1-1-2015

Instrumental Variable and Propensity Score Methods for Bias Adjustment in Non-Linear Models

Fei Wan
University of Pennsylvania, wanfei@mail.med.upenn.edu

Instrumental Variable and Propensity Score Methods for Bias Adjustment in Non-Linear Models

Abstract
Unmeasured confounding is a common concern when clinical and health services researchers attempt to estimate a treatment effect using observational data or randomized studies with non-perfect compliance. To address this concern, instrumental variable (IV) methods, such as two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI), have been widely adopted. In many clinical studies of binary and survival outcomes, 2SRI has been accepted as the method of choice over 2SPS but a compelling theoretical rationale has not been postulated.

First, We directly compare the bias in the causal hazard ratio estimated by these two IV methods. Under the potential outcome and principal stratification framework, we derive closed form solutions for asymptotic bias in estimating the causal hazard ratio among compliers for both the 2SPS and 2SRI methods by assuming survival time follows the Weibull distribution with random censoring. When there is no unmeasured confounding and no always takers, our analytic results show that 2SRI is generally asymptotically unbiased but 2SPS is not. However, when there is substantial unmeasured confounding, 2SPS performs better than 2SRI with respect to bias under certain scenarios. We use extensive simulation studies to confirm the analytic results from our closed-form solutions. We apply these two methods to prostate cancer treatment data from SEER-Medicare and compare these 2SRI and 2SPS estimates to results from two published randomized trials.

Next, we propose a novel two-stage structural modeling framework to understanding the bias in estimating the conditional treatment effect for 2SPS and 2SRI when the outcome is binary, count or time to event. Under this framework, we demonstrate that the bias in 2SPS and 2SRI estimators can be reframed to mirror the problem of omitted variables in non-linear models. We demonstrate that only when the influence of the unmeasured covariates on the treatment is proportional to their effect on the outcome that 2SRI estimates are generally unbiased for logit and Cox models. We also propose a novel dissimilarity metric to quantify the difference in these effects and demonstrate that with increasing dissimilarity, the bias of 2SRI increases in magnitude. We investigate these methods using simulation studies and data from an observational study of perinatal care for premature infants.

Last, we extend Heller and Venkatraman's covariate adjusted conditional log rank test by using the propensity score method. We introduce the propensity score to balance the distribution of covariates among treatment groups and reduce the dimensionality of covariates to fit the conditional log rank test. We perform the simulation to assess the performance of this new method and covariates adjusted Cox model and score test.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Epidemiology & Biostatistics

INSTRUMENTAL VARIABLE AND PROPENSITY SCORE METHODS FOR BIAS ADJUSTMENT IN NON-LINEAR MODELS

Fei Wan

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

Nandita Mitra

Associate Professor of Biostatistics

Co-Supervisor of Dissertation

Dylan Small

Professor of Statistics

Graduate Group Chairperson

John H. Holmes, Professor of Medical Informatics in Epidemiology

Dissertation Committee

Sharon Xiangwen Xie, Associate Professor of Biostatistics

Justin E. Bekelman, Assistant Professor of Radiation Oncology

INSTRUMENTAL VARIABLE AND PROPENSITY SCORE METHODS FOR BIAS ADJUSTMENT IN NON-LINEAR MODELS

ACKNOWLEDGEMENT

I would like to gratefully and sincerely thank two of my advisors: Drs. Nandita Mitra and Dylan Small, for their advising and guidance, understanding, and patience through every stage of my dissertation research. Without their generous support, it is unlikely for me to complete this work. I would like to express my special thanks to Dr. Mitra for her mentorship for almost 10 years. Only under her encouragement and generous help, I am able to fulfill my lifetime dream of completing a Ph.D degree at Penn.

I would like to express my gratitude to my dissertation committee members for their collaboration and flexibility with my dissertation meetings. I need to thank Dr. Justin Bekelman for his time and efforts on my first methodology paper. I like to thank Dr. Sharon Xie for her generous advices on my questions on survival models. I need further thank all the faculties and fellow students of Penn biostatistics program. Studying at Penn is truly unique educational experience for me.

Last but not least, I like to thank my parents, my wife, and my son Zhicheng Wan for their support to make the completion of this dissertation possible.

ABSTRACT

INSTRUMENTAL VARIABLE AND PROPENSITY SCORE METHODS FOR BIAS ADJUSTMENT IN NON-LINEAR MODELS

Fei Wan

Nandita Mitra

Dylan Small

Unmeasured confounding is a common concern when clinical and health services researchers attempt to estimate a treatment effect using observational data or randomized studies with non-perfect compliance. To address this concern, instrumental variable (IV) methods, such as two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI), have been widely adopted. In many clinical studies of binary and survival outcomes, 2SRI has been accepted as the method of choice over 2SPS but a compelling theoretical rationale has not been postulated.

First, We directly compare the bias in the causal hazard ratio estimated by these two IV methods. Under the potential outcome and principal stratification framework, we derive closed form solutions for asymptotic bias in estimating the causal hazard ratio among compliers for both the 2SPS and 2SRI methods by assuming survival time follows the Weibull distribution with random censoring. When there is no unmeasured confounding and no always takers, our analytic results show that 2SRI is generally asymptotically unbiased but 2SPS is not. However, when there is substantial unmeasured confounding, 2SPS performs better than 2SRI with respect to bias under certain scenarios. We use extensive simulation studies to confirm the analytic results from our closed-form solutions. We apply these two methods to prostate cancer treatment data from SEER-Medicare and compare these 2SRI and 2SPS estimates to results from two published randomized trials.

Next, we propose a novel two-stage structural modeling framework to understanding the bias in estimating the conditional treatment effect for 2SPS and 2SRI when the outcome is binary, count or time to event. Under this framework, we demonstrate that the bias in 2SPS and 2SRI estimators can be reframed to mirror the problem of omitted variables in non-linear models. We demonstrate that only when the influence of the unmeasured covariates on the treatment is proportional to their

effect on the outcome that 2SRI estimates are generally unbiased for logit and Cox models. We also propose a novel dissimilarity metric to quantify the difference in these effects and demonstrate that with increasing dissimilarity, the bias of 2SRI increases in magnitude. We investigate these methods using simulation studies and data from an observational study of perinatal care for premature infants.

Last, we extend Heller and Venkatraman's covariate adjusted conditional log rank test by using the propensity score method. We introduce the propensity score to balance the distribution of covariates among treatment groups and reduce the dimensionality of covariates to fit the conditional log rank test. We perform the simulation to assess the performance of this new method and covariates adjusted Cox model and score test.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

Evaluating the effectiveness of treatment and identifying the causal relationships between exposure and disease are critical objectives for clinical and health services researchers. The casual effects of a treatment can be rigorously defined under the potential outcomes framework (Holland, 1986; Rubin, 2005). Consider a case of a two arm trial that involves one active treatment and a control of no treatment. Let Z_i denote binary treatment status variable, where $Z_i = 1$ if subject i takes the active treatment and $Z_i = 0$ if subject i receives the control. $Y_i^{Z=1}$ is the potential outcome, on a continuous scale, when subject i receives the active treatment and $Y_i^{Z=0}$ is the potential outcome if subject i actually takes the control. The simple treatment effect for subject i is the difference between the two potential outcomes, defined as $Y_i^{Z=1} - Y_i^{Z=0}$. Clearly, only one potential outcome can be observed and the other one is often referred as "counter-factual" and written as $Y_i = Z_i Y_i^{Z=1} + (1 - Z_i) Y_i^{Z=0}$. Therefore, it is not possible to identify the casual effect for an individual because of this missing data issue (Rubin, 2005). However, the average casual effect $E(Y_i^{Z=1} - Y_i^{Z=0})$ for the population is identifiable from the data if certain assumptions are met.

Assume a binary treatment Z is randomized in the population and every subject complies with their assignment. Under the exchange-ability and consistency assumptions (Hernán and Robins, 2006a; Robins J.M. and Brumback, 2000), population average causal effect of treatment is consistently estimated by the difference between two group means,

$$\begin{aligned} E(Y_i^{Z=1} - Y_i^{Z=0}) &= E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0) \\ &= \bar{Y_1} - \bar{Y_0} \end{aligned}$$

Although randomized controlled trials (RCT) are considered as the standard methodology to investigate the causal effects of treatment because by design both observed and unobserved confounding would be controlled, non-compliance frequently occurs when subjects fail to adhere to the treatment assigned. Measured and unmeasured confounding factors may impact the outcome while causing non-compliance of treatment (Cai, Small, and Ten Have, 2011). When the outcomes of patients are compared by the actual treatment they receive, generally known as "as treated analysis", there

may exist prognostic factors that influence patients' compliance with treatment assignment. Thus, the estimator of treatment effect is biased when confounding factors are not fully measured and controlled for (Hernán and Hernández-Díaz, 2012). Alternatively, intention-to-treat analysis (ITT) is a widely accepted simple approach to non-compliance and subjects are analysed according to randomization scheme regardless of treatment actually received. Although the integrity of randomization is retained, ITT tends to underestimate treatment effects, and it measures the causal effects of treatment assignment, instead of effectiveness of treatment (Hernán and Hernández-Díaz, 2012).

Besides non-compliance problem, RCTs are subject to many other limitations, such as lack of generalizability, high cost, lengthy study period, ethic concerns, and difficulty in studying rare diseases, etc (Nallamothu and Hayward, 2008). When a RCT is not feasible, non-randomized observational studies are commonly used to examine the effectiveness of treatment or therapy in routine clinical practice. Compared to RCTs, well designed observational studies can provide more realistic results. Confounding, whether observed or not, is also the main problem of estimating the causal effects in observational studies. The traditional statistical methods, such as stratification, matching, multiple regression, and propensity score, have been used to reduce bias (Martens et al., 2006). These methods are valid under the assumption of no unmeasured confounding variables. In many cases, however, this assumption is very likely to be violated.

An alternative method that could potentially control for both measured and unmeasured confounding variables is the instrumental variable (IV) method. An IV has the following properties: (i) IV either correlates with or has causal effects on treatment or exposure; (ii) IV has no direct effects on outcome except its indirect effects through either treatment or exposure. (iii) there is no unmeasured confounding for the association between IV and outcome variable (Angrist, Imbens, and Rubin, 1996; Hernán and Robins, 2006b). Random assignment scheme in RCTs is an example of IV.

IV method has been widely used in econometrics, as an alternative to ordinary least square method, to obtain consistent parameter estimates in the presence of the endogenous regressors and it is also commonly referred as structural equation model or two-stage least squares (Angrist, Imbens, and Rubin, 1996; Martens et al., 2006). Because IV methods may consistently estimate the average causal effects of treatment even when unmeasured confounding is present (Newhouse and McClellan, 1998), the interests in applying IV technique in outcomes research have been grow-

ing. Terza, Basu, and Rathouz, 2008 made the extension of two IV based approaches, two-stage residual inclusion (2SRI) and two-stage predictor substitution (2SPS), to correcting for endogeneity bias in non-linear models for both binary and time to event outcome. When two IV methods are compared for their performance in estimating the conditional odds ratio or hazard ratio (on unmeasured confounding), they conclude that only 2SRI method produces consistent estimates. This finding rapidly increases the use of 2SRI method to control for unmeasured confounding in medical research (Hadley et al., 2010; Tan et al., 2012).

Under the frame work of potential outcomes, Angrist, Imbens, and Rubin, 1996 divided the subjects accordingly into four principal strata: 1) compliers, who always follow treatment assignment; 2) always takers, who always take the treatment; 3) defiers, who always take the opposite to treatment assignment; and 4) never takers, who never take treatment. They proved that IV estimator of two-stage least squares method consistently estimates a local averaged treatment effects (LATE) among compliers under five assumptions. The details of assumptions are discussed in Chapter 2. Under the same framework of potential outcomes and principal stratification, Cai, Small, and Ten Have, 2011 found analytically and by simulation that both 2SRI and 2SPS logistic regressions generated biased estimate of LATE among compliers. In chapter 2, under the same potential outcome and principal stratification framework, we derive closed form solutions for asymptotic bias in estimating the causal hazard ratio among compliers for both the 2SPS and 2SRI methods by assuming survival time follows the Weibull distribution with random censoring.

In chapter 3, we further assess the performance of 2 stage IV methods in estimating the conditional treatment effect given observed and unobserved covariates. For this purpose, we propose a novel two-stage structural modeling framework to accommodate one endogenous treatment variable and multiple unobserved covariates. This new framework is more relevant to clinical settings. Utilizing this framework, we demonstrate that the bias in 2SPS and 2SRI estimators can be reframed to mirror the problem of omitted variables in non-linear models. We demonstrate that only when the influence of the unmeasured covariates on the treatment is proportional to their effect on the outcome that 2SRI estimates are generally unbiased for logit and Cox models.

In contrast with instrumental variable method, Propensity score method is a common approach to control for confounding bias under no unmeasured confounder assumption. In chapter 4, we explore another use of propensity score method in estimating the conditional hazard ratio given

observed covariates. Covariates adjusted cox proportional hazard model is frequently used when proportional hazard assumption holds. Heller and Venkatraman, 2004 proposed a nonparametric covariate adjusted conditional log rank test to compare survival distributions among different treatment groups. This method is robust when proportional hazard assumption is violated and also does not require any independence assumption between treatment variable and covariates. However, their approach is valid for no more than three covariates. We use the propensity score to balance the distribution of covariates among treatment groups and reduce the dimensionality of covariates to circumvent the limitations of the conditional log rank test. We performed the simulation to assess the performance of this new method and covariates adjusted Cox model and score test.

CHAPTER 2

BIAS IN TWO STAGE INSTRUMENTAL VARIABLE METHODS

2.1. Introduction

Evaluating the effectiveness of treatment and identifying the causal relationship between exposure and disease are critical objectives for clinical and health services researchers. Confounding is often a concern when analyzing non-randomized observational studies and even randomized studies with non-compliance (Hernán and Robins, 2006b). Instrumental variable (IV) methods are increasingly being used in clinical comparative effectiveness studies to potentially control for both measured and unmeasured confounding. Angrist, Imbens, and Rubin, 1996 defined the IV for causal effects of treatment on outcome to be a variable satisfying the following five assumptions: i)The potential outcomes on one subject are unrelated with the particular assignment of treatment to the other subjects; ii) IV is randomly (or ignorably) assigned; iii) Any effect of IV on the outcome must be mediated by treatment received(the exclusion restriction);iv) IV has nonzero effect on treatment received; v) There are no defiers. (for details see section 2.2)

In a recent clinical study, we were interested in comparing the effectiveness of two treatments for prostate cancer in elderly men using SEER-Medicare, a large national observational database. Specifically, we planned to use IV methods to estimate the effect of the addition of external beam radiation therapy (EBRT) to androgen suppression therapy (ADT) in improving overall survival in men with locally advanced prostate cancer. We considered a commonly used IV in health services research: local area treatment patterns defined by the percentage of active treatment in hospital referral regions (HRR). This IV has been shown to capture regionally distinct structural variation in care (Bekelman et al., 2015). Such variation is not fully explained by patient characteristics. Further, this IV varies across HRRs and is strongly associated with treatment assignment. Finally, it is balanced across important observed prognostic factors. Although there is an extensive literature on the importance of choosing an appropriate instrument, less attention has been paid to using the appropriate modeling approach once an IV is selected.

Recently, there has been rapid uptake and widespread use of two IV based analytic approaches called two-stage residual inclusion (2SRI) and two-stage predictor substitution (2SPS)(Cai, Small,

and Ten Have, 2011; Terza, Basu, and Rathouz, 2008).These methods have been used to correct for bias due to endogeneity in non-linear models for both binary and time-to-event outcomes. Among these two IV approaches, 2SRI was shown to consistently estimate a conditional causal parameter under certain assumptions (Terza, Basu, and Rathouz, 2008) and has been adopted as the method of choice in clinical research studies involving survival outcomes(Gore et al., 2010; Hadley et al., 2010; Tan et al., 2012). The conditional causal parameter that Terza, Basu, and Rathouz, 2008 consider is only identified by making homogeneity assumptions that go beyond the five assumptions for a valid IV defined in the first paragraph. Angrist, Imbens, and Rubin, 1996 showed that under these five assumptions for a valid IV, the only treatment effect that is identified is the average treatment effect for the compliers, where the the compliers are the subjects who would take the treatment if encouraged to do so by the IV but would not take the treatment if not encouraged by the IV; this is called the local average treatment effect (LATE).In the context of a binary outcome, Cai, Small, and Ten Have, 2011 demonstrated that both the 2SRI and 2SPS methods generated biased estimates of LATE among compliers for binary outcome. In this paper, we focus on the properties of 2SPS and 2SRI as estimators of the LATE for time-to-event data.

Despite the fact that there is growing interest in applying two stage IV methods to time-to-event data, little is known about the potential bias of using such methods to estimate LATE among compliers. We derive closed form expressions of the bias and conduct extensive simulations to quantify this bias. We then apply both of the two-stage IV methods to our prostate cancer treatment data and compare them to the results from two published randomized clinical trials (Warde et al., 2011; Widmark et al., 2009)

2.2. Notation, Assumptions, Compliance Categories, and Model

2.2.1. Notation

Following the notation of Cai, Small, and Ten Have, 2011 and Nie, Cheng, and Small, 2011, an N-dimensional vector of binary IV is represented by $\underline{R}$. An IV value of 1 represents encouragement to receive the active treatment and 0 represents no encouragement to receive the active treatment. In a RCT setting, where the IV is the randomized assignment, then an IV value of 1 represents random assignment to treatment and 0 represents random assignment to control; in the prostate cancer observational study described in the introduction, an IV value of 1 represents a high local

area rate (above median) of adding EBRT to ADT and 0 represents a low local area rate (below the median) of adding EBRT to ADT. The ith element $R_i = 1$ implies that subject i is encouraged to receive the active treatment, whereas $R_i = 0$ indicates that subject i is not encouraged to receive the active treatment. Let $\underline{Z}^{\underline{R}}$ be an N-dimensional vector of potential treatment received given $\underline{R}$, and ith element $Z_i^{\underline{R}}$=1 indicates that subject i receives the active treatment and $Z_i^{\underline{R}}$=0 means that subject i receives the control under $\underline{R}$.

Similarly, we define $\underline{T}^{\underline{R},\underline{Z}}$ to be an N-dimensional vector of potential survival time under $\underline{R}$ and $\underline{Z}$, and ith element $T_i^{\underline{R},\underline{Z}}$ is the potential survival time for subject i under $\underline{R}$ and $\underline{Z}$. Let $\underline{L}^{\underline{R},\underline{Z}}$ to be an N-dimensional vector of potential censoring time under $\underline{R}$ and $\underline{Z}$, and ith element $L_i^{\underline{R},\underline{Z}}$ is the potential censoring time for subject i under $\underline{R}$ and $\underline{Z}$.

We define $\underline{Y}^{\underline{R},\underline{Z}}$=min$\{\underline{T}^{\underline{R},\underline{Z}}, \underline{L}^{\underline{R},\underline{Z}}\}$, the elementwise minimum of potential censoring and survival times,to be an N-dimensional vector of potential observed follow up time under $\underline{R}$ and $\underline{Z}$, and ith element $Y_i^{\underline{R},\underline{Z}}$ represents the potential follow up time for subject i under $\underline{R}$ and $\underline{Z}$. Let $\delta_i^{\underline{R},\underline{Z}} = I\{T_i^{\underline{R},\underline{Z}} \leq C_i^{\underline{R},\underline{Z}}\}$ indicates whether subject i is observed to terminate by failure ($\delta_i^{\underline{R},\underline{Z}} = 1$) or by censoring ($\delta_i^{\underline{R},\underline{Z}}$=0) given $\underline{R}$ and $\underline{Z}$. The vector $\underline{X}_i$ represents measured confounding variables for subject i.

2.2.2. Assumptions

The main assumptions we will make for causal modeling are the five assumptions made by Angrist et al. (Angrist, Imbens, and Rubin, 1996), and a random censoring assumption for the survival setting.

1) Stable Unit Treatment Value Assumption (SUTVA)(Rubin, 1986, 1990)

a. if R_i=R_i', then $Z_i^{\underline{R}} = Z_i^{\underline{R}'}$

b. if $R_i = R_i'$ and $Z_i = Z_i'$, then $Y_i^{\underline{R},\underline{Z}} = Y_i^{\underline{R}',\underline{Z}'}$

The SUTVA assumption says that the potential outcomes for subject i are not related with the treatment status of other subjects such that we can write $Z_i^{\underline{R}}$,$Y_i^{\underline{R},\underline{Z}}$, $T_i^{\underline{R},\underline{Z}}$, $L_i^{\underline{R},\underline{Z}}$,$\delta_i^{\underline{R},\underline{Z}}$ as $Z_i^{R_i}$,$Y_i^{R_i,Z_i}$,$T_i^{R_i,Z_i}$,$L_i^{R_i,Z_i}$, $\delta_i^{R_i,Z_i}$ respectively. The SUTVA assumption also implies the assumption of consistency, such that the value of the potential outcome given a treatment remains unchanged no matter what the treatment assignment mechanism is (Rubin, 1986)

2) Independence of the instrument $\underline{R}$ (Abadie, 2003):

Conditional on a vector of confounders $\underline{X}$, the random vector ($\underline{Y}^{\underline{R},\underline{Z}}, \underline{T}^{\underline{R},\underline{Z}}, \underline{L}^{\underline{R},\underline{Z}}, \underline{Z}^{\underline{R}}$) is independent of $\underline{R}$. In a randomized trial where R is the IV, the independence assumption holds without conditioning on $\underline{X}$.

3) Exclusion Restriction

$\forall \underline{Z}, \underline{R}$, and $\underline{R'}$, we have:

$\underline{T}^{\underline{R},\underline{Z}} = \underline{T}^{\underline{R'},\underline{Z}}$, $\underline{L}^{\underline{R},\underline{Z}} = \underline{L}^{\underline{R'},\underline{Z}}$, $\underline{Y}^{\underline{R},\underline{Z}} = \underline{Y}^{\underline{R'},\underline{Z}}$, This assumption implies that any effect of IV on potential outcomes must be through its effect on treatment actually received. Thus, we can write $T_i^{\underline{R},\underline{Z}}$, $L_i^{\underline{R},\underline{Z}}$,$Y_i^{\underline{R},\underline{Z}}$ as $T_i^{Z_i}$, $L_i^{Z_i}$,$Y_i^{Z_i}$ by combining the exclusion restriction and SUTVA assumptions.

4) Non-zero Average Causal Effect of $\underline{R}$ on $\underline{Z}$

$E[Z_i^1 - Z_i^0] \neq 0$

This assumption means the IV is correlated with treatment received.

5) Monotonicity (Imbens and Angrist, 1994)

$Z_i^1 \geq Z_i^0, \forall i \in N$

This assumption rules out the existence of defiers. No subject always does the opposite of the treatment assigned.

6) Independent censoring

The distribution of potential survival time $\underline{T}^{\underline{R},\underline{Z}}$ is independent of the distribution of potential censoring time $\underline{L}^{\underline{R},\underline{Z}}$.

2.2.3. Compliance Categories

Under the framework of principal stratification and potential outcomes (Angrist, Imbens, and Rubin, 1996; Rubin, 2005), subjects in a two-arm randomized trial can be categorized into 4 principal strata: Always takers (AT) are subjects who always take the treatment regardless of assignments ($Z^1 = 1, Z^0 = 1$); Compliers (C) are subjects who comply with their assignments($Z^1 = 1, Z^0 = 0$); Never takers (NT) are the subjects who never take the treatment no matter which group they are assigned to($Z^1 = 0, Z^0 = 0$); Defiers (D) are the subjects who take the treatment opposite of their assignments($Z^1 = 0, Z^0 = 1$).

2.2.4. Model

We first define the probability of receiving the treatment $Pr(R = 1) = r$, the probability of being a always taker $Pr(AT) = \rho_a$, and the probability of being a complier $Pr(C) = \rho_c$. We also define the probability of being a defier $Pr(D) = \rho_d$, but under the monotonicity assumption, there are no defiers so that $\rho_d = 0$. Hence, the probability of being a never taker $Pr(NT)$ is equal to $1 - \rho_a - \rho_c$.

We assume both potential censoring time and potential survival time follow the Weibull distribution with the same shape parameter α. The potential censoring time for the subjects in each principal strata follows $Weibull(\alpha, \lambda)$, and we define the parameters of the probability distribution of potential survival time for each principal strata as follows:

$$
\begin{aligned}
T^1|AT \sim Weibull(\alpha, \theta^1_{at}), \quad T^0|AT \sim Weibull(\alpha, \theta^0_{at}) \\
T^1|C \sim Weibull(\alpha, \theta^1_{c}), \quad T^0|C \sim Weibull(\alpha, \theta^0_{c}) \\
T^1|NT \sim Weibull(\alpha, \theta^1_{nt}), \quad T^0|NT \sim Weibull(\alpha, \theta^0_{nt})
\end{aligned}
$$

We also examined scenarios in which different shape parameters α's are assumed for the potential censoring time and the potential survival time. These details are given in Appendix E. The density of Weibull distribution is $f(t) = (\alpha/K)(t/K)^{K-1}exp(-(t/K)^{\alpha})$ and the hazard rate is $h(t) = \alpha K^{-\alpha}t^{\alpha-1}$. In the case of Weibull regression with covariates X, $K^{-\alpha}$ can be reparameterized as $exp(\beta X)$. The hazard rate for the compliers if treated is $h(T^1 = t|C) = \alpha t^{\alpha-1}(\theta^1_c)^{-\alpha}$. The hazard rate for the compliers if not treated is $h(T^0 = t|C) = \alpha t^{\alpha-1}(\theta^0_c)^{-\alpha}$. Hence, the log causal hazard ratio ϕ for the compliers is the difference between two log hazard rates:

$$
\begin{aligned}
\phi &= log[h(T^1 = t|C)] - log[h(T^0 = t|C)] \\
&= -\alpha(log(\theta^1_c) - log(\theta^0_c))
\end{aligned}
$$

2.3. Two Stage Predictor Substitution(2SPS)Method

The 2SPS method is frequently used and simple to implement (Terza, Basu, and Rathouz, 2008). In the first stage, the treatment received Z is regressed on the IV-treatment assignment R, and let

$P = \mathsf{E}(Z|R)$. In the second stage, a log linear model including P ,defined as:

$$log[h(Y|P)] = \eta + \xi P + log(h_0(y)), \quad h_0(Y) = \alpha y^{\alpha-1}$$

is fitted to estimate the coefficient ξ. This is 2SPS estimator of the log causal hazard ratio. We first derive a closed form expression to the probability limit of the maximal likelihood estimator (M.L.E) of ξ, then take the difference between this probability limit and true log causal parameter ϕ for the expression of the asymptotic bias of the 2SPS estimator as an estimator of the log causal hazard ratio for compliers.

2.3.1. Probability limit of M.L.E of causal parameter

Let $\hat{P}$ denote the predicted value from the estimated binary regression model. i.e., $\hat{P} = \hat{\mathsf{E}}(Z|R)$. When $\hat{P}$ is substituted for P, the second stage Weibull model becomes:

$$log[\lambda(Y|\hat{P})] = \eta^* + \xi^* \hat{P} + log(h_0^*(y))$$

Let $\hat{\xi}^*$ and $\hat{\xi}$ denote the estimators (M.L.E) of ξ^* and ξ respectively. As sample size $n \to \infty$, $\hat{P} \to P$, $\hat{\xi}^* \xrightarrow{p} \hat{\xi}$, and $\hat{\xi} \xrightarrow{p} \xi$. Therefore, $\hat{\xi}^* \xrightarrow{p} \xi$. To derive closed form expression for the asymptotic bias, we need to re-express ξ in terms of parameters specified in Section 2.2 under the principal stratification framework.

Only always takers receive the treatment when assigned to control($R = 0$). Both always takers and compliers take the treatment when assigned to treatment($R = 1$). Thus, it can be shown that (Cai, Small, and Ten Have, 2011):

$$p_0 = E(Z|R = 0) = \rho_a, \quad p_1 = E(Z|R = 1) = \rho_a + \rho_c$$

Since $P = \{p_0, p_1\}$ is an one-to-one transformation of $R = \{0, 1\}$, we have the following for the second stage Weibull regression:

$$\begin{aligned} log(h(Y|R = 0)) &= log(h(Y|P = p_0)) \\ &= \eta + \xi p_0 + log(h_0(y)) \end{aligned} \tag{2.1}$$

and,

$$\begin{aligned} log(h(Y|R=1)) &= log(h(Y|P=p_1)) \\ &= \eta + \xi p_1 + log(h_0(y)) \end{aligned} \tag{2.2}$$

Instead of working with a second stage model involving P, we can work with a model involving R instead. Solving (2.1) and (2.2), we have:

$$\xi = \frac{log(h(Y|R=1)) - log(h(Y|R=0))}{p_1 - p_0} \tag{2.3}$$

The log linear model including R assumes two underlying Weibull distributions of the same shape parameter α^*, $Weibull(\alpha^*, K_0)$ and $Weibull(\alpha^*, K_1)$, for subjects assigned to control ($R = 0$) and treatment ($R = 1$) respectively. Thus, (2.3) can be expressed as:

$$\xi = \frac{log(K_1^{-\alpha^*}) - log(K_0^{-\alpha^*})}{\rho_c}, \quad K_1^{-\alpha^*} = e^{\eta + \xi p_1}, \quad K_0^{-\alpha^*} = e^{\eta + \xi p_0} \tag{2.4}$$

It is worth noting that both follow up times of subjects assigned to control, denoted as $Y|R = 0$, and follow up times of subjects assigned to treatment, denoted as $Y|R = 1$, actually follow mixture distributions consisting of three different Weibull distributions. Details are given in Appendix A. However, the second stage Weibull model of 2SPS method imposes the two Weibull distributions, with the same shape parameter α^* but different scale parameters K_0, K_1, upon subjects assigned to treatment($R = 1$) or assigned to control($R = 0$) respectively. Thus, the M.L.E of α^*, K_0, K_1 are derived by maximizing the likelihood function $L_n(\alpha^*, K_0, K_1)$ that consists of products of two Weibull densities: $Weibull(\alpha^*, K_0)$ and $Weibull(\alpha^*, K_1)$.

