and a=0, and 2) the difference between the mean when a=0.52 and the mean of a=0.52 and a=0. Since it is difficult to get an appropriate effect size measure for each

[6] Using dummy coding, the regression coefficient can be interpreted as the mean difference between the target group (e.g., males) and the reference group (e.g., females). Using effect coding, the regression coefficient is the difference between the target group's mean (e.g., males) and the target and reference groups' mean (e.g., average of the males and females). The interpretation of the regression coefficient is clear using dummy coding but, less clear when effect coding is used.

joint test in logistic regression, only main or interaction effects that had $p < .01$ were focused.

7. Expected of Results

The performance across the different indirect effect test methods is of most interest. Based on results from previous simulation studies, the distribution of the product test, joint-significance test and the percentile bootstrap would performance best in terms of Type I error rate, statistical power and parameter coverage rate. Also, bootstrapping of the NIE effect would work as well as the three best performing methods. In contrast, Sobel test would produce too low Type I error rate with small effect size and small sample size. The biased-corrected bootstrap would produce higher than nominal (α = .05) Type I error rates in some conditions which also produces bogus high statistical power rate and lower than .95 parameter coverage rate.

The 2-way and 3-way interactions related to the IE factor show whether the indirect tests perform differently in different conditions of this study. Since this is the first study to investigate the survival-mediator model, the directions of the 2-way and 3-way interactions are exploratory in this study.

Although they are of less importance in this study, the following main effects were expected for the other factors in the study:

(1) Magnitude of the *a*-path and *b*-path: Both the magnitude of the *a*-path and the *b*-path were expected to have similar effects on the outcomes. The raw bias and the MSE were expected to be larger as the magnitude of the true path coefficients were larger. Relative bias and the sign of the raw bias were expected not to differ

as a function of the size of the path coefficients. Statistical power would be higher as the path coefficients had a larger effect size. The Type I error rate and parameter coverage rate would not be affected by the magnitude of the path coefficients.

(2) Magnitude of the *c'*-path: There would be no noticeable difference in all seven outcomes as a function of the *c'*-path size.

(3) Censoring proportion of the *M* variable: As the proposed survival-mediator model handles the censored data appropriately, no noticeable difference in all seven outcomes were expected.

(4) Sample size: Generally, the sample size affects both the accuracy (bias) of the estimate and the standard errors. Larger the sample size, smaller the bias and standard errors of the estimate. The raw bias, relative bias and MSE would be smaller as the sample size increased. Statistical power would increase with larger sample size. Type I error rate, parameter coverage rate and the sign of raw bias would be less affected by sample size.

III. RESULTS

The proposed single survival-mediator model converged in all conditions. There were no non-convergence or improper solutions while fitting the model to the datasets. The estimates of *a*, *b*, and *c'* were all in a reasonable range. The mean raw bias of all replications (72,000 replications) were 0.002 for the *a*-estimate, -0.000 for the *b*-estimate and 0.001 for the *c'*-estimate. None of the mean raw biases were over 0.020 in magnitude for any of the 72 study conditions. The maximum (in magnitude) mean raw bias for all three estimates was 0.018 for the *a*-estimate, -0.001 for the *b*-estimate and -0.012 for the *c'*-estimate[7]. The reasonable range of estimates of the *a*, *b* and *c'* parameters and convergence in all conditions validated data generation and that the proposed model worked as expected.

The following sections show the results for the average raw bias, average relative bias, sign of the raw bias, mean square error (MSE), Type I error rate, statistical power, and the parameter coverage rate, respectively. The first four outcomes show how the true parameter values are recovered by fitting the proposed single survival-mediator model. The last three outcomes show performance of the different indirect effect test methods. Since outcomes are not expected to change significantly with different values of the *c'*-path size and the censoring proportion, each of the tables and figures are presented with a specific fixed value of *c'*-path and/or censoring proportion.

[7] The condition that produced the maximum (in magnitude) raw bias was when *a*=0.52, *b*=0.39, *c'*=0.39, censoring proportion=0.3, and *n*=150 for the *a*-estimate; *a*=0.52, *b*=0.39, *c'*=0, censoring proportion=0.3, and *n*=150 for the *b*-estimate; and *a*=0, *b*=0, *c'*=0.39, censoring proportion=0, and *n*=150 for the *c'*-estimate.

1. Recovery of the True Parameter Values

The average raw bias, average relative bias, sign of the raw bias and the MSE are all related to how well the *ab*-product point estimate recovered the true parameter value. There were four indirect effect test methods that used the *ab* point estimate: Sobel test, distribution of the product test, percentile bootstrapping and bias-corrected bootstrapping. These methods differ in the way to evaluate whether the effect is significant or not, however the *ab* point estimate itself does not differ across the different methods. Therefore, the four methods produce the same values for the average raw bias, average relative bias, sign of the raw bias, and the MSE.

A. Average Raw Bias of the *ab* Estimate

Table III-1 and Figure III-1 show the average raw bias of the *ab* estimate when *c'* = 0. Table III-2 and Figure III-2 show the average raw bias when *c'* = 0.39. The average raw bias of all 72,000 observations was 0.000 with a standard deviation of 0.039. The ANOVA results for the average raw bias are shown in Table III-3. The ANOVA revealed that none of the main and interaction effects had an effect that exceeds partial $\widehat{\omega^2}$=.01.

B. Average Relative Bias of the *ab* Estimate

The average relative bias was only computed for conditions when the true *a* and *b* parameter values were non-zero. Table III-4 and Figure III-3 show the average relative bias of the *ab* estimate when *c'* = 0. Table III-5 and Figure III-4 show the average relative bias when *c'* = 0.39. The average relative bias of all 32,000 observations was 0.008 with a standard deviation of 0.580. As displayed in Tables III-4 and III-5, the minimum cell mean relative bias was -0.036 (*a*=0.21, *b*=0.14, *c'*=0, *n*=300 and censoring

proportion=0) and the maximum cell mean relative bias was 0.037 (*a*=0.52, *b*=0.39, *c'*=0.39, *n*=300 and censoring proportion=0). The minimum and maximum average relative biases were generally low, never exceeding 4% of the parameter value. Table III-6 shows the ANOVA results for the average relative bias. None of the main or interaction effects had a partial $\widehat{\omega^2}$ above .1.

C. Sign of the *ab* Raw Bias

Table III-7 and Figure III-5 show the sign of the *ab* raw bias when *c'* = 0. Table III-8 and Figure III-6 show the sign of the *ab* raw bias when *c'* = 0.39. Out of the 72,000 replications (1,000 replications for 72 (3x3x2x2x2) conditions), there were 59 datasets that produced estimates that were the same as the true population *ab* parameter. For a given condition, the maximum number of datasets that the estimate was equal to the true parameter value was 6 out of 1,000 replication datasets (this occurred for the following condition: *a*=0, *b*=0.39, *c'*=0.39, *n*=300, censoring proportion=0). Except of those 59 datasets, there was around a half and half split of the negative (36,020 or 50.07%) and positive (35,921 or 49.93%) raw biases. That is, almost half of the time the parameter estimate was smaller than the true parameter value and half of time the parameter estimate was larger than the true parameter value. Table III-9 shows the results of the logistic regression for the sign of the *ab* raw bias. A Wald statistic with $p < .01$ was considered as a meaningful effect. Logistic regression revealed that all of the effects had a *p*-value higher than .01.

D. Mean Squared Error

Table III-10 and Figure III-7 show the MSE of the *ab* estimate when *c'* = 0. Table III-11 and Figure III-8 show the MSE of the *ab* estimate when *c'* = 0.39. The average MSE across the 72,000 observations was 0.002 with a standard deviation of 0.004. As displayed in Tables III-10 and III-11, the maximum cell mean MSE was 0.006 (*a*=0, *b*=0.39, *c'*=0, *n*=150, censoring proportion=0.3) and there were multiple conditions (when either *a* or *b* or both *a* and *b* were zero) where the cell mean MSE was below 0.0001. Table III-12 shows the ANOVA result for the MSE. Effects that had a partial $\widehat{\omega^2}$ above .1 are meaningful. The ANOVA revealed that the size of the *b*-path had a significant effect on the MSE after controlling for all other effects (main effect of B: F = 9024.83, $p < .001$, partial $\widehat{\omega^2} = .20$). The MSE was essentially zero for conditions where *b* was zero. The MSE was smaller than 0.001 for small *b* (*b*=.14) and the MSE was above 0.002 for medium *b* (*b*=.39). MSE is a function of bias and the standard error of an estimate. The effect of factor B was very small for the average raw and relative biases. Therefore, the significant MSE difference between the different magnitudes of *b* originates from the standard errors of the *ab* indirect effect. A higher effect of the *b*-path introduced significant increase in the standard errors of the *ab* indirect effect.

2. Performance of the Six Indirect Effect Test Methods

Performance of the six different test methods (Sobel test, distribution of the product test, percentile bootstrapping, bias-corrected bootstrapping, joint-significance

test, and the bootstrap of the NIE) were evaluated by the Type I error rate, statistical power rate and parameter coverage rate.

A. Type I Error Rate

Separate tables and figures of Type I error rates are given for each combination of factor C (size of *c'*) and factor CP (censoring proportion). Table III-13 and Figure III-9 (*c'* = 0 and zero censoring); Table III-14 and Figure III-10 (*c'* = 0 and 0.3 censoring); Table III-15 and Figure III-11 (*c'* = 0.39 and zero censoring); and Table III-16 and Figure III-12 (*c'* = 0.39 and 0.3 censoring) show the Type I error rate for each of the indirect test methods. In each of the figures, the different indirect effect tests are represented by separate lines.

Results of the logistic regression of the Type I error rate are shown in Table III-17. Effects that are related to the IE factor and had a *p*-value lower than .01 are interpreted below. Logistic regression revealed that the AxIE (size of *a* x indirect effect test method) interaction was significant after controlling for all other effects, $\chi^2(10) = 46.86$, $p < .001$. The Type I error rate did not significantly differ across the six indirect effect test methods when the true value of *a* was zero. However, there was a significant difference in the Type I error rate when *a* was a small effect (*a*=0.21) or a medium effect (*a*=0.52). The Type I error rate difference between the bias corrected bootstrap method and the Sobel test was particularly large when *a* was a medium effect (*a*=0.52). While the distribution of the product method, percentile bootstrapping, joint significance test and the bootstrapping of the NIE produced a Type I error around the nominal level (α = .05) with small fluctuation across the other conditions, the Type I error rate for the bias

corrected bootstrap was high around .075 and the Type I error rate for the Sobel test was less than .025. This result supports the argument made by MacKinnon and his colleagues that the Type I error can be higher than the nominal level using the bias-corrected bootstrap method (Fritz et al., 2012) and also that the Sobel test produces lower than the nominal level Type I error (MacKinnon et al., 2002). Importantly, the results hold not only for continuous mediators and outcomes, but for the survival-mediator model proposed here, especially when the *a*-path gets larger.

Another significant effect was the BxIE (size of *b* x indirect effect test method) interaction after controlling for all other effects, $\chi^2(10) = 29.63$, $p < .01$. All of the indirect effect test methods produced a Type I error rate close to zero when *b* was a zero effect. However, when *b* was a small (*b*=0.14) or a medium (*b*=0.39) effect, the Type I error rate using the bootstrap method for the NIE was lower than the nominal level while the other methods produced around the nominal level Type I error rate. This result was prominently observed when the censoring proportion was 0.3. The bootstrap for the NIE displayed less than .025 Type I error rate while the other methods showed around .05 Type I error rate. The calculation of the NIE involves numerical integration across the time points available in the data. The high rate of censoring might cause less accurate computation of the NIE and thus cause low Type I error rates. More about calculation of the NIE will be described in the Discussion section below.

Other than the logistic regression results relevant to the IE factor, there was a main effect of the A factor (size of *a*) after controlling for all other effects, $\chi^2(2) = 546.08$, $p < .001$. The average Type I error rate was higher when *a* was a medium effect

(*a*=0.52; *Type I error rate*=.05) than when *a* was a zero effect (*a*=0; *Type I error rate*=.04) and when *a* was a small effect (*a*=0.21; *Type I error rate*=.01). Other significant 2-way interactions (BxSS) and 3-way interactions (BxCPxSS) are not discussed here since they were not focal variables of the study.

B. Statistical Power

Separate tables and figures of statistical power are given for each combination of factor C (size of *c'*) and factor CP (censoring proportion). Table III-18 and Figure III-13 (*c'* = 0 and zero censoring); Table III-19 and Figure III-14 (*c'* = 0 and 0.3 censoring); Table III-20 and Figure III-15 (*c'* = 0.39 and zero censoring); and Table III-21 and Figure III-16 (*c'* = 0.39 and 0.3 censoring) show the statistical power for each of the indirect test methods. In each of the figures, the different indirect effect tests are represented by separate lines. Note that the statistical power for the bias-corrected bootstrap method might be inflated because of the inflated Type I error rate.

Results of the logistic regression for statistical power are shown in Table III-22. Effects that are related to the IE factor and had a *p*-value lower than .01 are interpreted below. Logistic regression revealed that the CPxIE (censoring proportion x indirect test methods) interaction had a significant effect on statistical power while controlling for all other effects, $\chi^2(5) = 149.09$, $p < .001$. Comparing the Figures III-13 with III-14 and III-15 with III-16, the statistical power for the bootstrap of the NIE was constantly lower than the other methods when the censoring proportion was 0.3. There was somewhat less difference between the bootstrap of the NIE method and the other methods when the censoring proportion was zero. The lower statistical power for the bootstrap of the NIE

might be a function of the lower Type I error rate when the censoring proportion was 0.3. Again, this has to do with the numerical integration for computing the NIE and will be further discussed in the Discussion section below.

There was a significant main effect of factor A (size of *a*), $\chi^2(1) = 23236.06$, $p < .001$; significant main effect of factor CP (censoring proportion), $\chi^2(1) = 1259.79$, $p < .001$; a significant main effect of factor SS (sample size), $\chi^2(1) = 5396.18$, $p < .001$; and a significant main effect of factor IE (indirect effect test method), $\chi^2(5) = 450.38$, $p < .001$. The average statistical power was higher when the parameter value of *a* was medium (*statistical power* = .89) than small (*statistical power* = .29). The average statistical power was larger for the zero censoring condition (*statistical power* = .63) than the .3 censoring proportion condition (*statistical power* = .55). The average statistical power was higher for the 300 sample size condition (*statistical power* = .68) than the 150 sample size condition (*statistical power* = .50). Also, the average statistical power was lower for the bootstrap of the NIE method (statistical power = .53) than the other five methods which were similar to each other (statistical power = .60). Other significant 2-way interactions (AxCP; AxSS; BxCP; CPxSS) or 3-way interactions (AxBxCP; AxCPxSS; BxCPxSS) that are irrelevant with the IE factor is not discussed further since they were not focal variables of the study.

C. Parameter Coverage Rate

Separate tables and figures of parameter coverage rates for each combination of factor C (size of *c'*) and factor CP (censoring proportion) are presented. Table III-23 and Figure III-17 (*c'* = 0 and zero censoring); Table III-24 and Figure III-18 (*c'* = 0 and 0.3

censoring); Table III-25 and Figure III-19 (c' = 0.39 and zero censoring); and Table III-26 and Figure III-20 (c' = 0.39 and 0.3 censoring). In each of the figures, separate lines represent the different indirect effect tests. Note that only four lines (Sobel test, distribution of the product test, percentile bootstrapping and bias-corrected bootstrapping) are depicted in the figures. Parameter coverage rate is irrelevant for the joint-significant test because it does not quantify the indirect effect when testing the indirect effect. Also, parameter coverage rate was not computed for the bootstrap of the NIE method in this study because the true parameter value of the NIE is unknown.

Results of the logistic regression of the parameter coverage is shown in Table III-27. Effects that are related to the IE factor and had a *p*-value lower than .01 are interpreted below. Logistic regression revealed that there was a significant 3-way interaction among factor A (size of *a*), factor B (size of *b*) and factor IE (indirect effect test methods) after controlling for all other effects, $\chi^2(12) = 29.81$, $p < .01$. The difference among the four different indirect effect test methods was prominent when *a* was a medium effect (*a*=0.52) while *b* was a zero effect. The Sobel test produced a higher than nominal level (=.95) parameter coverage rate around .975 and the biased-corrected bootstrap method produced a lower than nominal level parameter coverage rate around .925 while the distribution of the product method and the percentile bootstrap methods on average produced about .95 parameter coverage rate. This supports the conclusion from the Type I error rates where the bias-corrected bootstrap method showed higher than nominal level Type I error rates and the Sobel test showed lower than nominal level Type I error rates. The difference among the different indirect effect test methods was smaller when *a* was a small effect (*a*=0.21) while holding *b* at a zero effect

and there was an even smaller difference between the methods when *a* and *b* were both zero effects. The difference between the indirect test methods were negligible when *b* was a small (*b*=0.14) or medium (*b*=0.39) effect regardless the size of *a* (*a*=0, *a*=0.21 or *a*=0.52). Moreover, the AxIE (size of *a* x indirect effect test method) interaction and the BxIE (size of *b* x indirect effect test method) interaction were both significant after controlling for all the other effects, $\chi^2(6) = 23.85$, $p < .01$ and $\chi^2(6) = 155.23$, $p < .001$, respectively. The maximum difference of the parameter coverage rate between the four indirect effect test methods became larger as the size of *a* increased. On average, while the order of the average parameter coverage rate did not change (Sobel test > distribution of the product test > percentile bootstrap > biased-corrected bootstrap) across the values of the *a*-coefficient, the maximum difference (Sobel test – biased corrected bootstrap) was 0.004, 0.011, and 0.022, respectively for a zero (*a*=0), small (*a*=0.21) and medium (*a*=0.52) sized *a*. Also, there was a noticeable difference in the parameter coverage rates between the four indirect effect test methods when *b* was a zero effect but, the difference between the indirect effect test methods were negligible when *b* was a small (*b*=0.14) or medium (*b*=0.39) effect. Again, across the values of *b*, the order of the average parameter coverage rate did not change (Sobel test > distribution of the product test > percentile bootstrap > biased-corrected bootstrap) but, on average, the maximum difference in the average parameter coverage rate between the Sobel test method and the biased-corrected bootstrap was 0.031 when *b* was a zero effect, 0.005 when *b* was a small effect (*b*=0.14) and 0.002 when *b* was a medium effect (*b*=0.39).

Logistic regression also revealed a significant interaction between the SS factor (sample size) and the IE factor (indirect effect test method) after controlling for all other

effects, $\chi^2(3) = 15.55$, $p < .01$. The maximum difference in the average parameter coverage rate between the four indirect effect test methods was between the Sobel test and the bias-corrected bootstrap (Sobel test > bias-corrected bootstrap) in both the 150 and 300 sample size conditions. The difference was slightly larger for the 150 sample size condition (Sobel test – bias-corrected bootstrap = 0.0133) than the 300 sample size condition (Sobel test – bias-corrected bootstrap = 0.0116).

Other than the logistic regression results related to the IE factor, the following main effects were significant after controlling for all other effects in the model: effect of the A factor (size of *a*), $\chi^2(2) = 472.50$, $p < .001$; effect of the B factor (size of *b*), $\chi^2(2) = 635.41$, $p < .001$; effect of the CP factor (censoring proportion). $\chi^2(1) = 8.55$, $p < .01$; effect of the SS factor (sample size), $\chi^2(1) = 36.58$, $p < .001$; and effect of the IE factor (indirect effect test methods), $\chi^2(3) = 160.41$, $p < .001$. The average parameter coverage rate decreased as a larger *a* was used, 0.966, 0.961, and 0.950 for *a*=0, *a*=0.21 (small) and *a*=0.52 (medium), respectively. The average parameter coverage rate was inflated when *b* was a zero effect (*parameter coverage rate*=.98) whereas, the average parameter coverage rate was equal at the nominal level of .95 when *b* was a small (*b*=0.14) or a medium (*b*=0.39) effect. The average parameter coverage rate was slightly higher when the censoring proportion was 0.3 (*parameter coverage rate*=0.960) than when the censoring proportion was zero (*parameter coverage rate*=0.958). The average parameter coverage rate was slightly higher when the sample size was 150 (*parameter coverage rate*=0.961) than when the sample size was 300 (*parameter coverage rate*=0.957). Finally, the average parameter coverage rate was 0.965 for the Sobel test, 0.960 for the

distribution of the product test, 0.958 for the percentile bootstrap method and 0.953 for the bias-corrected bootstrap method.

Other significant 2-way interactions (AxB; AxSS; BxSS; CxCP) or 3-way interactions (AxBxCP; AxCxCP; AxCPxSS; BxCxCP; BxCxSS) that are not relevant for the IE factor are not discussed further since they were not focal variables of the study.

IV. DISCUSSION

In psychological and health intervention studies, the duration of being in a specific state might be a short-term objective for which an intervention program aims. For example, the goal of a smoking cessation program is to keep the participants abstinent of tobacco as long as possible. The amount of time staying away from smoking can promote lung health measured by a spirometer (an instrument to measure the lung's air capacity) at a later time point. Survival mediation analysis can be used to answer research questions with these types of survival variables. In the smoking cessation program example, the mediator is a survival variable targeted by the intervention program to promote lung health in the future.

This dissertation focused on a single survival-mediator model (see Figure II-1) where the independent variable is a randomized treatment indicator, the outcome is a continuous variable and the mediator is a survival variable which measures both *when* the event had occurred and *whether* the event had been observed or not within the observation period. In this dissertation, a statistical model for the survival-mediator model was proposed. Also, a statistical issue (biased estimates in the Y-regression) of the model was discussed and a solution for the issue was proposed. Different methods to evaluate the mediated effect or in other words, the indirect effect were discussed and a mathematical derivation of the natural indirect effect (NIE) was shown. Finally, a simulation study was conducted to evaluate the performance of the different indirect effect test methods.

In the following sections, first, the statistical model for the survival-mediator model will be revisited with discussion about the underlying assumptions of the model

and the solution that was proposed for the *Y*-regression. Second, the indirect effect measures, the product of *a* and *b* (*ab*) and the NIE will be discussed focusing on the interpretation and statistical testing of the effects. Third, the results from the simulation study will be summarized and recommendations will be made for researchers. Also, limitations and future directions of the study will be discussed. Fourth, a made-up data example will be introduced to help interpret the estimates of the survival-mediator model. Last, final comments will be made about survival mediation analysis.

1. Statistical Model for the Survival-Mediator Model

The single survival-mediator model consists of two regression equations. The first one is the *M*-regression where the independent variable *X* predicts the survival variable, *M*. The Cox proportional hazards model (Cox, 1972) was proposed as the *M*-regression in this study. The Cox model is a widely used survival model where the hazard rate is modeled as a function of a baseline hazard rate (the hazard rate when all predictors in the model are zero) and an exponential function of a linear combination of a set of predictors (see Equation I.3.2 in Chapter I). The Cox model can be biased with violation of the proportional hazards assumption which means that there is no interaction between time and the predictors of the model. Figure IV-1 depicts the proportional hazards assumption. The proportion of two different hazard rates with different *X*-values becomes a subtractive function when the logarithm of the proportion is taken: $\log\left[\frac{h(t|X=0)}{h(t|X=1)}\right] = \log[h(t|X=0)] - \log[h(t|X=1)]$. In Figure IV-1, the logarithm of the proportion of the hazard rate of a person with *X*=1 to the hazard rate of a person with *X*=0 is a constant (a) across all time points. Although the proportional hazards assumption

can be stringent for some data, there are methods to relax the assumption by including the interaction term between time and the predictors in the model or using a stratified model (see Chapter I, 1. Survival Analysis, Section D.2.c for more details). Also, the advantage of using the Cox model is large enough to overcome its need of the stringent proportional hazards assumption. Even though the baseline hazard function is specified in the Cox model, the baseline hazard does not need to be estimated while estimating the regression parameters (*a*-parameter in the survival-mediator model).