Let $\hat{\alpha}^*$ denote the M.L.E of α^* and We set E($\frac{\partial \log(L_n(\alpha^*, \hat{K}_0(\alpha^*), \hat{K}_1(\alpha^*))}{\partial \alpha^*}$), the expectation of score equation derived from profile likelihood of α^*, equal to 0 and let $\widetilde{\alpha^*}$ be the solution. Under the assumptions stated in Section 2.2 and consistency of M.L.E, the probability limit of the estimator $\hat{\alpha}^*$ is $\widetilde{\alpha^*}$. Details are given in Appendix C. Once the parameters of the principal strata are defined, $\widetilde{\alpha^*}$ can be solved numerically using a root-finding algorithm such as the "bisection" method. Let $\hat{K_0}, \hat{K_1}$ be the M.L.Es of the two scale parameters K_0, K_1 respectively. After the value of $\widetilde{\alpha^*}$ is

determined, the probability limits of the estimators $\hat{K}_0, \hat{K}_1$ can be derived as follows:

$$\widetilde{K_0} = [\frac{1}{P(\delta=1|R=0)} \times \{\rho_a \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_n \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_c \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}]^{1/\widetilde{\alpha^*}} \tag{2.5}$$

and,

$$\widetilde{K_1} = [\frac{1}{P(\delta=1|R=1)} \times \{\rho_a \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_n \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_c \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}]^{1/\widetilde{\alpha^*}} \tag{2.6}$$

The detailed steps of the derivation of (2.5) and (2.6) are given in Appendix C. By substituting (2.5) and (2.6) into (2.4), we derive the expression of log causal hazard ratio ξ as the following:

$$\begin{aligned}\xi = \{&log([\frac{1}{P(\delta=1|R=1)} \times \{\rho_a \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_n \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_c \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}])^{-1} \\ &- log([\frac{1}{P(\delta=1|R=0)} \times \{\rho_a \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_n \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + \rho_c \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}])^{-1}\} \\ &\times \frac{1}{\rho_c}\end{aligned} \tag{2.7}$$

Thus, (2.7) is the closed-form expression of the probability limit of the log causal hazard ratio estimator $\hat{\xi}^*$ from the 2SPS Weibull model.

2.3.2. Bias analysis

The asymptotic bias of the causal parameter ξ of the 2SPS Weibull regression model is simply the difference between the true log causal hazard ratio ϕ and the derived closed form expression of ξ, such that

$$B_{2sps} = \xi + \alpha(log(\theta_c^1) - log(\theta_c^0)) \tag{2.8}$$

We can re-paramterize θ_{nt}^0 in (2.8) with one additional parameter $\Delta = -\alpha(log(\theta_{nt}^0) - log(\theta_c^0))$ as the following:

$$log(\theta_{nt}^0) = log(\theta_c^0) + \frac{\Delta}{\alpha} \tag{2.9}$$

Δ in (2.9) is the log hazard ratio between never takers and compliers given no treatment. It can be interpreted as the magnitude of the unmeasured confounding because the differences between principal strata are attributable to the unmeasured confounding (Cai, Small, and Ten Have, 2011). When $\Delta = 0$ or $\theta_{nt}^0 = \theta_c^0$,there is no unmeasured confounding.

We make the following observations about the bias of 2SPS method from (3.11): 1) When $\alpha = 1$ and we treat α^* as a known parameter and fix it at 1, that is the scenario when the survival outcomes of all principal strata follow exponential distributions and we also fit an exponential model in the second stage instead of estimating the shape parameter for a more general form of Weibull distribution; 2) When $\rho_c = 1$, every subject is a complier and (2.8) can be simplified as $\frac{1}{\alpha^*} - \frac{\gamma}{\alpha} - \psi(\frac{\alpha^*}{\alpha} + 1)\frac{1}{\alpha} = 0$. Then we have $\widetilde{\alpha^*} = \alpha$. Setting $\rho_c = 1$,$\rho_a = 0$, and $\rho_n = 0$, (2.8) becomes 0 so that bias $B_{2sps} = 0$ when a randomized controlled trial has perfect compliance; 3) When there is no causal effect ($\theta_c^1 = \theta_c^0$), all terms in (2.8) cancel out and we have $B_{2sps} = 0$; 4) When $\rho_a = 0$ and $\theta_c^0 = \theta_n^0$, there is no confounding because there are no always takers and never takers can't get treatment so that the confounding can only be attributable to the difference between never takers and compliers given no treatment(Cai, Small, and Ten Have, 2011). However, (2.8) can not be reduced to 0 under this setting so that the bias of 2SPS method B_{2sps} is generally not 0 even when there is no confounding. 5) λ, the scale parameter of the censoring distribution is involved in bias equation (2.9), which coincides with the results in Struthers and Kalbfleisch(Struthers and Kalbfleisch, 1986).

We can analyze how parameters influence the relationship between the magnitude of confounding and bias using derived closed form expression (2.9). For the purpose of demonstration only, here we create four scenarios in which there are no always takers. The results are revealed in Figure 2.1 (a)-(d).

In Figure 2.1, we can clearly see that the bias of the 2SPS method is not 0 when there is no confounding. The bias increases with the larger shape parameter α of the survival function (within each principal stratum).The bias is the smallest when we have an decreasing hazard rate ($\alpha < 1$) and the highest when we have an increasing hazard rate($\alpha > 1$). By comparing Figure 2.1 (a) and (b), we also observe that the bias decreases as the compliance rate increases from 0.5 to 0.8. When the scale parameter (θ_c) is smaller, the bias is also smaller (Figure 2.1 (a) vs. (c)). Although the probability of being randomly assigned to the treatment group is involved in computing the shape parameter of the second stage Weibull regression model, its effects on the bias are very small (compare Figure 2.1 (b) to (d)).

2.4. Two Stage Residual Inclusion(2SRI)Method

Similar to the 2SPS method, the 2SRI method involves two stage modeling (Terza, Basu, and Rathouz, 2008). In the first stage, we regress the treatment received Z on the IV-treatment assignment R and calculate the residual term $E = Z - \mathsf{E}(Z|R)$. In the second stage, we fit a log linear model on both treatment received variable Z and residual E as,

$$log(h(Y|Z,E))) = \lambda_0 + \lambda_1 Z + \lambda_2 E + log(h_0(y)), \quad h_0(Y) = \alpha y^{\alpha-1} \tag{2.10}$$

, to estimate the regression coefficient λ_1. This is 2SRI estimaor of the log causal hazard ratio. We derive the probability limit of the M.L.E of λ_1 first and then calculate the asymptotic bias by taking the difference between this probability limit of the estimator and true log causal hazard ratio among compliers.

2.4.1. Probability limit of M.L.E of causal parameter

As discussed in a previous study (Cai, Small, and Ten Have, 2011), (2.10) is not the true model for the hazard function $h(Y|Z,E)$. In fact the true model includes the interaction term between Z and E. However, deriving the closed-form expression for the probability limit of the estimator from (2.10) is very difficult when (2.10) is not the true model. With one additional assumption that there are no always takers, (2.10) becomes the true model. We derive a closed-form expression of the probability limit of the estimator of causal parameter λ_1 assuming that there are no always takers and thus (2.10) is the true model. Let $\hat{E}$ denote the residuals from the estimated binary regression model in the first stage. i.e., $\hat{E} = Z - \hat{\mathsf{E}}(Z|R)$. When $\hat{E}$ is substituted for E, (2.10) becomes:

$$log[h(Y|Z,\hat{E})] = \lambda_0^* + \lambda_1^* Z + \lambda_2^* \hat{E} + log(h_0^*(y))$$

Let $\hat{\lambda}_1^*$ and $\hat{\lambda}_1$ be the estimators (M.L.E) of λ_1^* and λ_1. As sample size $n \to \infty$, $\hat{E} \to E$, $\hat{\lambda}_1^* \xrightarrow{p} \hat{\lambda}_1$, and $\hat{\lambda}_1 \xrightarrow{p} \hat{\lambda}_1$. Thus, $\hat{\lambda}_1^* \xrightarrow{p} \lambda_1$. To derive a closed form expression for the asymptotic bias, we need to first re-express λ_1 in terms of the parameters specified in section 2.3 under the principal stratification framework.

As shown in a previous study (Cai, Small, and Ten Have, 2011), under the no always taker assumption, the first stage binary regression is $E(Z|R) = \rho_a + \rho_c R$ and residual term $E = Z - \mathsf{E}(Z|R)$, thus the residual term can be re-expressed as $E = Z - \rho_a - \rho_c R$. Since $\{Z, E\}$ has an one to one relationship with $\{Z, R\}$, we can establish the following equivalence between the model involving $\{Z, E\}$ and the model involving $\{Z, R\}$for the

second stage Weibull model:

$$\begin{aligned} log(h(Y|Z,E)) &= \lambda_0 + \lambda_1 Z + \lambda_2 E + log(h_0(y)) \\ &= \lambda_0 + \lambda_1 Z + \lambda_2 (Z - \rho_a - \rho_c R) + log(h_0(y)) \\ &= log(h(Y|Z,R)) \end{aligned} \tag{2.11}$$

Under the no always taker assumption, the second stage Weibull regression model defined by (2.10) assumes the three underlying Weibull distributions with the same shape parameter but different scale parameters for subjects in the three different subgroups: 1) $\sim Weibull(\alpha^*, K_0)$ for those who are assigned to treatment and receive the treatment actually ($Z = 1, R = 1$). Only compliers are in this group; 2) $\sim Weibull(\alpha^*, K_1)$ for those who are assigned to treatment but do not receive the treatment actually ($Z = 0, R = 1$), This group has only never takers; 3) $\sim Weibull(\alpha^*, K_2)$ for those who are assigned to control and do not receive the treatment ($Z = 0, R = 0$), both never takers and compliers are in this group. There are no subjects that are assigned to control but still take the active treatment ($Z = 1, R = 0$) under the assumption of no always takers. Thus, the M.L.E of α^*, K_0, K_1, K_2 are derived by maximizing the likelihood function $L_n(\alpha^*, K_0, K_1, K_2)$ that consists of products of three Weibull densities: $Weibull(\alpha^*, K_0)$, $Weibull(\alpha^*, K_1)$,and $Weibull(\alpha^*, K_2)$.

Let $\hat{\alpha}^*$ denote the M.L.E of α^* and set E($\frac{\partial \log(L_n(\alpha^*, \hat{K}_0(\alpha^*), \hat{K}_1(\alpha^*), \hat{K}_2(\alpha^*)))}{\partial \alpha^*}$),the expectation of score equation derived from profile likelihood of α^*, to 0 and let $\widetilde{\alpha^*}$ be the solution. Under the assumptions stated in section 2.2 and consistency of the M.L.E, the probability limit of the estimator $\hat{\alpha}^*$ is $\widetilde{\alpha^*}$. Details are given in Appendix D. With the parameters of principal strata defined, $\widetilde{\alpha^*}$ can be solved numerically using a root-finding algorithm. Let $\hat{K_0}, \hat{K_1}, \hat{K_2}$ be the M.L.Es of two scale parameters K_0, K_1, K_2. Once the value of $\widetilde{\alpha^*}$ is determined, we compute the probability limits of the estimators $\hat{K_0}, \hat{K_1}, \hat{K_2}$ as follows:

$$\widetilde{K_0} = [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha} + 1)[\frac{1}{\theta_c^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_c^1}{\lambda})^\alpha}}]^{1/\widetilde{\alpha^*}} \tag{2.12}$$

and

$$\widetilde{K_1} = [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha} + 1)[\frac{1}{\theta_{nt}^{0\alpha}} + \frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}}]^{1/\widetilde{\alpha^*}} \tag{2.13}$$

and

$$\widetilde{K_2} = [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\,\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_{nt} + \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_c}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}\rho_{nt} + \frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}\rho_c}]^{1/\widetilde{\alpha^*}} \tag{2.14}$$

The derivation of (2.12),(2.13) and (2.14) is detailed in Appendix D. Based on (2.11), we can establish the following three equations with all possible combination of values of Z and R excluding the always takers scenario(Z=1,R=0).

1) When Z=1 and R=1, there are only compliers in this subgroup.

$$\begin{aligned} & log(h(Y|Z=1,R=1)) = log(h(Y^{(1)}|Z=1,R=1)) \\ \rightarrow \quad & \lambda_0+\lambda_1+\lambda_2(1-\rho_C) = log(\widetilde{K_0}^{-\widetilde{\alpha^*}}) \\ & = log([\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_c^1}{\lambda})^\alpha}}]^{-1}) \end{aligned} \tag{2.15}$$

2) When Z=0 and R=1, there are only never takers in this subgroup.

$$\begin{aligned} & log(h(Y|Z=0,R=1)) = log(h(Y^{(0)}|Z=0,R=1)) \\ \rightarrow \quad & \lambda_0+\lambda_2(-\rho_C) = log(\widetilde{K_1}^{-\widetilde{\alpha^*}}) \\ & = log([\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\,\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}}]^{-1}) \end{aligned} \tag{2.16}$$

3) When Z=0 and R=0, there are mixture of both never takers and compliers in this subgroup.

$$\begin{aligned} & log(h(Y|Z=0,R=0)) = log(h(Y^{(0)}|Z=0,R=0)) \\ \rightarrow \quad & \lambda_0 = log(\widetilde{K_2}^{-\widetilde{\alpha^*}}) \\ & = log([\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\,\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_{nt} + \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_c}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}\rho_{nt} + \frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}\rho_c}]^{-1}) \end{aligned} \tag{2.17}$$

We then derive the closed form expression for the causal parameter λ_1 by solving (2.15),(2.16),and (2.17) for

λ_1 as follows:

$$
\begin{aligned}
\lambda_1 = log([&\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_c^1}{\lambda})^\alpha}}]^{-1}) \\
- log([&\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_{nt}+\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_c}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}\rho_{nt}+\frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}\rho_c}]^{-1}) \\
- \frac{1-\rho_C}{\rho_C}(log([&\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_{nt}+\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\rho_c}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}\rho_{nt}+\frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}\rho_c}]^{-1})) \\
- log([&\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}}]^{-1})
\end{aligned}
$$

2.4.2. Bias analysis

To compute asymptotic bias of the 2SRI method, we subtract the true log hazard ratio ϕ from the closed-form expression of λ_1.

$$B_{2SRI} = \lambda_1 + \alpha(log(\theta_c^1) - log(\theta_c^0)) \tag{2.18}$$

We can re-parameterize θ_{nt}^0 in (2.18) in the way as in Section 2.3 and let $\theta_{nt}^0 = \theta_c^0 e^{\frac{\Delta}{\alpha}}$. From the derived expression of asymptotic bias of 2SRI estimator, we can make the following observations: 1) When $\alpha = 1$, the survival outcome within a principal stratum follows an exponential distribution. If we treat α^* as known and set $\alpha^* = 1$, it means we fit an exponential regression model in the second stage; 2)When there is perfect compliance ($\rho_c = 1$), we have $B_{2SRI} = 0$. In this scenario, $\widetilde{\alpha^*} = \alpha$. By plugging $\rho_c = 1$ into (2.18), we can easily verify the results; 3) When there is no confounding ($\theta_c^0 = \theta_n^0$), $B_{2SRI} = 0$; 4) When there is no causal effect ($\theta_c^1 = \theta_c^0$), B_{2SRI} is not 0; 5) λ, the scale parameter of the censoring distribution is involved in bias equation (2.18), similar to the findings for 2SPS method.

We can analyze how parameters influence the relationship between the magnitude of confounding and bias from the 2SRI method using (2.18). Similar to the previous section, four scenarios were created assuming there are no always takers. The results are shown in Figure 2.2 (a)-(d). In Figure 2.2, it is apparent that the bias of the 2SRI method is 0 when there is no confounding. Intuitively, under the condition of no confounding, substituting the term of the estimated residuals in the second

stage survival model has no effect on the estimate of the causal parameter. By comparing Figure 2.2 (a) and (b), we also observe that the bias decreases as the compliance rate increases from 0.5 to 0.8. When the scale parameter (θ_c) is smaller, the bias tends to be smaller (Figure 2.2 (a) vs. (c)). The probability of being randomly assigned to the treatment group has very small impact on the bias(compare Figure 2.2 (b) to (d)).

2.5. Simulation

2.5.1. Simulation algorithm

We follow the five step algorithm used by Cai, Small, and Ten Have, 2011 to generate data for a simulation study. In the first step, a data set of N subjects is generated. Always takers, compliers, and never takers among these subjects are generated from a multinomial distribution with probabilities $\{\rho_a, \rho_c, \rho_n\}$. At the second step, treatment assignment status R is generated for each subject with probability $P(R = 1) = \rho_r$. Because outcome in the present study is time to event, we modified step 3 to generate potential survival time $\{T^0, T^1\}$and censoring time $\{L^0, L^1\}$ for each principal stratum based on the parameters $\theta^0_{at}, \theta^0_c, \theta^0_{nt}, \theta^1_{at}, \theta^1_c, \theta^1_{nt}, \lambda$. For instance, if a subject is a complier, the potential time to death under control T^0_c is generated from $weibull(\alpha, \theta^0_c)$ and the potential time to death under treatment T^1_c is generated from $weibull(\alpha, \theta^1_c)$. The potential censoring time $\{L^0_c, L^1_c\}$ are generated from $weibull(\alpha, \lambda)$. At step 4, we use compliance status (always taker, complier, or never taker) and treatment assignment status R to determine the treatment received status Z. For instance, if a subject is a complier and assigned to treatment group ($R = 1$), then $Z = 1$. If a subject is an always taker but assigned to the control group, then $Z = 0$. At step 5, the observed survival time and censoring time are generated as follows:

$T = T^1Z + T^0(1 - Z)$, and $L = L^1Z + L^0(1 - Z)$

and finally observed follow up time and censoring indicator are given as:

$Y = min(T, L)$, and $\delta = I(L \geq T)$

2.5.2. Simulation results

To demonstrate the consistency between the derived closed form expressions and the asymptotic biases from the 2SPS and 2SRI approaches under the assumption of no always takers($\rho_a = 0$), we ran the simulation 2000 times, with the sample size n=10000, according to the same parameter

settings presented in Figure 2.1 d) and Figure 2.2 d). Table 2.1 shows simulation results from 4 scenarios ($\alpha = 0.5, 1, 1.5, 2$). As shown in this table, the biases from simulated results are consistent with the values computed with the derived analytic formula for both the 2SPS and 2SRI Weibull models.We also considered 2SPS and 2SRI Cox models (the second stage regression is a Cox model instead of a Weibull model). The pattern of the biases from 2SPS and 2SRI Cox models remains the same as for the 2SPS and 2SRI Weibull models respectively. With decreasing hazard ($\alpha = 0.5$), the bias from using the 2SPS approach is smaller than the bias from the 2SRI approach. When the hazard is constant or increasing ($\alpha \geq 1$), the results are mixed. With stronger negative confounding, the 2SPS method produces smaller bias than the 2SRI method. However, with no confounding or stronger positive confounding, the 2SPS method produces larger bias than the 2SRI method.

To evaluate the performance of both 2SPS and 2SRI methods in the setting where there are always takers, we simulated the data with various combination of parameters based on the following settings: i) Shape parameter α varies among $\{0.5, 1, 2\}$, which represent decreasing, constant, and increasing hazard scenarios; ii)Probabilities of being always takers ρ_a and compliers ρ_c were set to 3 combinations: $\{0.2, 0.7\}$,$\{0.7, 0.2\}$, and $\{0, 0.5\}$. In this way, low, medium, and high levels of compliance were represented; iii) probability of being assigned to treatment ρ_r were set to $\{0.1, 0.5\}$ to reflect both new and relatively established treatments; iv) Scale parameter of censoring distribution were set to $\{0.5, 1, 2\}$; v) Each of the parameters $\theta_{at}^0, \theta_c^0, \theta_c^1$ was set to $\{0.5, 1, 3\}$ separately. Thus, 1458 possible combinations were created. For each setting, we generated 10,000 observations and fit the 2SPS and 2SRI models to the data. This process was repeated 2000 times.

The results are presented in Figure 2.3. The magnitude of bias increases with increasing magnitudes of unmeasured confounding. As the value of shape parameter α increases, the magnitude of bias increases. In the scenarios with decreasing hazard, the 2SPS method outperforms the 2SRI method. The 2SRI method tends to have larger asymptotic bias when the magnitude of unmeasured confounding is large. In the scenarios with constant hazard, the 2SPS method slightly outperforms the 2SRI method when the magnitude of unmeasured confounding is large. In the scenarios with increasing hazard, both approaches produce larger biases. The 2SRI method performs better when the magnitude of unmeasured confounding is small. When there are always takers, the 2SRI method could be biased even when there is no measured confounding. We also compared

the two methods using mean square error and the conclusions remain the same (2.4).

2.6. Seer-Medicare Prostate Cancer Study

Prostate cancer is the highest prevalence non-skin malignancy among American men (In 2011, there were an estimated 2,707,821 men living with prostate cancer in the United States. The number of deaths was 23.0 per 100,000 men per year). Unlike prostate cancers that are diagnosed at an early stage, locally advanced prostate cancer is associated with substantial morbidity and mortality. Radiation therapy is a common treatment for locally advanced prostate cancer. Two randomized trials recently demonstrated that radiation therapy reduces mortality for men with locally advanced tumors who also receive systemic androgen deprivation (Warde et al., 2011; Widmark et al., 2009). However, both trials excluded elderly patients and those with early stage, PSA-screen detected cancer and therefore had less generalizability, a common criticism of randomized evidence. Therefore, we applied two-stage IV methods to evaluate survival outcomes in locally advanced prostate cancer, assessing survival outcomes of androgen deprivation therapy with or without radiation therapy in comparison to the randomized trials.

We analyzed data from the Surveillance, Epidemiology and End Results (SEER)-Medicare database. The SEER-Medicare database links patient demographic and tumor-specific data collected by SEER cancer registries to Medicare claims for inpatient and outpatient care. We considered patients with prostate cancer diagnosed between January 1, 1995 and December 31, 2007 in SEER with follow up through December 31, 2010 in Medicare. The following patients were excluded: 1) older than age 85 ; 2) with unknown urban category; 3) in hospital referral regions (HRR) with less than 50 patients; 4) with unknown distance to the closest radiation facility; 5) patients who died within the first 9 months of the study. A total of 31,541 patients were selected and categorized as receiving androgen deprivation with or without radiation therapy.

The cohort was divided into the following three groups: 1) patients with American Joint Commission on Cancer (AJCC) Tumor stage (T-stage) of T2 or T3 and aged 65-75 (called the “RCT Cohort”). The patients in the “RCT Cohort“ are most comparable to the patients from the two randomized studies of androgen deprivation with or without radiation therapy (Warde et al., 2011; Widmark et al., 2009); 2) elderly patients under-represented or excluded from the published randomized trials with T-stage T2 or T3, aged 76-85 (called the “Elderly Cohort“); and 3) patients with early

stage, PSA-screen detected cancer with T-stage T1 disease who were excluded from the published randomized trials (called the “Screen-Detected Cohort“).

The study by Widmark et al., 2009 included men from 47 centers in Europe diagnosed between February, 1996 and December, 2002. 875 patients with locally advanced prostate cancer (T3; 78$\%$; prostate-specific antigen (PSA) $\leq$ 70 ng/mL; N0; M0) were enrolled. 439 patients were randomly assigned to androgen deprivation alone and the other 436 patients received androgen deprivation with radiation therapy. The study by Warde et. al. enrolled 1,205 patients with locally advanced (T3 or T4) prostate cancer, organ-confined disease (T2) with either PSA >40 ng/mL or PSA >20 ng/mL and a Gleason score of 8 or higher between 1995 and 2005. 1205 patients were randomly assigned to receive the androgen deprivation alone (n=602) or androgen deprivation with radiation therapy (n=603). The hazard ratios for overall mortality reported previously (Widmark et al., 2009) and (Warde et al., 2011) were 0.68 (95$\%$ CI 0.52—-0.89) and 0.77 (95$\%$ CI 0.61—-0.98). For ease of comparison, we combined the results of the randomized trials using weighted-average meta-analysis. The meta-analytic HR was 0.73 (0.61—-0.87).

To assess the effectiveness of androgen deprivation with or without radiation therapy in reducing overall mortality (death from any cause), we performed two-stage IV Weibull regression analysis (2SPS and 2SRI) using a local area treatment rate instrument and controlling for the propensity score. The local area treatment rate instrument was defined as the proportion of patients who received definitive treatment (surgery or radiation therapy) among all patients with prostate cancer in the hospital referral region (HRR) and we categorized this instrument into a binary variable according to its median. This IV measures the ‘aggressiveness’ of local area treatment and captures regionally distinct structural care variation not fully explained by patient characteristics. The IV was strongly associated with treatment assignment and balanced important prognostic factors (Bekelman et al., 2015). The propensity score model included potential confounding variables including age, race, ethnicity, clinical T stage, N stage, and World Health Organization tumor grade, 17 categories of co-morbid disease, urban residence, and census track median income.

As shown in Table 2.2, there is variability in the estimated HRs obtained from the 2SPS and 2SRI methods. We estimated the shape parameter $\alpha \approx 1.6$ from the data. Using Figure 3, we can see that the bias for both the 2SPS and 2SRI methods is the largest when we have an increasing hazard ($\alpha > 1$), even when the magnitude of unmeasured confounding is relatively small. When

the hazard function is a decreasing one ($\alpha < 1$), the 2SPS method produces more stable and less biased estimates than the 2SRI method. In this case, 2SPS may be a more appropriate approach to use. In the RCT Cohort, the estimated HRs (HR=0.96) from both IV methods are much larger than the meta-analytic HR from the two randomized studies. Note that the confidence intervals are also much larger in both IV analyses than in the original RCTs. In the published RCTs, the authors concluded that there was a statistically significant treatment effect (combined therapy is better) whereas from our IV analysis, we can't draw this conclusion. In the total study sample and separately in the RCT Cohort and the Screen-Detected Cohort, the two IV estimates are quite similar. However, for the Elderly Cohort, the estimate from the 2SPS method is different from the estimate from the 2SRI method.

2.7. Discussion

Many clinical and health services studies are using health care databases to compare the treatment effectiveness for drug and surgical therapies, but are prone to unmeasured confounding. Two stage IV methods have been gaining popularity among clinical researchers because these methods provide a relatively simple approach to analyzing survival outcome studies in the presence of unmeasured confounding. However, current knowledge about potential bias in estimating the log causal hazard ratio is limited. As demonstrated in our prostate cancer study, the large treatment effects estimated from two stage IV methods could be attributable to potential bias. We have derived closed-form expressions for the asymptotic bias of the 2SRI and 2SPS approaches assuming the survival times follow a Weibull distribution with shape parameter α and scale parameter K. We have demonstrated that these analytic results are consistent with our simulation results.

For binary outcomes, two previous studies (Cai, Small, and Ten Have, 2011; Ten Have, M, and M., 2003) demonstrated that the bias in the treatment effect estimated using the 2SRI approach increases as the magnitude of confounding increases. In this current work, we have shown analytically and by simulation that the 2SRI and 2SPS approaches are both biased in estimating the causal hazard ratio among compliers. In some situations when the hazard is decreasing (e.g among patients who have recently received a kidney transplantation), the 2SPS method is less biased than the 2SRI method and could be a more appropriate method to use. When the hazard is an increasing function, both IV methods may produce very large bias even under a moderate amount

of unmeasured confounding. In this case, we recommend exercising caution when interpreting results from two-stage IV survival models.

We have shown that even when all IV assumptions are met, both the 2SRI and the 2SPS methods could fail to consistently estimate the causal hazard ratio among compliers. Our analytic results for bias may help to guide researchers in deciding when the bias is likely to be reasonably small so that two stage IV methods may be reasonably applied. Furthermore, in a sensitivity analysis approach, one may estimate the shape parameter and the censoring proportion among patients assigned to treatment or control from the data. With the shape parameter and censoring proportions fixed based on our known data the level of the unmeasured confounding could be varied to examine how the estimates would change, as shown in Figures 1 and 2. Alternative methods include partial likelihood estimation (Cuzick et al., 2007).

2.8. Appendix

Appendix A: Mixture of Weibull Distributions

1) Prove the distribution function of observed survival time T conditional on random assignment R can be expressed as the following equations:

$$F(T|R=0) = 1 - (e^{-(\frac{t}{\theta^1_{AT}})^\alpha}\rho_A + e^{-(\frac{t}{\theta^0_{NT}})^\alpha}\rho_N + e^{-(\frac{t}{\theta^0_{C}})^\alpha}\rho_C) \tag{A.1}$$

and,

$$F(T|R=1) = 1 - (e^{-(\frac{t}{\theta^1_{C}})^\alpha}\rho_C + e^{-(\frac{t}{\theta^0_{NT}})^\alpha}\rho_N + e^{-(\frac{t}{\theta^1_{AT}})^\alpha}\rho_A) \tag{A.2}$$

In the above equations, AT represents always takers, C represents compliers, and NT represents never takers. Other definitions of parameters and distributions that are used in the proof are given

below:

$$R = \text{1 if assigned to treatment;0 if assigned to control}$$

$$Z = \text{1 if receives treatment; 0 if receives control}$$

$$\rho_r = P(R = 1)$$

$$\rho_A = P(AT)$$

$$\rho_C = P(C)$$

$$\rho_N = 1 - \rho_A - \rho_C$$

$$T^1 = \text{potential outcome for a patient under treatment}$$

$$T^0 = \text{potential outcome for a patient under control}$$

$$T^1|AT \sim weibull(\alpha, \theta^1_{AT})$$

$$T^1|C \sim weibull(\alpha, \theta^1_{C})$$

$$T^1|NT \sim weibull(\alpha, \theta^1_{NT})$$

$$T^0|AT \sim weibull(\alpha, \theta^0_{AT})$$

$$T^0|C \sim weibull(\alpha, \theta^0_{C})$$

$$T^0|NT \sim weibull(\alpha, \theta^0_{NT})$$

no defiers under monotonicity assumption

Proof:

$$
\begin{aligned}
F(T^{(1)}|Z=1,R=1) &= P(T^{(1)} \le t|Z=1,R=1) \\
&= \frac{P(T^{(1)} \le t, Z=1, R=1)}{P(Z=1,R=1)} \\
&= \frac{P(T^{(1)} \le t, AT, R=1) + P(T^{(1)} \le t, C, R=1)}{P(AT,R=1)+P(C,R=1)} \\
&= \frac{P(T^{(1)} \le t, AT)P(R=1) + P(T^{(1)} \le t, C)P(R=1)}{(P(AT)+P(C))P(R=1)} \qquad \because R \perp (T^{(1)}, T^{(0)}),
\end{aligned}
$$

$R\perp$principal strata

$$
\begin{aligned}
&= \frac{P(T^{(1)} \le t|AT)P(AT) + P(T^{(1)} \le t|C)P(C)}{P(AT)+P(C)} \\
&= (1 - e^{-(\frac{t}{\theta^1_{AT}})^{\alpha}}) \frac{P(AT)}{P(AT)+P(C)} \\
&+ (1 - e^{-(\frac{t}{\theta^1_{C}})^{\alpha}}) \frac{P(C)}{P(AT)+P(C)}
\end{aligned}
$$

$$
\begin{aligned}
F(T^{(0)}|Z=0,R=1) &= P(T^{(0)} \le t|Z=0,R=1) \\
&= \frac{P(T^{(0)} \le t, Z=0, R=1)}{P(Z=0,R=1)} \\
&= \frac{P(T^{(0)} \le t, NT, R=1)}{P(NT)P(R=1)} \\
&= P(T^{(0)} \le t|NT) \qquad \because R \perp (T^{(1)}, T^{(0)}), R\perp\text{principal strata} \\
&= (1 - e^{-(\frac{t}{\theta^0_{NT}})^{\alpha}})
\end{aligned}
$$

$F(T|R=1)$can be expressed as:

$$
\begin{aligned}
F(T|R=1) &= P(T \le t, Z=1|R=1) + P(T \le t, Z=0|R=1) \\
&= P(T \le t|Z=1,R=1)P(Z=1|R=1) \\
&+ P(T \le t|Z=0,R=0)P(Z=0|R=1) \\
&= P(T^{(1)} \le t|Z=1,R=1)P(Z=1|R=1) \\
&+ P(T^{(0)} \le t|Z=0,R=1)P(Z=0|R=1) \\
&= ((1-e^{-(\frac{t}{\theta^1_{AT}})^\alpha})\frac{P(AT)}{P(AT)+P(C)} + (1-e^{-(\frac{t}{\theta^1_{C}})^\alpha})\frac{P(C)}{P(AT)+P(C)})(P(AT)+P(C)) \\
&+ (1-e^{-(\frac{t}{\theta^0_{NT}})^\alpha})(P(NT)) \\
&= 1-(e^{-(\frac{t}{\theta^1_{C}})^\alpha}\rho_C + e^{-(\frac{t}{\theta^0_{NT}})^\alpha}\rho_N + e^{-(\frac{t}{\theta^1_{AT}})^\alpha}\rho_A)
\end{aligned}
$$

$$
\begin{aligned}
F(T^{(1)}|Z=1,R=0) &= P(T^{(1)} \le t|Z=1,R=0) \\
&= \frac{P(T^{(1)} \le t, Z=1, R=0)}{P(Z=1,R=0)} \\
&= \frac{P(T^{(1)} \le t, AT, R=0)}{P(AT,R=0)} \\
&= \frac{P(T^{(1)} \le t|AT)P(AT)P(R=0)}{P(AT)P(R=0)} \\
&= P(T^{(1)} \le t|AT) \\
&= 1-e^{-(\frac{t}{\theta^1_{AT}})^\alpha}
\end{aligned}
$$

$$
\begin{aligned}
F(T^{(0)}|Z=0,R=0) &= P(T^{(0)} \le t|Z=0,R=0) \\
&= \frac{P(T^{(0)} \le t, Z=0, R=0)}{P(Z=0,R=0)} \\
&= \frac{P(T^{(0)} \le t, NT, R=0) + P(T^{(0)} \le t, C, R=0)}{P(NT,R=0)+P(C,R=0)} \\
&= \frac{P(T^{(0)} \le t|NT)P(NT) + P(T^{(0)} \le t|C)P(C)}{P(NT)+P(C)} \\
&= (1-e^{-(\frac{t}{\theta^0_{NT}})^\alpha})\frac{P(NT)}{P(NT)+P(C)} \\
&+ (1-e^{-(\frac{t}{\theta^0_{C}})^\alpha})\frac{P(C)}{P(NT)+P(C)}
\end{aligned}
$$

$F(T|R=0)$can be expressed as:

$$
\begin{aligned}
F(T|R=0) &= P(T \le t, Z=1|R=0) + P(T \le t, Z=0|R=0) \\
&= P(T \le t|Z=1,R=0)P(Z=1|R=0) \\
&+ P(T \le t|Z=0,R=0)P(Z=0|R=0) \\
&= P(T^{(1)} \le t|Z=1,R=0)P(Z=1|R=0) \\
&+ P(T^{(0)} \le t|Z=0,R=0)P(Z=0|R=0) \\
&= ((1-e^{-(\frac{t}{\theta^1_{AT}})^\alpha})P(AT) \\
&+ [(1-e^{-(\frac{t}{\theta^0_{NT}})^\alpha})\frac{P(NT)}{P(NT)+P(C)} + (1-e^{-(\frac{t}{\theta^0_{C}})^\alpha})\frac{P(C)}{P(NT)+P(C)}](P(C)+P(NT)) \\
&= 1-(e^{-(\frac{t}{\theta^1_{AT}})^\alpha}\rho_A + e^{-(\frac{t}{\theta^0_{NT}})^\alpha}\rho_N + e^{-(\frac{t}{\theta^0_{C}})^\alpha}\rho_C)
\end{aligned}
$$

Appendix B: Proofs related with Derivation of Closed Form Solution

1) Assume survival time $T \sim Weibull(\alpha, K)$ and censoring time $L \sim Weibull(\alpha, \lambda)$. Let $Y = min(T, L)$ and $\delta = I(T \le L)$. Show that

$Y \sim Weibull(\alpha, (\frac{1}{\lambda^\alpha} + \frac{1}{K^\alpha})^{-1/\alpha})$

and,

$$
P(\delta = 1) = \frac{1}{1+\frac{K^\alpha}{\lambda^\alpha}} \tag{B.1}
$$

Proof:

$$
\begin{aligned}
P(Y \geq y) &= P(min(T, L) \geq y) \\
&= P(T \geq y, L \geq y) \\
&= \int_y^{+\infty} \alpha \frac{t^{\alpha-1}}{K^\alpha} exp(-(\frac{t}{K})^\alpha) dt \int_y^{+\infty} \alpha \frac{l^{\alpha-1}}{\lambda^\alpha} exp(-(\frac{l}{\lambda})^\alpha) dl \\
&= exp(-(\frac{y}{K})^\alpha) exp(-(\frac{y}{\lambda})^\alpha) \\
&= exp(-(\frac{y}{(\frac{1}{\lambda^\alpha} + \frac{1}{K^\alpha})^{-1/\alpha}})^\alpha)
\end{aligned}
$$

Thus,$Y \sim Weibull(\alpha, (\frac{1}{\lambda^\alpha} + \frac{1}{K^\alpha})^{-1/\alpha})$

$$
\begin{aligned}
P(\delta = 1) &= P(0 \leq T \leq L, 0 \leq L \leq \infty) \\
&= \int_0^{+\infty} \alpha \frac{l^{\alpha-1}}{\lambda^\alpha} exp(-(\frac{l}{\lambda})^\alpha) \int_0^{l} \alpha \frac{t^{\alpha-1}}{K^\alpha} exp(-(\frac{t}{K})^\alpha) dt dl \\
&= \int_y^{+\infty} \alpha \frac{l^{\alpha-1}}{\lambda^\alpha} exp(-(\frac{l}{\lambda})^\alpha)[1 - exp(-(\frac{l}{K})^\alpha)] dl \\
&= 1 - \int_y^{+\infty} \alpha \frac{l^{\alpha-1}}{\lambda^\alpha} exp(-(\frac{l}{\lambda})^\alpha) exp(-(\frac{l}{K})^\alpha) dl \\
&= 1 - \frac{1}{1 + (\lambda/K)^\alpha} \\
&= \frac{1}{1 + (\frac{K}{\lambda})^\alpha}
\end{aligned}
$$

2) Assume survival time T is a mixture of three Weibull distributions with Density $f(t) = \sum_{i=1}^3 p_i f(t_i)$. $T_1 \sim Weibull(\alpha, K_1)$, $T_2 \sim Weibull(\alpha, K_2)$, and $T_3 \sim Weibull(\alpha, K_3)$. The weights are p_1, p_2, p_3 and $\sum_{i=1}^3 p_i = 1$. The censoring time $L \sim Weibull(\alpha, \lambda)$. Let $Y = min(T, L)$ and $\delta = I(T \leq L)$. Show that

$$
P(\delta = 1) = p_1 \frac{1}{1 + \frac{K_1^\alpha}{\lambda^\alpha}} + p_2 \frac{1}{1 + \frac{K_2^\alpha}{\lambda^\alpha}} + p_3 \frac{1}{1 + \frac{K_3^\alpha}{\lambda^\alpha}} \tag{B.2}
$$

Proof:

$$
\begin{aligned}
P(\delta=1) &= P(\delta=1,G=1)+P(\delta=1,G=2)+P(\delta=1,G=3)\\
&= P(\delta=1|G=1)P(G=1)+P(\delta=1|G=2)P(G=2)+P(\delta=1|G=3)P(G=3)\\
&= p_1\frac{1}{1+\frac{K_1^\alpha}{\lambda^\alpha}}+p_2\frac{1}{1+\frac{K_2^\alpha}{\lambda^\alpha}}+p_3\frac{1}{1+\frac{K_3^\alpha}{\lambda^\alpha}}
\end{aligned}
$$

3) Given X follows a Weibull distribution (α^*, K). Show that

$$E(X^\alpha)=\Gamma(\frac{\alpha}{\alpha^*}+1)K^\alpha \tag{B.3}$$

Proof:

$$
\begin{aligned}
E(X^\alpha) &= \int X^\alpha \frac{\alpha^*}{K^{\alpha^*}}X^{\alpha^*-1}e^{-(\frac{X}{K})^{\alpha^*}}dx\\
&= \int y^{\frac{\alpha}{\alpha^*}}\frac{1}{K^{\alpha^*}}e^{-\frac{y}{K^{\alpha^*}}}dy && \text{Let } y=x^{\alpha^*}\\
&= \frac{1}{K^{\alpha^*}}\int y^{(\frac{\alpha}{\alpha^*}+1)-1}e^{-\frac{y}{K^{\alpha^*}}}dy\\
&= \frac{1}{K^{\alpha^*}}(K^{\alpha^*})^{(\frac{\alpha}{\alpha^*}+1)}\Gamma(\frac{\alpha}{\alpha^*}+1)\int \frac{1}{(K^{\alpha^*})^{(\frac{\alpha}{\alpha^*}+1)}\Gamma(\frac{\alpha}{\alpha^*}+1)}y^{(\frac{\alpha}{\alpha^*}+1)-1}e^{-\frac{y}{K^{\alpha^*}}}dy\\
&= \Gamma(\frac{\alpha}{\alpha^*}+1)K^\alpha
\end{aligned}
$$

4) Given X follows a Weibull distribution (α^*, K). Show that

$$E(log(X))=\frac{-\gamma}{\alpha^*}+log(K) \tag{B.4}$$

Proof:

$$
\begin{aligned}
E(log(X)) &= \int_0^\infty log(X)\frac{\alpha^*}{K^{\alpha^*}}X^{\alpha^*-1}e^{-(\frac{X}{K})^{\alpha^*}}dx \\
&= \int \frac{1}{\alpha^*}log(y)\frac{1}{K^{\alpha^*}}e^{-\frac{y}{K^{\alpha^*}}}dy && \text{Let } y = x^{\alpha^*} \\
&= \int \frac{1}{\alpha^*}(log(u)+\alpha^* log(K))e^{-u}du && \text{Let } y = uK^{\alpha^*} \\
&= \frac{1}{\alpha^*}\underbrace{\int log(u)e^{-u}du}_{-\gamma} + log(K)\int e^{-u}du && \text{Let } y = uK^{\alpha^*} \\
&= \frac{-\gamma}{\alpha^*} + log(K)
\end{aligned}
$$

5) Given X follows a Weibull distribution (α^*, K). Show that

$$E(X^\alpha log(X))\frac{1}{\alpha^*}\Gamma(\frac{\alpha}{\alpha^*}+1)(K^\alpha)(\psi(\frac{\alpha}{\alpha^*}+1)+\alpha^* log(K)) \tag{B.5}$$

Proof:

$$
\begin{aligned}
E(X^\alpha log(X)) &= \int_0^\infty X^\alpha log(X)\frac{\alpha^*}{K^{\alpha^*}}X^{\alpha^*-1}e^{-(\frac{X}{K})^{\alpha^*}}dx \\
&= \int y^{\frac{\alpha}{\alpha^*}}\frac{1}{\alpha^*}log(y)\frac{1}{K^{\alpha^*}}e^{-\frac{y}{K^{\alpha^*}}}dy && \text{Let } y = x^{\alpha^*} \\
&= \frac{1}{\alpha^*}\frac{1}{K^{\alpha^*}}\Gamma(\frac{\alpha}{\alpha^*}+1)(K^{\alpha^*})^{\frac{\alpha}{\alpha^*}+1} \\
&\underbrace{\int log(y)\frac{1}{\Gamma(\frac{\alpha}{\alpha^*}+1)(K^{\alpha^*})^{\frac{\alpha}{\alpha^*}+1}}y^{(\frac{\alpha}{\alpha^*}+1)-1}e^{-\frac{y}{K^{\alpha^*}}}dy}_{E(log(y))} && y \sim gamma(\frac{\alpha}{\alpha^*}+1, K^{\alpha^*}) \\
&= \frac{1}{\alpha^*}\frac{1}{K^{\alpha^*}}\Gamma(\frac{\alpha}{\alpha^*}+1)(K^{\alpha^*})^{\frac{\alpha}{\alpha^*}+1}(\psi(\frac{\alpha}{\alpha^*}+1)+\alpha^* log(K)) && \psi() \text{ is digamma function} \\
&= \frac{1}{\alpha^*}\Gamma(\frac{\alpha}{\alpha^*}+1)(K^\alpha)(\psi(\frac{\alpha}{\alpha^*}+1)+\alpha^* log(K))
\end{aligned}
$$

6) Let T_i denote the survival time and C_i denote the censoring time for subject i. T_i and C_i are independent. $T_i \sim weibull(\alpha, K)$, and $C_i \sim weibull(\alpha, \lambda)$. Let $Y_i = min(T_i, C_i)$ denote observed follow-up time and δ_i be the indicator variable $\delta_i = (T_i \leq C_i)$. Show that:

$$E(Y_i\delta_i) = E(Y_i)E(\delta_i) \tag{B.6}$$

Proof:

$$
\begin{aligned}
E(Y_i\delta_i) &= E(Y_i\delta_i(I(\delta_i=1)+I(\delta_i=0)) \\
&= E(Y_i\delta_i I(\delta_i=1)) + E(Y_i\delta_i I(\delta_i=0)) \\
&= E(T_i I(\delta_i=1)) \\
&= \int_0^\infty \int_0^\infty tI(\delta_i=1)f(t,c)dtdc \\
&= \int_0^\infty \int_0^\infty tI(t\le c)f_t(t)f_c(c)dtdc \\
&= \int_0^\infty \{\int_0^\infty I(t\le c)f_c(c)dc\}tf_t(t)dt \\
&= \int_0^\infty S_c(t)tf_t(t)dt \\
&= \int_0^\infty t\exp\left(-\frac{t^\alpha}{\lambda^\alpha}\right)\frac{\alpha}{K^\alpha}t^{\alpha-1}\exp\left(-\frac{t^\alpha}{K^\alpha}\right)\,\mathrm{d}t
\end{aligned}
$$

Let $K^* = (\frac{1}{\lambda^\alpha}+\frac{1}{K^\alpha})^{-1/\alpha}$ and use (B.1)

$$
\begin{aligned}
E(Y_i)E(\delta_i) &= (\frac{1}{1+\frac{K^\alpha}{\lambda^\alpha}})\int_0^\infty y\frac{\alpha}{K^{*\alpha}}y^{\alpha-1}\exp\left(-\frac{y^\alpha}{K^{*\alpha}}\right)dy \\
&= \int_0^\infty \frac{\alpha}{K^\alpha}y^{\alpha-1}\exp\left(-\frac{y^\alpha}{K^\alpha}\right)y\exp\left(-\frac{y^\alpha}{\lambda^\alpha}\right)dy
\end{aligned}
$$

Both $E(Y_i\delta_i)$ and $E(Y_i)E(\delta_i)$ have the same integral functions. Thus,

$$E(Y_i\delta_i) = E(Y_i)E(\delta_i)$$

Similarly, we can establish the following:

$$E(g(Y_i)\delta_i) = E(g(Y_i))E(\delta_i)$$

Appendix C: Derivation of probability limits of M.L.E of α, K_0, K_1 for 2SPS

Let $Y = min(T,C)$ be observed follow-up time and $\delta = I(T\le C)$ be the censoring time. The subjects are assigned to either treatment group ($R=1$) or control group ($R=0$). The distribution of each subgroup has different scale parameter K but the same shape parameter α^*. Thus, likelihood

function of observed follow up time Y can be written as:

$$L(y) = \prod_{i\in\{R=1\}}^{n_{R1}} [(\alpha^*/K_1)(y_i/K_1)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_1)^{\alpha^*})] \times \prod_{i\in\{R=0\}}^{n_{R0}} [(\alpha^*/K_0)(y_i/K_0)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_0)^{\alpha^*})]$$

For treatment assignment group and control assignment group, subjects are from compliers (c), never takers(nt),and always takers(at). Let n_{R_1}, n_{R_0} denote number of subjects assigned to treatment ($R = 1$) and control($R = 0$). Let $n_{R_1,at}, n_{R_1,nt}, n_{R_1,c}$ denote number of always takers, never takers, and compliers that are assigned to treatment group.$n_{R_1,at} + n_{R_1,nt} + n_{R_1,c} = n_{R_1}$. Let $n_{R_0,at}, n_{R_0,nt}, n_{R_0,c}$ denote number of always takers, never takers, and compliers, who are assigned to control group.$n_{R_0,at} + n_{R_0,nt} + n_{R_0,c} = n_{R_0}$. Therefore, the likelihood can be rewritten as:

$$\begin{aligned} L(y) = & \prod_{i\in\{R=1,at\}}^{n_{R_1,at}} [(\alpha^*/K_1)(y_i/K_1)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_1)^{\alpha^*})] \\ & \times \prod_{i\in\{R=1,c\}}^{n_{R_1,c}} [(\alpha^*/K_1)(y_i/K_1)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_1)^{\alpha^*})] \\ & \times \prod_{i\in\{R=1,nt\}}^{n_{R_1,nt}} [(\alpha^*/K_1)(y_i/K_1)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_1)^{\alpha^*})] \\ & \times \prod_{i\in\{R=0,at\}}^{n_{R_0,at}} [(\alpha^*/K_0)(y_i/K_0)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_0)^{\alpha^*})] \\ & \times \prod_{i\in\{R=0,c\}}^{n_{R_0,c}} [(\alpha^*/K_0)(y_i/K_0)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_0)^{\alpha^*})] \\ & \times \prod_{i\in\{R=0,nt\}}^{n_{R_0,nt}} [(\alpha^*/K_0)(y_i/K_0)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_0)^{\alpha^*})] \end{aligned}$$

Next, the log likelihood function is:

$$
\begin{aligned}
l(y) = & \sum_{i\in\{R=1,at\}}^{n_{R_1,at}} \delta_i\{log(\alpha^*) - log(K_1)) + (\alpha^* - 1)(log(y_i) - log(K_1))\} \\
& + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} \delta_i\{log(\alpha^*) - log(K_1)) + (\alpha^* - 1)(log(y_i) - log(K_1))\} \\
& + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} \delta_i\{log(\alpha^*) - log(K_1)) + (\alpha^* - 1)(log(y_i) - log(K_1))\} \\
& + \sum_{i\in\{R=1\}}^{n_{R_1}} -(y_i/K_1)^{\alpha^*} \\
& + \sum_{i\in\{R=0,at\}}^{n_{R_0,at}} \delta_i\{log(\alpha^*) - log(K_0)) + (\alpha^* - 1)(log(y_i) - log(K_0))\} \\
& + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} \delta_i\{log(\alpha^*) - log(K_0)) + (\alpha^* - 1)(log(y_i) - log(K_0))\} \\
& + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} \delta_i\{log(\alpha^*) - log(K_0)) + (\alpha^* - 1)(log(y_i) - log(K_0))\} \\
& + \sum_{i\in\{R=0\}}^{n_{R_0}} -(y_i/K_0)^{\alpha^*}
\end{aligned}
$$

To derive the M.L.E of K_0, K_1, take the first derivative of $l(y)$ with respect to K_0, K_1 and set score equation to 0, we have

$$
\hat{K_0} = [\frac{\sum_{i\in\{R=0\}}^{n_{R_0}} y_i^{\alpha^*}}{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i}]^{1/\alpha^*} \tag{C.1}
$$

and,

$$
\hat{K_1} = [\frac{\sum_{i\in\{R=1\}}^{n_{R_1}} y_i^{\alpha^*}}{\sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i}]^{1/\alpha^*} \tag{C.2}
$$

To derive the M.L.E of α^*, take the first derivative of $l(y)$ with respect to α^* and set score equation

to 0 and replace K_1, K_0 with the expressions (C.1) and (C.2), we have

$$
\begin{aligned}
0 &= \frac{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i}{\alpha^*} + \sum_{i\in\{R=0,at\}}^{n_{R_0,at}} \delta_i log(y_i) + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} \delta_i log(y_i) + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} \delta_i log(y_i) \\
&- \{ \sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i \} \frac{\sum_{i\in\{R=0\}}^{n_{R_0}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{R=0\}}^{n_{R_0}} y_i^{\alpha^*}} \\
&+ \frac{\sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i}{\alpha^*} + \sum_{i\in\{R=1,at\}}^{n_{R_1,at}} \delta_i log(y_i) + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} \delta_i log(y_i) + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} \delta_i log(y_i) \\
&- \{ \sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i \} \frac{\sum_{i\in\{R=1\}}^{n_{R_1}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{R=1\}}^{n_{R_1}} y_i^{\alpha^*}} \\
&= \frac{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i}{\alpha^*} + \sum_{i\in\{R=0,at\}}^{n_{R_0,at}} \delta_i log(y_i) + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} \delta_i log(y_i) + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} \delta_i log(y_i) \\
&- \{ \sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i \} \frac{\sum_{i\in\{R=0,at\}}^{n_{R_0,at}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{R=0,at\}}^{n_{R_0,at}} y_i^{\alpha^*} + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} y_i^{\alpha^*} + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} y_i^{\alpha^*}} \\
&+ \frac{\sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i}{\alpha^*} + \sum_{i\in\{R=1,at\}}^{n_{R_1,at}} log(y_i) + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} log(y_i) + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} log(y_i) \\
&- \{ \sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i \} \frac{\sum_{i\in\{R=1,at\}}^{n_{R_1,at}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{R=1,at\}}^{n_{R_1,at}} y_i^{\alpha^*} + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} y_i^{\alpha^*} + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} y_i^{\alpha^*}}
\end{aligned}
$$

M.L.E $\hat{\alpha}^*$ is the solution to the above equation. Next, divide both sides by total number of subject n, we have

$$
\begin{aligned}
0 &= \frac{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i / n_{R_0}}{\alpha^*} \frac{n_{R_0}}{n} + \{ \sum_{i\in\{R=0,at\}}^{n_{R_0,at}} \delta_i log(y_i) + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} \delta_i log(y_i) + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} \delta_i log(y_i) \} / n_{R_0} \frac{n_{R_0}}{n} \\
&- \frac{n_{R_0}}{n} \{ \sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i / n_{R_0} \} \frac{\{\sum_{i\in\{R=0,at\}}^{n_{R_0,at}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} y_i^{\alpha^*} log(y_i)\} / n_{R_0}}{\{\sum_{i\in\{R=0,at\}}^{n_{R_0,at}} y_i^{\alpha^*} + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} y_i^{\alpha^*} + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} y_i^{\alpha^*}\} / n_{R_0}} \\
&+ \frac{\sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i / n_{R_1}}{\alpha^*} \frac{n_{R_1}}{n} + \{ \sum_{i\in\{R=1,at\}}^{n_{R_1,at}} \delta_i log(y_i) + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} \delta_i log(y_i) + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} \delta_i log(y_i) \} / n_{R_1} \frac{n_{R_1}}{n} \\
&- \frac{n_{R_1}}{n} \{ \sum_{i\in\{R=1\}}^{n_{R_1}} \delta_i / n_{R_1} \} \frac{\{\sum_{i\in\{R=1,at\}}^{n_{R_1,at}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} y_i^{\alpha^*} log(y_i)\} / n_{R_1}}{\{\sum_{i\in\{R=1,at\}}^{n_{R_1,at}} y_i^{\alpha^*} + \sum_{i\in\{R=1,c\}}^{n_{R_1,c}} y_i^{\alpha^*} + \sum_{i\in\{R=1,nt\}}^{n_{R_1,nt}} y_i^{\alpha^*}\} / n_{R_1}}
\end{aligned}
$$

As $n_{R_1}, n_{R_0}, n_{R_1,at}, n_{R_1,nt}, n_{R_1,c}, n_{R_0,at}, n_{R_0,nt}, n_{R_0,c} \to \infty$, the score equation converges to the

following:

$$
\begin{aligned}
0 = {} & P(\delta = 1|R = 0)P(R = 0)/\alpha^* \\
& + P(R = 0)\{P(AT)E(\delta log(Y)|at, R = 0) + P(C)E(\delta log(Y)|c, R = 0) + P(NT)E(\delta log(Y)|nt, R = 0)\} \\
& - P(\delta = 1|R = 0)P(R = 0) \\
& \times \frac{\{P(AT)E(Y^{\alpha^*} log(Y)|at, R = 0) + P(C)E(Y^{\alpha^*} log(Y)|c, R = 0) + P(NT)E(Y^{\alpha^*} log(Y)|nt, R = 0)\}}{\{P(AT)E(Y^{\alpha^*}|at, R = 0) + P(C)E(Y^{\alpha^*}|c, R = 0)) + P(NT)E(Y^{\alpha^*}|nt, R = 0)\}} \\
& + P(\delta = 1|R = 1)P(R = 1)/\alpha^* \\
& + P(R = 1)\{P(AT)E(\delta log(Y)|at, R = 1) + P(C)E(\delta log(Y)|c, R = 1) + P(NT)E(\delta log(Y)|nt, R = 1)\} \\
& - P(\delta = 1|R = 1)P(R = 1) \\
& \times \frac{\{P(AT)E(Y^{\alpha^*} log(Y)|at, R = 1) + P(C)E(Y^{\alpha^*} log(Y)|c, R = 1) + P(NT)E(Y^{\alpha^*} log(Y)|nt, R = 1)\}}{\{P(AT)E(Y^{\alpha^*}|at, R = 1) + P(C)E(Y^{\alpha^*}|c, R = 1)) + P(NT)E(Y^{\alpha^*}|nt, R = 1)\}}
\end{aligned}
\tag{C.3}
$$

Use the results from Appendix B, we can derive the following:

$$
\begin{aligned}
E(\delta log(Y)|at, R=0) &= P(\delta=1|at, R=0)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha})) \\
&= \frac{1}{1+\frac{\theta_{at}^{1\ \alpha}}{\lambda^{\alpha}}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha}) \\
E(\delta log(Y)|nt, R=0) &= P(\delta=1|nt, R=0)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha})) \\
&= \frac{1}{1+\frac{\theta_{nt}^{0\ \alpha}}{\lambda^{\alpha}}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha})) \\
E(\delta log(Y)|c, R=0) &= P(\delta=1|c, R=0)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{c}^{0\alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha})) \\
&= \frac{1}{1+\frac{\theta_{c}^{0\alpha}}{\lambda^{\alpha}}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{c}^{0\alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha})) \\
E(Y^{\alpha^*} log(Y)|at, R=0) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*} log(Y)|nt, R=0) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*} log(Y)|c, R=0) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{c}^{0\alpha}} + \frac{1}{\lambda^{\alpha}}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{c}^{0\alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*}|at, R=0) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha} \\
E(Y^{\alpha^*}|nt, R=0) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha} \\
E(Y^{\alpha^*}|c, R=0) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{c}^{0\alpha}} + \frac{1}{\lambda^{\alpha}}]^{-\alpha^*/\alpha} \\
P(\delta=1|R=0) &= P_{at}\frac{1}{1+\frac{\theta_{at}^{1\ \alpha}}{\lambda^{\alpha}}} + P_{nt}\frac{1}{1+\frac{\theta_{nt}^{0\ \alpha}}{\lambda^{\alpha}}} + P_{c}\frac{1}{1+\frac{\theta_{c}^{0\alpha}}{\lambda^{\alpha}}}
\end{aligned}
$$

and,

$$
\begin{aligned}
E(\delta log(Y)|at, R=1) &= P(\delta=1|at, R=1)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
&= \frac{1}{1+\frac{\theta_{at}^{1\ \alpha}}{\lambda^\alpha}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
E(\delta log(Y)|nt, R=1) &= P(\delta=1|nt, R=1)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))) \\
&= \frac{1}{1+\frac{\theta_{nt}^{0\ \alpha}}{\lambda^\alpha}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
E(\delta log(Y)|c, R=1) &= P(\delta=1|c, R=1)(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{c}^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
&= \frac{1}{1+\frac{\theta_{c}^{1\alpha}}{\lambda^\alpha}}(\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{c}^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
E(Y^{\alpha^*} log(Y)|at, R=1) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*} log(Y)|nt, R=1) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*} log(Y)|c, R=1) &= (\psi(\frac{\alpha^*}{\alpha}+1) + \alpha log([\frac{1}{\theta_{c}^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{c}^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*}|at, R=1) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha} \\
E(Y^{\alpha^*}|nt, R=1) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha} \\
E(Y^{\alpha^*}|c, R=1) &= \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{c}^{1\alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha} \\
P(\delta=1|R=1) &= P_{at}\frac{1}{1+\frac{\theta_{at}^{1\ \alpha}}{\lambda^\alpha}} + P_{nt}\frac{1}{1+\frac{\theta_{nt}^{0\ \alpha}}{\lambda^\alpha}} + P_c\frac{1}{1+\frac{\theta_{c}^{1\alpha}}{\lambda^\alpha}}
\end{aligned}
$$

Let $\widetilde{\alpha^*}$ be the solution to the equation (C.3). By the consistency of M.L.E, Thus, we have $\hat{\alpha}^* \xrightarrow{P} \widetilde{\alpha^*}$

Next, substitute $\hat{\alpha}^*$ into equation (C.1)

$$
\begin{aligned}
\hat{K_0} &= [\frac{\sum_{i\in\{R=0\}}^{n_{R_0}} y_i^{\hat{\alpha}^*}}{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i}]^{1/\hat{\alpha}^*} \\
&= [\frac{n_{R_0}}{\sum_{i\in\{R=0\}}^{n_{R_0}} \delta_i}\{\sum_{i\in\{R=0,at\}}^{n_{R_0,at}} y_i^{\hat{\alpha}^*} + \sum_{i\in\{R=0,nt\}}^{n_{R_0,nt}} y_i^{\hat{\alpha}^*} + \sum_{i\in\{R=0,c\}}^{n_{R_0,c}} y_i^{\hat{\alpha}^*}\}/n_{R_0}]^{1/\hat{\alpha}^*}
\end{aligned}
$$