Another interesting model that can be used as the *M*-regression in a survival-mediator model is the accelerated failure time (AFT) model. The AFT model parametrizes the acceleration factor which shows how fast (or slow) the time scale is for a specific group of observations compared to another group of observations. The AFT model has its own advantages such that interpretation of the regression coefficients is straightforward, but one of the major difficulties with the AFT model is that the researcher has to specify a baseline survival function which can be complicated and might or might not fit the data well. See Chapter I, 1. Survival Analysis, Section D.1. for more details about the AFT model.

The other regression of the survival-mediator model is the *Y*-regression. In the proposed survival-mediator model of this dissertation, the *Y* variable is a continuous variable. Therefore, a conventional multiple regression where *X* and *M* are simultaneously predicting the *Y* variable can be used (see Equation I.2.2 in Chapter I, 2. Mediation Analysis). However, if the censored data in *M* is not treated adequately, the *Y*-regression estimates *b* (*M* predicting *Y* controlling for *X*) and *c'* (*X* predicting *Y* controlling for *M*) can be biased. Without any treatment, the last observed time point for

the censored data is used and it causes bias in the regression coefficients. See Chapter I, 3. Survival Mediation Analysis, Section B.2. for more details. This is an important issue especially because all conventional statistical packages such as SAS, SPSS or Mplus would easily fit the *Y*-regression to the data. The statistical programs do not know whether there are censored values in *M* and instead, the program would just treat the last available time point as a valid value of *M* and thus, introduce bias in the *Y*-regression. This was shown in the small simulation study conducted in Chapter I, 3. Survival Mediation Analysis, Section B.2.b. When the censored *M* values (the last available time) were used for the *Y*-regression, the *c'* estimate was biased and the Type I error rate of *c'* was inflated (see Table I-4 for results).

A proposed solution for the bias issue in the *Y*-regression is to treat the censored *M*-values as missing values. Then, assuming that the missing data mechanism is missing completely at random (MCAR), complete case analysis (a.k.a., listwise deletion) of the *Y*-regression produces unbiased estimates. This solution seems to work well, even better than the full information maximum likelihood (FIML) method which usually gives unbiased estimates as well as higher power than the listwise deletion method because it does not exclude observations from the analysis. Allison (2001) argued that the listwise deletion method is a robust method with violations of the MCAR assumption and even under the missing at random (MAR) assumption when only the predictor is missing but not the outcome missing in a regression analysis. In reality, the MAR assumption is more likely to be true than the MCAR assumption. Therefore, careful further investigation is needed to see how the missing data techniques (listwise deletion versus FIML) work under different assumptions (MAR or MCAR assumption) for the survival-

mediator model. Furthermore, in many cases, censoring may occur because the timing value itself is the cause of the censoring. That is, censoring may occur because the true time exceeds a specific given time point (i.e., the end of the observation). This missing data mechanism where the missingness is dependent on the variable's value itself is called not missing at random (NMAR). Future studies need to examine missing data techniques under the NMAR assumption.

2. Measures of the Indirect Effect

In mediation analysis with a single continuous mediator and a continuous outcome, the indirect effect is quantified as the product of *a* and *b*, *ab* or equivalently, $c - c'$. The equivalence of $ab = c - c'$ only holds for specific conditions: 1) when both the mediator and outcome are continuous variables and 2) the sample size for estimating the parameters in the *M*- and *Y*-regression are the same (MacKinnon, 2008). The equivalence between the two indirect effect measures does not hold for a survival-mediator model. The reason of the nonequivalence is related to the different uses of the survival mediator in the *M*- and *Y*-regression. For the *M*-regression, the outcome is the hazard rate. The *a*-coefficient, which is the log hazard rate change for a one-unit increase in the *X* variable (the difference between the treatment and control), is in the natural logarithm unit (refer to Equation I.1.28 of Chapter I). The exponent of *a* represents the hazard ratio which is a ratio of the hazard rate when $X=1$ to the hazard rate when $X=0$. In contrast, *M* is used as a continuous time variable in the *Y*-regression. The *b*-coefficient is the change in *Y* for a one-unit increase in *M* controlling for *X*. The discrepancy in the use of *M* in the *M*- and *Y*-regression no longer makes the *a* and *b* coefficients comparable in

units. A one-unit change of the hazard rate does not directly match on to a one-unit change in time. Therefore, the *ab*-product for the survival-mediator model is difficult to interpret. Nonetheless, the test of the *ab* effect is still effective as a test for the indirect effect. VanderWeele (2011) discusses a mediation model where the outcome variable is a survival variable and has mathematically proven that the test of the *ab* effect for the survival mediation model is a valid test for the indirect effect. The same principles apply to the survival-mediator model and therefore, although the estimate of the *ab* effect does not have a clearly meaningful interpretation, the test of it is still valid.

The $c - c'$ estimate for the indirect effect has not been discussed in this dissertation. The main reason why this estimate was not discussed was because previous research on mediation with binary variables have shown that the $c - c'$ estimate was not a good measure of the indirect effect and also the tests for the $c - c'$ estimate showed poor performance (MacKinnon & Dwyer, 1993; MacKinnon et al., 2002; MacKinnon et al., 2007). There are not many studies done on the $c - c'$ estimate with survival mediation models. Tein and MacKinnon (2003) showed that the *ab* estimate does not equal the $c - c'$ estimate for a mediation model with a survival outcome. However, it might be the case that for the survival-mediator model the $c - c'$ estimate performs differently than the previous studies. Future studies on the $c - c'$ estimate and on the relationship with the *ab* estimate for the survival-mediator model is needed.

Another estimate of the indirect effect that was also introduced in this dissertation for the survival-mediator model is the NIE. Following the mathematical derivation in Equation I.3.6 in Chapter I, the NIE for the survival-mediator model can be expressed as $b * \{\int S(M|X = 1)dm - \int S(M|X = 0)dm\}$. If there is no exposure-mediator (*XM*)

interaction and no other confounders in the model, the NIE should be directly related to the *ab* estimate. Since *b* is both multiplied by the remaining part of the *ab* and NIE estimates, the corresponding part of *a* in the NIE is the term in the curly brackets. If we go one step above from the final result of Equation I.3.6, the mathematical expression inside the curly brackets is $\{E[M|X=1]-E[M|X=0]\}$. Therefore, it is the mean time difference between the treatment and control group. One might question why not just calculate the mean time for the treatment and control group separately and take their difference instead of doing complicated calculations for the integrals? The answer to the question is that the expected value will be biased with censored data in *M*. The final mathematical expression of Equation I.3.6 gives unbiased estimates of the NIE because the survival function handles the censored data appropriately.

Unfortunately, the conditional survival functions in the final expression of the NIE is mathematically too complex to integrate across time. Instead, a numerical integration approach is used to approximate the integrals. Once given data, numerical integration using the trapezoidal rule[8] can be used to calculate the integrals of the survival functions. Also, the standard error of the NIE would be mathematically too complicated to derive and thus, a bootstrap approach was used to estimate confidence intervals for the NIE. As mentioned above, the corresponding part of the *a*-coefficient in the NIE is $\{E[M|X=1]-E[M|X=0]\}$. Although the actual quantity is estimated by $\{\int S(M|X=1)dm-\int S(M|X=0)dm\}$ using a numerical integration method, mathematically this value equals to the mean time difference between the treatment and

[8] The area under the curve can be estimated by calculating the area of a trapezoid for an interval. To increase accuracy of the estimate, multiple intervals are divided for the X-axis and the area of the trapezoids for each of the intervals are added up.

control group. Therefore, in contrast to the *ab* estimate, the parts in the NIE (*b* and the expression in curly brackets) are comparable in units (both are in time units). The NIE can be interpreted as the average change in *Y* between the treatment and the control group that is mediated by the time-to-event. An empirical example interpreting the NIE will be given in Section 4 of this chapter.

Last, the relationship between the *a*-parameter and the corresponding part in the NIE, $\{\int S(M|X=1)dm - \int S(M|X=0)dm\}$ is examined. Here, the comparison between the two quantities is empirical since we do not know the true value of the NIE and the NIE can only be calculated from the data. Data were generated following the steps described in Chapter II. One major difference was that only one replication for each condition was generated with a huge sample size of *n* = 1,000,000 to resemble the population. To be consistent, the other parameter values were fixed (*b* = 0.39, *c'* = 0.39, *n* = 1,000,000, and censoring proportion = 0.3)[9] while manipulating the values of the *a*-parameter at -0.82, -0.52, -0.21, 0, 0.21, 0.52, and 0.82. Figure IV-2 shows the empirical relationship between the *a*-estimate and the $\{\int S(M|X=1)dm - \int S(M|X=0)dm\}$. The relationship seems to be a nonlinear but it was unknown what the underlying true relationship would be. Further research on this subject is needed.

In conclusion, the NIE is the best measure of the indirect effect in a survival-mediator model. However, since the NIE is a complicated function, conventional statistical tests (*z* or *t* tests) are difficult to conduct. Bootstrapping of the NIE estimate is available but, as shown in the simulation study in this dissertation, the Type I error rate is

[9] There was a negligible difference in the estimates of *a* and $\{\int S(M|X=1)dm - \int S(M|X=0)dm\}$ when the *b*, *c'* or censoring proportion was modified to other values. Only the value of the *a*-parameter affected the estimates of $\{\int S(M|X=1)dm - \int S(M|X=0)dm\}$.

lower than the nominal level in some conditions and also bootstrapping of the NIE is statistically underpowered relative to the other indirect effect test methods especially when there is non-zero censoring in *M*. In contrast, the *ab*-estimate is difficult to interpret due to the incomparable units of the *a* and *b* parameters but, the advantage is that several tests are available and valid to evaluate the indirect effect. More research on the $c - c'$ estimate and its relationship with the other indirect effect measures for the survival-mediator model is needed.

3. Conclusions of the Simulation Study

A simulation study was conducted to assess performance of the different indirect effect test methods in the survival-mediator model. Following the model depicted in Figure II-1, 1,000 replication datasets were generated. The main factor of the simulation study was the six different indirect effect test methods and other data generating related factors that were manipulated were the magnitude of *a*, *b*, *c'*, the censoring proportion in *M*, and the sample size. For each of the 1,000 generated datasets, the Cox model for the *M*-regression was fitted to the data. For the *Y*-regression, an ordinary least squares (OLS) regression while treating the censoring in *M* as missing was fitted to the data. Type I error rate, statistical power, parameter coverage rate, raw bias, relative bias, negative or positive bias, and the mean squared error (MSE) were examined to see whether the proposed model fitted well to the data and to compare the performance of the different indirect effect test methods. ANOVA and logistic regression depending on the outcome (ANOVA for continuous outcomes and logistic regression for binary outcomes) were conducted to help interpret the results.

A. Summary of the Results

The results for the raw bias, relative bias and the sign of the raw bias consistently showed that the proposed survival-mediator model fitted the data well without major issues. For the three aforementioned outcomes, there were no substantial effects of the model generating factors by themselves or any of the 2- and 3-way interactions of the factors. This result supports that the proposed statistical model works well for the survival-mediator model across different conditions.

As the size of *b* increased, the MSE of the *ab* estimate increased substantially. The MSE captures two components: the bias of an estimate and the standard error of an estimate. The increased MSE as *b* increased came from the standard error of the *ab* estimate and not from the bias because the raw and relative bias showed no substantial difference as *b* increased.

Paralleling previous research (Fritz et al., 2012; MacKinnon et al., 2002) using a mediation model with a continuous mediator and outcome, the Type I error rate for the survival-mediator model was higher than the nominal level for the bias-corrected bootstrap method and lower than the nominal level for the Sobel test. In this simulation, this result was most clearly found when a=0.52 and b=0. Also, the Type I error rate using the bootstrap of the NIE method was lower than the nominal level, especially when a=0, b was either 0.14 or 0.39 and the censoring proportion was 0.3. The NIE was approximated using a numerical integration method. The accuracy of the numerical integration depends on how many time points are available in the data. The accuracy of the NIE increases if there are many time points and the accuracy of the NIE estimate decreases if there are fewer time points in the data. In Equation I.3.6, although the range

of the integrals is from zero to infinity, the area under the curve is only calculated for a finite range of time from zero to the maximum time in the data. Since the survival function is a decreasing function that converges to zero, the area under the curve after some time point becomes very small and thus it will make negligible difference in calculating the NIE. However, as the censoring proportion increases, the number of time points for the integration decreases and thus, the accuracy of the NIE estimate decreases. Inaccuracy in the NIE estimate might have caused the low Type I error rate. Another related result was that the bootstrap of the NIE method produced lower statistical power than the other methods, especially when the censoring proportion was 0.3. Again, the NIE estimate can be biased with censored data and this might have caused a low statistical power.

Finally, in some of the conditions, the parameter coverage rate was over the nominal level (=.95) for the Sobel test and under the nominal level for the bias-corrected bootstrap of *ab*. This was true only when b=0; the difference in the parameter coverage rate was most prominent when a=0.52. The difference between the methods was smaller when the size of a was smaller or when b was a non-zero effect regardless of the size of a. This result reflects the results from the Type I error rate. The bias-corrected bootstrap rejects the null hypothesis (ab=0) too often. This result is also reflected in the 95% confidence intervals that include the true parameter value (ab=0) than at lower than the expected 95% value. In contrast, the Sobel test accepts the null hypothesis too often which is also reflected in the 95% confidence intervals including the zero value too often.

B. Recommendations

Based on the simulation results and some other considerations that were previously mentioned, below are several recommendations when fitting a survival-mediator model.

1. The Cox model for the *M*-regression and the OLS regression with listwise deletion for the *Y*-regression are recommended statistical models to achieve unbiased estimates of the survival-mediator model. The simulation results support that these statistical models fit the data well except that the standard error of *ab* increased as an increase in the *b*-coefficient. Nonetheless, the estimates of the survival-mediator model were unbiased. There are limitations of the proposed statistical models which will be presented in the next section.

2. Cautions should be taken in making decisions about the indirect effect based on just the bias-corrected bootstrap method or the Sobel test. Both are statistical tests for the *ab* effect. As mentioned in Section 2 in the Discussion, the tests on the *ab* effect are valid tests of the indirect effect, although the *ab* estimate itself is difficult to interpret. However, as the simulation study showed and also as previous literature has shown, the bias-corrected bootstrap produces higher than nominal level Type I error rates and the Sobel test produces lower than nominal level Type I error rates. The recommended way to test the indirect effect would be to use multiple methods and then, compare the results. Researchers can make statistical decisions confidently if the multiple methods all agree.

3. Caution should be taken in making decisions about the indirect effect based on the bootstrap of the NIE method, especially when the censoring proportion is high. With highly censored data, the number of time points in the data might not be enough to

calculate an accurate estimate of the NIE. More importantly, there are no other statistical tests that can be conducted on the NIE except for bootstrapping methods. This restricts researchers in making statistical conclusions about the NIE for the survival-mediator model.

4. The NIE estimate should be reported and statistical tests conducted on the *ab* effect. The NIE for the survival-mediator model is interpretable but lacks a theoretically derived statistical testing method. On the other hand, the *ab* estimate is hard to interpret but a number of valid statistical tests are available. An example of the interpretation of the NIE estimate will be given in the next section for an empirical data example. Then, researchers can make statistical decisions by comparing conclusions from multiple statistical tests of the *ab* effect.

C. Limitations and Suggestions for Future Research

There has been some previous research on mediation analysis with a survival variable as an outcome (Tein & MacKinnon, 2003; VanderWeele, 2011). However, this is one of the first theoretical studies on mediation analysis with a survival variable as a mediator. In this dissertation, a statistical model that provides unbiased estimates of the survival-mediator was proposed. In addition, a clearly interpretable measure of the indirect effect, the NIE, was mathematically derived. Finally, a simulation study was conducted to examine the performance of the different indirect effect tests. This dissertation makes important contributions to the literature of survival mediation analysis. Nevertheless, there are several limitations of the dissertation and much more to be understood about the analysis of the survival-mediator model.

1. The MCAR assumption of the censored data: One of the important characteristics of survival data is censoring. Survival analysis where the outcome variable is a survival variable has been studied extensively over the past 40 years. In a model where the survival variable is used as an outcome (e.g., Cox model), appropriate estimators (e.g., maximum likelihood) have been developed to handle the censored data. However, little is known how the censored data should be handled when the survival variable is used as a predictor in the mediation model.

In this dissertation, the censoring in *M* was treated as missing values and assuming that the underlying missing data mechanism is MCAR, complete case analysis was utilized in the *Y*-regression. There might be a question whether the MCAR assumption actually holds for the simulation study conducted. The method that was used to generate the censoring data resembles how Mplus generates censored data (Muthén & Muthén, 1998-2012). In Mplus, the censoring process follows a random exponential distribution. Since an individual's censoring time is independently and randomly picked from the exponential distribution, the MCAR assumption is satisfied. The more interesting question is whether the MCAR assumption is reasonable for real data. In the real world, it is difficult to believe that the censoring process would be completely independent of all other variables in the study. Then, the MAR assumption seems more plausible for the censoring. Under the MAR assumption, the complete case analysis for the *Y*-regression might not be the best strategy to deal with the censoring data (Enders, 2010). On the other hand, Allison (2001) argued that the complete case analysis works well for regression analysis when only the predictor has missing data. It is yet unclear whether the complete case analysis will work well for the *Y*-regression of the survival-

mediator model even when MAR is the underlying missing data mechanism. Based on the small simulation study in Chapter 1, 3. Survival Mediation Analysis, Section B.2.b, the FIML method did not work well assuming MCAR. Theoretically, FIML should work as least as well as the complete case analysis assuming MCAR and FIML should work better than the complete case analysis assuming MAR as the underlying missing data mechanism. Future studies need to compare the complete case analysis versus the FIML method assuming either the MCAR censoring or the MAR censoring.

2. The relationship between the *ab* and NIE estimates: If the mathematical relationship between *ab* and the NIE were known, then we would be able to convert one estimate to another. The mathematical relationship will be challenging to determine because the function for the NIE is complex. Even using the simplest mathematical form for the survival function, the integral of the survival function is intractable. Instead, a numerical integration method was used to approximate the NIE. At this moment, the only way to investigate the relationship between the *ab* effect and the NIE is by inferring the relationship from empirical data. Figure IV-2 is one of the first attempts to infer the relationship between *a* and the corresponding part in the NIE. The relationship seems like a nonlinear function, but the actual nonlinear function is unknown. A more comprehensive study is needed to further investigate the nonlinear relationship.

3. More detailed or extreme levels of the data generation parameters: The data generation parameters that were manipulated in this dissertation were the size of *a*, *b*, *c'*, the censoring proportion of *M*, and the sample size. The size of the *a* and *b* parameters were manipulated to reflect zero, small and medium effect sizes and the size of *c'* was manipulated to reflect zero and medium effect sizes. Type I error rates were substantially

high for the bias-corrected bootstrap method and substantially low for the Sobel test when *a* was a medium effect and *b* was zero. There might be more interesting results with larger effects of the *a*, *b* and *c'* parameters. The censoring proportion was manipulated to be zero percent and 30%. With 30% censoring, the bootstrap of the NIE method started to show substantially lower statistical power than the other indirect effect test methods. Censoring proportions between zero and 30% need to be studied more carefully to see how robust the statistical power is for the bootstrap of the NIE method. Also, it might be intriguing to see the performance of the indirect effect test methods when the censoring proportion is higher than 30%. The sample sizes used for this study were 150 and 300. Sample sizes less than n=150 or greater than n=300 might be worth investigating (especially, $n < 150$) to assess how sample size affects the results. In sum, the conclusions made from the simulation studies are restricted to the levels that were used to generate the datasets. Therefore, a more thorough investigation is needed to generalize the results found from this study.

4. Other survival models for the *M*-regression: The Cox regression model was considered as the *M*-regression. The advantage of using the Cox model is that, although the baseline hazard rate is included in the mathematical expression of the Cox model, it need not to be specified or estimated when estimating the regression parameter that is the main interest of the mediation model. An alternative for the *M*-regression could be the AFT model. In contrast to the Cox model, the baseline survival function would need to be specified in order to estimate the regression parameter in the AFT model. The regression estimate can be biased if the baseline survival function is misspecified.

Nonetheless, the acceleration parameter of the AFT model offers an interesting interpretation and could be an estimand in which the researcher is particularly interested.

5. More work on interpreting empirical data: Much of this dissertation focused on theoretical issues for analysis of the survival-mediator model. An appropriate statistical model for the survival-mediator model was proposed and a new measure of the indirect effect was developed. A simulation study was conducted to examine the performance of the proposed method and to compare different statistical testing methods. However, no context was given to the model or the parameter estimates. Without a context, the estimates are less meaningful and the interpretation of the survival-mediator model is less clear. Much more work is needed to apply the proposed model to real data to fully understand the meaning of the survival-mediator model and its parameter estimates. To broaden the understanding of the survival-mediator model, I fit the model to a made-up data example and interpret its parameter estimates in the following section.

4. Constructed Empirical Example

To understand the meaning of the estimates of the survival-mediator model, I generated a random sample using the survival-mediator model (see Figure II-1). The parameter values that were selected are a=0.52, b=0.39, c'=0.39, λ_T=0.1, λ_C=0.043 (which generates a censoring proportion of approximately 0.3) and a sample size of 150. The survival-mediator model was fitted to the data and the results were interpreted. For meaningful interpretations, here is the description of the data.

A preventive intervention targeted to prevent child problem behaviors was randomly assigned to 150 families. Out of the 150 families, 74 people received the intervention and 76 received the control. One of the mediators that was of interest was the duration of coercive interactions between the parent and child. Prior literature has focused on the detrimental effects of coercive parent-child relations (Patterson, 1982). The intervention targets reducing the duration of coercive interactions. In turn, reduced coercive interactions are expected to reduce the child's future problem behaviors. To assess coercive interactions, a five-minute interaction task was given to parent-child dyads and the duration of a coercive interaction bout was measured in seconds. 23.33% (n=35) of the observations were censored because the participant's data were randomly lost after a certain time point. For example, the end time of the coercive interaction bout is unknown because the stopwatch malfunctioned after a certain time point. The child behavior checklist (CBCL) externalizing scale was used to measure child's problem behaviors one year after the interaction task.

The single survival-mediator model fitted to data produced parameter estimates of $\hat{a}$ = 0.44 (SE=0.20), $\hat{b}$ = 0.38 (SE=0.01), and $\widehat{c'}$ = 0.53 (SE=0.03). The hazard rate of ending the coercive bout was about 1.6 times (exp(.44)=1.56) more likely for the treatment group compared to the control group so there is evidence that the intervention reduced the time in coercive bouts. Furthermore, a one second increase in the duration of coercive interaction predicted an average of 0.38 points increase in the CBCL externalizing score after controlling for the treatment indicator. The predicted CBCL

internalizing score was 0.53 points larger on average for the treatment group than the control group while controlling for the coercive interaction duration.

The estimate of the indirect effect was $\widehat{ab}$ = .17 (*SE*=.01), $p < .001$, with *95% CI*=(.02, .32) using the distribution of the product method. One thousand bootstraps of the *ab* estimate resulted in a 95% confidence interval of (.03, .31). Since the intervention indicator predicted the hazard rate of ending the coercive bout (z=2.24 , p=.01), and the coercive bout duration predicted externalizing scores, controlling for the intervention indicator (z=26.32, p<.001), the joint-significance test was significant.

Finally, the natural indirect effect (NIE) was estimated as -1.38 for this dataset. The NIE is the difference in the average externalizing score between the treatment group and the control group that is produced through the mediator. On average, the treatment group had a 1.38 point lower externalizing score than the control group that was mediated by the duration in the coercive bout. For 1,000 bootstrap samples, the average NIE was -1.02 with a 95% bootstrap confidence interval of (-1.99, -0.09). The corresponding a estimate in the NIE ($\int_0^\infty S(M|X=1) - S(M|X=0)dm$) which shows the average coercive interaction duration difference between the treatment and control group was -3.11. Thus, on average, the treatment group ended the coercive interaction 3.11 seconds earlier than the control group.

5. Final Conclusions

Survival analysis integrates information from two variables: *when* the event occurs and *whether* the event occurs or not. If the event does not occur within the observation period, the event is termed a censored event. Censoring means that the event

did not occur yet, but there is a probability that the event will happen at some later time. The benefit of survival analysis is that the censoring probability is always considered and used in the analysis rather than being ignored. Events that occur in social sciences do not always lead to a terminal state (e.g., death) as in traditional survival analysis in medicine. Therefore, the survival variable can be used as a predictor of another outcome and/or as an outcome of another predictor. The current study examines such a model where the independent variable is a binary variable (e.g, intervention indicator) that predicts a survival mediator, which, in turn, predicts a continuous outcome.

A statistical model for the survival-mediator model was proposed in this study. The Cox model was used for the *M*-regression where the survival variable was used as an outcome and OLS regression was used for the *Y*-regression where the survival variable was used as a predictor. To deal with the censored observations in the *M* predictor in the *Y*-regression, the censored observations were treated as missing data and complete case analysis was used, assuming that the censoring mechanism was MCAR. Two different quantities for measuring the indirect effect were discussed. The NIE is the better measure of the indirect effect because it is clearly interpretable. However, due to its complex function the NIE does not have a theoretically derived statistical test. In contrast, the *ab* effect is difficult to interpret because of the inconsistent units of the *a* and *b* coefficients, but the statistical tests for the *ab* effect are valid tests of the indirect effect. A simulation study was conducted comparing different indirect effect test methods. The results showed that the Type I error rate was too high and parameter coverage was too low using the bias-corrected bootstrap method. In contrast, the Type I error rate was too low and the parameter coverage rate was too high using the Sobel test. Additionally, as

the censoring proportion in the data increased, the NIE estimate became inaccurate in some cases and the percentile bootstrap of the NIE method displayed low Type I error and low statistical power. Despite its limitations, this is the first study investigating the survival-mediator model. Much more work is required, but the proposed survival-mediator framework is a promising area of research that can answer unique research questions.

REFERENCES

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data* (Sage University paper series on quantitative applications in the social sciences, Number 07-046). Beverly Hills, CA: Sage.

Allison, P. D. (2001). *Missing data* (Sage university papers series on quantitative applications in the social sciences, 07-136). Thousand Oaks, CA: Sage.

Allison, P. D. (2010). *Survival Analysis Using SAS: A Practical Guide, Second Edition.* Cary, NC: SAS Institute Inc.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology, 51* (6), 1173-1182. doi: 10.1037/0022-3514.51.6.1173.

Beason, W., & Morgan, J. (1984). Glass Failure Prediction Model. *Journal of Structural Engineering, 110* (2), 197-212. doi: 10.1061/(ASCE)0733-9445(1984)110:2(197).

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine, 24* (11), 1713-1723. doi: 10.1002/sim.2059.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics, 30* (1), 89-99. doi: 10.2307/2529620.

Christensen, K. B., Labriola, M., Lund, T., & Kivimäki, M. (2008). Explaining the social gradient in long-term sickness absence: a prospective study of Danish employees. *Journal of Epidemiology Community Health. 62* (2), 181–183. doi: 10.1136/jech.2006.056135.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B, 34*, 187-220. doi: 10.2307/2985181.

Dagne, G. A., & Snyder, J. (2009). Bayesian Hierarchical Duration Model for Repeated Events: An Application to Behavioral Observations. *Journal of Applied Statistics, 36* (11), 1267-1279. doi: 10.1080/02664760802587032.

Dagne, G. A., & Snyder, J. (2011). Relationship of maternal negative moods to child emotion regulation during family interaction. *Development and Psychopathology, 23* (1), 211-223. doi: 10.1017/S095457941000074X.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistician Association, 76* (359), 312-319. doi: 10.1080/01621459.1977.10480613.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82* (397), 171-185. doi: 10.1080/01621459.1987.10478410.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41* (4), 1149-1160. doi: 10.3758/BRM.41.4.1149.

Finchman, M. (1989). Attendance makes the heart grow fonder: A hazard rate approach to modeling attendance. *Journal of Applied Psychology, 74* (2), 325-335. doi: 10.1037/0021-9010.74.2.325.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18* (3), 233-239. doi: 10.1111/j.1467-9280.2007.01882.x.

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47* (1), 61-87. doi: 10.1080/00273171.2012.640596.

Fulcher, I. R., Tchetgen Tchetgen, E., & Williams, P. L. (2016). Mediation analysis for censored survival data under an accelerated failure time model. *Harvard University Biostatistics Working Paper Series*. Working Paper 211.

Gardner, W. (1993). Hierarchical continuous-time sequential analysis: A strategy for clinical research. *Journal of Consulting and Clinical Psychology, 61* (6), 975-983. doi: 10.1037/0022-006X.61.6975.

Gardner, W., & Griffin, W. (1989). Methods for the analysis of parallel streams of continuously recorded social behaviors. *Psychological Bulletin, 105* (3), 446-455. doi : 10.1037/0033-2909.105.3.446.

Gelfand, L. A., MacKinnon, D. P., DeRubeis, R. J., & Baraldi, A. N. (2016). Mediation analysis with survival outcomes: Accelerated failure time vs. proportional hazards models. *Frontiers in Psychology, 7*, 1-10. doi: 10.3389/fpsyg.2016.00423.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. (2014). *Bayesian data analysis (3rd Ed.)*. Boca Raton, FL: Chapman and Hall/CRC.

Griffin, W., & Gardner, W. (1989). Analysis of behavioral durations in observational studies of social interaction. *Psychological Bulletin, 106* (3), 497-502. doi: 10.1037/0033-2909.106.3.497.

Hall, W. J., & Wellner, J. A. (1981). Mean residual life. In Csörgö, M., Dawson, D. A., Rao, J. N. K & Saleh, A. K. Md. E. (Eds.), *Statistics and Related Topics* (pp. 169-184). Amsterdam, Netherlands: North-Holland publishing company.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117* (1), 167-178. doi: 10.1037/0033-2909.117.1.167.

Hosmer, D. W., Lemeshow, S., May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd Ed.). Hoboken, NJ: John Wiley & Sons.

Hosmer, D. W., Lemeshow, S., Sturdivant, R. X. (2013). *Applied Logistic Regression (3rd ed.).* Hoboken, NJ: John Wiley & Sons.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York, NY: Springer.

Hsieh, F. Y., Lavori, P. W. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials, 21* (6), 552-560. doi: 10.1016/S0197-2456(00)00104-5.

Ibrahim, J. G., Chen, M-H., & Sinha, D. (2001). *Bayesian Survival Analysis*. New York, NY: Springer.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105* (4), 765-789. doi: 10.1017/S0003055411000414.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5* (5), 602-619. doi: 10.1177/0193841X8100500502.

Jung, S. Y., Rosenzweig, M., Linkov, F., Brufsky, A., Weissfeld, J. L., Sereika, S. M. (2011). Comorbidity as a mediator of survival disparity between younger and older women diagnosed with metastatic breast cancer. *Hypertension, 59* (2), 205–211. doi: 10.1161/HYPERTENSIONAHA.111.171736.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data (2nd ed.).* Hoboken, NJ: John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53* (282), 457-481. doi: 10.1080/01621459.1958.10501452.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature, 26* (2), 646-679.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data*. New York, NY: Springer.

Lacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology, 22* (4), 582-594. doi: 10.1016/j.jcps.2012.03.006.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrika, 47* (4), 939-956. doi: 10.2307/1914140.

Lancaster, T., & Nickell, S. (1980). The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society, Series A (General), 143 (2),* 141-165. doi: 10.2307/2981986.

Lange, T., & Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology, 22* (4), 575–581. doi: 10.1097/EDE.0b013e31821c680c.

Lavori, P. W., Keller, M. B., & Klerman, G. L. (1984). Relapse in affective disorders: A reanalysis of the literature using life table methods. *Journal of Psychiatric Research, 18* (1), 13-25. doi: 10.1016/0022-3956(84)90045-1.

MacKinnon, D, P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17* (2), 144-158. doi: 10.1177/0193841X9301700202.

MacKinnon, D. P., Lockwood, C. M., & Hoffman, J. M. (1998). *A new method to test for mediation*. Paper presented at the annual meeting of the Society for Prevention Research, Park City, UT.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39* (1), 1-24. doi: 10.1207/s15327906mbr3901_4.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39* (3), 384-389. doi: 10.3758/BF03193007.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7* (1), 83-104. doi: 10.1037/1082-989X.7.1.83.

MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials, 4* (5), 499-513. doi: 10.1177/1740774507083434.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22* (4), 719-748. doi: 10.1093/jnci/22.4.719.

Masyn, K. (2014). Discrete-time survival analysis in prevention science In Sloboda, Z. & Petras, H. (Eds.), *Defining Prevention Science* (pp. 513-536). New York, NY: Springer.

Meeker, W. Q., & Escobar, L. A. (1994). An algorithm to compute the cdf of the product of two normal random variables. *Communications in Statistics: Simulation and Computation, 23* (1), 271-280. doi: 10.1080/03610919408813168.

Mills, M. (2011). *Introducing survival analysis and event history analysis*. Thousand Oaks, CA: SAGE.

Muldowney, P., Ostaszewki, K., & Wojdowski, W. (2012). The Darth Vader Rule. *Tatra Mountains Mathematical Publications, 52*, 53-63. doi: 10.2478/v10127-012-0025-9.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. *Manuscript submitted for publication*.

Muthén, B., & Asparouhov, T. (2014). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling, 22* (1), 12-23. doi: 10.1080/10705511.2014.935843.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (pp. 411-420). San Francisco, CA: Morgan Kaufmann.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society, Series A (General), 135* (2), 185-207. doi: 10.2307/2344317.

Rice, J., Reich, T., Andreasen, N. C., Endicott, J., Van Eerdewegh, M., Fishman, R., Hirschfeld, R., & Klerman, G. L. (1987). The familial transmission of bipolar illness. *Archives of General Psychiatry, 44* (5), 441-447. doi: 10.1001/archpsyc.1987.01800170063009.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66* (5), 681-701. doi: 10.1037/h0037350.

SAS Institute Inc. (2014). *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

Shafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7* (2), 147-177. doi: 10.1037/1082-989X.7.2.147.

Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review, 49* (2), 261-272. doi: 10.2307/2095575.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110* (2), 268-290. doi: 10.1037/0033-2909.110.2.268.

Singer, J. D., & Willett, J. B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18* (2), 155-195. doi: 10.2307/1165085.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: modeling change and event occurrence*. New York, NY: Oxford University Press.

Singer, J. D., Fuller, B., Keiley, M. K., & Wolf, A. (1998). Early child-care selection: Variation by geographic location, maternal characteristics, and family structure. *Developmental Psychology, 34* (5), 1129-1144. doi: 10.1037/0012-1649.34.5.1129.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312. doi: 10.2307/270723.

Stevens, V. J., & Hollis, J. F. (1989). Preventing smoking relapse using an individually tailored skills training technique. *Journal of Consulting and Clinical Psychology, 57* (3), 420-424. doi: 10.1037/0022-006X.57.3.420.

Stoolmiller, M. (2014). *Using multivariate multilevel survival analysis to study reliability and change in hazard rates of emotions derived from parent-child dyadic social interaction.* Presented at the Dynamic Mediation workshop in Tempe, AZ.

Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes using multilevel survival analysis. *Psychological Methods, 11* (2), 164-177. doi: 10.1037/1082-989X.11.2.164.

Stoolmiller, M., & Snyder, J. (2013). Embedding multilevel survival models of dyadic social interaction in structural equation models: Hazard rates as both outcomes and predictors. *Journal of Pediatric Psychology, 39* (2), 222-232. doi: 10.1093/jpepsy/jst076.

Tein, J-Y., & MacKinnon, D. P. (2003). Estimating mediated effects with survival data in H. Yanai, A. O. Rikkyo, K. Shigemasu, Y. Kano, & J. H. Meulman (Eds.) New *Developments on Psychometrics* (pp. 405-512). Tokyo, Japan: Springer-Verlag Tokyo Inc.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43* (3), 692-700. doi: 10.3758/s13428-011-0076-x.

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology, 22* (4), 582-585. doi: 10.1097/EDE.0b013e31821db37e.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface, 2* (4), 457-468. doi: 10.4310/SII.2009.v2.n4.a7.

Zhai, G., & Lin, H. (2004). Controller failure time analysis for symmetric control systems. *International Journal of Control, 77* (6), 598-605. doi: 10.1080/00207170410001703232.

Table I-1.

Example of Calculating the Kaplan-Meier Estimate

Time to death (t_i)	Number survived (N_{t_i})	Number died (E_{t_i})	Number censored (C_{t_i})	Kaplan-Meier estimate ($\hat{S}(t)$)
t_0 = 0	100	0	0	1.00
t_1 = 2	100	1	0	0.99
t_2 = 10	99	8	1	0.91
t_3 = 15	90	30	5	0.61

Table I-2.

Comparison of the Accelerated Failure Time Model and Cox Proportional Hazards Model

	Accelerated failure time model	Cox proportional hazards model
Focused dependent variable	Survival function	Hazard function
Functional form	Parametric	Semi-parametric
Assumption	Distribution of the survival function	Proportional hazards assumption
Estimator	Maximum likelihood	Partial maximum likelihood
Advantages	Easy interpretation	No need to specify a baseline hazard function
Disadvantages	True distribution of the survival function can be misspecified	The proportional hazards assumption can be violated.

Table I-3.

Made-up Example to Illustrate Partial Likelihood

Event j	Individual i	Time (years) t_i	Event (death) δ_i
1	79	5	1
.	25	6	0
2	67	7	1
.	43	7	0
3	8	11	1
.	17	11	0
…	…	…	…

Table I-4.

Simulation Results Using Four Different Methods to Deal with the Censoring Predictor Issue in the Y-regression

Method	Average *b* (*S.E.*)	Statistical Power of *b*	Average *c'* (*S.E.*)	Type-I error of *c'*
1. True *M*	.14 (.01)	100%	-.001 (.06)	5%
2. Censored *M*	.14 (.14)	100%	-.11 (.08)	24.5%
3. FIML	.17 (.10)	12.6%	-.02 (.27)	0%
4. Complete Case	.14 (.01)	100%	-.001 (.07)	5%

Note. b is the regression coefficient for *M* and *c'* is the regression coefficient for *X* in the *Y*-regression.

FIML = Full Information Maximum Likelihood.

Table III-2.

Average Raw Bias of the ab Estimate when c' = 0

a	*b*	Censoring Proportion	Sample Size	Average Raw Bias
0	0	0	150	0.0000
0	0.14	0	150	-0.0001
0	0.39	0	150	0.0047
0.21	0	0	150	0.0000
0.21	0.14	0	150	0.0000
0.21	0.39	0	150	0.0029
0.52	0	0	150	-0.0001
0.52	0.14	0	150	0.0005
0.52	0.39	0	150	0.0009
0	0	.3	150	0.0000
0	0.14	.3	150	-0.0009
0	0.39	.3	150	0.0011
0.21	0	.3	150	0.0001
0.21	0.14	.3	150	0.0009
0.21	0.39	.3	150	- 0.0029
0.52	0	.3	150	-0.0004
0.52	0.14	.3	150	0.0000
0.52	0.39	.3	150	-0.0014
0	0	0	300	0.0000
0	0.14	0	300	-0.0001
0	0.39	0	300	0.0004
0.21	0	0	300	0.0000
0.21	0.14	0	300	-0.0003
0.21	0.39	0	300	0.0006
0.52	0	0	300	0.0002
0.52	0.14	0	300	0.0010
0.52	0.39	0	300	-0.0011
0	0	.3	300	0.0000
0	0.14	.3	300	-0.0007
0	0.39	.3	300	-0.0030
0.21	0	.3	300	-0.0001
0.21	0.14	.3	300	0.0005
0.21	0.39	.3	300	-0.0001
0.52	0	.3	300	0.0000
0.52	0.14	.3	300	-0.0004
0.52	0.39	.3	300	0.0002

Table III-3.

Average Raw Bias of the ab Estimate when c' = 0.39

a	*b*	Censoring Proportion	Sample Size	Average Raw Bias
0	0	0	150	0.0000
0	0.14	0	150	-0.0012
0	0.39	0	150	0.0011
0.21	0	0	150	0.0000
0.21	0.14	0	150	-0.0006
0.21	0.39	0	150	0.0019
0.52	0	0	150	-0.0002
0.52	0.14	0	150	0.0014
0.52	0.39	0	150	0.0039
0	0	.3	150	0.0000
0	0.14	.3	150	-0.0003
0	0.39	.3	150	0.0049
0.21	0	.3	150	-0.0001
0.21	0.14	.3	150	-0.0001
0.21	0.39	.3	150	0.0013
0.52	0	.3	150	-0.0006
0.52	0.14	.3	150	-0.0010
0.52	0.39	.3	150	0.0074
0	0	0	300	0.0000
0	0.14	0	300	-0.0005
0	0.39	0	300	-0.0029
0.21	0	0	300	0.0000
0.21	0.14	0	300	0.0007
0.21	0.39	0	300	0.0008
0.52	0	0	300	0.0000
0.52	0.14	0	300	0.0007
0.52	0.39	0	300	0.0039
0	0	.3	300	-0.0001
0	0.14	.3	300	0.0003
0	0.39	.3	300	0.0025
0.21	0	.3	300	0.0001
0.21	0.14	.3	300	0.0004
0.21	0.39	.3	300	0.0009
0.52	0	.3	300	0.0004
0.52	0.14	.3	300	0.0007
0.52	0.39	.3	300	0.0040

Table III-3.

Analysis of Variance Results for the Average Raw Bias of the ab Estimate

Source	*DF*	Type III *SS*	Mean Square	*F* Value	*Pr* > *F*	Partial $\widehat{\omega^2}$
A	2	0.0055	0.0027	1.86	0.16	0.0000
B	2	0.0282	0.0141	9.61	<.0001	0.0002
C	1	0.0108	0.0108	7.38	0.01	0.0001
CP	1	0.0003	0.0003	0.18	0.67	0.0000
SS	1	0.0030	0.0030	2.03	0.15	0.0000
AxB	4	0.0076	0.0019	1.30	0.27	0.0000
AxC	2	0.0091	0.0045	3.09	0.05	0.0001
AxCP	2	0.0012	0.0006	0.42	0.66	0.0000
AxSS	2	0.0048	0.0024	1.63	0.20	0.0000
BxC	2	0.0208	0.0104	7.09	0.00	0.0002
BxCP	2	0.0001	0.0000	0.03	0.97	0.0000
BxSS	2	0.0137	0.0068	4.67	0.01	0.0001
CxCP	1	0.0115	0.0115	7.83	0.01	0.0001
CxSS	1	0.0001	0.0001	0.04	0.84	0.0000
CPxSS	1	0.0012	0.0012	0.80	0.37	0.0000
AxBxC	4	0.0159	0.0040	2.71	0.03	0.0001
AxBxCP	4	0.0091	0.0023	1.55	0.19	0.0000
AxBxSS	4	0.0089	0.0022	1.51	0.20	0.0000
AxCxCP	2	0.0061	0.0031	2.09	0.12	0.0000
AxCxSS	2	0.0005	0.0002	0.17	0.85	0.0000
AxCPxSS	2	0.0002	0.0001	0.08	0.93	0.0000
BxCxCP	2	0.0188	0.0094	6.41	0.00	0.0002
BxCxSS	2	0.0012	0.0006	0.42	0.66	0.0000
BxCPxSS	2	0.0013	0.0006	0.43	0.65	0.0000
CxCPxSS	1	0.0007	0.0007	0.45	0.50	0.0000

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size.

Table III-4.

Average Relative Bias of the ab Estimate when c' = 0

a	*b*	Sample Size	Censoring Proportion	Relative Bias
0.21	0.14	150	0	0.0351
0.21	0.39	150	0	0.0074
0.52	0.14	150	0	0.0042
0.52	0.39	150	0	0.0298
0.21	0.14	300	0	-0.0356
0.21	0.39	300	0	-0.0003
0.52	0.14	300	0	-0.0069
0.52	0.39	300	0	-0.0102
0.21	0.14	150	.3	0.0075
0.21	0.39	150	.3	0.0135
0.52	0.14	150	.3	- 0.0056
0.52	0.39	150	.3	0.0178
0.21	0.14	300	.3	-0.0016
0.21	0.39	300	.3	-0.0059
0.52	0.14	300	.3	0.0011
0.52	0.39	300	.3	0.0351

Table III-5.

Average Relative Bias of the ab Estimate when c' = 0.39

a	*b*	Sample Size	Censoring Proportion	Relative Bias
0.21	0.14	150	0	-0.0189
0.21	0.39	150	0	0.0232
0.52	0.14	150	0	0.0197
0.52	0.39	150	0	0.0192
0.21	0.14	300	0	-0.0019
0.21	0.39	300	0	0.0165
0.52	0.14	300	0	-0.0133
0.52	0.39	300	0	0.0367
0.21	0.14	150	.3	0.0225
0.21	0.39	150	.3	0.0101
0.52	0.14	150	.3	0.0093
0.52	0.39	150	.3	0.0191
0.21	0.14	300	.3	0.0146
0.21	0.39	300	.3	0.0111
0.52	0.14	300	.3	0.0094
0.52	0.39	300	.3	0.0196

Table III-6.