Asymptotically, it converges to

$$\hat{K_0} \to [\frac{1}{P(\delta=1|R=0)}\{P_{at}E(Y_{at,0}^{\widetilde{\alpha^*}}) + P_{nt}E(Y_{nt,0}^{\widetilde{\alpha^*}}) + P_cE(Y_{c,0}^{\widetilde{\alpha^*}})\}]^{1/\widetilde{\alpha^*}}$$
$$= [\frac{1}{P(\delta=1|R=0)}\times$$
$$\{P_{at}\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + P_{nt}\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + P_c\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}]^{1/\widetilde{\alpha^*}}$$

Similarly, $\hat{K_1}$ converges to

$$\hat{K_1} \to [\frac{1}{P(\delta=1|R=1)}\times$$
$$\{P_{at}\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{at}^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + P_{nt}\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha} + P_c\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\}]^{1/\widetilde{\alpha^*}}$$

Appendix D: Derivation of probability limits of M.L.E of α, K_0, K_1, K_2 for 2SRI

Under the no AT assumption, we can find an expression for λ_1 as follows. The first stage regression can be re-expressed as following:

$$E(Z|R) = \rho_{AT} + \rho_C R$$
$$E = Z - E(Z|R)$$
$$= Z - \rho_A - \rho_C R$$

Note that Z, E and Z, R are one-to-one correspondence. Knowing Z, E will let us know Z, R and vice versa. Under no always taker assumption, we observe three subgroups 1)$Z = 1, R = 1$. Only compliers in this group; 2) $Z = 0, R = 1$, Only never takers in this group; 3)$Z = 0, R = 0$, both never takers and compliers in this group. There are no patients that are assigned to control but still takes on active treatment ($Z = 1, R = 0$). For the 3 subgroups, essentially we are fitting 3 Weibull distributions with the same shape parameter α^* and 3 different shape parameter K_0, K_1, K_2 with Weibull regression model: $logh(t) = \lambda_0 + \lambda_1 Z + \lambda_2 E$

The likelihood function is:

$$
\begin{aligned}
L(y) = & \prod_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} [(\alpha^*/K_0)(y_i/K_0)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_0)^{\alpha^*})] \\
& \times \prod_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} [(\alpha^*/K_1)(y_i/K_1)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_1)^{\alpha^*})] \\
& \times \prod_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} [(\alpha^*/K_2)(y_i/K_2)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_2)^{\alpha^*})] \\
& \times \prod_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} [(\alpha^*/K_2)(y_i/K_2)^{\alpha^*-1}]^{\delta_i}[exp(-(y_i/K_2)^{\alpha^*})]
\end{aligned}
$$

The log likelihood is:

$$
\begin{aligned}
l(y) = & \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i\{log(\alpha^*) - log(K_0)) + (\alpha^*-1)(log(y_i) - log(K_0))\} + \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} -(y_i/K_0)^{\alpha^*} \\
& + \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i\{log(\alpha^*) - log(K_1)) + (\alpha^*-1)(log(y_i) - log(K_1))\} + \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} -(y_i/K_1)^{\alpha^*} \\
& + \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i\{log(\alpha^*) - log(K_2)) + (\alpha^*-1)(log(y_i) - log(K_2))\} + \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} -(y_i/K_2)^{\alpha^*} \\
& + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i\{log(\alpha^*) - log(K_2)) + (\alpha^*-1)(log(y_i) - log(K_2))\} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} -(y_i/K_2)^{\alpha^*}
\end{aligned}
$$

Take the first derivative of $l(y)$ with respective to K_0, K_1, K_2 respectively and set score equation to 0, then we have

$$
\hat{K_0} = [\frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} y_i^{\alpha^*}}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i}]^{1/\alpha^*} \tag{D.1}
$$

$$
\hat{K_1} = [\frac{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} y_i^{\alpha^*}}{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i}]^{1/\alpha^*} \tag{D.2}
$$

$$\hat{K_2} = [\frac{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*}}{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i}]^{1/\alpha^*} \tag{D.3}$$

Take the first derivative of $l(y)$ with respective to α^* and replace K_0, K_1, K_2 with expression (D.1) , (D.2) , (D.3), then we have:

$$\begin{aligned}
\frac{dlog(L(y))}{d\alpha} &= \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i\{\frac{1}{\alpha^*} + log(y_i)\} - \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i \frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*}} \\
&+ \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i\{\frac{1}{\alpha^*} + log(y_i)\} - \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i \frac{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*}} \\
&+ \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i\{\frac{1}{\alpha^*} + log(y_i)\} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i\{\frac{1}{\alpha^*} + log(y_i)\} \\
&- (\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i) \frac{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*}} \\
&= \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i log(y_i) - \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i \frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*}} \\
&+ \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i log(y_i) - \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i \frac{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*}} \\
&+ \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i log(y_i) + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i log(y_i) \\
&- (\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i) \frac{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*} log(y_i)}{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*}} \\
&= 0
\end{aligned}$$

M.L.E $\hat{\alpha}^*$ is the solution to the above score equation. Next, divide the equation by total sample size n,

$$
\begin{aligned}
0 = & \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i \frac{1}{\alpha^*}/n_{Z1,R1,c} \times \frac{n_{Z1,R1,c}}{n} + \sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i log(y_i)/n_{Z1,R1,c} \times \frac{n_{Z1,R1,c}}{n} \\
& - \Big(\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} \delta_i/n_{Z1,R1,c} \times \frac{n_{Z1,R1,c}}{n} \Big) \frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*} log(y_i)/n_{Z1,R1,c}}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} (y_i)^{\alpha^*}/n_{Z1,R1,c}} \\
& + \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i \frac{1}{\alpha^*}/n_{Z0,R1,nt} \times \frac{n_{Z0,R1,nt}}{n} + \sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i log(y_i)/n_{Z0,R1,nt} \times \frac{n_{Z0,R1,nt}}{n} \\
& - \Big(\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i/n_{Z0,R1,nt} \times \frac{n_{Z0,R1,nt}}{n} \Big) \frac{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*} log(y_i)/n_{Z0,R1,nt}}{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} (y_i)^{\alpha^*}/n_{Z0,R1,nt}} \\
& + \Big\{ \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i log(y_i) \Big\}/n_{Z0,R0,nt} \times \frac{n_{Z0,R0,nt}}{n} \\
& + \Big\{ \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i \frac{1}{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i log(y_i) \Big\}/n_{Z0,R0,c} \times \frac{n_{Z0,R0,c}}{n} \\
& - \Big(\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i \Big)/n_{Z0,R0} \times \frac{n_{Z0,R0}}{n} \\
& \times \frac{\{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} log(y_i) + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*} log(y_i)\}/n_{Z0,R0}}{\{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\alpha^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\alpha^*}\}/n_{Z0,R0}}
\end{aligned}
$$

As sample sizes in each principal strata $\to \infty$, the score equation will converge to:

$$
\begin{aligned}
0 = & \frac{1}{\alpha^*} P(\delta=1|Z=1,R=1)P(Z=1,R=1) + E(\delta log(y)|Z=1,R=1)P(Z=1,R=1) \\
& - P(\delta=1|Z=1,R=1)P(Z=1,R=1)\frac{E(Y^{\alpha^*} log(Y)|Z=1,R=1))}{E(Y^{\alpha^*}|Z=1,R=1)} \\
& + \frac{1}{\alpha^*} P(\delta=1|Z=0,R=1)P(Z=0,R=1) + E(\delta log(y)|Z=0,R=1)P(Z=0,R=1) \\
& - P(\delta=1|Z=0,R=1)P(Z=0,R=1)\frac{E(Y^{\alpha^*} log(Y)|Z=0,R=1))}{E(Y^{\alpha^*}|Z=0,R=1)} \\
& + \frac{1}{\alpha^*} P(\delta=1|Z=0,R=0,nt)P(Z=0,R=0,nt) + E(\delta log(y)|Z=0,R=0,nt)P(Z=0,R=0,nt) \\
& + \frac{1}{\alpha^*} P(\delta=1|Z=0,R=0,c)P(Z=0,R=0,c) + E(\delta log(y)|Z=0,R=0,c)P(Z=0,R=0,c) \\
& - \Big(\frac{P_{nt}}{P_{nt}+P_c} P(\delta=1|Z=0,R=0,nt) + \frac{P_c}{P_{nt}+P_c} P(\delta=1|Z=0,R=0,c)\Big) \\
& \times (P(Z=0,R=0,nt) + P(Z=0,R=0,c)) \\
& \times \frac{P(nt)E(Y^{\alpha^*} log(Y)|Z=0,R=0,nt) + P(c)E(Y^{\alpha^*} log(Y)|Z=0,R=0,c)}{P(nt)E(Y^{\alpha^*}|Z=0,R=0,nt) + P(c)E(Y^{\alpha^*}|Z=0,R=0,c)}
\end{aligned}
\tag{D.4}
$$

where,

$$
\begin{aligned}
P(\delta = 1|Z = 1, R = 1) &= \frac{1}{1 + (\frac{\theta_c^1}{\lambda})^\alpha} \\
P(Z = 1, R = 1) &= P(C, R = 1) \\
&= P_c P(R = 1) \\
E(\delta log(y)|Z = 1, R = 1) &= \frac{1}{1 + \frac{\theta_c^{1\,\alpha}}{\lambda^\alpha}} (\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_c^{1\,\alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha})) \\
E(Y^{\alpha^*} log(Y)|Z = 1, R = 1) &= (\psi(\frac{\alpha^*}{\alpha} + 1) + \alpha log([\frac{1}{\theta_c^{1\,\alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha} + 1)[\frac{1}{\theta_c^{1\,\alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha} \\
E(Y^{\alpha^*}|Z = 1, R = 1) &= \Gamma(\frac{\alpha^*}{\alpha} + 1)[\frac{1}{\theta_c^{1\,\alpha}} + \frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha} \\
P(\delta = 1|Z = 0, R = 1) &= \frac{1}{1 + (\frac{\theta_{nt}^0}{\lambda})^\alpha} \\
P(Z = 0, R = 1) &= P(nt, R = 1) \\
&= P_{nt} P(R = 1) \\
E(\delta log(y)|Z = 0, R = 1) &= \frac{1}{1 + \frac{\theta_{nt}^{0\ \alpha}}{\lambda^\alpha}} (\frac{-\gamma}{\alpha} + log([\frac{1}{\theta_{nt}^{0\ \alpha}} + \frac{1}{\lambda^\alpha}]^{-1/\alpha}))
\end{aligned}
$$

$$E(Y^{\alpha^*}log(Y)|Z=0,R=1) = (\psi(\frac{\alpha^*}{\alpha}+1)+\alpha log([\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha}$$

$$E(Y^{\alpha^*}|Z=0,R=1) = \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}$$

$$P(\delta=1|Z=0,R=0,nt) = \frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}$$

$$P(Z=0,R=0,nt) = P(nt,R=0)$$

$$= P_{nt}P(R=0)$$

$$E(\delta log(y)|Z=0,R=0,nt) = \frac{1}{1+\frac{\theta_{nt}^{0\ \alpha}}{\lambda^\alpha}}(\frac{-\gamma}{\alpha}+log([\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-1/\alpha}))$$

$$E(Y^{\alpha^*}log(Y)|Z=0,R=0,nt) = (\psi(\frac{\alpha^*}{\alpha}+1)+\alpha log([\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha}$$

$$E(Y^{\alpha^*}|Z=0,R=0,nt) = \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}$$

$$P(\delta=1|Z=0,R=0,c) = \frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}$$

$$P(Z=0,R=0,c) = P(c,R=0)$$

$$= P_cP(R=0)$$

$$E(\delta log(y)|Z=0,R=0,c) = \frac{1}{1+\frac{\theta_c^{0\alpha}}{\lambda^\alpha}}(\frac{-\gamma}{\alpha}+log([\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-1/\alpha}))$$

$$E(Y^{\alpha^*}log(Y)|Z=0,R=0,c) = (\psi(\frac{\alpha^*}{\alpha}+1)+\alpha log([\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-1/\alpha}))\Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}\frac{1}{\alpha}$$

$$E(Y^{\alpha^*}|Z=0,R=0,c) = \Gamma(\frac{\alpha^*}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\alpha^*/\alpha}$$

$\widetilde{\alpha^*}$ is the solution to the equation (D.4). Thus, $\hat{\alpha}^* \to \widetilde{\alpha^*}$. Probability limits of M.L.E of K_0 can be derived as following:

$$\hat{K_0} = [\frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} y_i^{\hat{\alpha}^*}}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}}\delta_i}]^{1/\hat{\alpha}^*}$$

$$= [\frac{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}} y_i^{\hat{\alpha}^*}/n_{Z1,R1,c}}{\sum_{i\in\{Z=1,R=1,c\}}^{n_{Z1,R1,c}}\delta_i/n_{Z1,R1,c}}]^{1/\hat{\alpha}^*}$$

$$\to [\frac{E(y_i^{\widetilde{\alpha^*}}|Z=1,R=1,c)}{P(\delta=1|Z=1,R=1,c)}]^{1/\widetilde{\alpha^*}}$$

$$= [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{1\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_c^1}{\lambda})^\alpha}}]^{1/\widetilde{\alpha^*}}$$

Similarly, for K_1, K_2,

$$\hat{K_1} = [\frac{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} y_i^{\hat{\alpha}^*}}{\sum_{i\in\{Z=0,R=1,nt\}}^{n_{Z0,R1,nt}} \delta_i}]^{1/\hat{\alpha}^*}$$
$$\rightarrow [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}}]^{1/\widetilde{\alpha^*}}$$

$$\hat{K_2} = [\frac{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} y_i^{\hat{\alpha}^*} + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} y_i^{\hat{\alpha}^*}}{\sum_{i\in\{Z=0,R=0,nt\}}^{n_{Z0,R0,nt}} \delta_i + \sum_{i\in\{Z=0,R=0,c\}}^{n_{Z0,R0,c}} \delta_i}]^{1/\hat{\alpha}^*}$$
$$\rightarrow [\frac{\Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_{nt}^{0\ \alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\frac{P_{nt}}{P_{nt}+P_c} + \Gamma(\frac{\widetilde{\alpha^*}}{\alpha}+1)[\frac{1}{\theta_c^{0\alpha}}+\frac{1}{\lambda^\alpha}]^{-\widetilde{\alpha^*}/\alpha}\frac{P_c}{P_{nt}+P_c}}{\frac{1}{1+(\frac{\theta_{nt}^0}{\lambda})^\alpha}\frac{P_{nt}}{P_{nt}+P_c} + \frac{1}{1+(\frac{\theta_c^0}{\lambda})^\alpha}\frac{P_c}{P_{nt}+P_c}}]^{1/\widetilde{\alpha^*}}$$

Appendix E: Assumption of the same shape parameter for survival and censoring distributions

In section 2 of the manuscript, we made the assumption that both time to event and censoring time have the same shape parameter so that close form solution could be derived. To evaluate the potential impact on the bias when the time to event and censoring time have two different shape parameters and the assumption is violated, we re-evaluated the scenario in the table 1 with the shape parameter $\alpha = 0.5$. We set the shape parameter of censoring distribution to be 1.2 and compared the differences. We found that the differences in bias of 2SPS between two scenarios ranges from 0.01 to 0.018 (δ varies from -2 to 2). For 2SRI approach, the differences ranges from 0.001 to 0.13. These differences are attributable to the different censoring proportions between two scenarios. The shape of relationship between bias and δ remains approximately unchanged(data not shown). It should be noted that under the assumption of having the same shape parameters for both survival time and censoring time, the maximum likelihood estimator based on the survival likelihood that does not incorporate the assumption of the shape parameters being the same is not fully efficient.

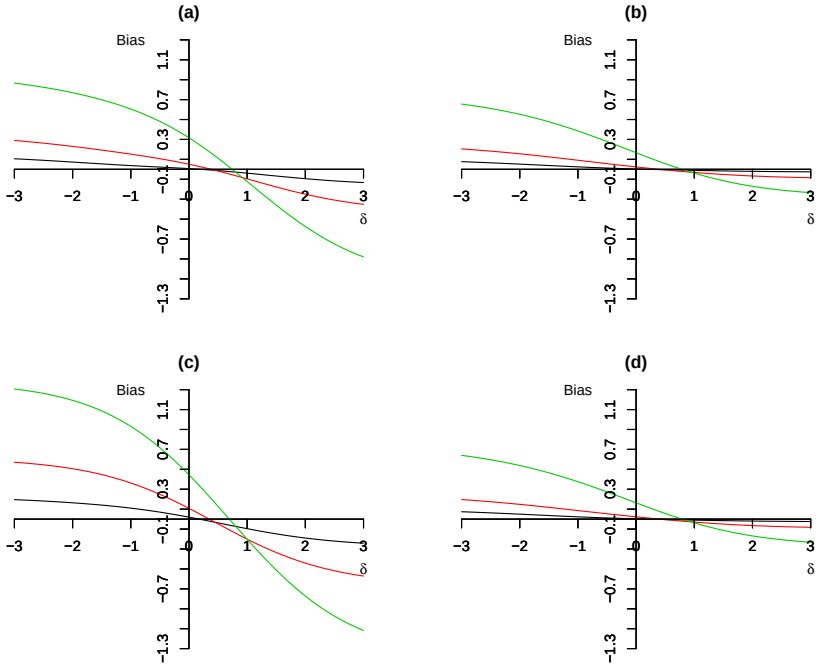


Figure 2.1: Plot of bias against magnitude of unmeasured confounding Δ using 2SPS method:(a)$P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.5$, $\theta_c^1=3.33$, $\theta_c^0=1.67$.(b)$P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.8$, $\theta_c^1=3.33$, $\theta_c^0=1.67$.(c) $P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.5$, $\theta_c^1=33.3$, $\theta_c^0=16.7$. (d) $P(R=1)=0.5$, $\rho_a=0$, $\rho_c=0.8$, $\theta_c^1=3.33$, $\theta_c^0=1.67$. The different colour of solid line corresponds to different shape parameter: black ($\alpha=0.5$),red ($\alpha=1$),and green ($\alpha=2$).

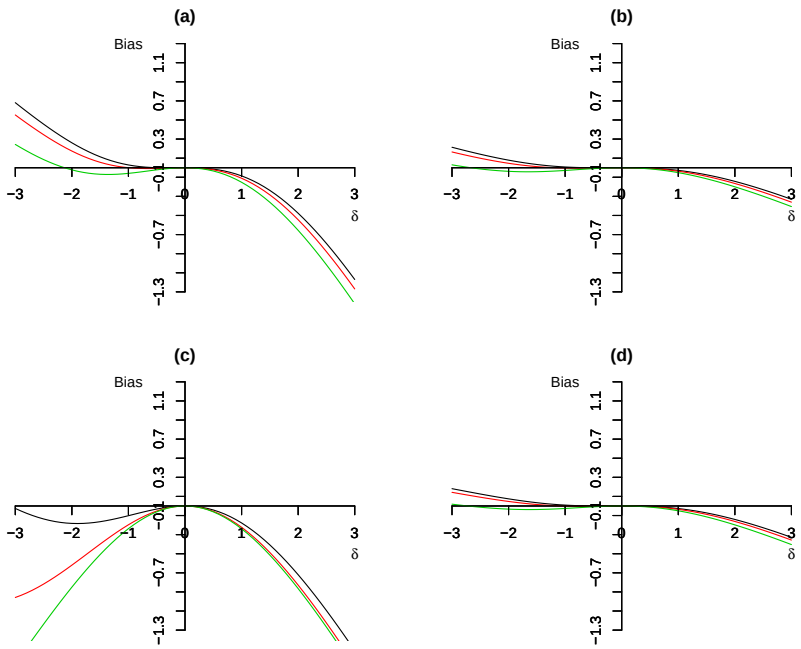


Figure 2.2: Plot of bias against magnitude of unmeasured confounding Δ using 2SRI method:(a)$P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.5$, $\theta_c^1=3.33$, $\theta_c^0=1.67$.(b)$P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.8$, $\theta_c^1=3.33$, $\theta_c^0=1.67$.(c) $P(R=1)=0.8$, $\rho_a=0$, $\rho_c=0.5$, $\theta_c^1=33.3$, $\theta_c^0=16.7$. (d) $P(R=1)=0.5$, $\rho_a=0$, $\rho_c=0.8$, $\theta_c^1=3.33$, $\theta_c^0=1.67$. The different colour of solid line corresponds to different shape parameter: black ($\alpha=0.5$),red ($\alpha=1$),and green ($\alpha=2$).

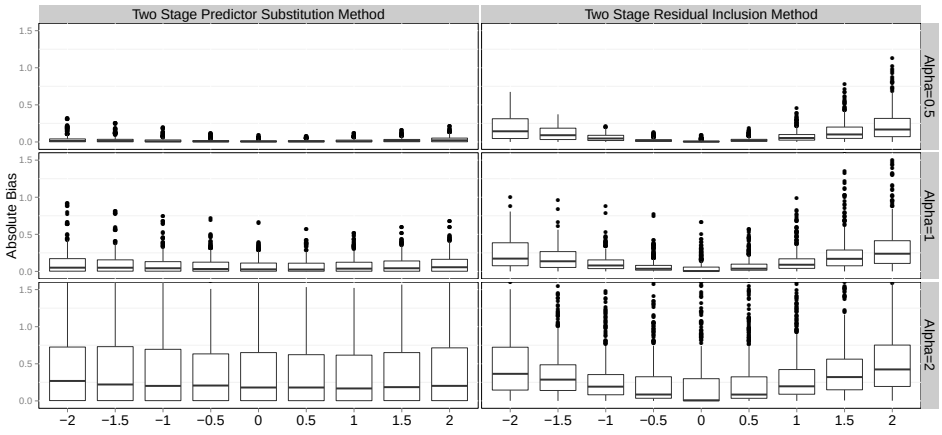


Figure 2.3: Absolute bias in estimating log causal hazard ratio using two stage IV methods (X-axis is the magnitude of confounding Δ , Y-axis is the absolute bias). For 2SRI method or 2SPS method, the biases computed for each of 1458 possible scenarios were grouped by the magnitude of shape parameter α (decreasing hazard for $\alpha=0.5$, constant hazard for $\alpha=1$, and increasing hazard for $\alpha=2$) and the magnitude of confounding Δ (larger values represent lager confounding effects and 0 represents no confounding).

Table 2.1: Bias in estimating log causal hazard ratio parameter ($\rho_a = 0, \rho_c = 0.5, \rho_r = 0.8, \theta_c^1 = 3.33, \theta_c^0 = 1.67$)

α	δ	$Bias_{2sps}^{analytic}$	$Bias_{2sps}^{AFT}$	$Bias_{2sps}^{Cox}$	$Bias_{2sri}^{analytic}$	$Bias_{2sri}^{AFT}$	$Bias_{2sri}^{Cox}$
0.5	2	-0.094	-0.093	-0.091	-0.477	-0.476	-0.476
	1.5	-0.067	-0.068	-0.064	-0.238	-0.239	-0.235
	1	-0.039	-0.040	-0.039	-0.086	-0.087	-0.086
	0.5	-0.013	-0.016	-0.012	-0.015	-0.018	-0.014
	0	0.007	0.009	0.007	0.000	0.002	0.000
	-0.5	0.023	0.020	0.026	0.000	-0.003	-0.001
	-1	0.038	0.037	0.051	0.029	0.028	0.029
	-1.5	0.055	0.053	0.075	0.114	0.112	0.108
	-2	0.073	0.074	0.101	0.261	0.263	0.236
1	2	-0.250	-0.253	-0.247	-0.545	-0.550	-0.544
	1.5	-0.177	-0.175	-0.177	-0.285	-0.284	-0.284
	1	-0.096	-0.093	-0.097	-0.110	-0.107	-0.112
	0.5	-0.017	-0.020	-0.018	-0.022	-0.025	-0.023
	0	0.051	0.053	0.055	0.000	0.002	0.000
	-0.5	0.107	0.106	0.116	-0.007	-0.008	-0.009
	-1	0.152	0.153	0.177	0.000	0.000	0.002
	-1.5	0.193	0.191	0.232	0.057	0.053	0.055
	-2	0.230	0.232	0.280	0.175	0.176	0.157
1.5	2	-0.422	-0.423	-0.418	-0.605	-0.607	-0.602
	1.5	-0.285	-0.285	-0.284	-0.326	-0.325	-0.326
	1	-0.132	-0.133	-0.134	-0.133	-0.134	-0.134
	0.5	0.019	0.023	0.021	-0.028	-0.027	-0.029
	0	0.153	0.152	0.159	0.000	-0.004	0.000
	-0.5	0.261	0.266	0.274	-0.015	-0.012	-0.015
	-1	0.345	0.342	0.376	-0.030	-0.033	-0.027
	-1.5	0.412	0.412	0.461	-0.005	-0.008	-0.002
	-2	0.467	0.468	0.531	0.078	0.075	0.068
2	2	-0.574	-0.578	-0.571	-0.656	-0.656	-0.653
	1.5	-0.359	-0.360	-0.357	-0.362	-0.361	-0.359
	1	-0.122	-0.124	-0.122	-0.152	-0.153	-0.152
	0.5	0.111	0.115	0.112	-0.034	-0.032	-0.036
	0	0.317	0.320	0.324	0.000	0.003	0.002
	-0.5	0.481	0.479	0.494	-0.022	-0.026	-0.026
	-1	0.605	0.605	0.636	0.059	-0.059	-0.056
	-1.5	0.698	0.701	0.747	-0.069	-0.069	-0.063
	-2	0.769	0.770	0.833	-0.023	-0.024	-0.026

$Bias_{2sps}^{analytic}$ - bias computed using analytic formula derived for 2SPS method; $Bias_{2sps}^{AFT}$-bias computed via simulation for 2SPS Weibull accelerated failure time model;$Bias_{2sps}^{Cox}$-bias computed via simulation for 2SPS Cox model; $Bias_{2sri}^{analytic}$ - bias computed using analytic formula derived for 2SRI method; $Bias_{2sri}^{AFT}$-bias computed via simulation for 2SRI Weibull accelerated failure time model;$Bias_{2sri}^{Cox}$-bias computed via simulation for 2SRI Cox model;

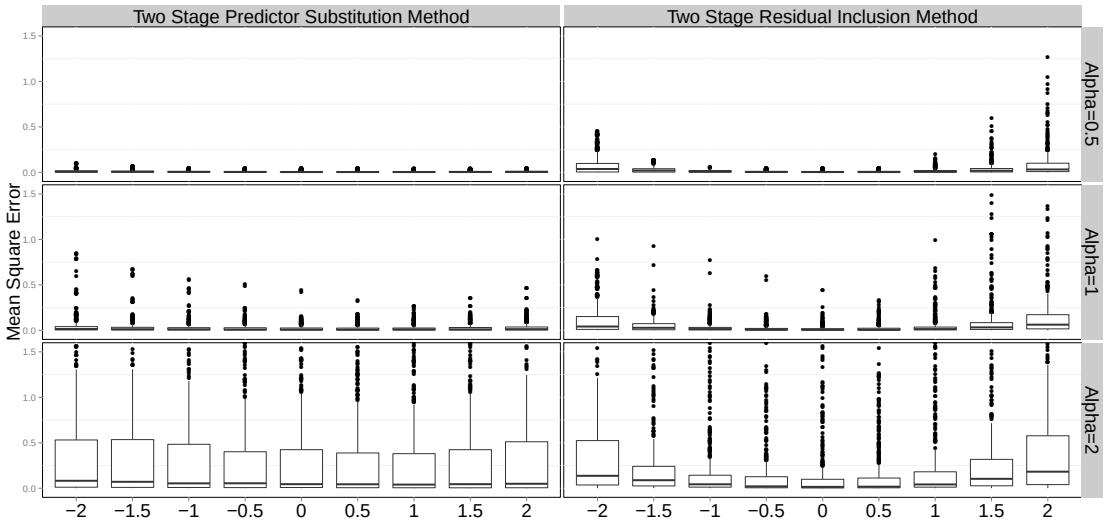


Figure 2.4: Mean square error in estimating log causal hazard ratio using two stage IV methods (X-axis is the magnitude of confounding Δ , Y-axis is the Mean Square Error).For 2SRI method or 2SPS method, the mean square error computed for each of 1458 possible scenarios were grouped by the magnitude of shape parameter α (decreasing hazard for $\alpha = 0.5$, constant hazard for $\alpha = 1$, and increasing hazard for $\alpha = 2$) and the magnitude of confounding Δ (larger values represent lager confounding effects and 0 represents no confounding).

Table 2.2: Bias in estimating causal hazard ratio parameter for prostate cancer study

Outcome	Group	IV_{2sri}	IV_{2sps}
All cause mortality	Total (n=31541)	0.57(0.17-1.06)	0.59(0.19-1.09)
	RCT Cohort (n=12924)	0.96(0.18-5.81)	0.97(0.18-5.94)
	Elderly Cohort (n=14340)	0.74(0.20-1.83)	0.96(0.26-2.35)
	Screen-Detected Cohort (n=4277)	0.34(0.02-2.99)	0.35(0.03-3.22)

CHAPTER 3

A GENERAL FRAMEWORK FOR ASSESSING BIAS IN TWO-STAGE INSTRUMENTAL VARIABLE MODELS

3.1. Introduction

The presence of unmeasured confounding may pose challenges when clinical and health service researchers attempt to estimate the causal effect of an exposure on study outcomes using observational data. This causal effect is often represented by a parameter associated with the exposure variable in a structural model including outcome Y, exposure D, observed covariates $\mathbf{X}$, unobserved covariates $\mathbf{U}$. The conditional causal parameter can be interpreted as the change in an outcome per unit change in the exposure variable while keeping all other measured or unmeasured covariates constant. When $\mathbf{U}$ is omitted from the model, the resulting estimator is generally biased and inconsistent in estimating the true causal parameter, even when observed covariates $\mathbf{X}$ are controlled for. In epidemiology this bias is termed "confounding bias" and it is referred to as "endogeneity" in economics.

Instrumental variables (IVs) are routinely used to account for the unmeasured confounding bias in observational studies. A valid IV must correlate with exposure, and all the effect of IV on outcome must be mediated via exposure (known as “exclusion restriction”). IV methods are well developed in the context of continuous outcomes and linear models. The two stage predictor substitution (2SPS) method, generally known as two stage least squares (2SLS) for continuous outcomes, is one of the most widely used approaches. The fitted value from the first stage linear regression of the exposure on the IV is used to replace the exposure in the second stage linear outcome model. The 2SPS estimator consistently estimates the causal effect of the exposure given observed and unobserved covariates. Two stage residual inclusion (2SRI) method is an equally popular alternative. It includes the residual from the first stage regression as an additional covariate, in conjunction with the exposure, in the second stage regression model. The resulting 2SRI estimator is also a consistent estimator of conditional causal parameter. However, clinical and health service researchers are often interested in evaluating discrete or time to event data. The extensions of the two stage IV approaches to generalized linear models and survival models have been proposed in

a straightforward approach (Terza, Basu, and Rathouz, 2008). In correspondence with each type of outcome, the second stage linear model is simply replaced with the corresponding nonlinear models (i.e, Logistic, Poisson, and Cox proportional hazard models).