Analysis of Variance Results for the Average Relative Bias of the ab Estimate

Source	*DF*	Type III *SS*	Mean Square	*F* Value	*Pr* > *F*	Partial $\widehat{\omega^2}$
A	1	0.0015	0.0015	0.00	0.95	0.0000
B	1	0.1122	0.1122	0.33	0.56	0.0000
C	1	0.6679	0.6679	1.99	0.16	0.0000
CP	1	0.1331	0.1331	0.40	0.53	0.0000
SS	1	0.0090	0.0090	0.03	0.87	0.0000
AxB	1	0.0395	0.0395	0.12	0.73	0.0000
AxC	1	0.1908	0.1908	0.57	0.45	0.0000
AxCP	1	0.0233	0.0233	0.07	0.79	0.0000
AxSS	1	0.0272	0.0272	0.08	0.78	0.0000
BxC	1	0.8837	0.8837	2.63	0.10	0.0001
BxCP	1	0.1919	0.1919	0.57	0.45	0.0000
BxSS	1	0.1968	0.1968	0.59	0.44	0.0000
CxCP	1	0.0561	0.0561	0.17	0.68	0.0000
CxSS	1	0.0844	0.0844	0.25	0.62	0.0000
CPxSS	1	0.1326	0.1326	0.39	0.53	0.0000
AxBxC	1	0.0064	0.0064	0.02	0.89	0.0000
AxBxCP	1	1.5902	1.5902	4.73	0.03	0.0001
AxBxSS	1	0.0080	0.0080	0.02	0.88	0.0000
AxCxCP	1	0.0026	0.0026	0.01	0.93	0.0000
AxCxSS	1	0.1093	0.1093	0.33	0.57	0.0000
AxCPxSS	1	0.0134	0.0134	0.04	0.84	0.0000
BxCxCP	1	0.7067	0.7067	2.10	0.15	0.0000
BxCxSS	1	0.5482	0.5482	1.63	0.20	0.0000
BxCPxSS	1	0.1757	0.1757	0.52	0.47	0.0000
CxCPxSS	1	0.1410	0.1410	0.42	0.52	0.0000

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size.

Table III-7.

Sign of the ab Raw Bias when c' = 0

a	*b*	Censoring Proportion	Sample Size	Negative sign ($\widehat{ab} < ab$)	Exact recover ($\widehat{ab} = ab$)
0	0	0	150	0.484	0.001
0	0.14	0	150	0.500	0.001
0	0.39	0	150	0.473	0.002
0.21	0	0	150	0.498	0.002
0.21	0.14	0	150	0.503	0.000
0.21	0.39	0	150	0.484	0.000
0.52	0	0	150	0.510	0.000
0.52	0.14	0	150	0.504	0.000
0.52	0.39	0	150	0.505	0.000
0	0	.3	150	0.497	0.000
0	0.14	.3	150	0.526	0.002
0	0.39	.3	150	0.506	0.001
0.21	0	.3	150	0.477	0.001
0.21	0.14	.3	150	0.492	0.000
0.21	0.39	.3	150	0.520	0.000
0.52	0	.3	150	0.497	0.001
0.52	0.14	.3	150	0.508	0.000
0.52	0.39	.3	150	0.504	0.000
0	0	0	300	0.499	0.002
0	0.14	0	300	0.517	0.003
0	0.39	0	300	0.481	0.003
0.21	0	0	300	0.511	0.000
0.21	0.14	0	300	0.494	0.000
0.21	0.39	0	300	0.508	0.000
0.52	0	0	300	0.502	0.000
0.52	0.14	0	300	0.473	0.000
0.52	0.39	0	300	0.520	0.000
0	0	.3	300	0.517	0.002
0	0.14	.3	300	0.519	0.003
0	0.39	.3	300	0.514	0.001
0.21	0	.3	300	0.522	0.000
0.21	0.14	.3	300	0.512	0.000
0.21	0.39	.3	300	0.509	0.000
0.52	0	.3	300	0.505	0.000
0.52	0.14	.3	300	0.516	0.000
0.52	0.39	.3	300	0.500	0.000

Table III-8.

Sign of the ab Raw Bias when c' = 0.39

a	*b*	Censoring Proportion	Sample Size	Negative sign ($\widehat{ab} < ab$)	Exact recover ($\widehat{ab} = ab$)
0	0	0	150	0.518	0.000
0	0.14	0	150	0.521	0.001
0	0.39	0	150	0.483	0.004
0.21	0	0	150	0.512	0.000
0.21	0.14	0	150	0.503	0.000
0.21	0.39	0	150	0.487	0.000
0.52	0	0	150	0.514	0.000
0.52	0.14	0	150	0.485	0.000
0.52	0.39	0	150	0.484	0.000
0	0	.3	150	0.506	0.001
0	0.14	.3	150	0.501	0.000
0	0.39	.3	150	0.476	0.002
0.21	0	.3	150	0.496	0.001
0.21	0.14	.3	150	0.525	0.000
0.21	0.39	.3	150	0.496	0.000
0.52	0	.3	150	0.536	0.000
0.52	0.14	.3	150	0.522	0.000
0.52	0.39	.3	150	0.463	0.000
0	0	0	300	0.543	0.005
0	0.14	0	300	0.522	0.000
0	0.39	0	300	0.523	0.006
0.21	0	0	300	0.480	0.001
0.21	0.14	0	300	0.479	0.000
0.21	0.39	0	300	0.501	0.000
0.52	0	0	300	0.508	0.000
0.52	0.14	0	300	0.481	0.000
0.52	0.39	0	300	0.469	0.000
0	0	.3	300	0.513	0.003
0	0.14	.3	300	0.484	0.004
0	0.39	.3	300	0.491	0.004
0.21	0	.3	300	0.488	0.002
0.21	0.14	.3	300	0.481	0.000
0.21	0.39	.3	300	0.486	0.000
0.52	0	.3	300	0.467	0.000
0.52	0.14	.3	300	0.497	0.000
0.52	0.39	.3	300	0.472	0.000

Table III-9.

Logistic Regression Results for the Sign of the ab Raw Bias

Effect	DF	Wald χ^2	Pr > χ^2
A	2	3.84	0.15
B	2	5.75	0.06
C	1	1.99	0.16
CP	1	0.20	0.65
SS	1	0.00	0.99
AxB	4	3.88	0.42
AxC	2	3.56	0.17
AxCP	2	0.35	0.84
AxSS	2	5.98	0.05
BxC	2	5.31	0.07
BxCP	2	2.17	0.34
BxSS	2	3.65	0.16
CxCP	1	4.57	0.03
CxSS	1	3.96	0.05
CPxSS	1	0.53	0.47
AxBxC	4	5.41	0.25
AxBxCP	4	5.29	0.26
AxBxSS	4	1.82	0.77
AxCxCP	2	7.11	0.03
AxCxSS	2	3.26	0.20
AxCPxSS	2	0.64	0.73
BxCxCP	2	0.65	0.72
BxCxSS	2	2.99	0.22
BxCPxSS	2	0.56	0.76
CxCPxSS	1	1.86	0.17

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size.

Table III-10.

Mean Squared Error of the ab Estimate when c' = 0

a	*b*	Censoring Proportion	Sample Size	Mean Squared Error
0	0	0	150	0.0000
0	0.14	0	150	0.0005
0	0.39	0	150	0.0040
0.21	0	0	150	0.0000
0.21	0.14	0	150	0.0006
0.21	0.39	0	150	0.0044
0.52	0	0	150	0.0000
0.52	0.14	0	150	0.0006
0.52	0.39	0	150	0.0048
0	0	.3	150	0.0000
0	0.14	.3	150	0.0008
0	0.39	.3	150	0.0063
0.21	0	.3	150	0.0000
0.21	0.14	.3	150	0.0008
0.21	0.39	.3	150	0.0059
0.52	0	.3	150	0.0001
0.52	0.14	.3	150	0.0008
0.52	0.39	.3	150	0.0057
0	0	0	300	0.0000
0	0.14	0	300	0.0003
0	0.39	0	300	0.0022
0.21	0	0	300	0.0000
0.21	0.14	0	300	0.0003
0.21	0.39	0	300	0.0020
0.52	0	0	300	0.0000
0.52	0.14	0	300	0.0003
0.52	0.39	0	300	0.0023
0	0	.3	300	0.0000
0	0.14	.3	300	0.0004
0	0.39	.3	300	0.0031
0.21	0	.3	300	0.0000
0.21	0.14	.3	300	0.0004
0.21	0.39	.3	300	0.0031
0.52	0	.3	300	0.0000
0.52	0.14	.3	300	0.0004
0.52	0.39	.3	300	0.0028

Table III-11.

Mean Squared Error of the *ab* Estimate when *c'* = 0.39

a	*b*	Censoring Proportion	Sample Size	Mean Squared Error
0	0	0	150	0.0000
0	0.14	0	150	0.0006
0	0.39	0	150	0.0041
0.21	0	0	150	0.0000
0.21	0.14	0	150	0.0006
0.21	0.39	0	150	0.0043
0.52	0	0	150	0.0000
0.52	0.14	0	150	0.0006
0.52	0.39	0	150	0.0051
0	0	.3	150	0.0000
0	0.14	.3	150	0.0009
0	0.39	.3	150	0.0060
0.21	0	.3	150	0.0000
0.21	0.14	.3	150	0.0008
0.21	0.39	.3	150	0.0052
0.52	0	.3	150	0.0001
0.52	0.14	.3	150	0.0008
0.52	0.39	.3	150	0.0060
0	0	0	300	0.0000
0	0.14	0	300	0.0003
0	0.39	0	300	0.0022
0.21	0	0	300	0.0000
0.21	0.14	0	300	0.0003
0.21	0.39	0	300	0.0022
0.52	0	0	300	0.0000
0.52	0.14	0	300	0.0003
0.52	0.39	0	300	0.0025
0	0	.3	300	0.0000
0	0.14	.3	300	0.0004
0	0.39	.3	300	0.0029
0.21	0	.3	300	0.0000
0.21	0.14	.3	300	0.0004
0.21	0.39	.3	300	0.0029
0.52	0	.3	300	0.0000
0.52	0.14	.3	300	0.0004
0.52	0.39	.3	300	0.0029

Table III-12.

Analysis of Variance Results for the Mean Squared Error of the ab Estimate

Source	*DF*	Type III *SS*	Mean Square	*F* Value	*Pr* > *F*	Partial $\widehat{\omega^2}$
A	2	0.0001	0.0001	5.57	0.00	0.00
B	2	0.2102	0.1051	9024.83	<.0001	0.20
C	1	0.0000	0.0000	0.02	0.89	0.00
CP	1	0.0031	0.0031	266.00	<.0001	0.00
SS	1	0.0169	0.0169	1447.27	<.0001	0.02
AxB	4	0.0001	0.0000	3.01	0.02	0.00
AxC	2	0.0001	0.0000	3.09	0.05	0.00
AxCP	2	0.0002	0.0001	8.41	0.00	0.00
AxSS	2	0.0001	0.0000	2.54	0.08	0.00
BxC	2	0.0000	0.0000	0.06	0.94	0.00
BxCP	2	0.0038	0.0019	161.39	<.0001	0.00
BxSS	2	0.0223	0.0111	956.22	<.0001	0.03
CxCP	1	0.0000	0.0000	3.20	0.07	0.00
CxSS	1	0.0000	0.0000	0.15	0.70	0.00
CPxSS	1	0.0003	0.0003	28.12	<.0001	0.00
AxBxC	4	0.0001	0.0000	3.00	0.02	0.00
AxBxCP	4	0.0004	0.0001	8.03	<.0001	0.00
AxBxSS	4	0.0001	0.0000	2.06	0.08	0.00
AxCxCP	2	0.0000	0.0000	0.79	0.45	0.00
AxCxSS	2	0.0000	0.0000	1.06	0.34	0.00
AxCPxSS	2	0.0001	0.0001	4.77	0.01	0.00
BxCxCP	2	0.0001	0.0000	3.81	0.02	0.00
BxCxSS	2	0.0000	0.0000	0.57	0.57	0.00
BxCPxSS	2	0.0003	0.0002	14.87	<.0001	0.00
CxCPxSS	1	0.0000	0.0000	0.39	0.53	0.00

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size.

Table III-13.

Type I Error Rate when c' = 0 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0	0	150	0.000	0.000	0.001	0.005	0.001	0.000
0	0.14	150	0.045	0.045	0.050	0.049	0.045	0.032
0	0.39	150	0.033	0.034	0.034	0.034	0.034	0.020
0.21	0	150	0.001	0.008	0.011	0.029	0.015	0.008
0.52	0	150	0.016	0.047	0.055	0.089	0.048	0.051
0	0	300	0.001	0.004	0.003	0.009	0.006	0.002
0	0.14	300	0.052	0.054	0.055	0.055	0.054	0.043
0	0.39	300	0.058	0.058	0.058	0.057	0.058	0.048
0.21	0	300	0.006	0.018	0.021	0.035	0.021	0.020
0.52	0	300	0.028	0.053	0.057	0.082	0.053	0.057

*NIE=natural indirect effect.

Table III-14.

Type I Error Rate when c' = 0 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0	0	150	0.000	0.004	0.005	0.008	0.004	0.000
0	0.14	150	0.045	0.053	0.056	0.059	0.053	0.013
0	0.39	150	0.053	0.053	0.056	0.054	0.053	0.015
0.21	0	150	0.003	0.004	0.008	0.021	0.006	0.006
0.52	0	150	0.015	0.042	0.046	0.084	0.045	0.034
0	0	300	0.000	0.001	0.001	0.004	0.001	0.001
0	0.14	300	0.052	0.052	0.052	0.056	0.052	0.023
0	0.39	300	0.065	0.065	0.061	0.061	0.065	0.025
0.21	0	300	0.002	0.008	0.011	0.024	0.016	0.008
0.52	0	300	0.014	0.046	0.055	0.082	0.049	0.051

*NIE=natural indirect effect

Table III-15.

Type I Error Rate when c' = 0.39 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0	0	150	0.000	0.001	0.001	0.004	0.003	0.001
0	0.14	150	0.054	0.056	0.054	0.060	0.056	0.045
0	0.39	150	0.042	0.042	0.046	0.045	0.042	0.035
0.21	0	150	0.000	0.005	0.013	0.027	0.009	0.009
0.52	0	150	0.011	0.034	0.044	0.070	0.042	0.043
0	0	300	0.001	0.001	0.001	0.005	0.002	0.001
0	0.14	300	0.044	0.048	0.046	0.047	0.048	0.035
0	0.39	300	0.059	0.059	0.060	0.064	0.059	0.047
0.21	0	300	0.003	0.015	0.016	0.040	0.021	0.015
0.52	0	300	0.028	0.057	0.057	0.073	0.057	0.057

*NIE=natural indirect effect

Table III-16.

Type I Error Rate when c' = 0.39 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0	0	150	0.000	0.000	0.001	0.002	0.000	0.000
0	0.14	150	0.044	0.050	0.056	0.059	0.050	0.022
0	0.39	150	0.046	0.048	0.050	0.047	0.048	0.019
0.21	0	150	0.001	0.005	0.006	0.019	0.007	0.003
0.52	0	150	0.004	0.030	0.041	0.065	0.040	0.025
0	0	300	0.000	0.003	0.004	0.008	0.004	0.001
0	0.14	300	0.045	0.046	0.045	0.043	0.046	0.019
0	0.39	300	0.041	0.041	0.043	0.040	0.041	0.015
0.21	0	300	0.003	0.009	0.009	0.030	0.016	0.005
0.52	0	300	0.025	0.052	0.059	0.082	0.057	0.054

*NIE=natural indirect effect

Table III-17.

Logistic Regression Results for Type I Error Rate

Effect	*DF*	Wald χ^2	*Pr* > χ^2
A	2	546.0848	<.0001
B	2	0.5476	0.76
C	1	0.4554	0.50
CP	1	0.0022	0.96
SS	1	0.0066	0.94
IE	5	8.4081	0.14
AxC	2	2.2963	0.32
AxCP	2	5.6643	0.06
AxSS	2	4.0955	0.13
AxIE	10	46.8555	<.0001
BxC	2	2.6681	0.26
BxCP	2	1.5908	0.45
BxSS	2	14.5291	0.00
BxIE	10	29.6309	0.00
CxCP	1	2.3466	0.13
CxSS	1	3.3801	0.07
CxIE	5	1.7475	0.88
CPxSS	1	0.2705	0.60
CPxIE	5	4.6687	0.46
SSxIE	5	1.9774	0.85
AxCxCP	2	0.4145	0.81
AxCxSS	2	2.6942	0.26
AxCxIE	10	5.4494	0.86
AxCPxSS	2	2.8557	0.24
AxCPxIE	10	5.7113	0.84
AxSSxIE	10	2.8513	0.98
BxCxCP	2	7.5859	0.02
BxCxSS	2	5.6407	0.06
BxCxIE	10	2.2371	0.99
BxCPxSS	2	10.8727	0.00
BxCPxIE	10	2.4670	0.99
BxSSxIE	10	3.9207	0.95
CxCPxSS	1	1.8862	0.17
CxCPxIE	5	0.1564	1.00
CxSSxIE	5	4.0367	0.54
CPxSSxIE	5	1.5944	0.90

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size; IE=indirect effect test method.

Table III-18.

Statistical Power when c' = 0 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0.21	0.14	150	0.238	0.243	0.245	0.253	0.243	0.192
0.21	0.39	150	0.250	0.251	0.252	0.251	0.251	0.212
0.52	0.14	150	0.874	0.884	0.880	0.877	0.884	0.844
0.52	0.39	150	0.871	0.871	0.867	0.862	0.871	0.831
0.21	0.14	300	0.443	0.450	0.441	0.447	0.450	0.397
0.21	0.39	300	0.428	0.428	0.440	0.432	0.428	0.398
0.52	0.14	300	0.992	0.992	0.992	0.992	0.992	0.992
0.52	0.39	300	0.997	0.997	0.996	0.996	0.997	0.992

*NIE=natural indirect effect

Table III-19.

Statistical power when c' = 0 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0.21	0.14	150	0.181	0.202	0.207	0.217	0.202	0.091
0.21	0.39	150	0.165	0.165	0.178	0.173	0.165	0.086
0.52	0.14	150	0.742	0.761	0.758	0.775	0.761	0.583
0.52	0.39	150	0.752	0.754	0.762	0.758	0.754	0.587
0.21	0.14	300	0.326	0.333	0.333	0.345	0.333	0.215
0.21	0.39	300	0.336	0.336	0.337	0.340	0.336	0.231
0.52	0.14	300	0.968	0.971	0.969	0.972	0.971	0.939
0.52	0.39	300	0.971	0.971	0.976	0.976	0.971	0.936

*NIE=natural indirect effect

Table III-20.

Statistical Power when c' = 0.39 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0.21	0.14	150	0.228	0.237	0.249	0.240	0.237	0.185
0.21	0.39	150	0.251	0.251	0.246	0.246	0.251	0.199
0.52	0.14	150	0.878	0.884	0.881	0.888	0.884	0.850
0.52	0.39	150	0.865	0.865	0.861	0.859	0.865	0.822
0.21	0.14	300	0.452	0.457	0.457	0.448	0.457	0.414
0.21	0.39	300	0.430	0.430	0.439	0.431	0.430	0.407
0.52	0.14	300	0.994	0.995	0.994	0.995	0.995	0.991
0.52	0.39	300	0.995	0.995	0.995	0.995	0.995	0.991

*NIE=natural indirect effect

Table III-21.

Statistical Power when c' = 0.39 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel	Distribution of the product	Percentile Bootstrap	Bias Corrected Bootstrap	Joint Significance	Percentile Bootstrap of NIE*
0.21	0.14	150	0.171	0.195	0.209	0.217	0.195	0.088
0.21	0.39	150	0.171	0.172	0.166	0.168	0.172	0.078
0.52	0.14	150	0.725	0.750	0.759	0.771	0.750	0.581
0.52	0.39	150	0.793	0.795	0.793	0.787	0.795	0.624
0.21	0.14	300	0.320	0.329	0.341	0.345	0.329	0.226
0.21	0.39	300	0.339	0.341	0.333	0.339	0.341	0.224
0.52	0.14	300	0.966	0.969	0.970	0.968	0.969	0.942
0.52	0.39	300	0.976	0.976	0.974	0.974	0.976	0.937

*NIE=natural indirect effect

Table III-22.

Logistic Regression Results for Statistical Power

Effect	*DF*	Wald χ^2	*Pr* > χ^2
A	1	23236.06	<.0001
B	1	1.18	0.28
C	1	0.11	0.74
CP	1	1259.79	<.0001
SS	1	5396.18	<.0001
IE	5	450.38	<.0001
AxB	1	0.67	0.41
AxC	1	0.34	0.56
AxCP	1	254.24	<.0001
AxSS	1	1472.86	<.0001
AxIE	5	7.53	0.18
BxC	1	1.73	0.19
BxCP	1	19.56	<.0001
BxSS	1	0.37	0.54
BxIE	5	4.46	0.49
CxCP	1	1.10	0.29
CxSS	1	0.00	0.96
CxIE	5	0.06	1.00
CPxSS	1	90.24	<.0001
CPxIE	5	149.09	<.0001
SSxIE	5	2.11	0.83
AxBxC	1	1.02	0.31
AxBxCP	1	38.66	<.0001
AxBxSS	1	0.07	0.79
AxBxIE	5	3.67	0.60
AxCxCP	1	1.29	0.26
AxCxSS	1	0.73	0.39
AxCxIE	5	0.15	1.00
AxCPxSS	1	72.42	<.0001
AxCPxIE	5	1.00	0.96
AxSSxIE	5	10.58	0.06
BxCxCP	1	4.66	0.03
BxCxSS	1	0.93	0.33
BxCxIE	5	2.99	0.70
BxCPxSS	1	19.18	<.0001
BxCPxIE	5	1.98	0.85
BxSSxIE	5	3.82	0.58
CxCPxSS	1	0.44	0.51
CxCPxIE	5	0.13	1.00
CxSSxIE	5	0.53	0.99
CPxSSxIE	5	4.99	0.42

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size; IE=indirect effect test method.

Table 23.

Parameter ab Coverage Rate when c' = 0 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel Test	Distribution of the product	Percentile Bootstrapping	Bias Corrected Bootstrapping
0	0	150	1.000	1.000	0.999	0.995
0	0.14	150	0.955	0.955	0.950	0.951
0	0.39	150	0.967	0.966	0.966	0.966
0.21	0	150	0.999	0.992	0.989	0.971
0.21	0.14	150	0.950	0.947	0.947	0.947
0.21	0.39	150	0.957	0.954	0.956	0.956
0.52	0	150	0.984	0.953	0.945	0.911
0.52	0.14	150	0.956	0.955	0.950	0.948
0.52	0.39	150	0.939	0.939	0.942	0.947
0	0	300	0.999	0.996	0.997	0.991
0	0.14	300	0.948	0.946	0.945	0.945
0	0.39	300	0.942	0.942	0.942	0.943
0.21	0	300	0.994	0.982	0.979	0.965
0.21	0.14	300	0.948	0.947	0.942	0.945
0.21	0.39	300	0.954	0.953	0.956	0.954
0.52	0	300	0.972	0.947	0.943	0.918
0.52	0.14	300	0.949	0.948	0.947	0.947
0.52	0.39	300	0.949	0.948	0.948	0.947

Table 44.

Parameter ab Coverage Rate when c' = 0 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel Test	Distribution of the product	Percentile Bootstrapping	Bias Corrected Bootstrapping
0	0	150	1.000	0.996	0.995	0.992
0	0.14	150	0.955	0.947	0.944	0.941
0	0.39	150	0.947	0.947	0.944	0.946
0.21	0	150	0.997	0.996	0.992	0.979
0.21	0.14	150	0.949	0.946	0.943	0.935
0.21	0.39	150	0.949	0.949	0.951	0.945
0.52	0	150	0.985	0.958	0.954	0.916
0.52	0.14	150	0.960	0.956	0.959	0.955
0.52	0.39	150	0.962	0.961	0.957	0.958
0	0	300	1.000	0.999	0.999	0.996
0	0.14	300	0.948	0.948	0.948	0.944
0	0.39	300	0.935	0.935	0.939	0.939
0.21	0	300	0.998	0.992	0.989	0.976
0.21	0.14	300	0.948	0.946	0.940	0.942
0.21	0.39	300	0.942	0.941	0.940	0.935
0.52	0	300	0.986	0.954	0.945	0.918
0.52	0.14	300	0.952	0.953	0.954	0.952
0.52	0.39	300	0.957	0.957	0.951	0.958

Table 25.

Parameter ab Coverage Rate when c' = 0.39 and Censoring Proportion = 0

a	*b*	Sample Size	Sobel Test	Distribution of the product	Percentile Bootstrapping	Bias Corrected Bootstrapping
0	0	150	1.000	0.999	0.999	0.996
0	0.14	150	0.946	0.944	0.946	0.940
0	0.39	150	0.958	0.958	0.954	0.955
0.21	0	150	1.000	0.995	0.987	0.973
0.21	0.14	150	0.950	0.949	0.943	0.946
0.21	0.39	150	0.946	0.946	0.951	0.945
0.52	0	150	0.989	0.966	0.956	0.930
0.52	0.14	150	0.950	0.948	0.941	0.946
0.52	0.39	150	0.939	0.937	0.935	0.937
0	0	300	0.999	0.999	0.999	0.995
0	0.14	300	0.956	0.952	0.954	0.953
0	0.39	300	0.941	0.941	0.940	0.936
0.21	0	300	0.997	0.985	0.984	0.960
0.21	0.14	300	0.946	0.946	0.946	0.946
0.21	0.39	300	0.943	0.943	0.945	0.941
0.52	0	300	0.972	0.943	0.943	0.927
0.52	0.14	300	0.955	0.956	0.955	0.951
0.52	0.39	300	0.938	0.938	0.935	0.935

Table 26.

Parameter ab Coverage Rate when c' = 0.39 and Censoring Proportion = 0.3

a	*b*	Sample Size	Sobel Test	Distribution of the product	Percentile Bootstrapping	Bias Corrected Bootstrapping
0	0	150	1.000	1.000	0.999	0.998
0	0.14	150	0.956	0.950	0.944	0.941
0	0.39	150	0.954	0.952	0.950	0.953
0.21	0	150	0.999	0.995	0.994	0.981
0.21	0.14	150	0.957	0.951	0.950	0.943
0.21	0.39	150	0.964	0.964	0.965	0.965
0.52	0	150	0.996	0.970	0.959	0.935
0.52	0.14	150	0.949	0.952	0.956	0.951
0.52	0.39	150	0.951	0.949	0.950	0.944
0	0	300	1.000	0.997	0.996	0.992
0	0.14	300	0.955	0.954	0.955	0.957
0	0.39	300	0.959	0.959	0.957	0.960
0.21	0	300	0.997	0.991	0.991	0.970
0.21	0.14	300	0.946	0.946	0.942	0.942
0.21	0.39	300	0.958	0.957	0.953	0.951
0.52	0	300	0.975	0.948	0.941	0.918
0.52	0.14	300	0.958	0.953	0.953	0.952
0.52	0.39	300	0.954	0.954	0.950	0.951

Table III-27.

Logistic Regression Results for Parameter *ab* Coverage Rate

Effect	*DF*	Wald χ^2	*Pr* > χ^2
A	2	472.50	<.0001
B	2	635.41	<.0001
C	1	4.91	0.03
CP	1	8.55	0.00
SS	1	36.58	<.0001
IE	3	160.41	<.0001
AxB	4	622.19	<.0001
AxC	2	5.87	0.05
AxCP	2	6.12	0.05
AxSS	2	11.33	0.00
AxIE	6	23.85	0.00
BxC	2	4.98	0.08
BxCP	2	4.15	0.13
BxSS	2	30.98	<.0001
BxIE	6	155.23	<.0001
CxCP	1	9.42	0.00
CxSS	1	0.49	0.48
CxIE	3	0.30	0.96
CPxSS	1	1.28	0.26
CPxIE	3	3.57	0.31
SSxIE	3	15.55	0.00
AxBxC	4	4.45	0.35
AxBxCP	4	24.12	<.0001
AxBxSS	4	11.85	0.02
AxBxIE	12	29.81	0.00
AxCxCP	2	11.59	0.00
AxCxSS	2	6.69	0.04
AxCxIE	6	0.27	1.00
AxCPxSS	2	14.20	0.00
AxCPxIE	6	1.19	0.98
AxSSxIE	6	1.27	0.97
BxCxCP	2	13.22	0.00
BxCxSS	2	13.34	0.00
BxCxIE	6	0.62	1.00
BxCPxSS	2	1.38	0.50
BxCPxIE	6	2.36	0.88
BxSSxIE	6	15.68	0.02
CxCPxSS	1	0.73	0.39
CxCPxIE	3	0.11	0.99
CxSSxIE	3	0.27	0.97
CPxSSxIE	3	0.40	0.94

A=size of *a*; B=size of *b*; C=size of *c'*; CP=censoring proportion; SS=sample size; IE=indirect effect test method.

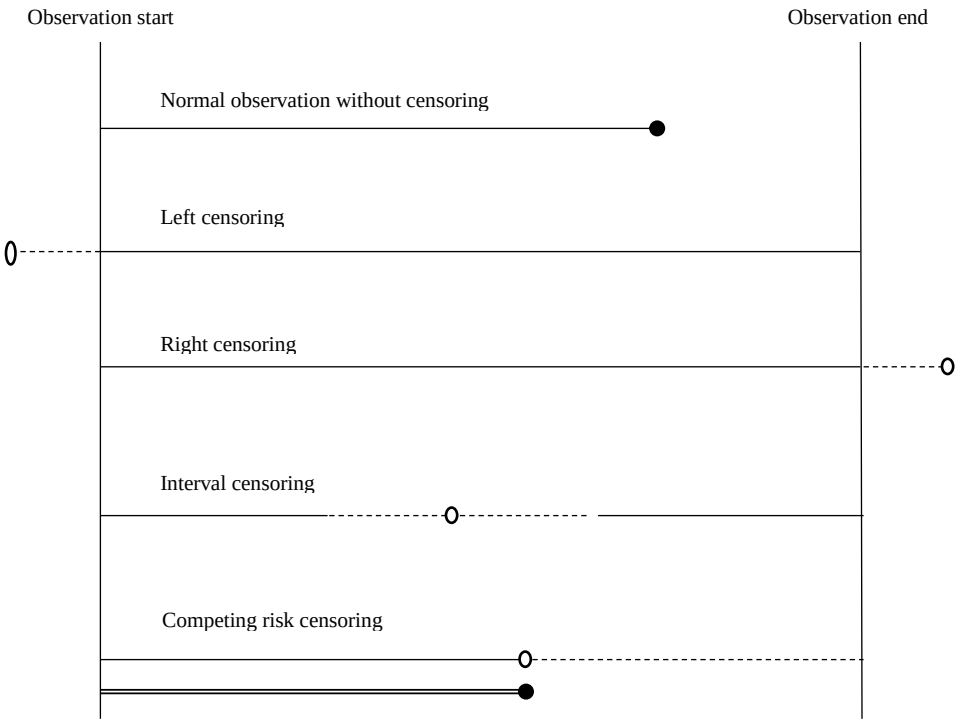


Figure I-1. Four Different Types of Censoring. The single solid line represents time within the observation period for an event of interest, the dashed line represents the time outside the observation for an event of interest, and the double solid line represents the time of a competing event. Filled dots represent an event that has been observed and the hollow dots represent unobserved (censored) events.

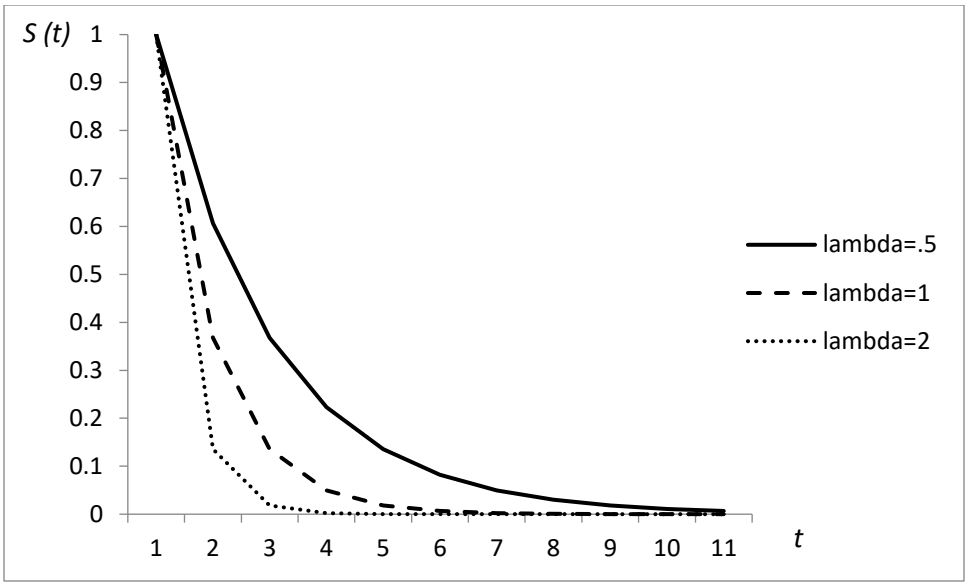


Figure I-2. An Example of the Exponential Survival Function when λ= .5, 1.0 and 2.

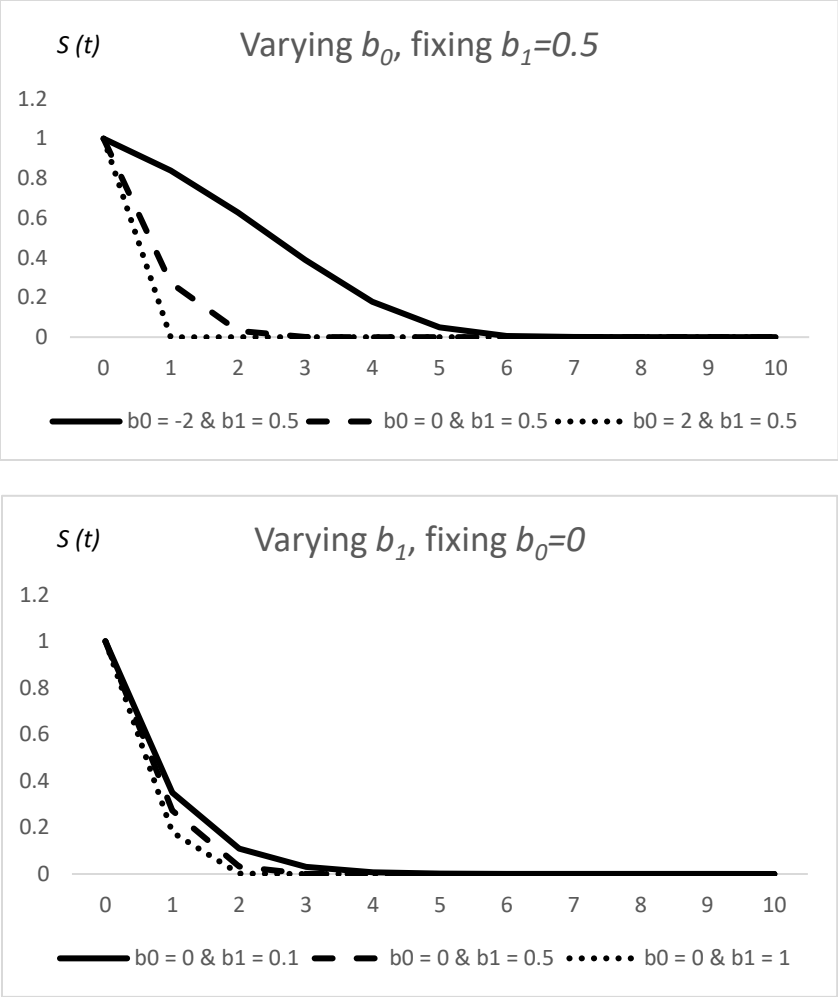
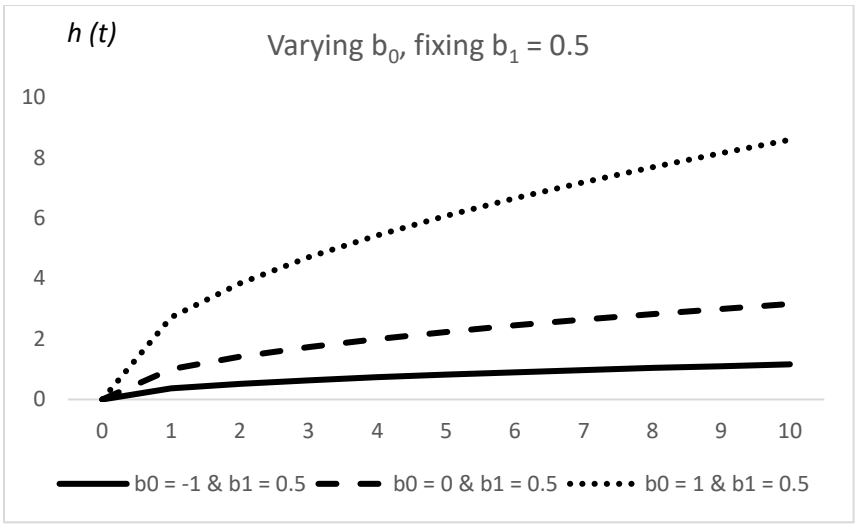


Figure I-3. An Example of the Gompertz Survival Function Varying Parameters Values One at a Time



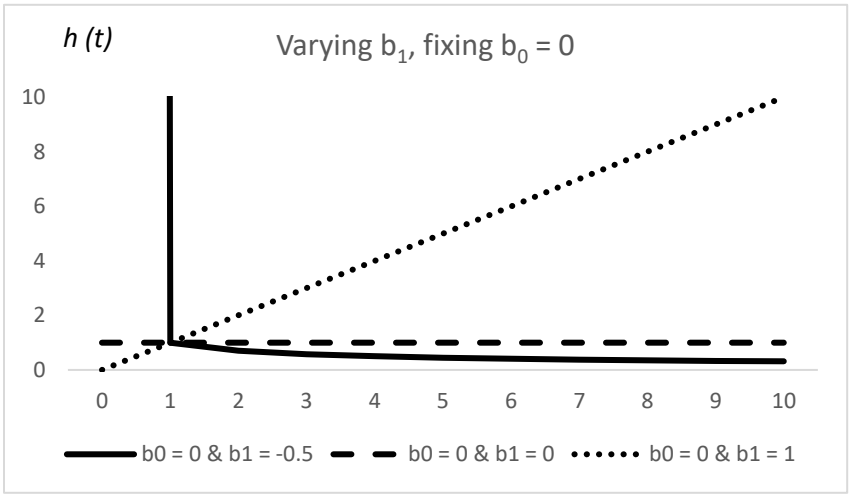


Figure I-4. An Example of the Weibull Hazard Function Varying Parameter Values One at a Time

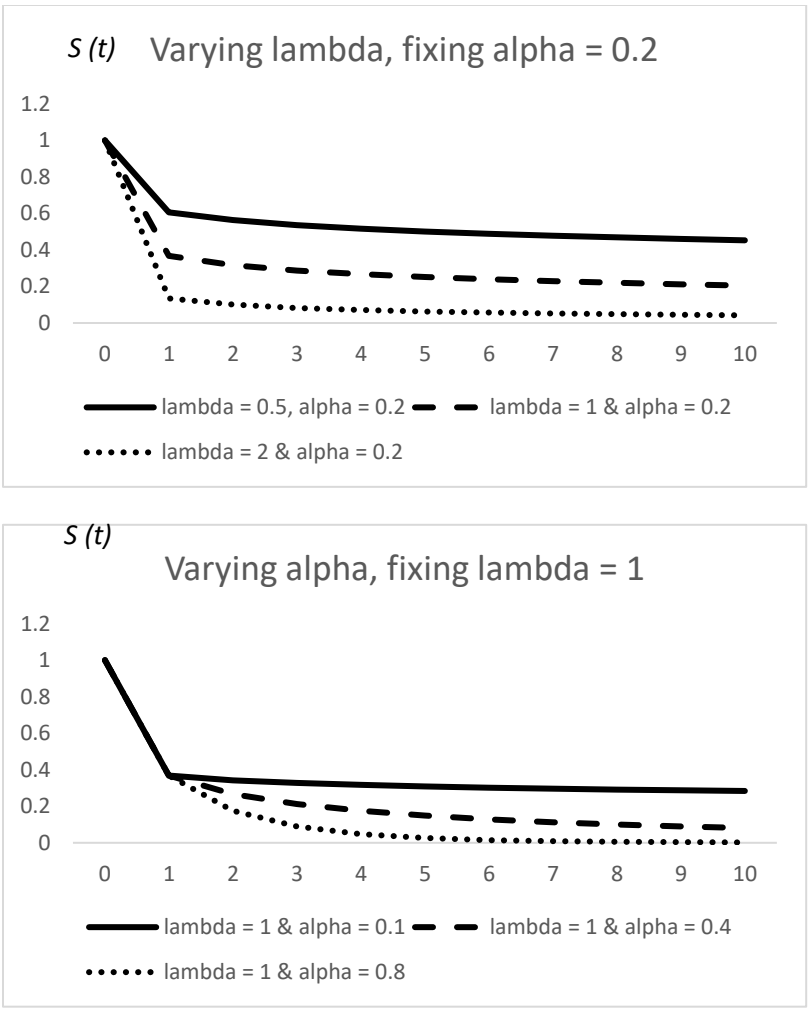


Figure I-5. An Example of the Weibull Survival Function Varying Parameter Values One at a Time

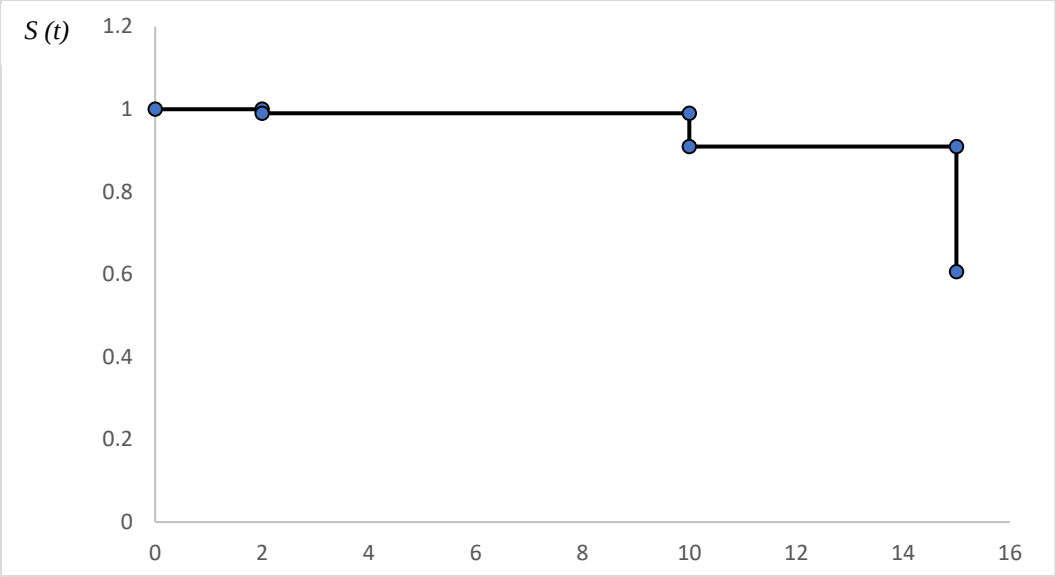


Figure I-6. An Example of the Kaplan-Meier Estimator. The numbers used to plot this figure are from Table I-1.

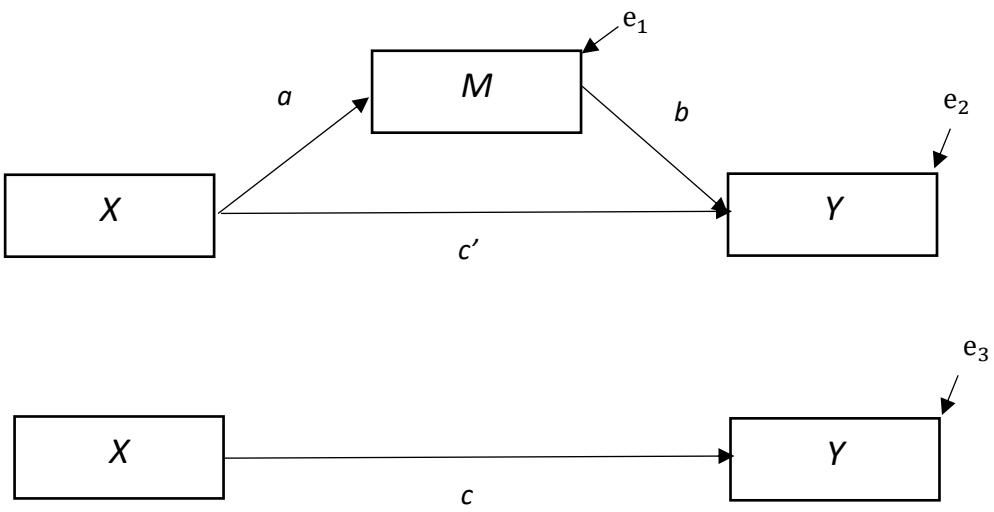


Figure I-7. The Single Mediator Model.

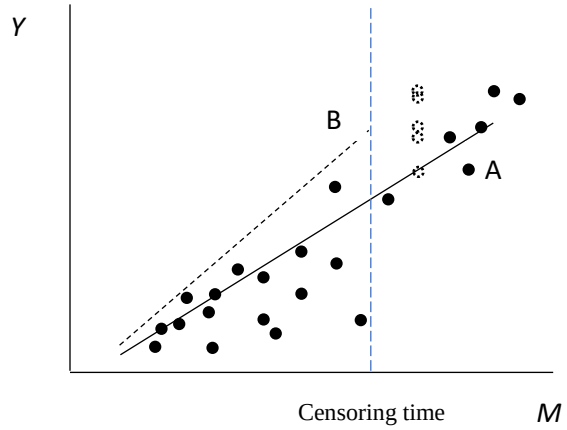


Figure I-8. A Hypothetical Example of the Regression Bias when Censored Values are Used for *M*. The solid dots represent the data with true values. The hollow dots are *Y* values at censored *M* values. Line-A represents the regression line when the true values of *M* are used in the *Y*-regression and line-B represents the regression line when *M* is censored.