However, the consistency of the two stage IV estimators in the context of nonlinear models remains unclear. Terza, Basu, and Rathouz, 2008 constructed a two stage nonlinear modeling framework to account for endogeneity. Within this framework the 2SRI method consistently estimates the causal effect of endogenous exposure variables but the 2SPS estimator is biased and inconsistent. Because of these findings, the 2SRI approach has been advocated as the method of choice in clinical studies involving discrete and survival outcomes (Gore et al., 2010; Hadley et al., 2010; Mortensen et al., 2014; Tan et al., 2012). Wang, 2012 investigated the relationship between pharmacokinetics, a measure of drug exposure, and the risk of adverse events with Poisson regression models and dose level as the IV. The author demonstrated the consistency of the 2SPS and 2SRI estimators analytically and with simulation. IV methods have become increasingly popular in epidemiological literature with Mendalian randomization as well. However, both the 2SRI and 2SPS approaches have been demonstrated to be biased in estimating phenotype-disease log odds ratio, in which dichotomous gene is IV and the exposure is a continuous phenotype (Burgess, 2013; Palmer et al., 2008). Such biases increase with the increasing magnitude of unmeasured confounding.

Confounding also occurs in randomized trials when patients fail to comply with the treatment assignment. The reasons for their compliance with treatment or not may impact the outcome as well. For example, patients with poor prognosis may have worse outcome and they may be more (or less) compliant than patients with good prognosis. Nagelkerke et al., 2000 proposed a 2SRI type estimator for treatment effects in a two-arm randomized trial with non-compliance for nonlinear models. The residual is first estimated from a regression model of the actual treatment received on treatment assignment indicator. Treatment assignment is a perfect IV because it satisfies the core assumptions of IV (Baiocchi, Cheng, and Small, 2014). The estimated residual and treatment indicator are included in the second stage nonlinear model to estimate treatment effects. However, the bias of this 2SRI estimator increases with increasing confounding (Ten Have, Joffe, and Cary, 2003). When treatment effect is heterogeneous and under certain assumptions, Angrist, Imbens, and Rubin, 1996 proved that the 2SLS estimator converges to the local average treatment effect (LATE) among compliers within potential outcomes and principal stratification framework. Complier

is a sub-population of patients that always comply with the treatment assigned. For binary outcome and time to event outcome, Cai, Small, and Ten Have, 2011 and Wan et al., 2015 showed analytically and via simulation that both the 2SPS and 2SRI methods are biased in estimating causal odds ratios and causal hazard ratios among compliers, respectively.

In light of the increasing interests in applying 2SPS and 2SRI methods in clinical and health service research for discrete or time to event data and conflicting conclusions in the current literature, the purpose of this paper is to investigate the consistency of the 2SPS and 2SRI estimators in three commonly used nonlinear models. We propose a new two stage modeling framework that is more relevant to clinical settings. Under this framework, we demonstrate that the bias in 2SPS and 2SRI estimators can be transformed into the problem of omitted variables in non-linear models. We then perform comprehensive simulation to assess the bias of 2SPS and 2SRI. Finally, we analyze infant birth defect data using the 2SPS and 2SRI.

3.2. Notations, Assumptions, and Framework

3.2.1. The nonlinear model framework in current literature

The causal relationship between endogenous variables and discrete or time-to-event outcome variables is often formulated within structural equation modeling frameworkTerza, Basu, and Rathouz, 2008. Such models are widely used in economics. Conditional on observed exogenous variables and unobserved confounding variables, the true nonlinear model that explain the causal effects of covariates on the outcome variable Y is assumed to have the following functional form

$$Y = f(\beta_e^{\mathsf{T}}\mathbf{X_e} + \beta_o^{\mathsf{T}}\mathbf{X_o} + \beta_u^{\mathsf{T}}\mathbf{X_u}) + e, \quad \mathbb{E}(e|\mathbf{X}_e, \mathbf{X_o}, \mathbf{X_u}) = 0. \tag{3.1}$$

where $f(\cdot)$ is a known nonlinear function, $\mathbf{X_e} =< x_{e_1}, x_{e_2}, ..., x_{e_p} >$ is a $p-$vector of endogenous exposure variables and $\beta_e =< \beta_{e_1}, \beta_{e_2}, ..., \beta_{e_p} >$ is a $p-$vector of parameters of interest measuring the causal effects of exposure variables $\mathbf{X_e}$ on outcome Y conditional on observed and unobserved covariates, $\mathbf{X_o} =< x_{o_1}, x_{o_2}, ..., x_{o_k} >$ is a $k-$vector of observed covariates, $\mathbf{X_u} =< x_{u_1}, x_{u_2}, ..., x_{u_p} >$ is a $p-$vector of unobserved covariates that are correlated with endogenous variables $\mathbf{X_e}$. β_o and β_u are $k-$vector and $p-$vector of regression parameters for $\mathbf{X_o}$ and $\mathbf{X_u}$, respectively. e is a random error that is not correlated with $\{\mathbf{X_e}, \mathbf{X_o}, \mathbf{X_u}\}$.

Simply fitting a regression model of Y on observed variables $\mathbf{X_e}$ and $\mathbf{X_o}$ may result in a biased estimate of β_e because of endogeneity. To formalize the relationship between $\mathbf{X_e}$ and $\mathbf{X_u}$ and to reveal how IV can be used to correct endogeneity bias, the following set of nonlinear auxiliary (or reduced form) equations are defined for each pair of an endogenous exposure variable and an unmeasured covariate:

$$x_{e_s} = g_s(\alpha_0^{\mathsf{T}}\mathbf{X_o} + \alpha_1^{\mathsf{T}}\mathbf{R}) + x_{u_s} \quad \text{for} \quad s = 1, 2, 3, ..., p \tag{3.2}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{R} =< r_1, r_2, ..., r_m >$ is a $m-$vector of IVs and $\alpha_1 =< \alpha_{1,1}, \alpha_{1,2}, ..., \alpha_{1,m} >$ is a $m-$vector coefficient vector($m \geq p$). Under the nonlinear model framework specified by equations (3.1) and (3.2), Terza, Basu, and Rathouz, 2008 concluded that 2SRI produces consistent estimates of conditional causal parameters β_e for nonlinear outcome models.

In clinically relevant settings, however, only one endogenous exposure variable but multiple unmeasured covariates are often involved. The modeling framework defined by equations (3.1) and (3.2) requires an inherent assumption that the effects of unmeasured covariates on the outcome variable are proportional to their effects on the endogenous exposure variable. Only under this strict assumption the consistency of the 2SRI estimator can be established. In the next section, we propose a new framework to assess the bias of two stage IV models in estimating the conditional treatment effect of an exposure variable when such an assumption is violated.

3.2.2. A new nonlinear model framework for assessing bias

In this new framework, we let D represent a continuous exposure variable, $\mathbf{X} =< x_1, x_2, ..., x_k >$ be a $k-$vector of observable covariates, and $\mathbf{U} =< u_1, u_2, ..., u_p >$ represent a $p-$vector of unobservable covariates that may be correlated with D. For simplicity and without loss of generality, we assume both $\mathbf{X}$ and $\mathbf{U}$ are all standardized with zero mean and one unit standard deviation. Let $\eta(\cdot)$ denote the linear predictor in a regression model including exposure variable D, measured covariates $\mathbf{X}$, and unmeasured covariates $\mathbf{U}$, and is written as

$$\eta = \beta_0 + \beta_1 D + \beta_{\mathbf{2}}^{\mathbf{T}}\mathbf{X} + \underbrace{\beta_{\mathbf{3}}^{\mathbf{T}}\mathbf{U}}_{\varepsilon_1} \tag{3.3}$$

β_1 is the parameter of interest and represents the treatment effect of exposure variable D on the outcome, conditional on measured covariates $\mathbf{X}$ and unmeasured covariates $\mathbf{U}$. We assume that there is no interaction between the treatment effect and the level of measured and unmeasured covariates. $\beta_{\mathbf{2}}$ and $\beta_{\mathbf{3}}$ are $k-$vector and $p-$vector of regression coefficients for $\mathbf{X}$ and $\mathbf{U}$, respectively. In particular, We define $\beta_3 = k\mathbf{b}_3$, where k is a constant and $\mathbf{b}_3 =< b_{3,1}, b_{3,2}, ..., b_{3,p} >$ is a $p-$dimensional normalized unit vector such that $\|b_3\| = 1$. It should be noted that $k \neq 0$ and thus $\beta_3 \neq \mathbf{0}$ because some of unmeasured covariates must be correlated with outcome, otherwise model (3.3) has no endogeneity problem, and conditional treatment effect, measured by "β_1", can be estimated consistently by controlling for observed covariates $\mathbf{X}$. We also define $\varepsilon_1 = \beta_3^T \mathbf{U} = k \sum_{i=1}^m b_{3,i} u_i$ as an implicit error term for model (3.3).

When outcome Y is binary or count, the nonlinear outcome model (3.1) becomes

$$G(\mathbb{E}(Y|D, \mathbf{X}, \mathbf{U})) = \eta(D, \mathbf{X}, \mathbf{U})$$

where $G(\cdot)$ is a known link function, e.g., logit link for binary data and log link for count data.

When outcome Y is time to event, the nonlinear outcome model (1) can be represented by a log hazard function

$$\log\{\lambda(Y = t|D, \mathbf{X}, \mathbf{U})\} = \eta(D, \mathbf{X}, \mathbf{U})$$

where λ is hazard function. In this case, β_0 in equation (3.3) becomes $\beta_0(t) = \log(\lambda_0(t))$, the log baseline hazard function.

Suppose we have an IV R, the linear treatment model in the form of equation (3.2) is formulated as

$$D = \alpha_0 + \alpha_1 R + \alpha_{\mathbf{2}}^{\mathbf{T}} \mathbf{X} + \underbrace{\alpha_{\mathbf{3}}^{\mathbf{T}} \mathbf{U}}_{\varepsilon_0} \tag{3.4}$$

where $\alpha_{\mathbf{2}}$ and $\alpha_{\mathbf{3}}$ are $k-$vector and $p-$vector of coefficients for $\mathbf{X}$ and $\mathbf{U}$, respectively. We then define $\alpha_3 = l\mathbf{a}_3$, where l is a constant and $\mathbf{a}_3 =< a_{3,1}, a_{3,2}, ..., a_{3,p} >$ is a $p-$normalized vector such that $\|a_3\| = 1$, and the implicit error $\varepsilon_0 = \alpha_3^T \mathbf{U} = k \sum_{i=1}^m a_{3,i} u_i$. The treatment model (3.4) is a standard specification for a continuous endogenous variable in the two stage IV modeling

framework. It should be emphasized that IV R is assumed to be a univariate variable throughout the paper only for simplicity reason but this assumption can be relaxed to accommodate multiple IVs.

The two stage nonlinear modeling framework defined by equations (3.3) and (3.4) possesses the following properties: (P.1) If $b_{3,i} = 0$ and $a_{3,i} \neq 0$ for some $i \in \{1, 2, ..., p\}$, then unobserved variable u_i is an independent predictor or measurement error of the treatment model (3.4). That is, u_i impacts D but not Y; (P.2) If $b_{3,i} \neq 0$ and $a_{3,i} = 0$ for some $i \in \{1, 2, ..., p\}$, then unobserved variable u_i is an independent predictor or measurement error of the outcome model (3.3). u_i impacts Y but not D; (P.3) If $b_{3,i} \neq 0$ and $a_{3,i} \neq 0$ for some $i \in \{1, 2, ..., p\}$, then unobserved variable u_i is a confounder of the association between outcome Y and treatment D.

Within this two stage modeling framework, we require the following standard IV assumptions: (C.1) $R \perp \mathbf{U}|\mathbf{X}$, IV R is independent of $\mathbf{U}$ conditional on $\mathbf{X}$; (C.2) R correlates with D sufficiently conditional on $\{\mathbf{X}, \mathbf{U}\}$. That is, R is a strong IV ; (C.3) $\mathbb{E}(\mathbf{U}|\mathbf{X}) = \mathbb{E}(\mathbf{U}) = 0$, unmeasured covariates $\mathbf{U}$ is mean independent of observed covariates $\mathbf{X}$.

Under (C.1), we can infer that any linear combination of unmeasured covariates $\mathbf{U}$ is independent of R conditional on measured covariates $\mathbf{X}$ so that $\beta_{\mathbf{3}}^{\mathbf{T}}\mathbf{U} \perp R|\mathbf{X}$ in model (3.3) and $\alpha_{\mathbf{3}}^{\mathbf{T}}\mathbf{U} \perp R|\mathbf{X}$ in model (3.4). Thus, we have cov$(\varepsilon_1, R) = 0$ and cov$(\varepsilon_0, R) = 0$, two classic assumptions for IV R under a structural model framework. The conditional independence of IV R and error term ε_1 ensures the exclusion restriction assumption is satisfied. That is, there is no direct effect of IV R on the outcome and all of its effect on the outcome has to go through exposure D. Assumption (C.2) assures that there is a strong correlation between IV R and exposure D and thus the possibility of the bias from a weak IV is excluded. The mean independence assumption specified in (C.3) suggests that observed covariates $\mathbf{X}$ are exogenous in models (3.3) and (3.4) because we have $\mathbb{E}(\beta_3^{\mathsf{T}}\mathbf{U}|\mathbf{X}) = 0$ and $\mathbb{E}(\alpha_3^{\mathsf{T}}\mathbf{U}|\mathbf{X}) = 0$. In next section, we demonstrate that under this new framework, the 2SPS and 2SRI estimators are consistent when outcome is continuous but such consistency does not necessarily hold for nonlinear models, even with a valid IV.

3.3. Bias analysis

In this section, we utilize the two stage modeling framework proposed in section (3.2.2) to assess the potential bias in estimating "conditional" treatment effect, represented by β_1 in model (3.3), using the 2SPS and 2SRI approaches. We first transform bias problems for 2SPS and 2SRI in nonlinear models into bias problems of omitting variables in nonlinear models in section(3.3.1) and section(3.3.2). We investigate the relationship between omitted terms of 2SPS and 2SRI in section (3.3.3). Next, in section(3.3.4) we specifically evaluate the performance of 2SPS and 2SRI in Poisson, Logistic, and Cox proportional hazard models. In section(3.3.5) two metrics are proposed to quantify the relationship between unmeasured covariates and the magnitude of bias for 2SPS and 2SRI, respectively.

3.3.1. Two stage predictor substitution method

First substitute the exposure variable D defined by equation (3.4) into equation (3.3)

$$\begin{aligned}\eta &= \beta_0 + \beta_1(\alpha_0 + \alpha_1 R + \alpha_2^{\mathsf{T}}\mathbf{X} + \alpha_{\mathbf{3}}^{\mathsf{T}}\mathbf{U}) + \beta_2^{\mathsf{T}}\mathbf{X} + \beta_{\mathbf{3}}^{\mathsf{T}}\mathbf{U} \\ &= \beta_0 + \beta_1(\alpha_0 + \alpha_1 R + \alpha_2^{\mathsf{T}}\mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X} + \varepsilon \end{aligned} \tag{3.5}$$

where the implicit error $\varepsilon = (\beta_1\alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}})\mathbf{U}$.

Under assumptions (C.1) and (C.3), taking the expectation of both sides of model (3.4) conditional on R and $\mathbf{X}$ gives

$$\mathbb{E}(D|R, \mathbf{X}) = \alpha_0 + \alpha_1 R + \alpha_2^{\mathsf{T}}\mathbf{X}$$

Then equation (5) becomes

$$\eta = \beta_0 + \beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X} + \varepsilon \tag{3.6}$$

In equation (3.6), $\mathbb{E}(D|R, \mathbf{X})$ is a function of IV R and $\mathbf{X}$, and ε is a linear combination of unmeasured covariates $\mathbf{U}$. Thus, $\mathbb{E}(D|R, \mathbf{X}) \perp \varepsilon|\mathbf{X}$ by assumption (C.1).

2SPS first estimates the predicted value of D using a linear regression model including R and $\mathbf{X}$

only,

$$\hat{D} = \hat{\alpha}_0 + \hat{\alpha}_1 R + \hat{\alpha}_2^{\mathsf{T}} \mathbf{X}$$

This $\hat{D}$ consistently estimate $\mathbb{E}(D|R, X)$ because OLS estimators $\{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2\}$ are unbiased and consistent. Therefore, in the scenario where the outcome Y is continuous, equation (3.6) can be approximated asymptotically by

$$Y = \beta_0^* + \beta_1^* \hat{D} + \beta_2^{*\mathsf{T}} \mathbf{X} + \varepsilon$$

As $n \to \infty$, $\hat{D} \xrightarrow{p} \mathbb{E}(D|R, X)$, and $\beta_1^* \xrightarrow{p} \beta_1$. Conditional on $\mathbf{X}$, $\hat{D}$ is independent of ε asymptotically. The 2SPS estimator $\hat{\beta}_1^* \xrightarrow{p} \beta_1^*$, and so $\hat{\beta}_1^* \xrightarrow{p} \beta_1$. Therefore, 2SPS consistently estimates the causal parameter β_1 when outcome is continuous.

However, when outcome is not continuous and a nonlinear outcome model needs to be fitted in the second stage, the consistency of the 2SPS estimator may fail. For example, when Y is binary or time to event outcome, the second stage outcome model of the 2SPS method, according to equation (3.6), can be written as,

$$G(\mathbb{E}(Y|R, \mathbf{X}, \mathbf{U})) = \beta_0 + \beta_1 \mathbb{E}(D|R, \mathbf{X}) + \beta_2^{\mathsf{T}} \mathbf{X} + \varepsilon \tag{3.7}$$

where $G(\cdot)$ is logit link function, or

$$\text{log}\{\lambda(Y = t|R, \mathbf{X}, \mathbf{U})\} = \beta_0(t) + \beta_1 \mathbb{E}(D|R, \mathbf{X}) + \beta_2^{\mathsf{T}} \mathbf{X} + \varepsilon \tag{3.8}$$

where $\lambda(\cdot)$ is the hazard function and $\beta_0(t) = \text{log}(\lambda_0(t))$ is the baseline log hazard function. But since unmeasured confounders $\mathbf{U}$ are not observed and ε is omitted from models (3.7) and (3.8), one is forced to fit a reduced model

$$G(\mathbb{E}(Y|R, \mathbf{X})) = \tilde{\beta}_0 + \tilde{\beta}_1 \mathbb{E}(D|R, \mathbf{X}) + \tilde{\beta}_2^{\mathsf{T}} \mathbf{X} \tag{3.9}$$

for generalized linear models,or

$$\log\{\lambda(Y=t|R,\mathbf{X})\} = \tilde{\beta}_0(t) + \tilde{\beta}_1\mathbb{E}(D|R,\mathbf{X}) + \tilde{\beta}_2^{\mathsf{T}}\mathbf{X} \tag{3.10}$$

for hazard models. This reduced model can be interpreted as a marginal model with respect to omitted error term ε. The "marginal" causal parameter $\tilde{\beta}_1$ in models (3.9) and (3.10) may differ from the "conditional" parameter β_1 in models (3.7) and (3.8) (Gail, Wieand, and Piantadosi, 1984; Zeger, Liang, and Albert, 1988). This phenomenon is also known as noncollapsibility. Thus, when predicted values $\hat{D}$ replaces $\mathbb{E}(D|R,\mathbf{X})$ in models (3.9) and (3.10), the corresponding models become

$$G(\mathbb{E}(Y|R,\mathbf{X})) = \tilde{\beta}_0^* + \tilde{\beta}_1^*\hat{D} + \tilde{\beta}_2^{*\mathsf{T}}\mathbf{X} \tag{3.11}$$

or,

$$\log\{\lambda(Y=t|R,\mathbf{X})\} = \tilde{\beta}_0^*(t) + \tilde{\beta}_1^*\hat{D} + \tilde{\beta}_2^{*\mathsf{T}}\mathbf{X} \tag{3.12}$$

As $n \to \infty$, the 2SPS estimator $\widehat{\tilde{\beta}_1^*}$ consistently estimate $\tilde{\beta}_1^*$. Also, When $\hat{D} \xrightarrow{p} D$, $\tilde{\beta}_1^* \xrightarrow{p} \tilde{\beta}_1$, and then $\widehat{\tilde{\beta}_1^*} \xrightarrow{p} \tilde{\beta}_1$. However, the 2SPS estimator $\widehat{\tilde{\beta}_1^*}$ may fail to estimate the conditional causal parameter β_1 consistently because the marginal parameter $\tilde{\beta}_1$ could differ from β_1 depending on the collapsibility of the outcome models.

3.3.2. Two stage residual inclusion method

Similarly, under the same two stage framework defined by equations (3.3) and (3.4), the 2SRI estimator could be biased as well. First, we decompose β_3, coefficient vector for unmeasured covariates in equation (3.3), into two orthogonal components: one along with the direction of α_3, coefficient vector of unmeasured covariates in treatment model (3.4), and the other vector orthogonal to α_3 (Figure 2.1),

$$\beta_3 = \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3 + (\beta_3 - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3)$$

Re-express $\beta_3^\intercal \mathbf{U}$ term in equation (3.3) as

$$\beta_3^\intercal \mathbf{U} = \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal \mathbf{U} + (\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal)\mathbf{U} \tag{3.13}$$

Equation (3.13) can be interpreted as a least square projection of $\beta_3^\intercal \mathbf{U}$ onto $\alpha_3^\intercal \mathbf{U}$. $\frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)$ is the regression coefficient. The second term is the residual (details see Appendix A). It is noteworthy that $\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal = \beta_3^\intercal \sin(\theta)$.

Substituting equation (3.13) into equation (3.3), we get

$$\eta = \beta_0 + \beta_1 D + \boldsymbol{\beta_2^\intercal} \mathbf{X} + \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal \mathbf{U} + (\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal)\mathbf{U} \tag{3.14}$$

A simple manipulation of equation (3.4) leads to,

$$\boldsymbol{\alpha_3^\intercal} \mathbf{U} = D - (\alpha_0 + \alpha_1 R + \boldsymbol{\alpha_2^\intercal} \mathbf{X}) = D - \mathbb{E}(D|R, \mathbf{X})$$

which can be interpreted as a "residual" term from the first stage linear treatment model. Therefore, Equation (3.14) can be written as,

$$\eta = \beta_0 + \beta_1 D + \beta_2^\intercal \mathbf{X} + \gamma_1 \delta + \nu \tag{3.15}$$

where

$$\begin{aligned} \delta &= D - \mathbb{E}(D|R, \mathbf{X}) \\ \gamma_1 &= \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta) \\ \nu &= (\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal)\mathbf{U} \end{aligned}$$

In equation (3.15), D is a function of R, $\mathbf{X}$, and $\alpha_3^\intercal \mathbf{U}$, and the implicit error ν is a linear combination of unmeasured confounders $\mathbf{U}$. Given the fact that $\delta = \alpha_3^\intercal \mathbf{U}$, it is easy to show $D \perp \nu | \{\mathbf{X}, \delta\}$ by condition (C.1). Exposure D is exogenous in model (3.15) once observed covariates $\mathbf{X}$ and the residual δ from treatment model are controlled for.

The 2SRI approach first estimates the residual δ using a linear regression model including R and

$\mathbf{X}$ only,

$$\hat{\delta} = D - \hat{D} = D - (\hat{\alpha}_0 + \hat{\alpha}_1 R + \hat{\alpha}_2^{\intercal}\mathbf{X})$$

This $\hat{\delta}$ consistently estimate δ because of consistency of OLS estimators $\{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2\}$. When the outcome Y is continuous, equation (3.15) can be approximated asymptotically by

$$Y = \beta_0^* + \beta_1^* D + \beta_2^* \mathbf{X} + \gamma_1^* \hat{\delta} + \nu \tag{3.16}$$

As $n \to \infty$, $\hat{\delta} \xrightarrow{p} \delta$, and $\beta_1^* \xrightarrow{p} \beta_1$. Conditional on $\mathbf{X}$ and $\hat{\delta}$, the exposure variable D is independent of ν asymptotically in model (3.16). Thus, $\hat{\beta}_1^* \xrightarrow{p} \beta_1^*$, and $\hat{\beta}_1^* \xrightarrow{p} \beta_1$. Therefore, the 2SRI estimator $\hat{\beta}_1^*$ consistently estimate the parameter of interest β_1 when the outcome model is a linear model.

However, like the 2SPS estimator, when the outcome is not continuous and a nonlinear outcome model is fitted in the second stage, the consistency of the 2SRI estimator is also questionable. For example, when Y is binary or time to event outcome, the second stage outcome model of 2SRI approach, according to equation (3.15), can be written as,

$$G(\mathbb{E}(Y|D, R, \mathbf{X}, \mathbf{U})) = \beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu \tag{3.17}$$

where $G(\cdot)$ is logit link function, or

$$\text{log}\{\lambda(Y = t|D, R, \mathbf{X}, \mathbf{U})\} = \beta_0(t) + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu \tag{3.18}$$

where $\lambda(\cdot)$ is the hazard function and $\beta_0(t) = \text{log}(\lambda_0(t))$ is the baseline log hazard function. But since unmeasured confounders $\mathbf{U}$ are not observed and ν term is omitted from models (3.17) and (3.18),the resulting "marginal" models are

$$G(\mathbb{E}(Y|D, R, \mathbf{X})) = \tilde{\beta}_0 + \tilde{\beta}_1 D + \tilde{\beta}_2 \mathbf{X} + \tilde{\gamma}_1 \delta \tag{3.19}$$

, or

$$\text{log}\{\lambda(Y = t|D, R, \mathbf{X})\} = \tilde{\beta}_0(t) + \tilde{\beta}_1 D + \tilde{\beta}_2 \mathbf{X} + \tilde{\gamma}_1 \delta \tag{3.20}$$

The marginal causal parameter of $\tilde{\beta}_1$ in models (3.19) and (3.20) may differ from β_1 in models (3.17) and (3.18) (Gail, Wieand, and Piantadosi, 1984; Zeger, Liang, and Albert, 1988). When the estimated residual $\hat{\delta}$ replaces δ in models (3.19) and (3.20), the corresponding models are

$$G(\mathbb{E}(Y|D,R,\mathbf{X})) = \tilde{\beta}_0^* + \tilde{\beta}_1^* D + \tilde{\beta}_2^* \mathbf{X} + \tilde{\gamma}_1^* \hat{\delta} \tag{3.21}$$

, or

$$\log\{\lambda(Y=t|D,R,\mathbf{X})\} = \tilde{\beta}_0^* + \tilde{\beta}_1^* D + \tilde{\beta}_2^* \mathbf{X} + \tilde{\gamma}_1^* \hat{\delta} \tag{3.22}$$

As $n \to \infty$, $\hat{\delta} \xrightarrow{p} \delta$ and $\tilde{\beta}_1^* \xrightarrow{p} \tilde{\beta}_1$. The 2SRI estimator $\widehat{\tilde{\beta}_1^*}$ consistently estimate $\tilde{\beta}_1^*$, and $\tilde{\beta}_1$, but $\widehat{\tilde{\beta}_1^*}$ may not be a consistent estimator of β_1 because $\tilde{\beta}_1$ could be different from β_1.

3.3.3. Relationship between omitted terms in 2SPS and 2SRI

As discussed in sections (3.3.1) and (3.3.2), the omitted error term ε of 2SPS is $(\beta_1\alpha_3^\intercal + \beta_3^\intercal)\mathbf{U}$ and the omitted error term ν of 2SRI is $(\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal)\mathbf{U}$. The norm of the coefficient for ε has the following relationship with the norm of the coefficient for ν

$$\|\beta_1\alpha_3^\intercal + \beta_3^\intercal\|^2 = \|\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal\|^2 + \|(\beta_1 + \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta))\alpha_3^\intercal\|^2$$

Then, we have

$$\|\beta_1\alpha_3^\intercal + \beta_3^\intercal\| \geq \|\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal\|$$

The details are in Appendix A.

3.3.4. Two stage Poisson,Logistic,and Survival IV Models

As revealed in section (3.3.1) and (3.3.2), 2SPS decomposes the endogenous exposure variable D into two components: 1) the endogenous component $\beta_1\alpha_3^\intercal\mathbf{U}$, which is integrated into the error term; 2) the exogenous component $\mathbb{E}(D|R,\mathbf{X})$, which is exogenous to the error term. On the other hand, 2SRI decomposes the unmeasured component $\beta_3^\intercal\mathbf{U}$ into two orthogonal components: 1) the estimable component $\alpha_3^\intercal\mathbf{U}$, which is the residual from the first stage treatment model; 2) the

composite error term ν. Controlling for the estimable component, IV, and observed covariates, the exposure variable is exogenous to this error term. Thus, omitting these endogenous error terms does not cause any bias in estimating the conditional causal parameter when we apply 2SPS or 2SRI to linear models. However, a "marginal" model, resulting from omitting composite error terms, may have different parameters from the corresponding the "conditional" model for nonlinear models (Zeger, Liang, and Albert, 1988). The term "bias" for two stage IV methods is defined as $\beta_1 - \tilde{\beta}_1$, the difference between conditional causal parameter β_1 and marginal causal parameter $\tilde{\beta}_1$ to which two stage IV estimators converges to when composite error term ε or ν is ignored. Next, we use approaches adopted by Lin, Psaty, and Kronmal, 1998 and Mitra and Heitjan, 2007 to examine the bias of 2SPS and 2SRI in Poisson, logistic, and Cox proportional hazard models.

(1) *Poisson model:* Suppose that Y is a response variable taking integer values $0, 1, 2, ...N$ and $Y|\{D, \mathbf{X}, \mathbf{U}\} \sim Pois(e^{\eta})$, where η is the linear predictor defined by equation (3.3). The marginal form of 2SPS Poisson model is

$$\mathbb{E}(Y|R, \mathbf{X}) = \exp(\beta_0 + \beta_1 \mathbb{E}(D|R, \mathbf{X}) + \beta_2^{\mathsf{T}} \mathbf{X}) c(\mathbf{X})$$

where $c(\mathbf{X}) = \int_{-\infty}^{\infty} \exp(\varepsilon) dF(\varepsilon|\mathbf{X})$, a function of $\mathbf{X}$ only (details in Appendix B). Both the marginal Poisson model and conditional Poisson model have the same parameter β_1.