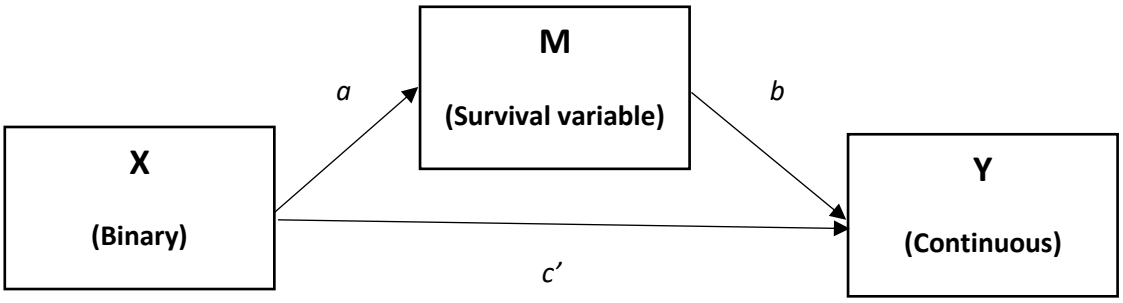


Figure II-1. Single Survival-Mediator Model Used in the Simulation Study

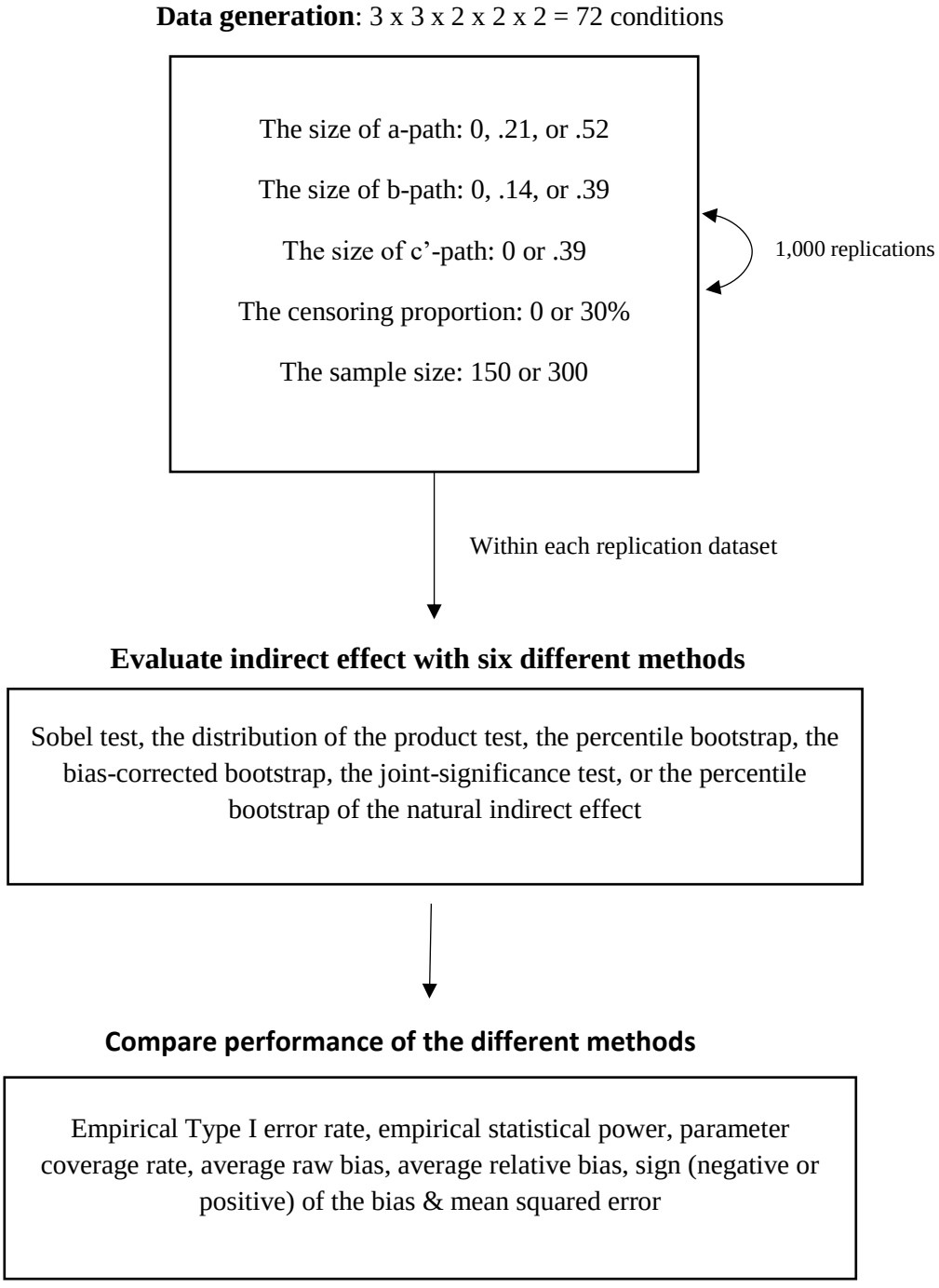


Figure II-2. Flow Chart of the Simulation Study Procedure

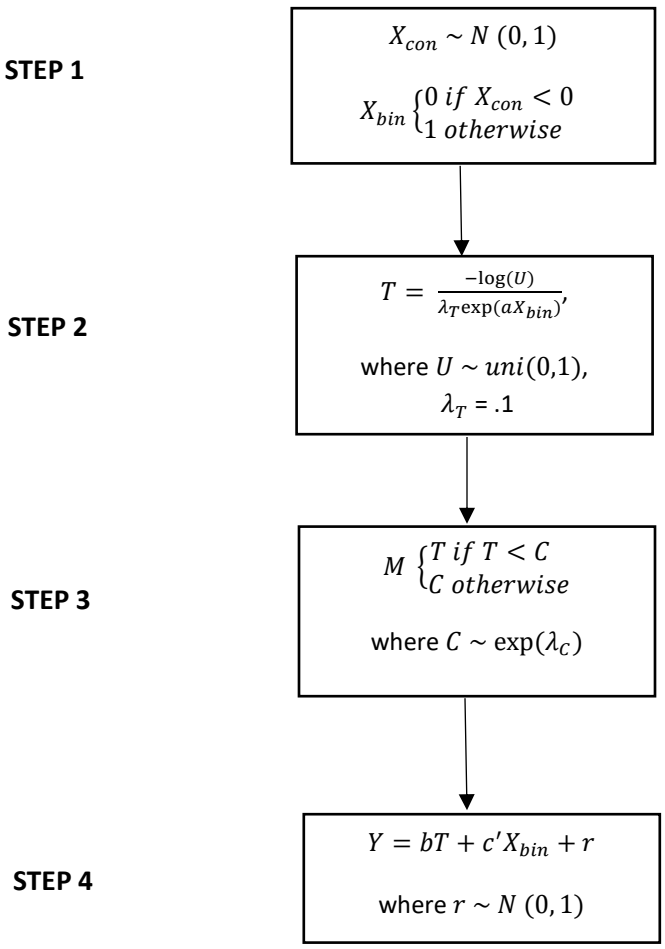


Figure II-3. Flow Chart of the Data Generation Process

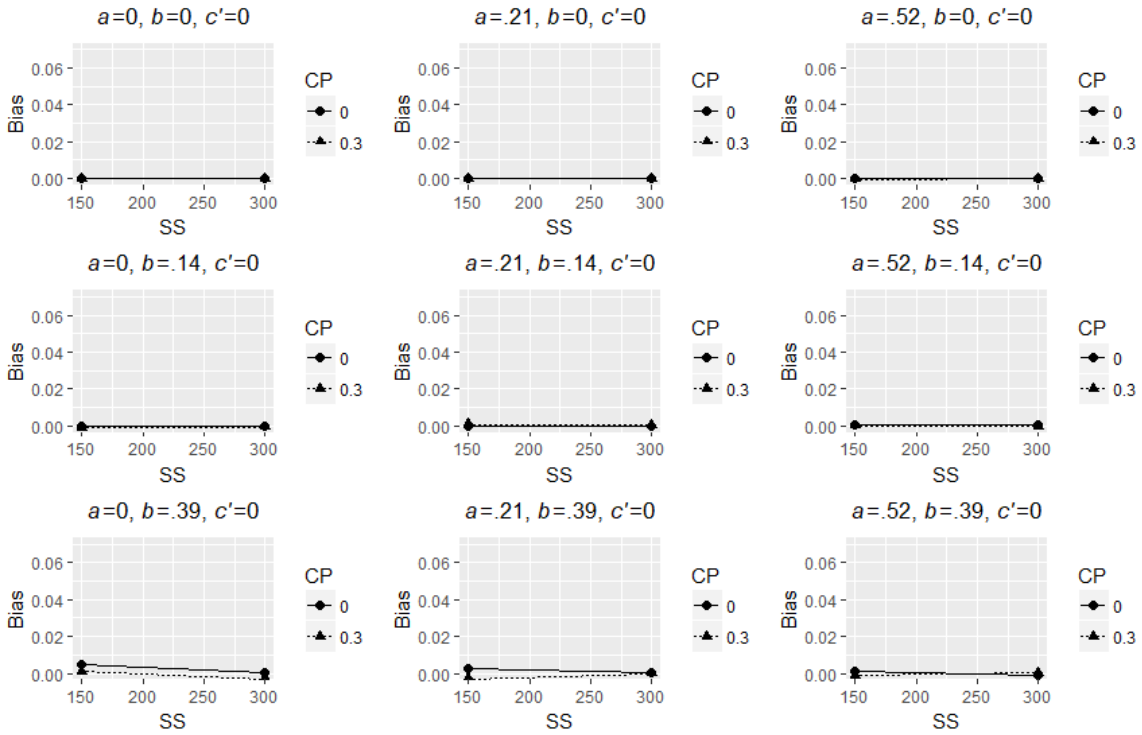


Figure III-1. Average Raw Bias of the *ab* Estimate when *c'* = 0. "Bias" on the y-axis denotes the average raw bias, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

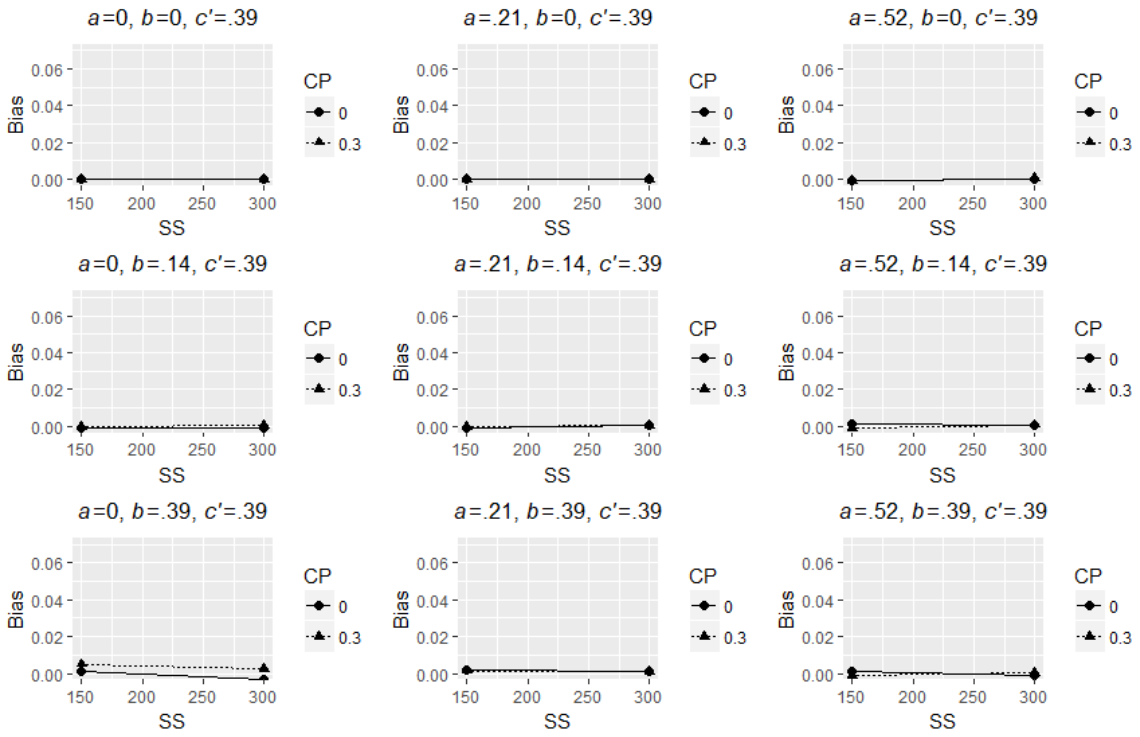


Figure III-2. Average Raw Bias of the *ab* Estimate when *c'* = 0.39. "Bias" on the y-axis denotes the average raw bias, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

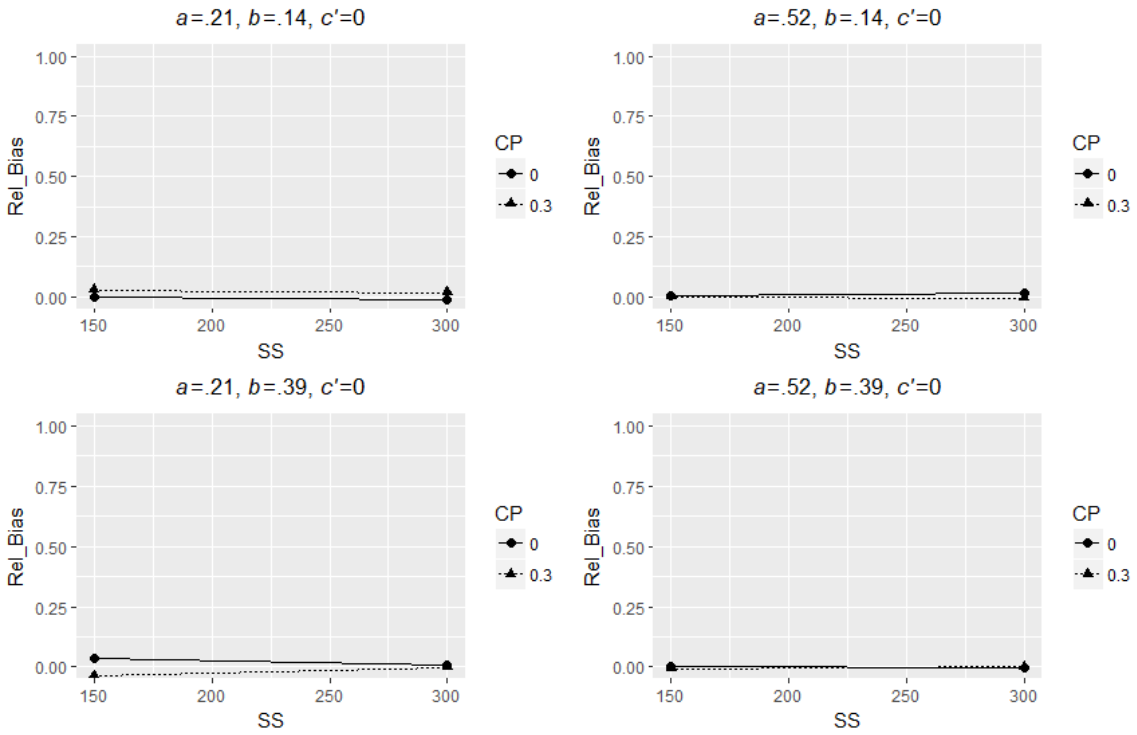


Figure III-3. Average Relative Bias of the *ab* Estimate when *c'*=0. "Rel_Bias" on the y-axis denotes the average relative bias, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

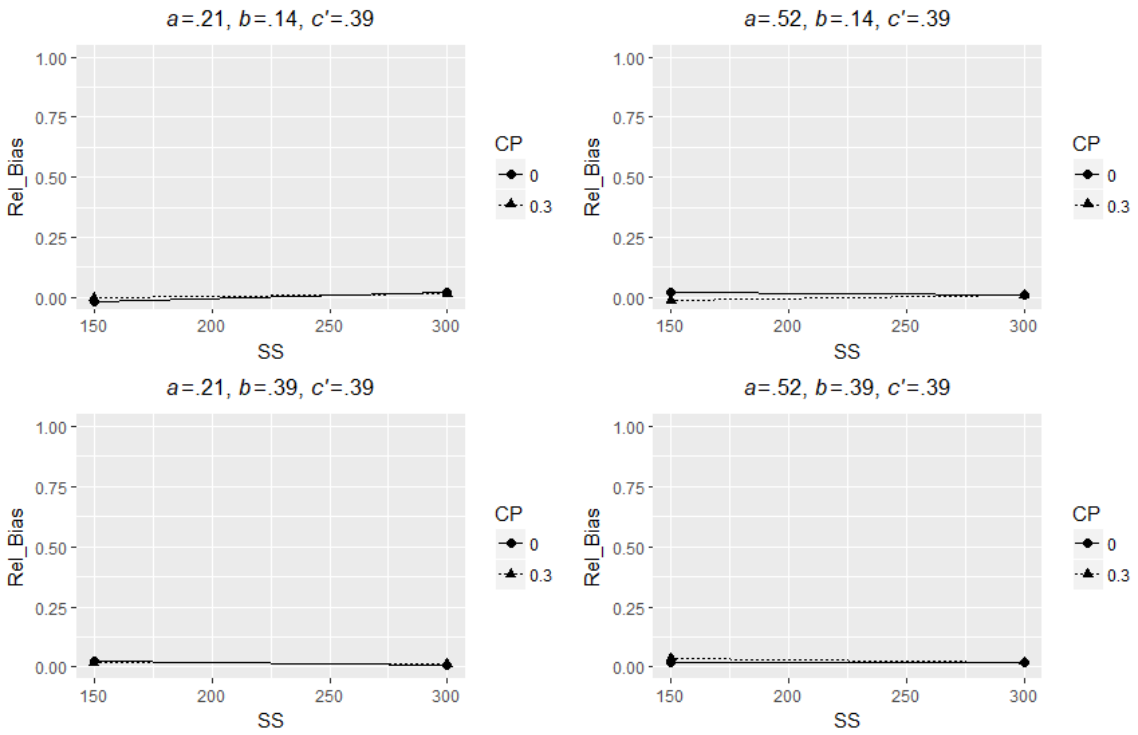


Figure III-4. Average Relative Bias of the *ab* Estimate when *c'* = 0.39. "Rel_Bias" on the y-axis denotes the average relative bias, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

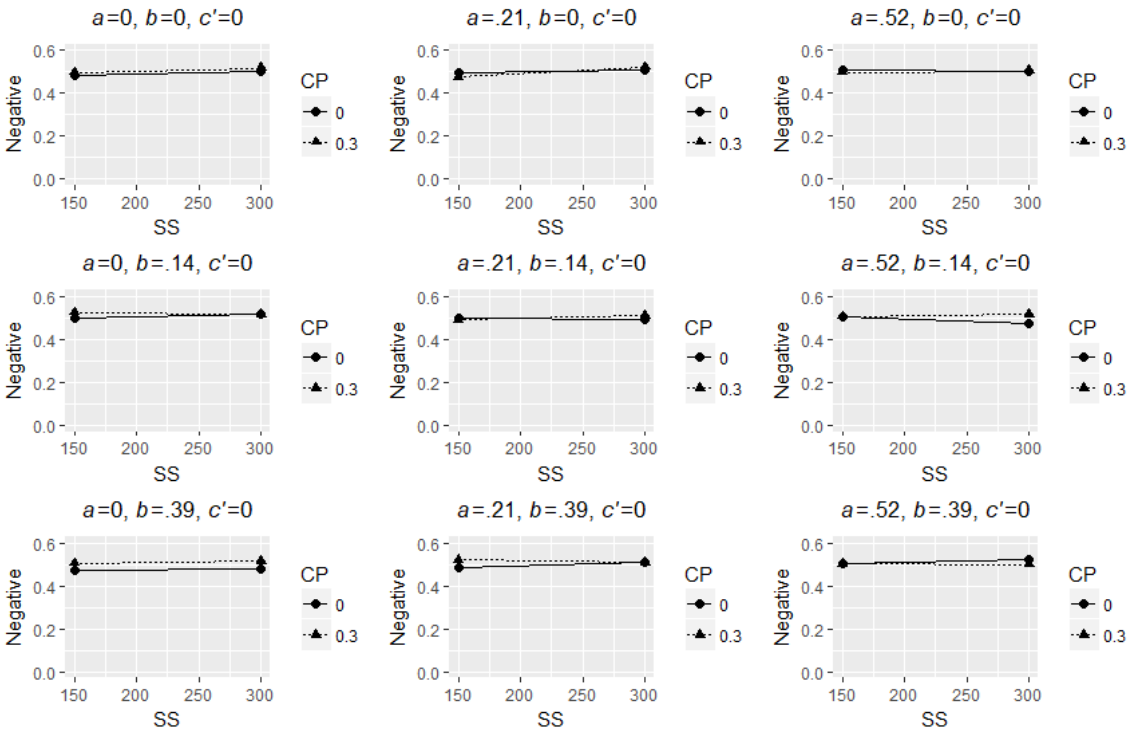


Figure III-5. Sign of the *ab* Raw Bias when *c'*=0. "Negative" on the y-axis denotes the proportion of negative raw biases (estimate of *ab* < true *ab*), "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

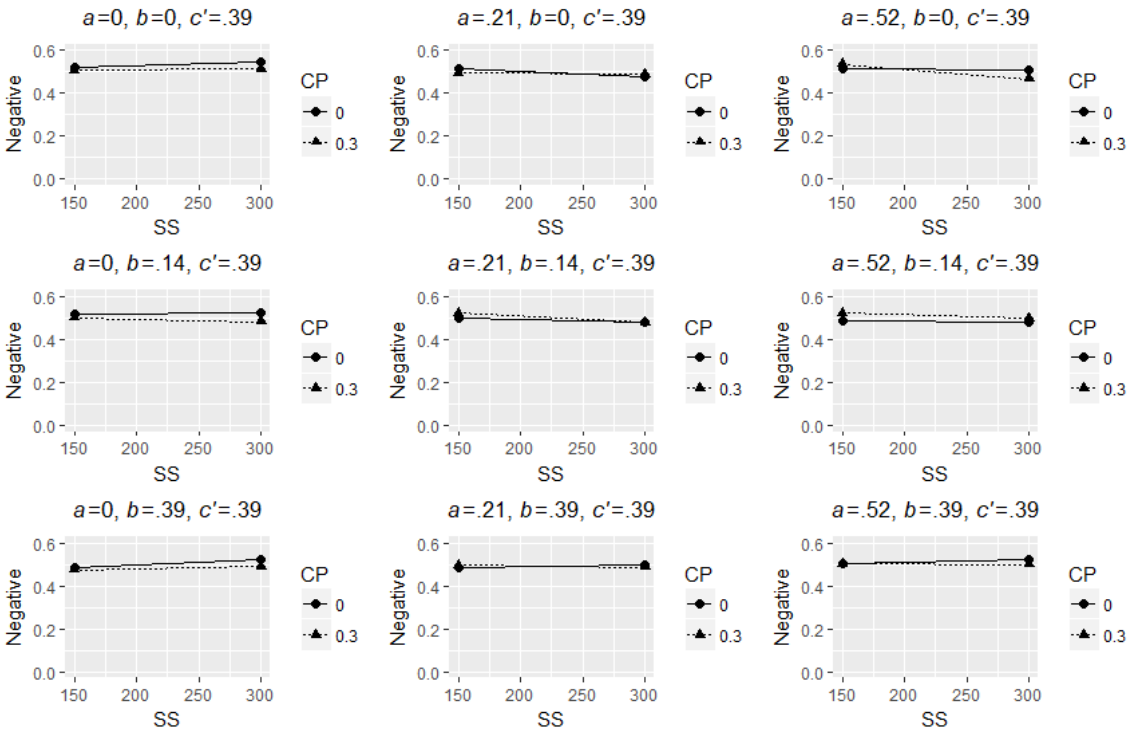


Figure III-6. Sign of the *ab* Raw Bias when *c'* = 0.39. "Negative" on the y-axis denotes the proportion of negative raw biases (estimate of *ab* < true *ab*), "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