The marginal model of 2SRI Poisson model is

$$\mathbb{E}(Y|D, R, \mathbf{X}) = \exp(\beta_0 + \beta_1 D + \beta_2^{\mathsf{T}} \mathbf{X} + \gamma_1 \delta) c(\mathbf{X}, \delta)$$

where $c(\mathbf{X}, \delta) = \int_{-\infty}^{\infty} \exp(\nu) dF(\nu|\mathbf{X}, \delta)$. It is a function of $\mathbf{X}$ and δ. The marginal model and conditional model have the same causal parameter β_1 associated with exposure D (details in Appendix B). Therefore, both 2SPS and 2SRI Poisson models yield consistent estimates of conditional treatment effect.

(2) *Logistic outcome model*. Now assume Y is a response variable taking binary values of 0 or 1, and $Y|\{D, \mathbf{X}, \mathbf{U}\} \sim Bernoulli(\frac{e^{\eta}}{1+e^{\eta}})$, where $\eta(\cdot)$ is linear predictor defined by equation (3.3).

The marginal form of 2SRI logistic model is

$$\text{logit}\{\mathbb{P}(Y=1|D,R,\mathbf{X})\} = \beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + g(D,R,\mathbf{X})$$

where

$$g(D,R,\mathbf{X}) = \log\frac{\mathbb{A}}{\mathbb{B}}$$

and,

$$\mathbb{A} = \int_{-\infty}^{\infty} \frac{\exp(\nu)}{1+\exp(\beta_0+\beta_1 D+\beta_2\mathbf{X}+\gamma_1\delta)+\nu)} dF(\nu|\delta,\mathbf{X})$$
$$\mathbb{B} = \int_{-\infty}^{\infty} \frac{1}{1+\exp(\beta_0+\beta_1 D+\beta_2\mathbf{X}+\gamma_1\delta)+\nu)} dF(\nu|\delta,\mathbf{X})$$

The marginal form of the 2SPS logit outcome model (3.7) is

$$\text{logit}\{\mathbb{P}(Y=1|D,R,\mathbf{X})\} = \beta_0 + \beta_1 \mathbb{E}(D|R,\mathbf{X}) + \beta_2 \mathbf{X} + g(R,\mathbf{X})$$

The derived bias term $g(R,\mathbf{X})$ for 2SPS logistic model is

$$g(R,\mathbf{X}) = \log\frac{\mathbb{A}}{\mathbb{B}}$$

where

$$\mathbb{A} = \int_{-\infty}^{\infty} \frac{\exp(\varepsilon)}{1+\exp(\beta_0+\beta_1 \mathbb{E}(D|R,\mathbf{X})+\beta_2\mathbf{X}+\varepsilon)} dF(\varepsilon|\mathbf{X})$$
$$\mathbb{B} = \int_{-\infty}^{\infty} \frac{1}{1+\exp(\beta_0+\beta_1 \mathbb{E}(D|R,\mathbf{X})+\beta_2\mathbf{X}+\varepsilon)} dF(\varepsilon|\mathbf{X})$$

When there is no treatment effect ($\beta_1 = 0$), both 2SRI and 2SPS are consistent. When the treatment effect is not null but coefficients for implicit error terms (ν or ε) are zero, both 2SRI and 2SPS are also consistent. That is, $\beta_1\alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}} = 0$ for 2SPS and $\beta_3^{\mathsf{T}} - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^{\mathsf{T}} = 0$ for 2SRI (details in Appendix C). A simple form of the relationship between marginal and conditional causal parameters can be approximated by the following expressions if we can

assume that $R \perp \mathbf{U}$ without conditional on $\mathbf{X}$ Zeger, Liang, and Albert, 1988,

$$\tilde{\beta}_1 \approx \frac{\beta_1}{\sqrt{1 + c^2 \|\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal\|^2}}$$

for 2SRI, and

$$\tilde{\beta}_1 \approx \frac{\beta_1}{\sqrt{1 + c^2 \|\beta_1\alpha_3^\intercal + \beta_3^\intercal\|^2}}$$

for 2SPS, where $c = 16\sqrt{3}/(15\pi)$. As the magnitudes of the effects of omitted error terms, measured by $\|\beta_3^\intercal - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^\intercal\|$ or $\|\beta_1\alpha_3^\intercal + \beta_3^\intercal\|$ for 2SRI or 2SPS respectively, increases, the values of marginal parameters $\tilde{\beta}_1$ shrink towards the null effect. The bias $\tilde{\beta}_1 - \beta_1$ is a function of treatment effect β_1.

(3) *Cox proportional hazard model*. Let Y be the time to event. Given exposure D, measured covariates $\mathbf{X}$ and unmeasured covariates $\mathbf{U}$, the marginal form of 2SRI Cox proportional hazard model (3.20) can be written as

$$\lambda(t|D, \delta, \mathbf{X}) = \lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta)g(t, D, R, \mathbf{X})$$

where bias term

$$g(t, D, R, \mathbf{X}) = \frac{\int_{-\infty}^{\infty} \exp(\nu)(\exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)dF(\nu|R, \mathbf{X})}{\int_{-\infty}^{\infty} \exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)dF(\nu|R, \mathbf{X})}$$

and the marginal form of 2SPS Cox proportional hazard model (3.8) can be written as

$$\lambda(t|R, \mathbf{X}) = \lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X})g(t, R, \mathbf{X})$$

where bias term

$$g(t, R, \mathbf{X}) = \frac{\int_{-\infty}^{\infty} \exp(\varepsilon)(\exp(-\Lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X} + \varepsilon)dF(\varepsilon|R, \mathbf{X})}{\int_{-\infty}^{\infty} \exp(-\Lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X} + \varepsilon)dF(\varepsilon|R, \mathbf{X})}$$

When there is no treatment effect ($\beta_1 = 0$), both 2SRI and 2SPS are consistent. When the

treatment effect is not null but coefficients for implicit error terms (ν or ε) are zero, both 2SRI and 2SPS are consistent (details in Appendix D).

It is difficult to derive a simple closed form expression of $\tilde{\beta}_1$ in terms of β_1 for Cox model with censoring. If we assume there is no observed covariate and by the results of Lin, Logan, and Henley, 2013, $\tilde{\beta}_1$ is the solution to the following score equation:

$$\begin{aligned} 0 &= \mathbb{S}(\tilde{\beta}_1, \beta_1, \phi) \\ &= \mathbb{E}_{obs}\left[D - \frac{\mathbb{E}_{DU}\left\{ e^{\tilde{\beta}_1 D} e^{-H_0(t) e^{\beta_1 D + \phi \mathbf{U}}} D C(t) \right\}}{\mathbb{E}_{DU}\left\{ e^{\tilde{\beta}_1 D} e^{-H_0(t) e^{\beta_1 D + \phi \mathbf{U}}} C(t) \right\}} \right] \end{aligned}$$

where $\phi = \beta_1 \alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}}$ for 2SPS or $\phi = \beta_3^{\mathsf{T}} - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^{\mathsf{T}}$ for 2SRI. $C(t)$ is censoring distribution. $\mathbb{E}_{obs}$ is the mean over subjects with events. $\mathbb{E}_{DU}$ is the expectation with respect to exposure variable D and unmeasured covariates $\mathbf{U}$.

3.3.5. Dissimilarity metric

The magnitude of coefficients for omitted terms is one important factor influencing the bias (Neuhaus and Jewell, 1993). $\|\beta_1 \alpha_3^{\mathsf{T}} + \beta_3\|$ is the size of coefficient for 2SPS. For 2SRI, the coefficient vector $\beta_3^{\mathsf{T}} - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^{\mathsf{T}}$ is equal to $\beta_3^{\mathsf{T}}\sin(\theta)$. The metric $\|\beta_3^{\mathsf{T}}\|\sin(\theta)$ reveals two sources of biases for 2SRI: the magnitude of effects of unmeasured covariates on outcome, represented by $\|\beta_3^{\mathsf{T}}\|$, and the "dissimilarity" between coefficient vectors α_3 and β_3, measured by $\sin(\theta)$. The $\sin(\cdot)$ function measures the dissimilarity between two coefficient vectors and it satisfies the three criteria for a standard distance measure: i) non-negativity and identity of indiscernibles; ii) symmetry; iii) triangle inequality (Details see Appendix A). Larger values of this metric suggests two vectors are more dissimilar.

3.4. Simulation

3.4.1. Simulation algorithm

As discussed in previous sections, various factors may impact the bias in estimating conditional treatment effect when applying 2SPS and 2SRI in nonlinear models, such as size of treatment effect, magnitude of unmeasured confounding, dissimilarity between the two coefficient vectors of unmeasured covariates, and censoring proportion etc. We design a comprehensive simulation

study, according to the following steps, to assess the impact of these factors:

(I) We set the size of treatment effect at three levels: $\beta_1 = \{0, 0.4, 0.8\}$.

(II) IV $R \sim N(0,1)$, each of four unmeasured variables $u_i \sim N(0,1)$ for $i \in \{1,2,3,4\}$. In treatment model (3.4), the coefficients for IV and unmeasured covariates are determined in such way that the desired strong association between IV and treatment variable are achieved. We use the explained proportion of variation, $\mathbb{R}^2 = \text{var}(\alpha_1 R)/\text{var}(\alpha_1 R + \sum_{i=1}^{4} \alpha_{3,i} u_i)$, to measure the strength of IV because this measure is not variable with changing sample size. We let $\alpha_0 = 1.2$ and $\alpha_1 = 2$. The coefficient vector for unmeasured variables is $\alpha_3 = l \times a_3$, where $a_3 = \{a_{3,1}, a_{3,2}, a_{3,3}, a_{3,4}\}$ is a unit vector. The explained proportion of variation by IV R is $4/(4 + 1 * l^2)$. We fix $l = 1$, $\mathbb{R}^2$ attributable to IV is $\sim 80\%$. In outcome model (3.3), coefficient vector for unmeasured covariates is $\beta_3 = k \times b_3$, where $b_3 = \{b_{3,1}, b_{3,2}, b_{3,3}, b_{3,4}\}$ is a unit vector and k is chosen to be 0.5, 1, or 2, representing low, medium, and high levels of effects of unmeasured covariates on outcome. Each element in vectors a_3 and b_3 are sampled randomly from $\{0, 1, \ldots, 8, 9\}$ first and then normalized. The signs of each coefficient are generated randomly from $\sim Bernoulli(p = 0.5)$.

(III) For each pair of coefficient vectors α_3 and β_3, we simulated a sample of 10000 observations for each type of outcomes (binary, count, and time to event) using equations (3.3) and (3.4). Specifically, count data are generated using Poisson distribution $\sim P(\mu = \exp(\eta))$, where η is the linear predictor defined in equation (3.3). Binary data are generated from $\sim Bernoulli(p = \frac{\exp(\eta)}{1+\exp(\eta)})$. Time to event data are simulated using $\sim Weibull(\alpha, \exp(\eta))$. Censoring time is also generated using $\sim Weibull(\alpha, c)$. Shape parameter α are set at $\{0.5, 1, 1.5\}$, representing decreasing, constant, and increasing hazard scenarios. Value of scale parameter c is chosen to yield $0\%, 45\%$, and 65% censoring rate.

(IV) For each combination of influencing factors, we generate 2000 pairs of coefficient vectors. For each pair, we simulate a data with 10000 observations. We use the 2SRI and 2SPS methods to estimate conditional treatment effect β_1 on each simulated data. The process is repeated for 1000 times and these 1000 estimates of treatment effect are averaged to compute the 2SRI and 2SPS estimators.

3.4.2. Simulation results

Under the null hypothesis of no treatment effect, 2SRI estimators are unbiased in all scenarios for Poisson, logistic and Cox proportional hazard models (Supplementary Figure 3.7,3.8, and 3.9). When treatment effect is not null, the results are mixed. For Poisson models, 2SRI produces consistent estimates of the conditional treatment effect for most scenarios. When two vectors are highly dissimilar and unmeasured effects are high, there are some minor bias(Figure 3.2). For logit models, the bias of 2SRI tends to increase as treatment effect size and dissimilarity metric increases. The magnitude of the increasing trend is magnified by the strength of effect of unmeasured covariate on outcome. Stronger unmeasured effect, which ranges from 0.5 to 1.5, larger increase in bias of 2SRI (Figure 3.3). Figures 3.4 and 3.4 reveals the results for Cox model. The bias of 2SRI increases with increasing dissimilarity. Higher unmeasured effect, larger bias (Figure 3.4). Larger size of treatment effect, larger bias (Figure 3.5). Bias of 2SRI is the highest with increasing hazard function. Censoring provides some protective effect. 2SRI in Censored data is less biased than in non-censored data. 2SPS exibits similar trends with its norm $\|\beta_1\alpha_3 + \beta_3\|$ (Data not shown). 2SRI estimators are less biased and have more variability than 2SPS estimators (Figure 3.6).

3.5. Discussion

Because of their simplicity, two stage IV methods are very popular approaches to control for unmeasured confounding among health service researchers. The conclusions from previous studies are in conflicts (Burgess, 2013; Cai, Small, and Ten Have, 2011; Palmer et al., 2008; Wan et al., 2015). To comprehensively evaluate the consistency of 2SPS and 2SRI in nonlinear models, We proposed a new two stage modeling framework which accommodates clinical settings that often involve single exposure variable and multiple unmeasured covariates. Within this framework, we demonstrate that bias problems of 2SRI and 2SPS can be reduced to the extensively studied bias problems from omitting variables in non-linear models (Gail, Wieand, and Piantadosi, 1984; Lin, Psaty, and Kronmal, 1998). Instead of estimating the conditional causal parameter, the two stage IV estimators converge to a marginal causal parameter when the composite error terms, consisting of unmeasured covariates, are not accounted for. The magnitude of bias can be assessed by comparing the difference between conditional causal parameter and marginal one. When treatment effect is null, both 2SPS and 2SRI are unbiased. However, when treatment effect is not null, we

can not simply extend 2SRI or 2SPS to logistic or Cox models without producing biased estimates. The attenuation of estimated conditional treatment effect suggests the 2SPS or 2SRI estimators are biased towards null hypothesis based on simulation and analytic results.

We further revealed that the bias of the 2SRI estimator is also attributable to the dissimilarity between the effects of unmeasured covariates on the outcome and their effects on the treatment. The more similar between their effects on treatment and outcome, the less biased the 2SRI estimator is. The consistency of the 2SRI estimator is only established when the effects on the outcome and the treatment of unmeasured covariates are proportional to each other. However, this assumption is too strict to hold in real settings.

The framework and findings proposed in the current study may be helpful to address a wide range problem with applications of two stage IV methods in estimating the conditional treatment effect. In this paper, we have used this framework to explain the conflicting conclusions when applying two stage IV methods to different types of outcomes in the current literature. This framework may also be used as guidance to evaluate the alternative two stage regression model based IV approaches.

3.6. Appendix

Appendix A: Proofs for Coefficient Vectors

(A.1): *A new dissimilarity measure between two vectors* Let x and y denote the two $k-$element vectors, and θ be the angle between two vectors. The sine dissimilarity, sin(θ), between two vectors is defined by their cross product and magnitudes as

$$\sin(\theta) = \frac{x \times y}{\|x\|\|y\|}$$

This sine dissimilarity metric satisfies the three criteria for common distance measure: 1) non-negativity and identity of indiscernibles; 2) symmetry; 3) triangle inequality. Proofs are given as follows:

1. *Non-negativity and identity of indiscernibles:* Distance is positive between two different points, and is zero precisely from a point to itself. That is, d(x,y) $\geq$ 0, and d(x,y) = 0 if and only if x = y.

Proof:

Non-negativity:

Let θ denote the angle between vectors x, y and $0 \leq \theta \leq \pi$

$$\begin{aligned} \mathsf{d}(x, y) &= \sin(\theta) \\ &\geq 0 \quad \forall\ \theta \in [0, \pi] \end{aligned}$$

Identity of indiscernibles:

When x and y are in the same or opposite directions, $y = \pm x$, two vectors are considered to be similar under both situations "equivalently". That is, when the angle between two vectors are 0 and $\sin(0) = 0$, x and y are similar. when the angle between two vectors are π and $\sin(\pi) = 0$, x and y are similar. If x and y are similar (the angle θ between two vectors is either 0 or π), $\mathsf{d}(x, y) = \sin(\theta) = 0$.

2. *Symmetry:* the distance between x and y is the same in either direction. That is, $\mathsf{d}(x, y) = \mathsf{d}(y, x)$.

   Proof:

   Let θ denote the angle between vectors x, y and $0 \leq \theta \leq \pi$

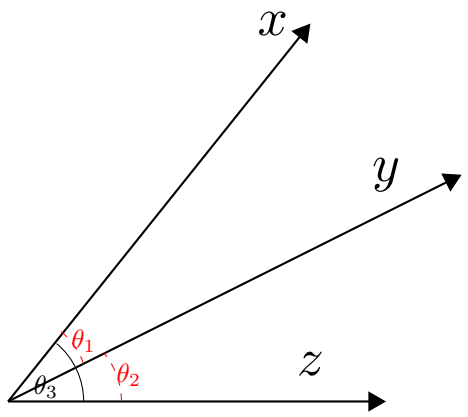
   $$\begin{aligned} \mathsf{d}(x, y) &= \sin(\theta) \\ &= \mathsf{d}(y, x) \end{aligned}$$

3. *Triangle inequality:* the distance between two points is the shortest distance along any path. That is, $\mathsf{d}(x, z) \leq \mathsf{d}(x, y) + \mathsf{d}(y, z)$, $\forall x, y, z$.

   Proof:

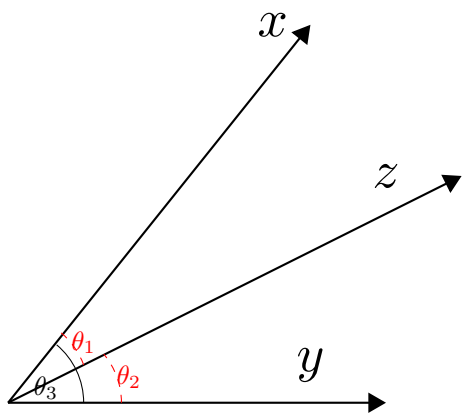
   Let θ_1 denote the angle between vectors x, y, θ_2 denote the angle between vectors y, z, and θ_3 denote the angle between vectors x, z. $0 \leq \theta_i \leq \pi, \forall i \in 1, 2, 3$

(3.1) When $\theta_3 = \theta_1 + \theta_2$



$$\begin{aligned} \mathsf{d}(x,z) &= \sin(\theta_3) \\ &= \sin(\theta_1 + \theta_2) \\ &= \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2) \\ &\leq \sin(\theta_1) + \sin(\theta_2) = \mathsf{d}(x,y) + \mathsf{d}(y,z) \end{aligned}$$

(3.2) When $\theta_1 = \theta_3 - \theta_2$



$$
\begin{aligned}
\mathsf{d}(x,z) &= \mathsf{sin}(\theta_1) \\
&= \mathsf{sin}(\theta_3 - \theta_2) \\
&= \mathsf{sin}(\theta_3)\mathsf{cos}(\theta_2) - \mathsf{cos}(\theta_3)\mathsf{sin}(\theta_2) \\
&\leq \mathsf{sin}(\theta_3) + \mathsf{sin}(\theta_2) = \mathsf{d}(x,y) + \mathsf{d}(y,z)
\end{aligned}
$$

(A.2): *Least square projection of* $\beta_3^T\mathbf{U}$ *onto* $\alpha_3^T\mathbf{U}$

Let β_3 and α_3 denote the coefficient vectors for unmeasured covariates $\mathbf{U}$ in outcome model (3.3) and treatment model (3.4). Let θ denote the angle between the two vectors.

$$
\beta_3^\mathsf{T}\mathbf{U} = \frac{\|\beta_3\|\mathsf{cos}(\theta)}{\|\alpha_3\|}\alpha_3^\mathsf{T}\mathbf{U} + (\beta_3^\mathsf{T} - \|\beta_3\|\mathsf{cos}(\theta)\frac{\alpha_3^\mathsf{T}}{\|\alpha_3\|})\mathbf{U} \tag{A.1}
$$

Show the equation (A.1) is a least square projection of $\beta_3^\mathsf{T}\mathbf{U}$ onto $\alpha_3^\mathsf{T}\mathbf{U}$.

Proof:

When unmeasured confounders $\mathbf{U}$ (normalized) are independent, let ρ denote the coefficient from least square regression of $\beta_3^\mathsf{T}\mathbf{U}$ on $\alpha_3^\mathsf{T}\mathbf{U}$. Let $\mathbf{\Sigma}$ denote the variance and covariance matrix of $\mathbf{U}$. In this case, $\mathbf{\Sigma}$ is an identity matrix denoted as $\mathbf{I}$.

$$
\begin{aligned}
\rho &= \frac{\mathsf{cov}(\beta_3^\mathsf{T}\mathbf{U}, \alpha_3^\mathsf{T}\mathbf{U})}{\mathsf{var}(\alpha_3^\mathsf{T}\mathbf{U})} \\
&= \frac{\beta_3 \cdot \alpha_3}{\|\alpha_3\|^2} \quad \because \Sigma = \mathbf{I} \\
&= \frac{\|\beta_3\|\|\alpha_3\|\mathsf{cos}(\theta)}{\|\alpha_3\|^2} \\
&= \frac{\|\beta_3\|\mathsf{cos}(\theta)}{\|\alpha_3\|}
\end{aligned}
$$

When unmeasured confounders $\mathbf{U}$ (normalized) are not independent, equation (A.1) is a least square projection of $\beta_3^\mathsf{T}\mathbf{U}$ onto $\alpha_3^\mathsf{T}\mathbf{U}$ by ignoring the correlation structure.

(A.3): Proposition: $\|\beta_1\alpha_3^\mathsf{T} + \beta_3^\mathsf{T}\| \geq \|\beta_3^\mathsf{T} - \|\beta_3\|\mathsf{cos}(\theta)\frac{\alpha_3^\mathsf{T}}{\|\alpha_3\|}\|$

proof:

$$
\begin{aligned}
\|\beta_1\alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}}\|^2 =& < \beta_1\alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}}, \beta_1\alpha_3^{\mathsf{T}} + \beta_3^{\mathsf{T}} > \\
=& < (\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}} + (\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|}), \\
& (\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}} + (\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|}) > \\
=& \|(\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|})\|^2 + \|(\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}}\|^2 \\
& + < (\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}}, \beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|} > \\
& + < \beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|}, (\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}} > \\
=& \|\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|}\|^2 + \|(\beta_1 + \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|})\alpha_3^{\mathsf{T}}\|^2 + 0 + 0 \\
\geq& \|\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|}\|^2
\end{aligned}
$$

In third step, use the fact $\beta_3^{\mathsf{T}} - \frac{\|\beta_3\|}{\|\alpha_3\|}\cos(\theta)\alpha_3^{\mathsf{T}}$ is perpendicular to α_3^{T}

Q.E.D

Appendix B: Bias analysis for omitting a random term in two stage IV Poisson model

(1) *Two stage predictor substitution Poisson model* From equation (3.7), the conditional expectation of Y given IV R, observed covariates $\mathbf{X}$, and unmeasured covariates $\mathbf{U}$ can be derived as

$$\mathbb{E}(Y|R, \mathbf{X}, \mathbf{U}) = \exp(\beta_0 + \beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X} + \varepsilon) \tag{B.1}$$

Let $F(\varepsilon|R, \mathbf{X})$ be the conditional distribution of ε given R and $\mathbf{X}$. Under condition (C.1), $\varepsilon \perp R|\mathbf{X}$, thus $F(\varepsilon|R, \mathbf{X})$ can be simplified as $F(\varepsilon|\mathbf{X})$. Given this conditional model (B.1), the

marginal model $\mathbb{E}(Y|R,\mathbf{X})$ is

$$\begin{aligned}\mathbb{E}(Y|R,\mathbf{X}) &= \int_{-\infty}^{\infty} \exp(\beta_0 + \beta_1 \mathbb{E}(D|R,\mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X} + \varepsilon) dF(\varepsilon|\mathbf{X}) \\ &= \exp(\beta_0 + \beta_1 \mathbb{E}(D|R,\mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X}) \int_{-\infty}^{\infty} \exp(\varepsilon) dF(\varepsilon|\mathbf{X}) \\ &= \exp(\beta_0 + \beta_1 \mathbb{E}(D|R,\mathbf{X}) + \beta_2^{\mathsf{T}}\mathbf{X}) c(\mathbf{X})\end{aligned}$$

where $c(\mathbf{X}) = \int_{-\infty}^{\infty} \exp(\varepsilon) dF(\varepsilon|\mathbf{X})$, a function of $\mathbf{X}$ only. Both the marginal Poisson model and conditional Poisson model have the same parameter β_1.

(2) *Two stage residual inclusion Poisson model* From equation (3.5), the conditional expectation of outcome Y given treatment D, IV R, observed covariates $\mathbf{X}$, and unmeasured covariates $\mathbf{U}$ is

$$\mathbb{E}(Y|D,R,\mathbf{X},\mathbf{U}) = \exp(\beta_0 + \beta_1 D + \beta_2^{\mathsf{T}}\mathbf{X} + \gamma_1 \delta + \nu) \tag{B.2}$$

Let $F(\nu|D,\mathbf{X},\delta)$ be the conditional distribution of ν given exposure D, measured confounders $\mathbf{X}$, and residual δ. Under condition (C.1), $F(\nu|D,\mathbf{X},\delta)$ can be simplified as $F(\nu|\mathbf{X},\delta)$ because $D \perp \nu|\{\mathbf{X},\delta\}$. Given this conditional model (B.2), the marginal model $\mathbb{E}(Y|D,R,\mathbf{X})$ is

$$\begin{aligned}\mathbb{E}(Y|D,R,\mathbf{X}) &= \int_{-\infty}^{\infty} \exp(\beta_0 + \beta_1 D + \beta_2^{\mathsf{T}}\mathbf{X} + \gamma_1 \delta + \nu) dF(\nu|\mathbf{X},\delta) \\ &= \exp(\beta_0 + \beta_1 D + \beta_2^{\mathsf{T}}\mathbf{X} + \gamma_1 \delta) \int_{-\infty}^{\infty} \exp(\nu) dF(\nu|\mathbf{X},\delta) \\ &= \exp(\beta_0 + \beta_1 D + \beta_2^{\mathsf{T}}\mathbf{X} + \gamma_1 \delta) c(\mathbf{X},\delta)\end{aligned}$$

where $c(\mathbf{X},\delta) = \int_{-\infty}^{\infty} \exp(\nu) dF(\nu|\mathbf{X},\delta)$. It is a function of $\mathbf{X}$ and δ. The marginal model and conditional model have the same causal parameter β_1 associated with exposure D, but the parameters for $\mathbf{X}$ and δ, β_2^{T} and γ, are different between two models.

For both 2SPS and 2SRI Poisson models, if we can relax condition (C.1) and assume that $R \perp \mathbf{U}$ without conditional on $\mathbf{X}$, $c(\mathbf{X})$ and $c(\mathbf{X},\delta)$ become constants and thus two approaches estimate all parameters consistently.

Appendix C: Bias analysis for omitting a random term in two stage IV logistic model

(1) *Two stage residual inclusion:* In section (3.3.2), the second stage logistic model of binary outcome Y conditional on treatment D, instrumental variable R, observed covariates $\mathbf{X}$,and unmeasured covariates $\mathbf{U}$ is

$$\text{logit}\{\mathbb{P}(Y=1|D,R,\mathbf{X},\mathbf{U})\} = \beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu \tag{C.1}$$

where

$$\begin{aligned}
\delta &= D - \mathbb{E}(D|R,\mathbf{X}) \\
\gamma_1 &= \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|} \\
\nu &= (\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|})\mathbf{U}
\end{aligned}$$

Given this logistic model (C.1), the conditional probability of Y given $\mathbf{X}, R, D$ is

$$\begin{aligned}
\mathbb{P}(Y=1|D,R,\mathbf{X}) = \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta)\times \\
\int_{-\infty}^{\infty} \frac{\exp(\nu)}{1+\exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu)} dF(\nu|D,\delta,\mathbf{X})
\end{aligned} \tag{C.2}$$

where $F(\nu|D,\delta,\mathbf{X})$ is the conditional distribution of ν given D, δ, and $\mathbf{X}$. Under condition (C.1), $F(\nu|D,\delta,\mathbf{X})$ can be simplified as $F(\nu|\delta,\mathbf{X})$ because $D \perp \nu|\{\delta,\mathbf{X}\}$. When ν term is omitted from model (C.1), model (C.1) usually does not reduce to a logistic model as follows

$$\text{logit}\{\mathbb{P}(Y=1|D,R,\mathbf{X})\} = \beta_0^* + \beta_1^* D + \beta_2^* \mathbf{X} + \gamma_1^* \delta \tag{C.3}$$

However, it is still of practical importance to investigate how well model (C.3) may approximate model (C.2). Use the same approach by (Lin et al, 1998), we can show that the model (C.2) is

$$\text{logit}\{\mathbb{P}(Y=1|D,R,\mathbf{X})\} = \beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + g(D,R,\mathbf{X}) \tag{C.4}$$

where

$$g(D, R, \mathbf{X}) = \log \frac{\mathbb{A}}{\mathbb{B}} \tag{C.5}$$

and,

$$\mathbb{A} = \int_{-\infty}^{\infty} \frac{\exp(\nu)}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta) + \nu)} dF(\nu|\delta, \mathbf{X})$$
$$\mathbb{B} = \int_{-\infty}^{\infty} \frac{1}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta) + \nu)} dF(\nu|\delta, \mathbf{X})$$

The Proof of (C.5) is given below.

Proof:

First, from equations (C.2) and (C.4), we can establish

$$\frac{\exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + g(D, R, \mathbf{X}))}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + g(D, R, \mathbf{X}))} =$$
$$\exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta \times$$
$$\int_{-\infty}^{\infty} \frac{\exp(\nu)}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu)} dF(\nu|\delta, \mathbf{X})$$

Next, divide the common factor term on both sides,

$$\frac{\exp(g(D, R, \mathbf{X}))}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta) \exp(g(D, R, \mathbf{X}))} =$$
$$\int_{-\infty}^{\infty} \frac{\exp(\nu)}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu)} dF(\nu|\delta, \mathbf{X})$$

Last, re-express the above equation to derive the following expression for $g(D, R, X)$

$$g(D, R, \mathbf{X}) = \log \frac{\mathbb{A}}{\mathbb{B}}$$

where

$$\mathbb{A} = \int_{-\infty}^{\infty} \frac{\exp(\nu)}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu)} dF(\nu|\delta, \mathbf{X})$$
$$\mathbb{B} = \int_{-\infty}^{\infty} \frac{1}{1 + \exp(\beta_0 + \beta_1 D + \beta_2 \mathbf{X} + \gamma_1 \delta + \nu)} dF(\nu|\delta, \mathbf{X})$$

From equation (C.5), it is easy to verify that the 2SRI logistic model produces the consistent estimate under two conditions.

(1) when $\beta_1 = 0$, $g(D, R, \mathbf{X})$ contains no D term. so when the treatment effects is null, there is no difference between the causal parameters in the marginal and conditional models.

(2) when $\beta_3 - k\cos(\theta)a_3 = 0$, $\nu = 0$, and we have $g(D, R, X) = 0$. Thus, model (C.4) with ν term omitted has exactly the same causal parameter β_1 as the model (C.1).

(2) *Two stage predictor substitution.* In section (3.3.1), the second stage logistic model of binary outcome Y conditional on instrumental variable R, observed covariates $\mathbf{X}$,and unmeasured covariates $\mathbf{U}$ is

$$\text{logit}\{\mathbb{P}(Y = 1|R, \mathbf{X}, \mathbf{U}) = \beta_0 + \beta_1 \mathbb{E}(D|R, \mathbf{X}) + \beta_2 \mathbf{X} + \varepsilon$$

where $\varepsilon = (\beta_1 \alpha_3 + \beta_3)\mathbf{U}$. From condition (C.1), $\varepsilon \perp \mathbb{E}(D|R, \mathbf{X})|\mathbf{X}$.

Similarly, we can write a model form with a bias term when ε term is omitted as

$$\text{logit}\{\mathbb{P}(Y = 1|D, R, \mathbf{X})\} = \beta_0 + \beta_1 \mathbb{E}(D|R, \mathbf{X}) + \beta_2 \mathbf{X} + g(R, \mathbf{X})$$

The derived bias term $g(R, \mathbf{X})$ for two stage predictor substitution logistic model is

$$g(R, \mathbf{X}) = \log \frac{\mathbb{A}}{\mathbb{B}}$$

where

$$\mathbb{A} = \int_{-\infty}^{\infty} \frac{\exp(\varepsilon)}{1+\exp(\beta_0+\beta_1\mathbb{E}(D|R,\mathbf{X})+\beta_2\mathbf{X}+\varepsilon)} dF(\varepsilon|\mathbf{X})$$
$$\mathbb{B} = \int_{-\infty}^{\infty} \frac{1}{1+\exp(\beta_0+\beta_1\mathbb{E}(D|R,\mathbf{X})+\beta_2\mathbf{X}+\varepsilon)} dF(\varepsilon|\mathbf{X})$$

The 2SPS logistic model produces the consistent estimate under the two conditions:

(1) when $\beta_1 = 0$, $g(R, \mathbf{X})$ contains no $\mathbb{E}(D|R, \mathbf{X})$ term. So when the treatment effects is null, there is no difference between the causal parameters in the marginal and conditional models.

(2) when $\beta_1\alpha_3 + \beta_3 = 0$, we have $g(R, \mathbf{X}) = 0$.

Appendix D: Bias analysis for omitting a random term in two stage IV Cox proportional hazard model

Let T denote the time to event. Using the results from sections (3.3.1) and (3.3.2), we can specify the conditional hazard function of T under the second stage Cox proportional model for both 2SPS and 2SRI methods separately. The corresponding bias of the 2SPS and 2SRI methods can be assessed using the techniques laid out in (Lin et al, 1998).

(1) *Two stage residual inclusion:* Conditional on treatment D,instrumental variable R, observed covariates $\mathbf{X}$, and unmeasured covariates $\mathbf{U}$, the log hazard function of T under the 2SRI Cox proportional model is

$$\log\{\lambda(t|D, R, \mathbf{X}, \mathbf{U})\} = \beta_0(t) + \beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu \tag{D.1}$$

where

$$\begin{aligned}
\beta_0(t) &= \log(\lambda_0(t)) \quad \text{unspecified baseline log hazard function} \\
\delta &= D - \mathbb{E}(D|R, \mathbf{X}) \\
\gamma_1 &= \frac{\|\beta_3\|\cos(\theta)}{\|\alpha_3\|} \\
\nu &= (\beta_3^{\mathsf{T}} - \|\beta_3\|\cos(\theta)\frac{\alpha_3^{\mathsf{T}}}{\|\alpha_3\|})\mathbf{U}
\end{aligned}$$

When covariates $\mathbf{U}$ are unmeasured and ν term is omitted, we assume the log hazard function of T conditional on $\{D, R, \mathbf{X}\}$ is

$$\log\{\lambda(t|D, R, \mathbf{X})\} = \beta_0^*(t) + \beta_1^* D + \beta_2^* \mathbf{X} + \gamma_1^* \delta \tag{D.2}$$

where

$$\beta_0^*(t) = \log(\lambda_0^*(t)) \quad \text{unspecified baseline log hazard function}$$

To assess the potential difference between β_1 and β_1^*, we first denote $F(\nu|D, R, \mathbf{X})$ as the conditional distribution function of ν given D,R, and $\mathbf{X}$. Under assumption (C.1), $\nu \perp D|\{R, \mathbf{X}\}$. Thus, the distribution function can be simplified as $F(\nu|R, \mathbf{X})$. We let $f(t|\cdot)$ and $S(t|\cdot)$ be the conditional density and survival functions of time to event T. Then $\lambda(t|D, R, \mathbf{X})$ in model (D.2) can be expressed as

$$\begin{aligned}
\lambda(t|D, R, \mathbf{X}) &= \frac{f(t|D, R, \mathbf{X})}{S(t|D, R, \mathbf{X})} \\
&= \frac{\int_{-\infty}^{\infty} f(t|D, R, \mathbf{X}, \nu) dF(\nu|R, \mathbf{X})}{\int_{-\infty}^{\infty} S(t|D, R, \mathbf{X}, \nu) dF(\nu|R, \mathbf{X})}
\end{aligned} \tag{D.3}$$

Given model (D.1),

$$\int_{-\infty}^{\infty} f(t|D,R,\mathbf{X},\nu)dF(\nu|D,R,\mathbf{X}) = \int_{-\infty}^{\infty} \lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)$$
$$\times \exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)$$
$$dF(\nu|R,\mathbf{X})$$
$$\int_{-\infty}^{\infty} S(t|D,R,\mathbf{X},\nu)dF(\nu|D,R,\mathbf{X}) = \int_{-\infty}^{\infty} \exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)$$
$$dF(\nu|R,\mathbf{X})$$

where cumulative hazard function $\Lambda_0(t)$ is defined as

$$\Lambda_0(t) = \int_{-\infty}^{\infty} \lambda_0(u)du$$

Then, equation (D.3) becomes

$$\lambda(t|D,\delta,\mathbf{X}) = \lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta)g(t,D,R,\mathbf{X}), \tag{D.4}$$

where

$$g(t,D,R,\mathbf{X}) = \frac{\int_{-\infty}^{\infty}\exp(\nu)(\exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)dF(\nu|R,\mathbf{X})}{\int_{-\infty}^{\infty}\exp(-\Lambda_0(t)\exp(\beta_1 D + \beta_2\mathbf{X} + \gamma_1\delta + \nu)dF(\nu|R,\mathbf{X})} \tag{D.5}$$

In equation (D.4), when $\beta_3 - k\cos(\theta)a_3 = 0$, $\nu = 0$ and $g(t,D,R,\mathbf{X}) = 1$. When $\beta_1 = 0$, $g(t,D,R,\mathbf{X})$ does not contain exposure D. Thus, β_1^* in model (D.2) and β_1 in model (D.1) are the same. Under these two conditions, the 2SRI Cox proportional hazard model estimates the conditional causal parameter β_1 consistently.

(2) *Two stage predictor substitution:* Under the 2SPS Cox proportional hazard model, the log hazard function of T conditional on instrumental variable R, observed covariates $\mathbf{X}$, and unmeasured covariates $\mathbf{U}$

$$\log\{\lambda(t|R,\mathbf{X}),\mathbf{U}\} = \beta_0(t) + \beta_1\mathbb{E}(D|R,\mathbf{X}) + \beta_2\mathbf{X} + \varepsilon, \tag{D.6}$$

where

$$\beta_0(t) = \log(\lambda_0(t)) \quad \text{unspecified baseline log hazard function}$$
$$\varepsilon = (\beta_1\alpha_3 + \beta_3)\mathbf{U}$$

When covariates $\mathbf{U}$ are not measured and ε is omitted, the conditional hazard function of T, conditional on $\{R, \mathbf{X}\}$, is defined as

$$\log\{\lambda(t|R, \mathbf{X})\} = \beta_0^*(t) + \beta_1^*\mathbb{E}(D|R, \mathbf{X}) + \beta_2^*\mathbf{X} \tag{D.7}$$

where

$$\beta_0^*(t) = \log(\lambda_0^*(t)) \quad \text{unspecified baseline log hazard function}$$

As in previous section, the bias term $g(t, R, \mathbf{X})$ for the 2SPS Cox proportional hazard model can be derived as

$$\lambda(t|R, \mathbf{X}) = \lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X})g(t, R, \mathbf{X}) \tag{D.8}$$

where

$$g(t, R, \mathbf{X}) = \frac{\int_{-\infty}^{\infty} \exp(\varepsilon)(\exp(-\Lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X} + \varepsilon)dF(\varepsilon|R, \mathbf{X})}{\int_{-\infty}^{\infty} \exp(-\Lambda_0(t)\exp(\beta_1\mathbb{E}(D|R, \mathbf{X}) + \beta_2\mathbf{X} + \varepsilon)dF(\varepsilon|R, \mathbf{X})} \tag{D.9}$$

where $F(\varepsilon|R, X)$ is the conditional distribution function.

In equation (D.8), when $\beta_3\alpha_3 + \beta_3 = 0$, $\nu = 0$ and $g(t, R, \mathbf{X}) = 1$. When $\beta_1 = 0$, $g(t, R, \mathbf{X})$ does not contain exposure D. Thus, β_1^* in model (D.7) and β_1 in model (D.6) are the same. Under these two conditions, the 2SPS Cox proportional hazard model estimates the conditional causal parameter β_1 consistently.

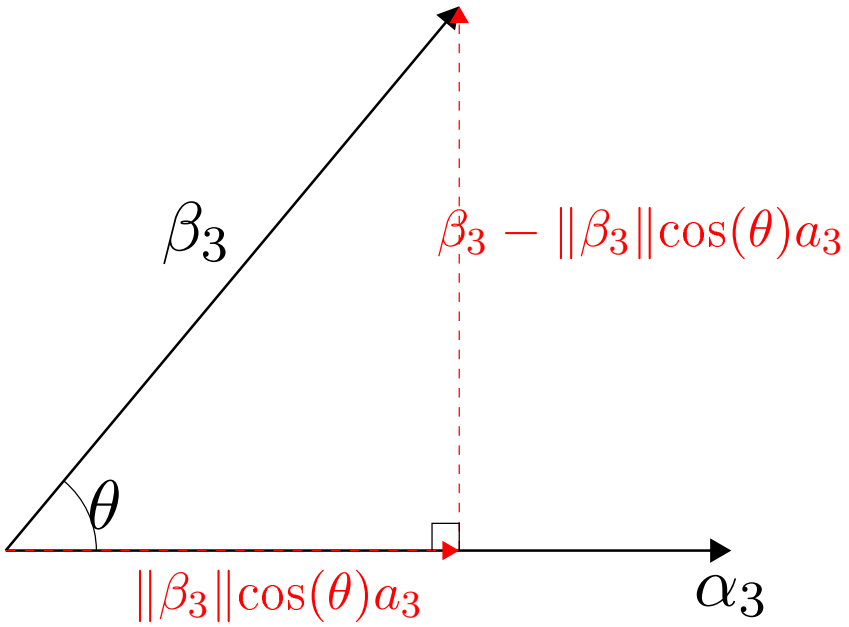


Figure 3.1: Decomposing β_3 into two orthogonal components

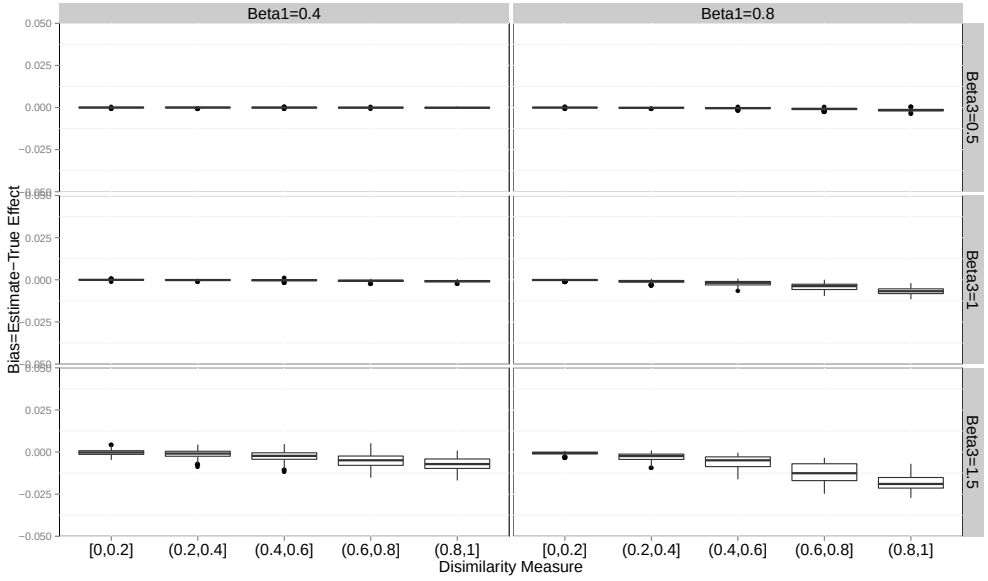


Figure 3.2: Boxplot of 2SRI Poisson model estimates when treatment effect is nonzero. β_1 is treatment effect; β_3 is effect of unmeasured covariates on outcome. Bias is the difference between estimates and true treatment effect

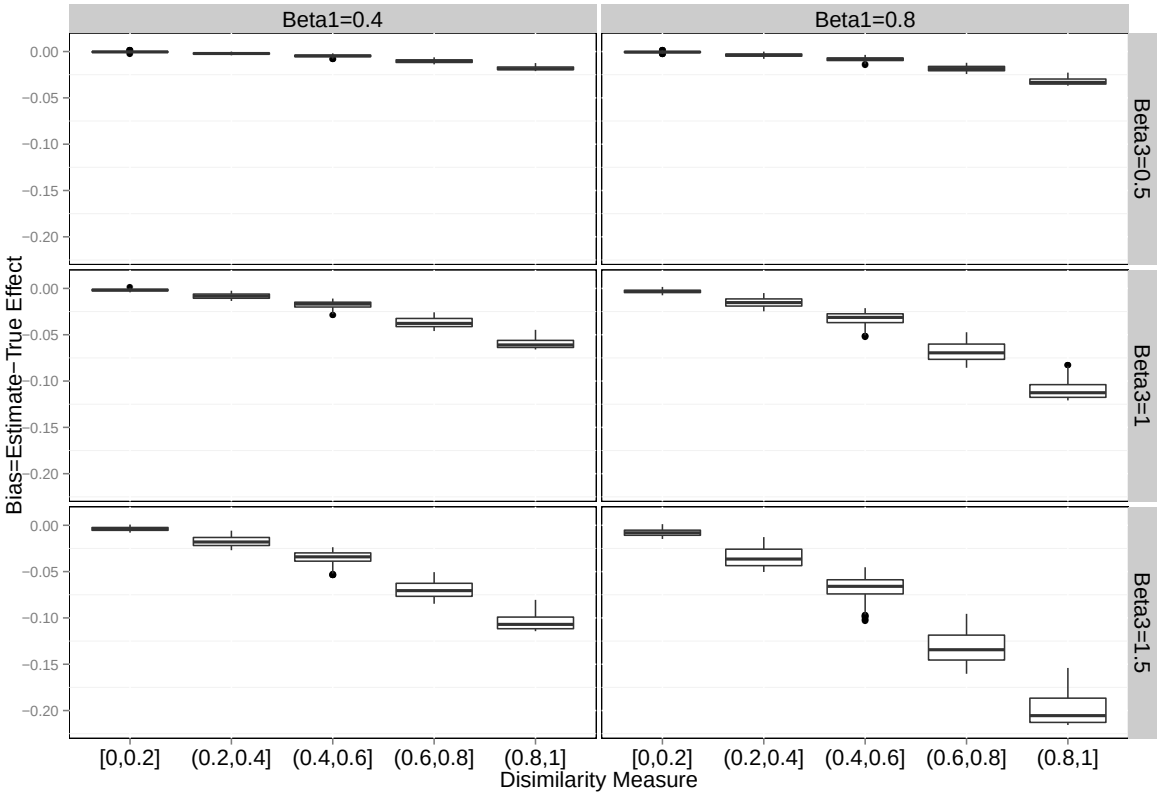


Figure 3.3: Boxplot of 2SRI logistic model estimates when treatment effect is nonzero. β_1 is treatment effect; β_3 is effect of unmeasured covariates on outcome. Bias is the difference between estimates and true treatment effect.

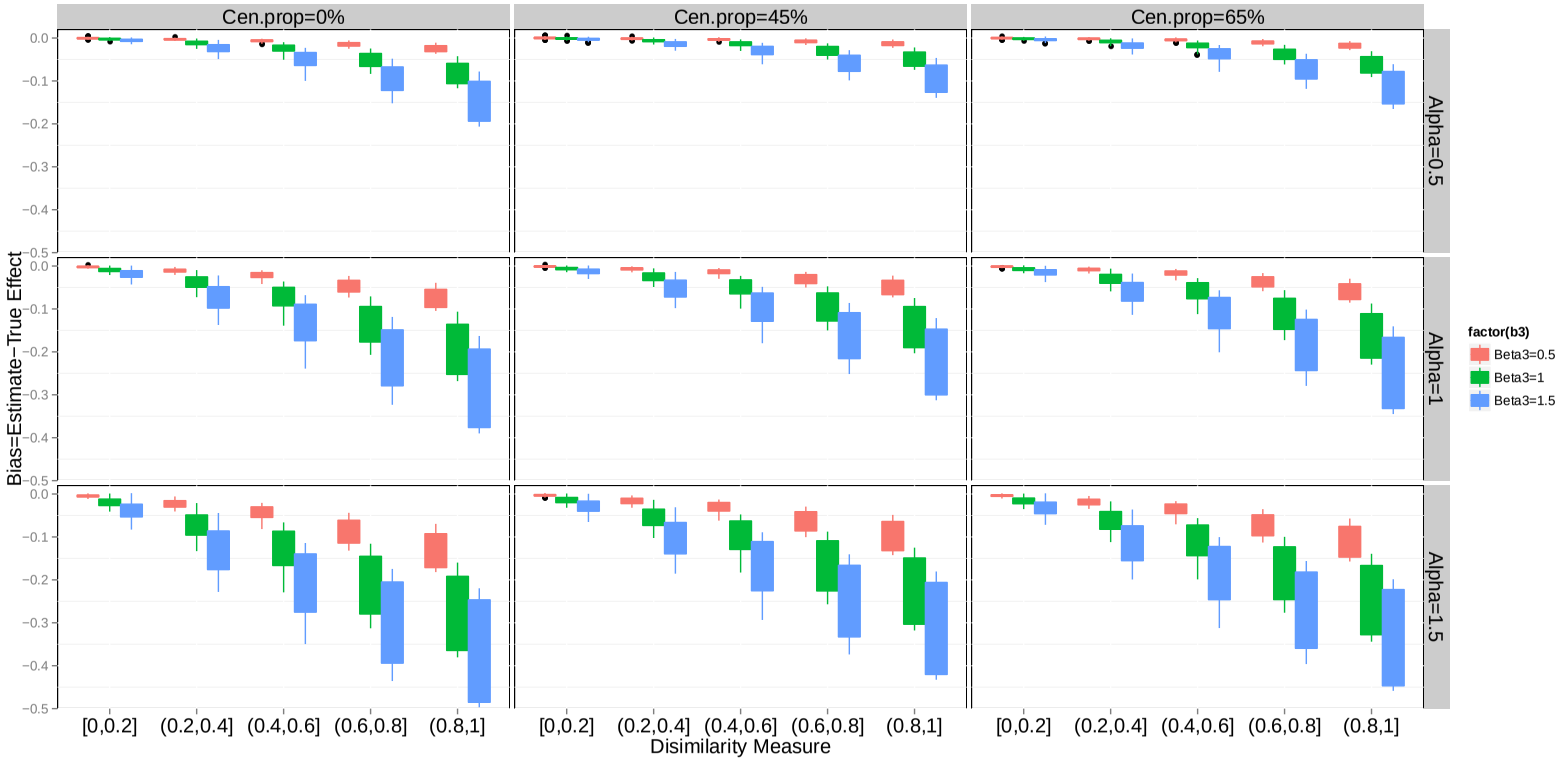


Figure 3.4: Boxplot of 2SRI Cox model estimates when treatment effect is nonzero. β_1 is treatment effect; β_3 is effect of unmeasured covariates on outcome. Bias is the difference between estimates and true treatment effect. Red colored box is for low level effect of unmeasured covariates ($\beta_3 = 0.5$). Green colored box is for medium level effect of unmeasured covariates ($\beta_3 = 1$). Blue colored box is for high level effect of unmeasured covariates ($\beta_3 = 1.5$)

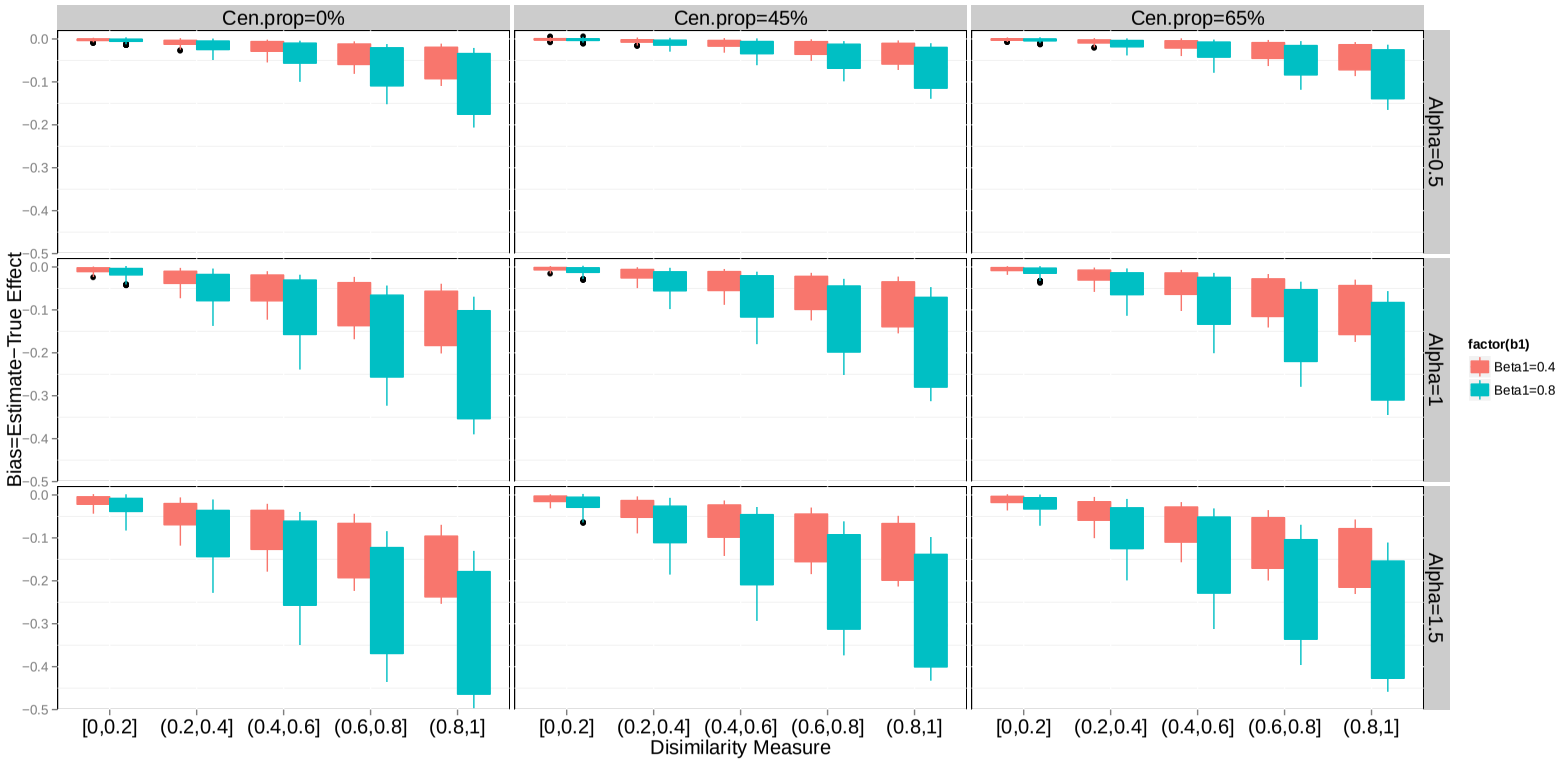


Figure 3.5: Boxplot of 2SRI Cox model estimates when treatment effect is nonzero. β_1 is treatment effect; β_3 is effect of unmeasured covariates on outcome. Bias is the difference between estimates and true treatment effect.Red colored box is for smaller size of treatment effect ($\beta_1 = 0.4$). Green colored box is for medium level effect of unmeasured covariates ($\beta_1 = 0.8$)

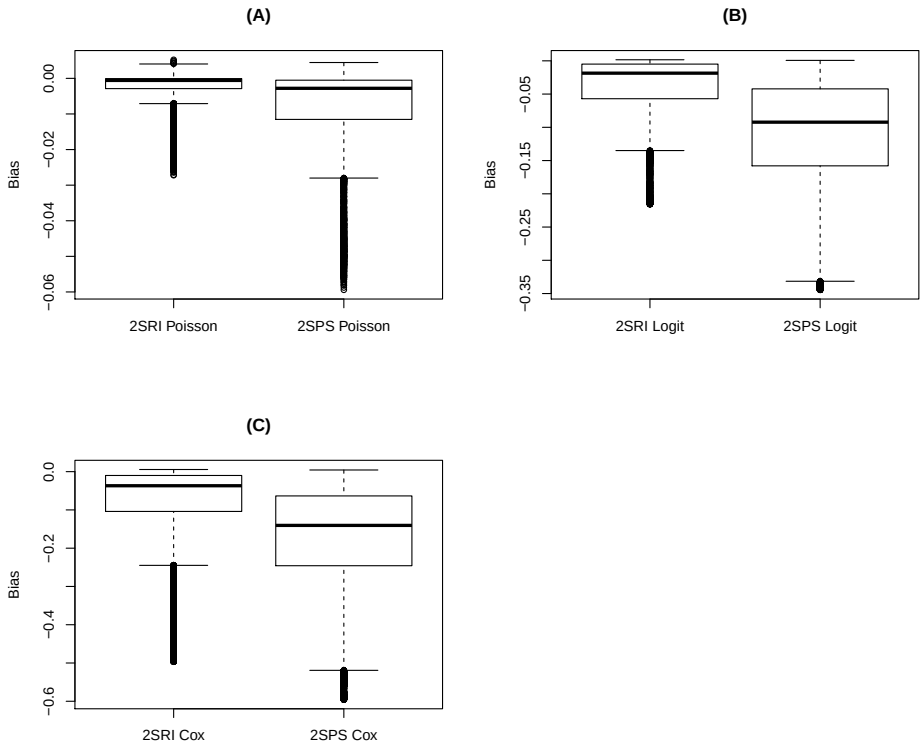


Figure 3.6: Boxplots of 2SRI and 2SPS comparisons for Poisson, Logit, and Cox models

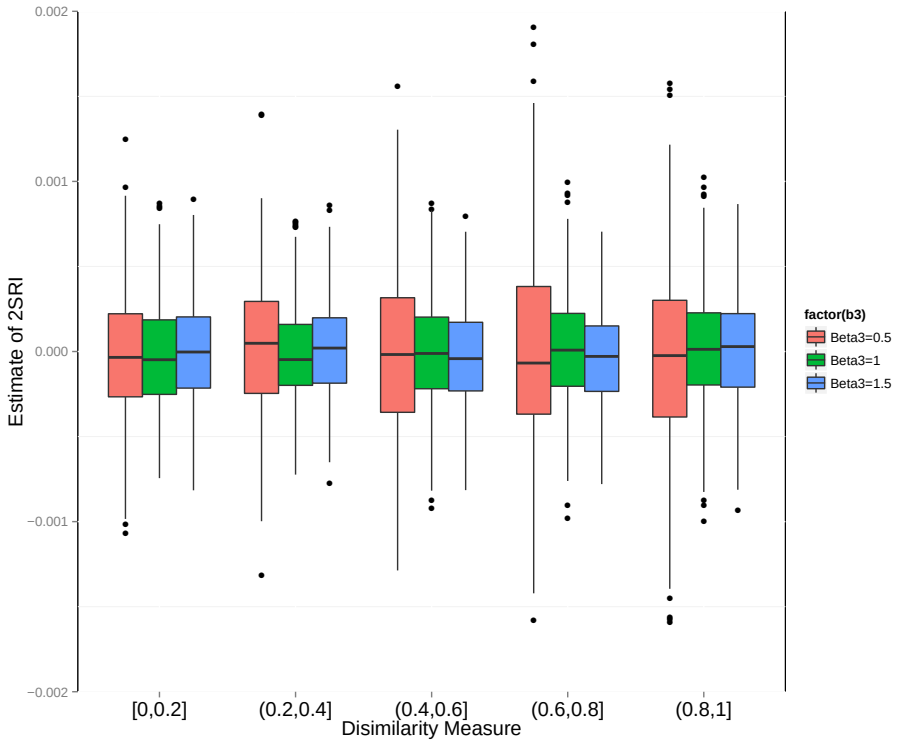


Figure 3.7: Boxplot of 2SRI Poisson model estimates when treatment effect is zero. Color red- effects of unmeasured covariates are low ($\beta_3 = 0.5$); Color green-effects of unmeasured covariates are medium ($\beta_3 = 1$); Color blue-effects of unmeasured covariates are high ($\beta_3 = 1.5$)