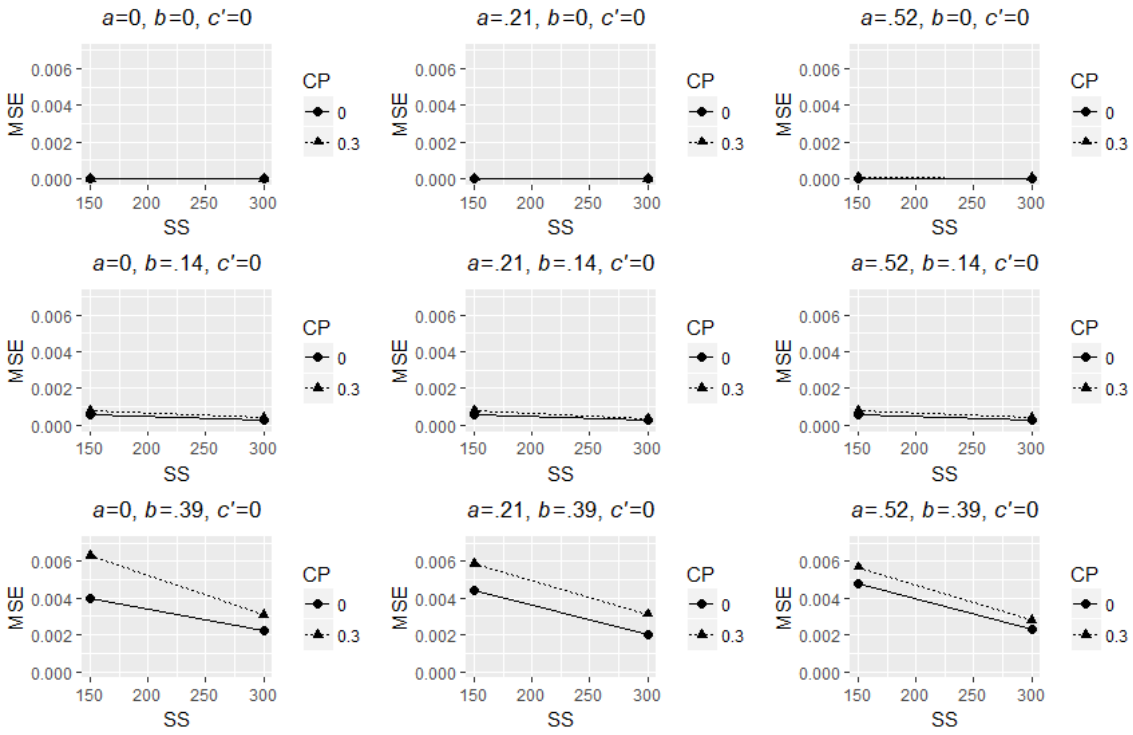


Figure III-7. The Mean Squared Error of the *ab* Estimate when *c'*=0. "MSE" on the y-axis denotes the mean squared error, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

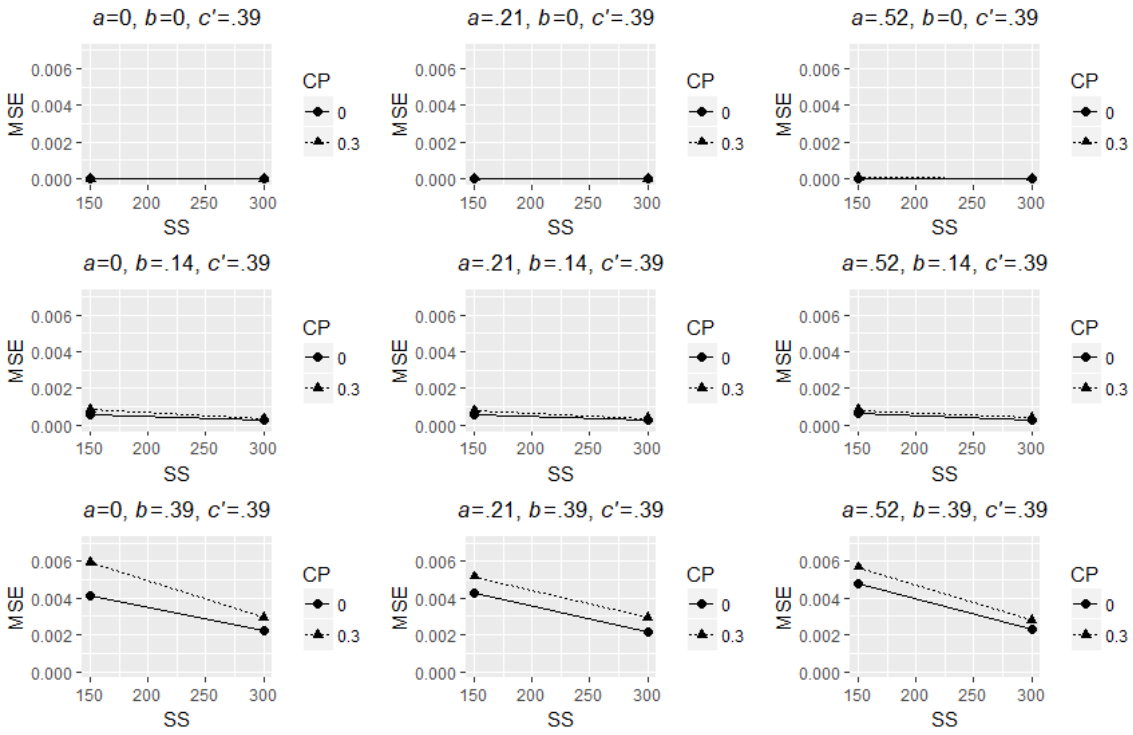


Figure III-8. The Mean Squared Error of the *ab* Estimate when *c'* = 0.39. "MSE" on the y-axis denotes the mean squared error, "SS" on the x-axis denotes sample size and "CP" on the legend denotes censoring proportion.

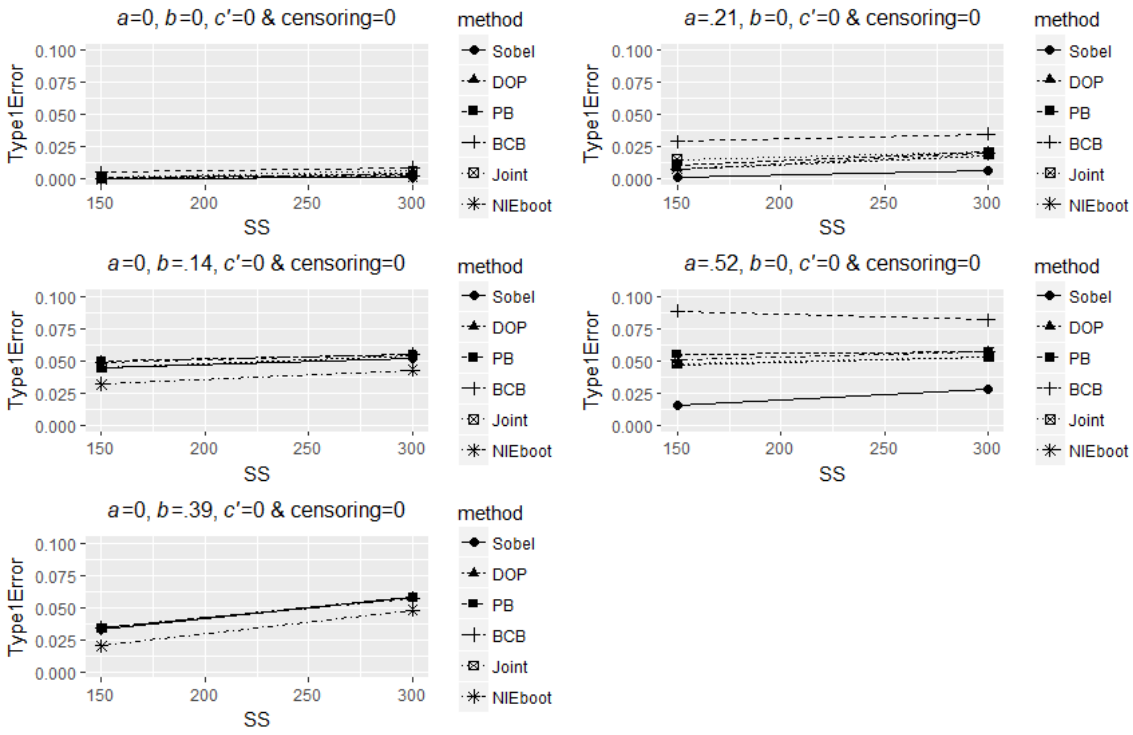


Figure III-9. Type I Error Rate when *c'* = 0 and Censoring Proportion = 0. "Type1Error" on the y-axis denotes the Type I error rate, "SS" on the x-axis denotes sample size and "method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

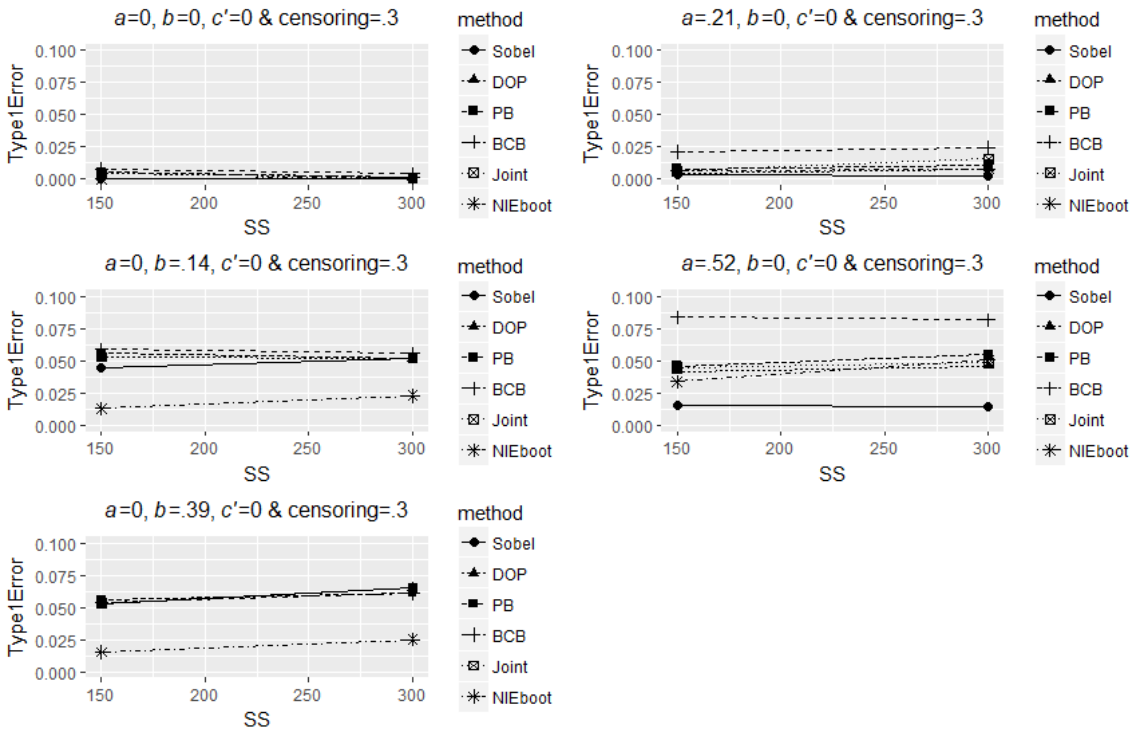


Figure III-10. Type I Error Rate when *c'* = 0 and Censoring Proportion = 0.3.

"Type1Error" on the y-axis denotes the Type I error rate, "SS" on the x-axis denotes sample size and "method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

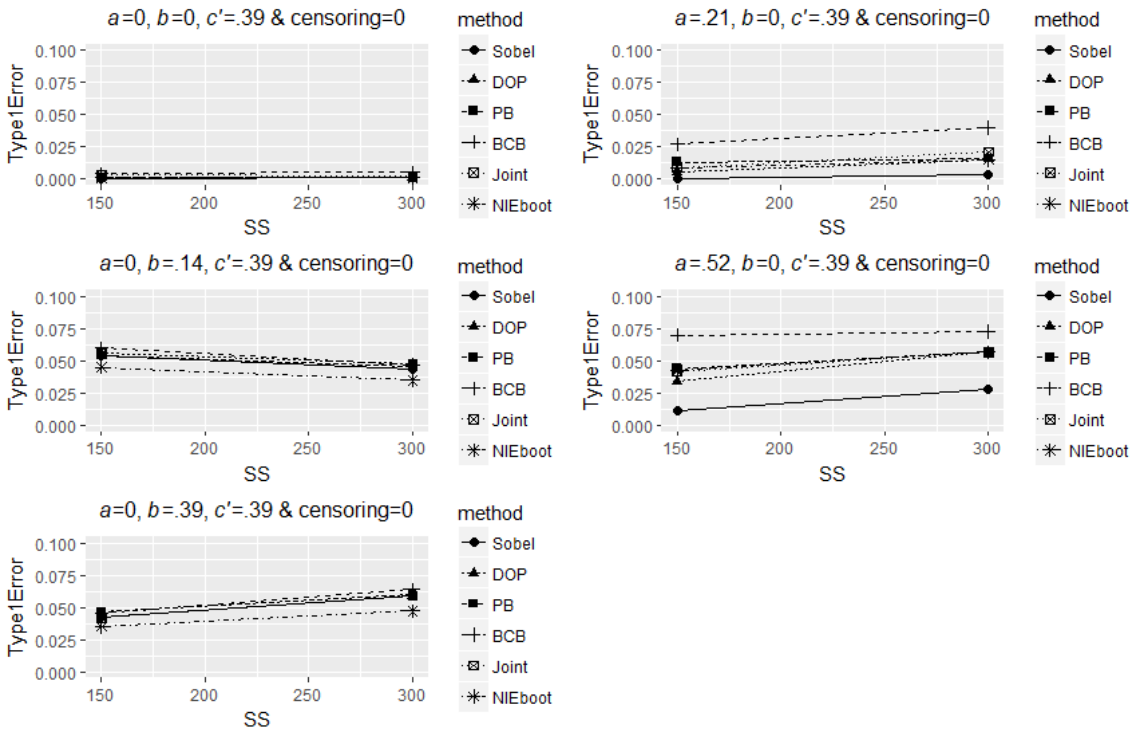


Figure III-11. Type I Error Rate when *c'* = 0.39 and Censoring Proportion = 0. "Type1Error" on the y-axis denotes the Type I error rate, "SS" on the x-axis denotes sample size and "method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

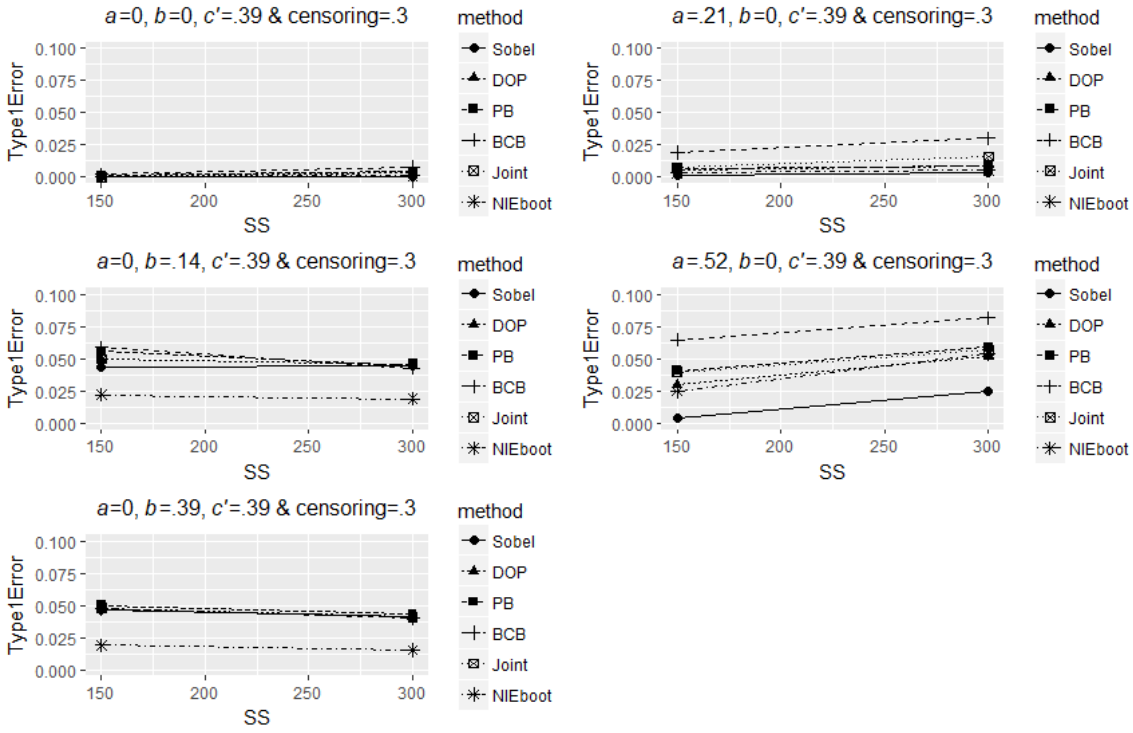


Figure III-12. Type I Error Rate when *c'* = 0.39 and Censoring Proportion = 0.3. "Type1Error" on the y-axis denotes the Type I error rate, "SS" on the x-axis denotes sample size and "method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

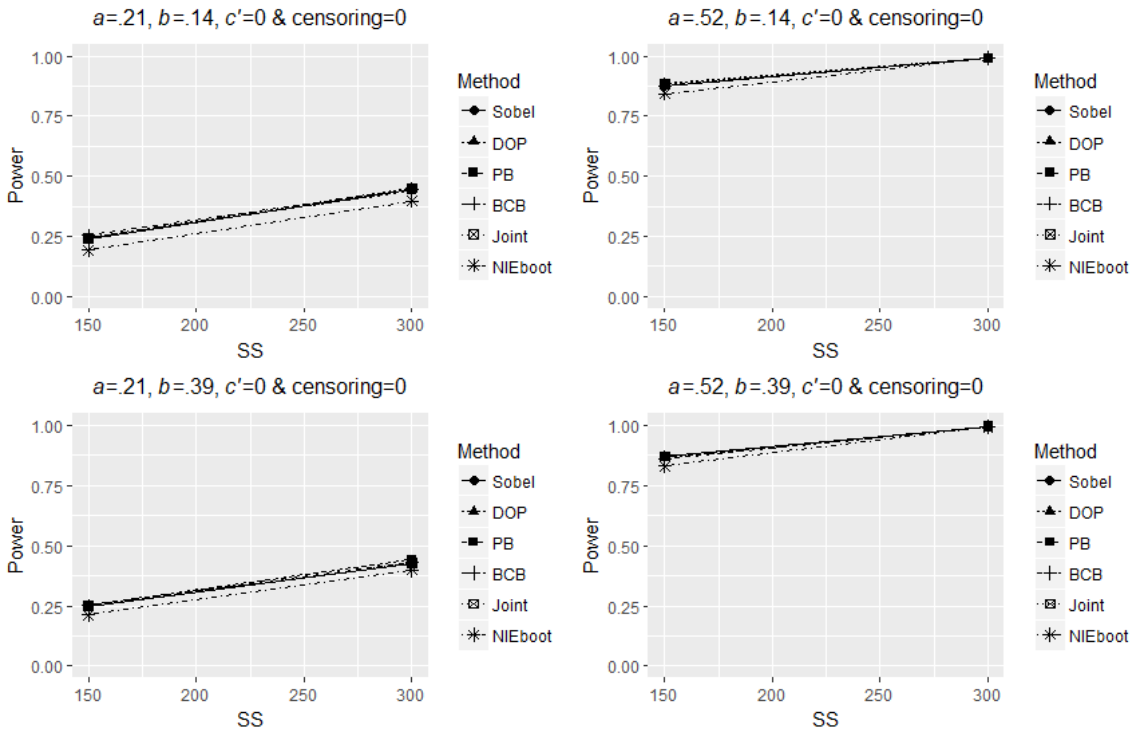


Figure III-13. Statistical Power when *c'* = 0 and Censoring Proportion = 0. "Power" on the y-axis denotes the statistical power, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

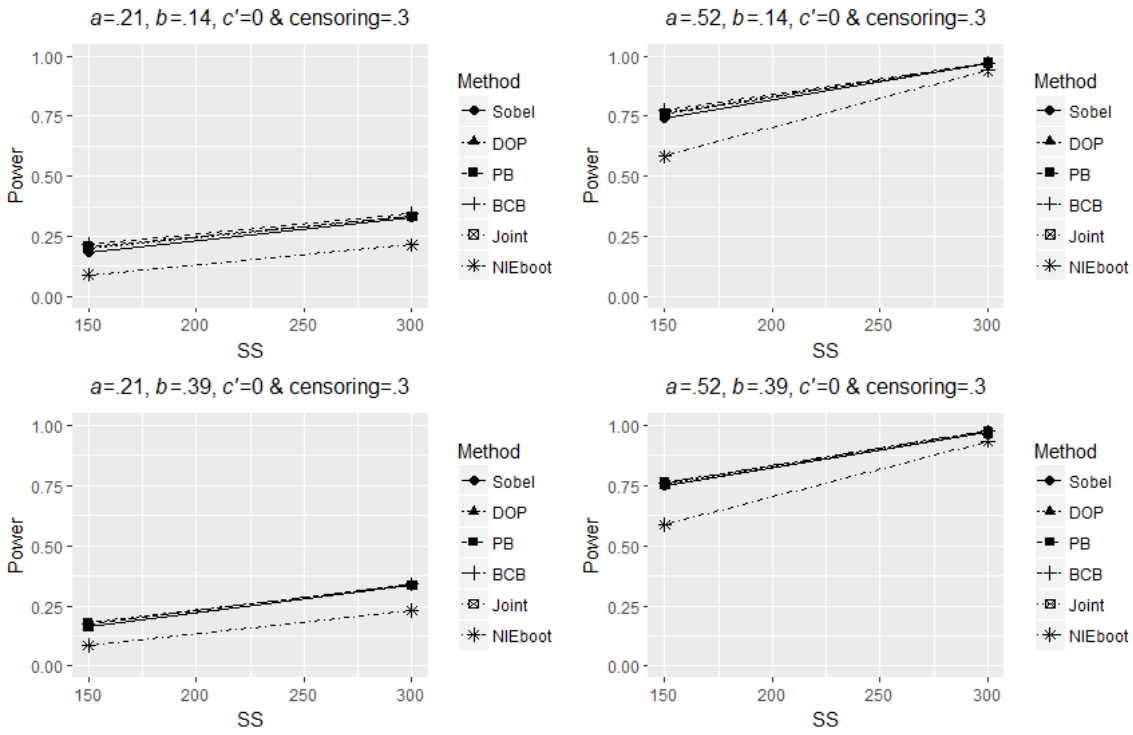


Figure III-14. Statistical Power when *c'* = 0 and Censoring Proportion = 0.3. "Power" on the y-axis denotes the statistical power, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

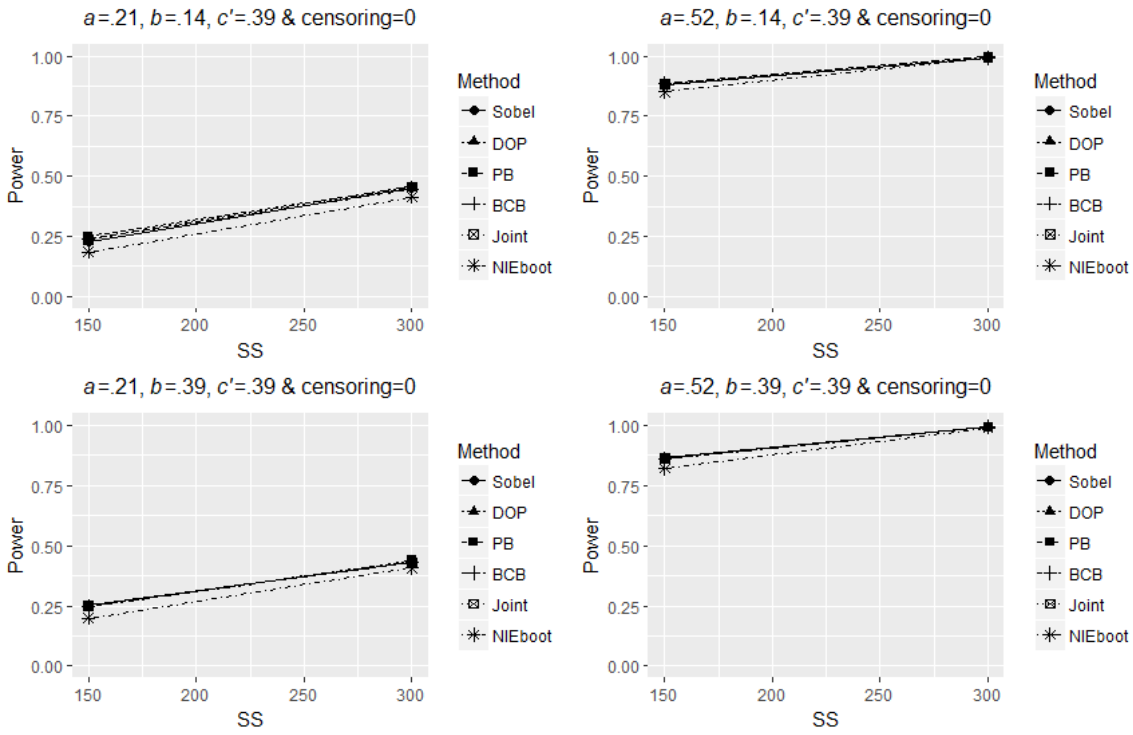


Figure III-15. Statistical Power when *c'* = 0.39 and Censoring Proportion = 0. "Power" on the y-axis denotes the statistical power, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

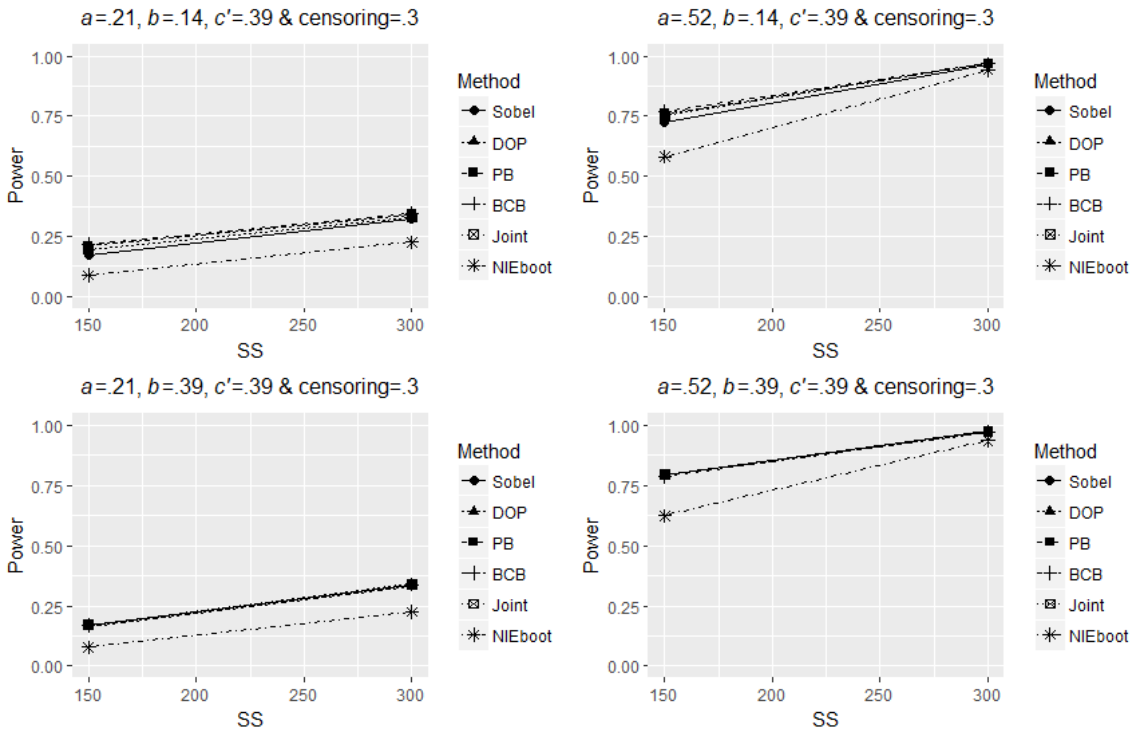


Figure III-16. Statistical Power when *c'* = 0.39 and Censoring Proportion = 0.3. "Power" on the y-axis denotes the statistical power, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; "BCB"=bias-corrected bootstrap of the *ab*-estimate; "Joint"=the joint significance test; and "NIEboot"=percentile bootstrap of the natural indirect effect.

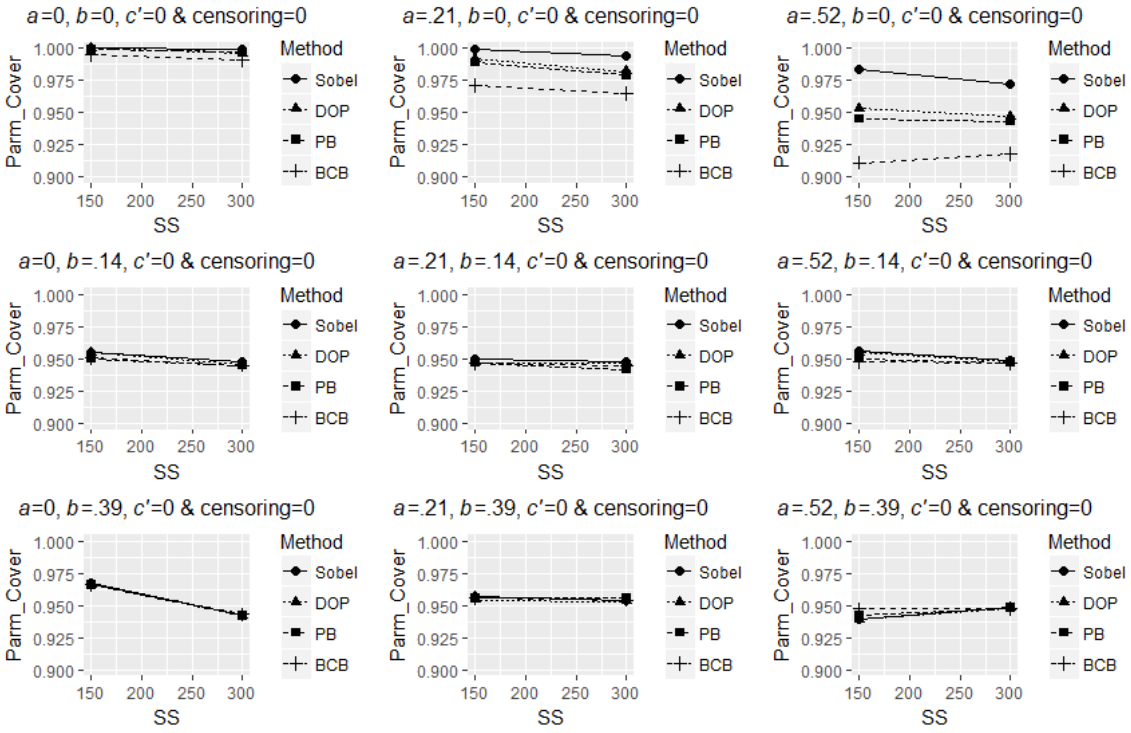


Figure III-17. Parameter *ab* Coverage Rate when *c'* = 0 and Censoring Proportion = 0. "Parm_Cover" on the y-axis denotes the parameter coverage rate, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; and "BCB"=bias-corrected bootstrap of the *ab*-estimate.

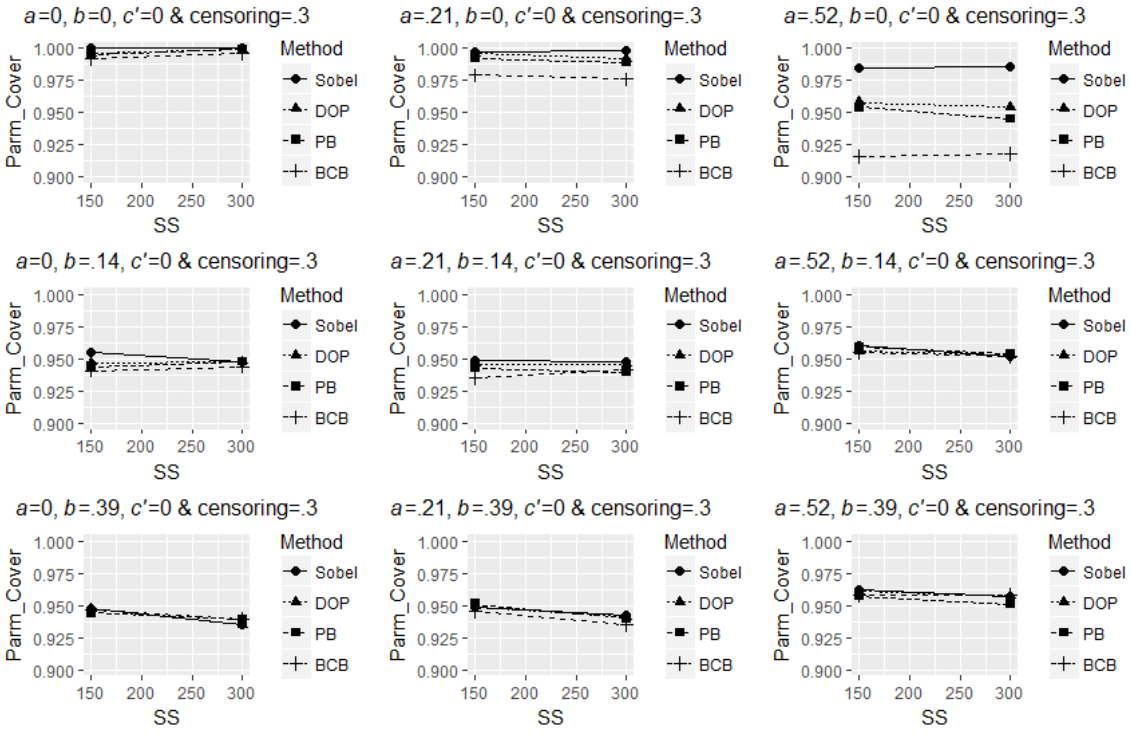


Figure III-18. Parameter ab Coverage Rate when *c'* = 0 and Censoring Proportion = 0.3. "Parm_Cover" on the y-axis denotes the parameter coverage rate, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; and "BCB"=bias-corrected bootstrap of the *ab*-estimate.

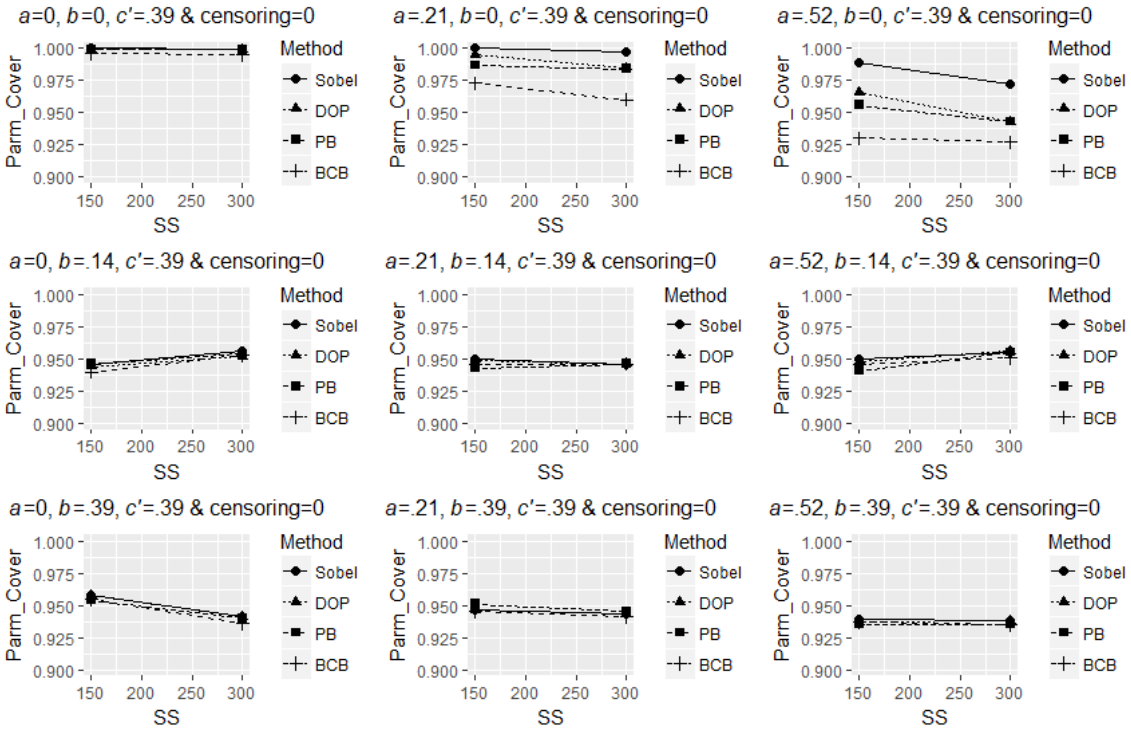


Figure III-19. Parameter *ab* Coverage Rate when *c'* = 0.39 and Censoring Proportion = 0. "Parm_Cover" on the y-axis denotes the parameter coverage rate, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; and "BCB"=bias-corrected bootstrap of the *ab*-estimate.

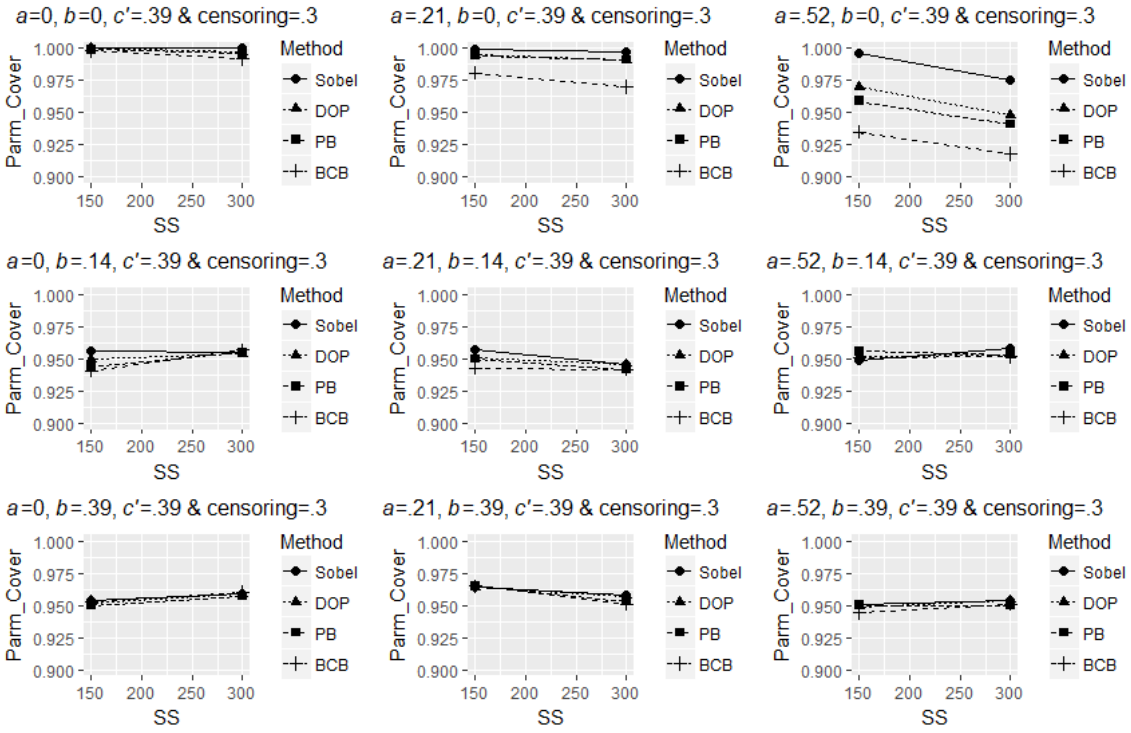


Figure III-20. Parameter *ab* Coverage Rate when *c'* = 0.39 and Censoring Proportion = 0.3. "Parm_Cover" on the y-axis denotes the parameter coverage rate, "SS" on the x-axis denotes sample size and "Method" on the legend title denotes the different indirect effect test methods: "Sobel"=the Sobel test; "DOP"=distribution of the product test; "PB"=percentile bootstrap of the *ab*-estimate; and "BCB"=bias-corrected bootstrap of the *ab*-estimate.

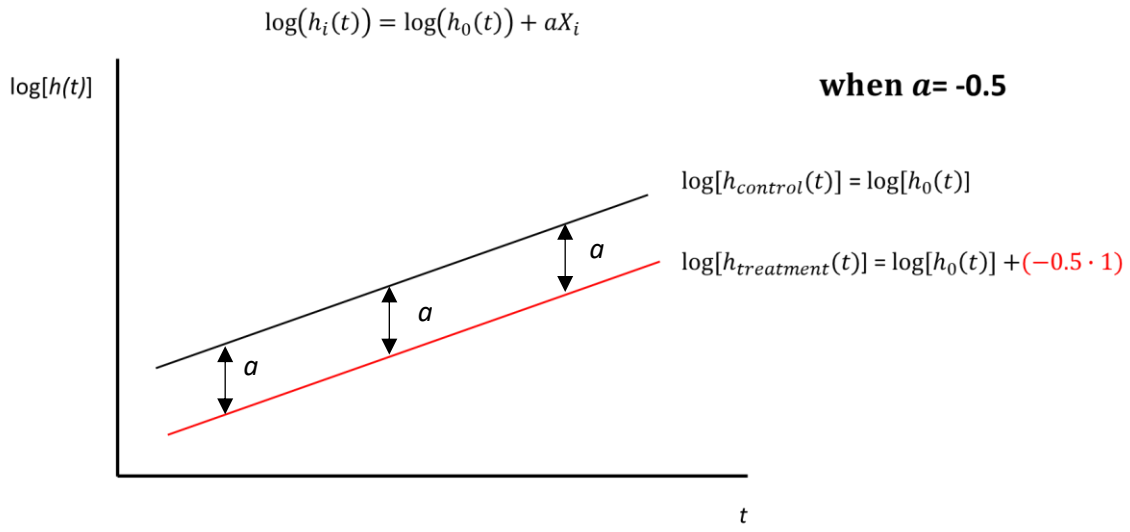


Figure IV-1. An Illustration of the Proportional Hazards Assumption of the Cox Model. The logarithm of the hazard rate is given at the Y-axis and time is given at the X-axis. The black regression line represents the regression when X=0 and the red line represents the regression when X=1. The proportional hazards assumption is shown by the difference between the two regression lines is a constant (a) across all time points.

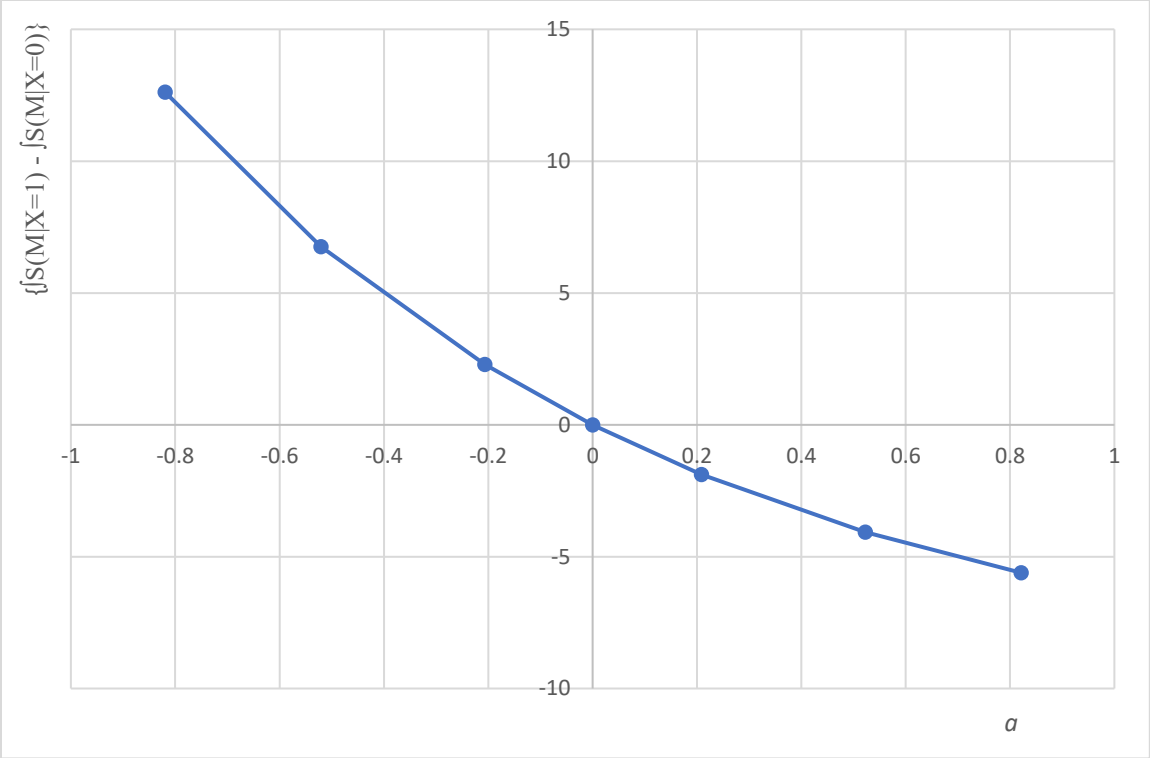


Figure IV-2. The Empirical Relationship Between the *a* Estimate and the Corresponding Part, $\{\int S(M|X=1) - \int S(M|X=0)\}$ in the Natural Indirect Effect. All other parameter values were fixed the same (*b*=0.39, *c'*=0.39, *n*=1,000,000, and censoring proportion=0.3) while only the *a*-parameter values were manipulated at -0.82, -0.52, -0.21, 0, 0.21, 0.52 and 0.82. There was a (hard to define) nonlinear relationship between the *a*-estimate and $\{\int S(M|X=1) - \int S(M|X=0)\}$.