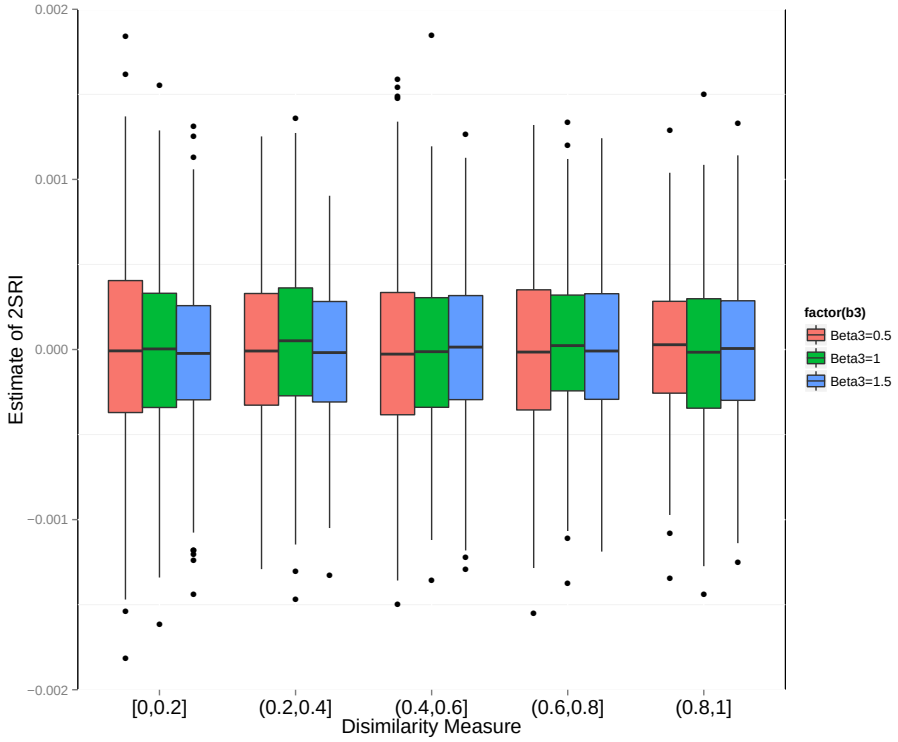


Figure 3.8: Boxplot of 2SRI logistic model estimates when treatment effect is zero. Color red- effects of unmeasured covariates are low ($\beta_3 = 0.5$); Color green-effects of unmeasured covariates are medium ($\beta_3 = 1$); Color blue-effects of unmeasured covariates are high ($\beta_3 = 1.5$)

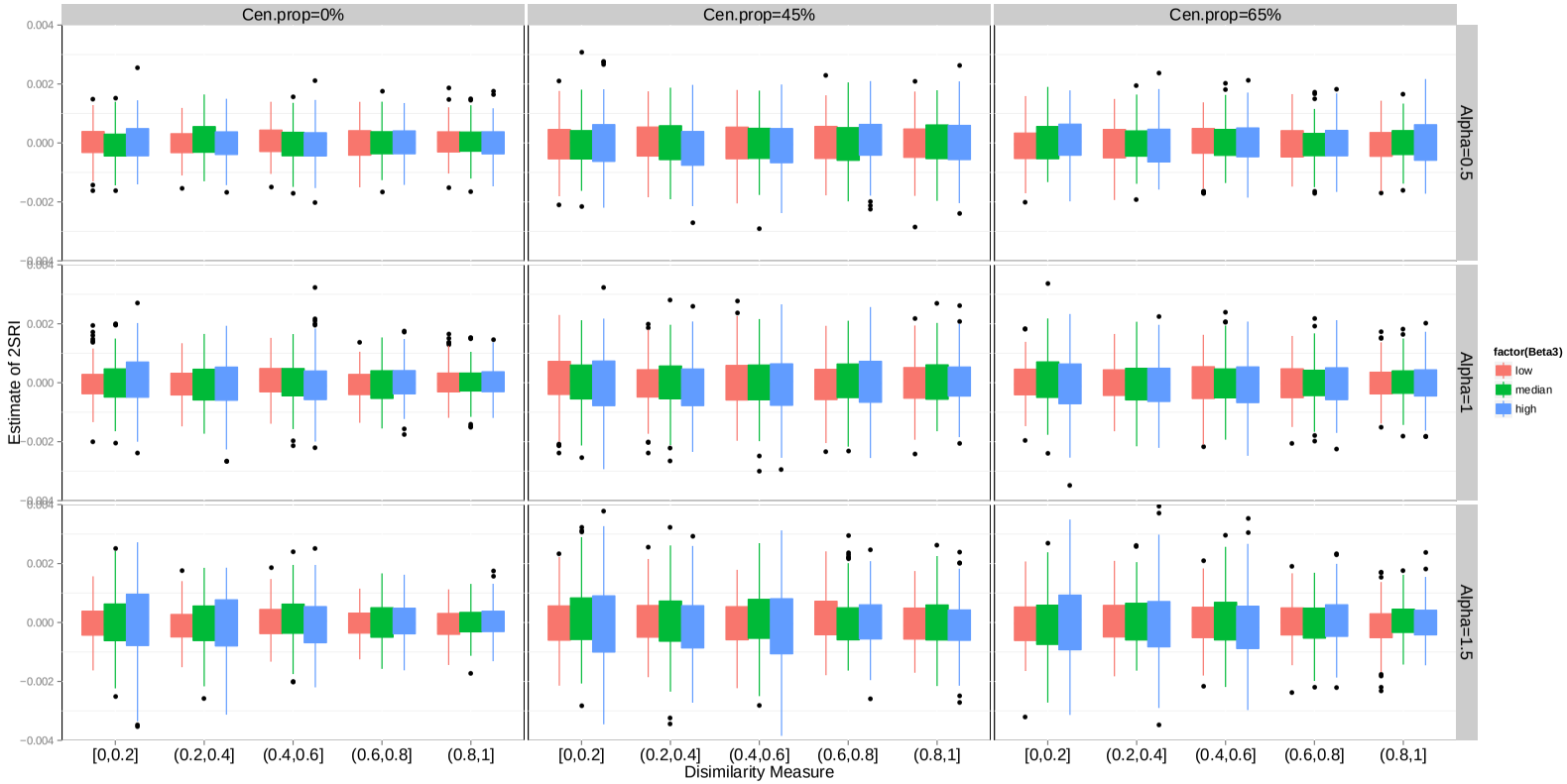


Figure 3.9: Boxplot of 2SRI Cox model estimates when treatment effect is zero. Color red- effects of unmeasured covariates are low ($\beta_3 = 0.5$); Color green-effects of unmeasured covariates are medium ($\beta_3 = 1$); Color blue-effects of unmeasured covariates are high ($\beta_3 = 1.5$)

CHAPTER 4

A CONDITIONAL LOG RANK TEST ADJUSTED WITH PROPENSITY SCORE TO COMPARE SURVIVAL DISTRIBUTIONS

4.1. Introduction

Confounding is a major problem arising from non-randomized observational studies, in which an extraneous variable may correlate with both treatment variable and outcome variable at the same time. The confounding could bias the estimate of the treatment effect if it is not controlled for. A common approach to correct for the biased estimate of the treatment effect resulting from confounding is to include both confounding variables and treatment variable in the same regression model. When encountering right censored time to event data, clinical researchers routinely resort to the multivariable Cox proportional hazard model.

The Cox proportional hazards model assumes that the relationship of the hazard rate at time t given the covariates X and treatment indicator Z can be specified as the following log linear functional form

$$\lambda(t|Z=z, X=x) = \lambda_0(t)\text{exp}(\beta z + \alpha^T x) \tag{4.1}$$

where λ_0 is the unspecified baseline hazard, Z denotes binary treatment variable (Z=1 if subject receives treatment, Z=0 if subject receives the placebo), X is a p-dimensional vector consisting of measured confounding variables $(X_1, X_2, ..., X_p)^T$ at the baseline, β is the regression coefficient associated with Z, α is a p-dimensional vector of regression coefficients for X. The underlying proportional hazard assumption requires that for any two sets of variables (z_1, x_1) and (z_0, x_0), the ratio of two hazard functions $\frac{\lambda(t|Z=z_1, X=x_1)}{\lambda(t|Z=z_0, X=x_0)}$ is time invariant.

we maximize the following partial likelihood to get the maximum partial likelihood estimator $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$,

$$L(\beta, \alpha) = \prod_{i=1}^{n} [\frac{\text{exp}(\beta z_i + \alpha^T x_i)}{\sum_{j \in R_i} \text{exp}(\beta z_i + \alpha^T x_i)}]^{\delta_i} \tag{4.2}$$

where R_i is the set of subjects at risk at time t and δ_i indicates a failure or not at time t.

The parameter of interest β measures the "treatment effect" of Z. Thus, we could test the null hypothesis of no treatment effect $H_0 : \beta = 0$ using the covariates adjusted Cox score test. When the model (4.1) is correctly specified, the adjusted Cox score test is

$$[U_\beta(0, \hat{\alpha}(0))]^T [J^{\beta\beta}(0, \hat{\alpha}(0))][U_\beta(0, \hat{\alpha}(0))] \sim \chi^2_{\alpha,1}$$

where $\hat{\alpha}(0)$ is the restricted m.l.e of α given $\beta = 0$, and

$$U_\beta(\beta, \alpha) = \frac{\partial \text{log}(L(\beta, \alpha))}{\partial \beta},$$
$$J^{\beta\beta}(\beta, \alpha) = \frac{\partial^2 \text{log}(L(\beta, \alpha))}{\partial^2 \alpha} / (\frac{\partial^2 \text{log}(L(\beta, \alpha))}{\partial^2 \beta} \frac{\partial^2 \text{log}(L(\beta, \alpha))}{\partial^2 \alpha} - (\frac{\partial^2 \text{log}(L(\beta, \alpha))}{\partial \beta \partial \alpha})^2)$$

When the model (4.1) is misspecified, such as some variables are omitted from the model, Lin and Wei, 1989 proposed a robust sandwich estimator for the variance of proportional hazard score test statistics. With this robust variance estimator used, the score test is still valid under model misspecification if the treatment variable Z is independent of covariates X.

Kong and Slud, 1997 extended Lin and Wei's approach under a more general condition that the treatment variable Z is independent of covariates X for subjects at risk at any given time t and a more general functional forms of X.

Heller and Venkatraman, 2004 proposed a covariate adjusted non-parametric test for comparing survival distributions among multiple groups. The validity of this non-parametric test does not require either proportional hazard assumption or any independence assumption between treatment variable Z and covariates X. The use of kernel smoothing to estimate the expectation of treatment indicator Z conditional on covariate X among subjects at risk at each time t requires that the number of covariates could not exceed 3 due to the curse of dimensionality.

To extend Heller and Venkatraman's conditional log rank test, we introduced the propensity score to balance the unbalanced distribution of covariates X among treatment groups and reduce the dimensionality of covariate X for kernel smoothing. In section 2, we discussed the extension of conditional log rank test using propensity score. In section 3, simulation studies were performed

to compare the size and power of log rank, adjusted Cox score test, Lin and Wei's robust score test, Kong and Slud's robust score test, and the extended conditional log rank test under various scenarios. Section 4 provides a data example. Section 5 discuss the findings.

4.2. Test Statistic

4.2.1. Notation and Assumption

For the ith subject, let S_i denote the survival time and C_i represent censoring time. $T_i = \min(S_i, C_i)$ is the observed survival time. δ_i is censoring indicator, which is equal to 1 if $S_i \leq C_i$ and 0 otherwise. $N_i(t) = I(T_i \leq t, \delta_i = 1)$ is the counting process that count the number of observed events (0 or 1) for the ith subject, and $Y_i(t) = I(T_i \geq t)$ is the at-risk process, which indicates whether the ith subject is at risk right before time t.

We make the following assumptions: i) let X_i denote the p-dimensional vector of confounding variables $(X_{i1}, X_{i2}, ..., X_{ip})^T$ observed prior to treatment assignment for the ith subject, and there is no other unmeasured confounding variables at baseline; ii) let $e(X)$ denote a known propensity score function. The distribution of covariates X conditional on $e(X)$ between two treatment groups (Z=1 vs. Z=0) are the same: $X \perp Z|e(X)$. Let $b(X)$ be a function of X and finer than $e(X)$. iii) the random vectors $(S_i, C_i, Z_i, X_i), i = 1, 2, 3....n$ are independently and identically distributed. iv) the failure time S and censoring time C are independent conditional on treatment variable Z and covariates X.

4.2.2. Propensity score

4.2.3. Constructing the Test

The null hypothesis that the conditional hazard functions of two treatment groups given measured covariates X are the same is:

$$\lambda_0(t|x) = \lambda_1(t|x) \qquad \forall t, x$$

Following the steps listed by Heller and Venkatraman, 2004, we first have that

$$\begin{aligned}\mathsf{E}(\mathsf{d}N(t)|\text{past}) &= \mathsf{P}(\mathsf{d}N(t)=1|\text{past}) \\ &= \lambda(t)\mathsf{d}tI(T\geq t) \\ &= \lambda(t)Y(t)\mathsf{d}t\end{aligned}$$

Thus, under the independence assumption of the failure time and censoring time conditional on covariates (X, Z), Heller and Venkatraman, 2004 defined the relationship between counting process, at risk process, and the covariates by

$$\mathsf{E}(\mathsf{d}N(t)|Y(t),X,Z) = Y(t)Z\lambda_1(t|X)\mathsf{d}t + Y(t)(1-Z)\lambda_0(t|X)\mathsf{d}t \tag{4.3}$$

where $\lambda_j(t|x)$ represents the conditional hazard for a subject with group variable $z = j$. Based on the expression (1), the counting process for the treatment group ($Z = 1$) can be defined by

$$\mathsf{E}(Z\mathsf{d}N(t)|Y(t),X) = Y(t)\mathsf{E}(Z|Y(t),X)\lambda_1(t|X)\mathsf{d}t$$

Next,take the expectation of both sides (with respect to Z) under the null hypothesis that $\lambda_0(t|X) = \lambda_1(t|X) = \lambda(t|X)$, we have

$$\begin{aligned}\mathsf{E}_Z(\mathsf{E}(Z\mathsf{d}N(t)|Y(t),X,Z)) &= \mathsf{E}_Z(Y(t)Z\lambda_1(t|X)\mathsf{d}t|Y(t),X) \\ &= Y(t)\lambda(t|X)\mathsf{E}(Z|Y(t),X)\mathsf{d}t \quad \text{under } H_0\end{aligned} \tag{4.4}$$

Since $\mathsf{E}(Z|Y(t),X)$ is a function of $(Y(t),X)$, equation (2) shows that under H_0 the counting process is independent of group assignment. Heller and Venkatraman, 2004 constructed a non-parametric test using empirical estimates the left- and right-hand s ides of equation (2) and specifically, conditional expectation $\mathsf{E}(Z\mathsf{d}N(t)|Y(t)=1,X=x_i)$ is constructed non-parametrically using kernel smoothing function

$$\mathsf{E}(Z|Y(t)=1,X=x_i) = \frac{\sum_j Y_j(t)z_jK_g(x_j,x_i)}{\sum_j Y_j(t)K_g(x_j,x_i)} \tag{4.5}$$

when X is p dimensional covariate vector ($p > 1$), the multivariate kernel function K_g is defined as

$$K_g(u) = \prod_{l=1}^{p} g_l^{-1} k(g_l^{-1} u_l)$$

where g is p dimensional vector of bandwidth controlling the degree of smoothness for each element of covariate vector X.

However, the multivariate kernel smoothing suffers from the curse of dimensionality when p≥ 3. That is, the number of neighboring data points around any value X in a higher dimensional space will be very small, unless the sample size is extremely large (Hastie,2010).

Based on theorem (2) from Rosenbaum and Rubin, 1983, $b(X)$ is a 1-dimensional balancing score function of p-dimensional baseline covariates X such that

$$X \perp Z | b(X)$$

and,

$$E(Z|Y(t_0) = 1, X = x_i) = E(Z|Y(t_0) = 1, X = b_i(x))$$

where t_0 is the baseline or the study beginning time point.

Under independence censoring assumption, the patients censored at time t_c have the same risk as the patients not censored at the same time. Thus, it can be inferred that there are no unmeasured factors correlating with both Z and T during the study time period. It then follows

$$\mathsf{E}(Z|Y(t) = 1, X = x_i) = \mathsf{E}(Z|Y(t) = 1, X = b_i(x)) \tag{4.6}$$

for each time point t. Because $e(X)$ itself is the coarsest balancing score function, the equation (2.4) remains valid with $e(X)$ replacing $b(X)$. Therefore, equation (2.3) can be re-expressed as

$$\mathsf{E}(Z|Y(t) = 1, X = e_i(x)) = \frac{\sum_j Y_j(t) z_j K_g(e_j(x), e_i(x))}{\sum_j Y_j(t) K_g(e_j(x), e_i(x))} \tag{4.7}$$

Then from equation (2) and using the fact that

$$\mathsf{E}(dN(t)|X) = \lambda(t|X)Y(t)\mathsf{d}t$$

a conditional log rank test statistics S_n adjusting for balancing score function $e(x)$ can be derived as

$$S_n = \sum_i \int z_i \mathsf{d}N_i(t) - \sum_i \int \frac{\sum_j Y_j(t) z_j K_g(e_j(x), e_i(x))}{\sum_j Y_j(t) K_g(e_j(x), e_i(x))} \mathsf{d}N_i(t) \tag{4.8}$$

This statistics was able to be re-expressed as a function of difference in the estimated conditional hazards between treatment and control groups (Heller and Venkatraman, 2004). It follows from theorem 1 (Heller and Venkatraman, 2004) that under the null hypothesis that $\lambda_0(t|x) = \lambda_1(t|x) = \lambda(t|x)$,

$$n^{-1/2} S_n \to N(0, V)$$

where asymptotic variance V is estimated consistently by

$$\sum_i \hat{v}_i i^2 + \sum_{i \neq j} \hat{v}_{ij}(2\hat{v}_{ii} + 2\hat{v}_{jj} + \hat{v}_{ij} + \hat{v}_{ji}) + \sum_{i \neq j \neq l} \hat{v}_{ij}(\hat{v}_{ij}\hat{v}_{il} + \hat{v}_{ij}\hat{v}_{li} + \hat{v}_{ij}\hat{v}_{jl} + \hat{v}_{ij}\hat{v}_{lj})$$

and

$$\hat{v}_{ij} = \delta_i(z_i - \frac{\alpha_1(x_i, t_i)}{\alpha_0(x_i, x_i)} - \frac{I(t_j \geq t_i) K_b(x_j, x_i)}{\alpha_0(x_i, t_i)}(z_j - \frac{\alpha_1(x_i, t_i)}{\alpha_0(x_i, t_i)}))$$

and $\alpha_j(x,t) = \alpha_{j1}(x,t) + \alpha_{j0}(x,t), j = 0, 1$ The propensity score function $e(X)$ for baseline covariates X can be estimated consistently under correct model specification using logistic regression model

$$e(x) = P(Z = 1|X) = \frac{e^{\beta X}}{1 + e^{\beta X}}$$

4.3. Simulation study

In this section, we performed a series of simulations to evaluate the type 1 error rate and the statistical power for the log rank test, the covariates adjusted Cox score test, Lin and Wei robust method, Kong and Slud robust method, and the propensity score adjusted conditional log rank test. We considered the following factors in the simulation study design:

a. Proportional hazards assumption holds or not. The association between time to event T and covariates (Z, X) can be expressed by a log linear model $\log(T|Z = z, X = x) = \beta z + \alpha^T x + \epsilon$. When the error term $\epsilon \sim$ standard extreme value distribution, the proportional hazard assumption holds. When the error term $\epsilon \sim N(0, \sigma^2)$, the proportional hazard assumption will be violated.

b. The strength of association between confounders X and treatment indicator Z. We used different values of $R^2_{\text{ps}} =$ (0.25,0.5,0.75) from propensity score model $\text{logit}(Z|X = x) = l^T x$ to specify the low, medium, and high level of association between confounders X and treatment indicator Z. The R^2_{ps} can be computed by tuning parameter γ as the following (Heller and Venkatraman, 2004):

$$R^2_{\text{ps}} = \text{var}(\gamma l^T x)/(\text{var}(\gamma l^T x) + \pi^2)$$

c. The strength of association between confounders X and time to event T. Similarly, we also used 3 different values of $R^2_{\text{cox}} =$ (0.25,0.5,0.75) to specify the low, medium, and high level of association between confounders X and T. The R^2_{cox} can be computed by tuning parameter ϕ as the following (Heller and Venkatraman, 2004):

$$R^2_{\text{cox}} = \text{var}(\phi \alpha^T x)/(\text{var}(\phi \alpha^T x) + \text{var}(\epsilon))$$

d. Two different levels of censoring proportions (25%,50%) were specified.

e. Two different proportions of treated subjects (vs. subjects in placebo group): 30% and 50% were specified. This can be done by tuning the intercept term in the propensity score model.

Thus, we have 72 different combination of scenarios. For each scenario, we generate the data in the following steps:

1. 10 confounders X were simulated. $X_1 \sim$ N(10,2), $X_2 \sim$ N(30,5), $X_3 \sim 10 + 3 *$ Uniform$(0, 1)$, $X_4 \sim$ Bernoulli$(p = 0.5)$, $X_5 \sim$ Pois$(\mu = 10)$, $X_6 \sim$ log Normal$(\mu = 4.16, \sigma^2 = 4.96)$, $X_7 \sim$ exp$(\lambda = 1.5)$, $X_8 \sim 12 +$ Gamma$(k = 9, r = 2)$, $X_9 \sim 3 + 0.5$Pois$(\mu = 10) + 1.5N(0, 1)$, and $X_{10} \sim 12.5 + 0.8N(0, 1) + 2$Uniform$(0, 1)$;

2. The binary treatment indicator Z_i for the ith subject is generated using Bernoulli($P_i(Z_i|X_i)$), where $P_i(Z_i|X_i)$ is determined by covariates X_i through a logit function: logit$(P(Z_i = 1|X_i = x_i)) = l^T * x_i$. X_i is a 10-dimensional vector and l^T is set to (l_0,0.16,0.65,-0.58,1.68,-0.4,0.09,-0.2,-0.69,0.28,0.74). When l_0 is -10.5, we have $\sim 50\%$ subjects receiving treatment. When l_0 is -13.8, we have $\sim 30\%$ subjects receiving treatment.

3. The time to event T_i for the ith subject was generated from the log linear model log$(T_i|Z_i = z_i, X_i = X_i) = \beta z_i + \alpha^T x_i + \epsilon$. $\beta = 0$ when we evaluate the type I error rate and $\beta = 0.7(OR = 2)$ when we evaluate the power. α^T is set to (-0.2,0.013,-0.13,0.09,-0.07,0.06,0.22,0.19,0.11,-0.19). The censoring time C_i is generated from uniform distribution (0,c). The values of c were determined for the desired censoring proportion 25% and 50%.

We repeat the same process for 1500 times. In each replication, the sample size was set to 600. The simulation results were presented in Figure 4.1 and 4.2. The results are mixed for type I error rate and power. Figure 4.1 reveals that the propensity score adjusted log rank test outperform all other methods and retain its nominal level in most simulation scenarios. However, propensity score adjusted conditional log rank test only performs better than Kong and Slud's robust score test in terms of power.

4.4. Discussion

When clinical researchers compare the survival distributions between two treatment groups in a observational study, a Cox proportional hazard model including the binary treatment indicator and observed covariates is commonly used when proportional hazard assumption is reasonable. When such assumption is not valid, Lin and Wei, 1989 and Kong and Slud, 1997 have proposed covariates adjusted robust score statistics but under the assumption that the treatment group indicator is independent of all other covariates, which makes their tests not suitable for observational studies because independence occurs mainly in randomized studies. Heller and Venkatraman, 2004 pro-

posed a non-parametric conditional log rank test for right censored survival data, which does not require the proportional hazard assumption and the independence assumption between treatment group variable and all other covariates. Although this method is desirable for comparing treatment effect in observation studies, the authors used kernel smoothing to estimate the expectation of treatment variable conditional on covariates among the subjects at risk and in such way, covariates are incorporated into the test statistics. Due to the curse of dimensionality on non-parametric smoothing methods such as kernel smoothing. There is a limitation on the number of covariates that can be controlled for and no more than three variables) are allowed for a satisfactory performance.

In this paper, we proposed a simple extension to the conditional log rank test by using propensity score as a summary score of all the confounders. In this way, we reduce the dimension of the covariates and apply the kernel smoothing on the scalar propensity score. Thus, limitation imposed on conditional log rank is removed. The propensity score adjusted conditional log rank could be an robust alternative to confounders adjusted Cox proportional hazard models when the proportional hazard assumption does not hold.

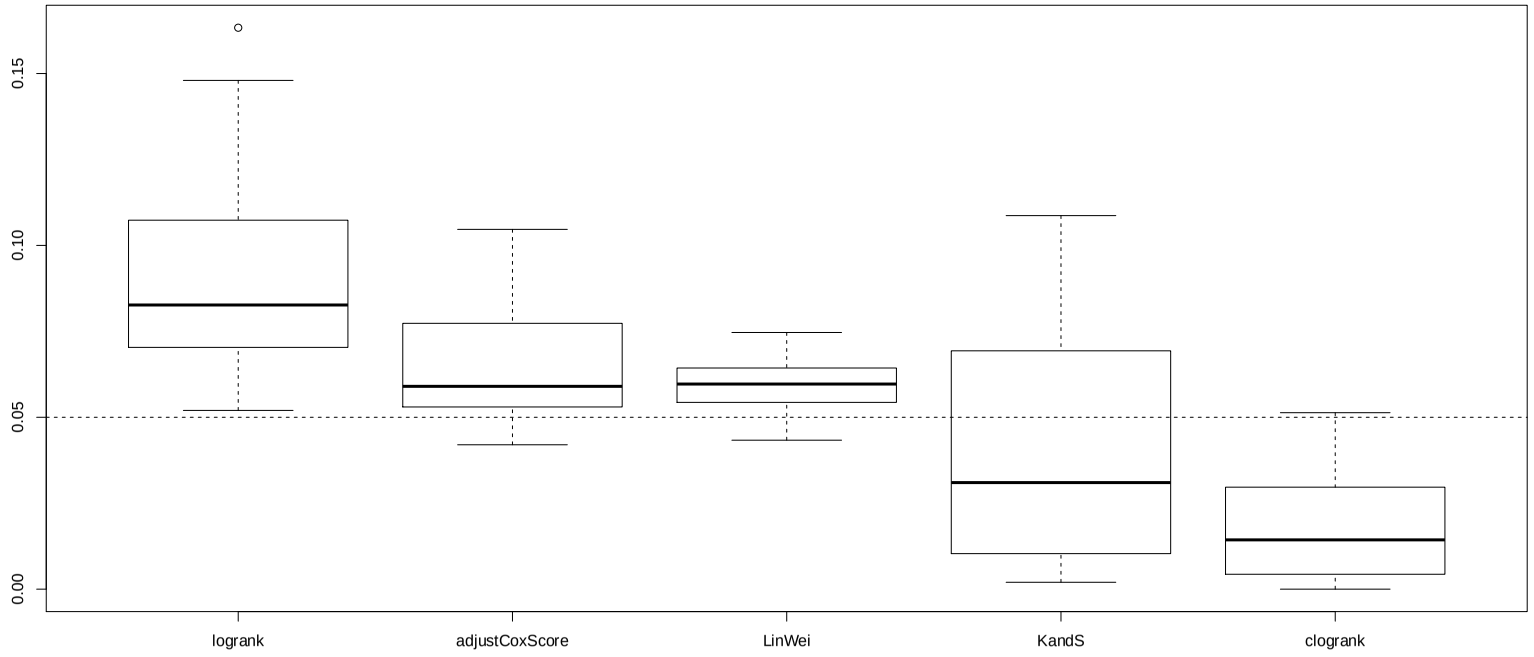


Figure 4.1: Boxplot of type I error rates for unadjusted log rank test, score test based on Cox proportional hazard model, robust score tests (lin and Wei, Kong and Slud), and propensity score adjusted log rank test

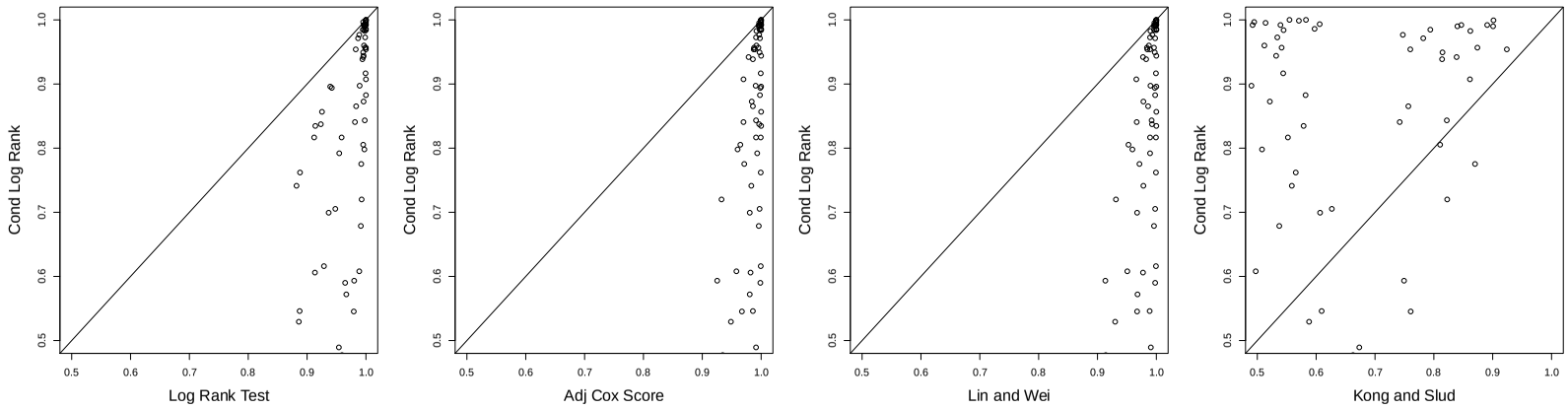


Figure 4.2: Scatter plots of the power for the propensity score adjusted log rank test against unadjusted log rank test, score test based on Cox proportional hazard model, robust score tests (lin and Wei, Kong and Slud)

CHAPTER 5

DISCUSSION

This dissertation made several contributions to the causal inference, particularly two stage IV methods. In chapter 2 we investigated the consistency of 2SPS and 2SRI methods in estimating the causal hazard ratio among compilers. Under principal stratification and potential outcome framework, we assume the potential survival time in each compliance group (e.g. always taker, never taker, complier) follows Weibull distribution with the same shape parameter but different scale parameters for receiving active treatment or control. For 2SPS, the true local treatment effect is the log causal hazard hazard among complier, which is defined as difference in log scale parameters when treated and not treated among compliers, weighted by negative shape parameter. we first showed that the second stage Weibull model including predicted treatment status is equivalent to the Weibull model including binary treatment assignment status (IV) because predicted treatment received status has a one-to-one relationship with IV. This re-parameterization does not change values of linear predictor of a Weibull model. Next, we showed that the 2SPS estimator is equal to the difference between log hazard among subjects assigned to treatment (R=1) and log hazard among subjects assigned to control (R=0), divided by proportion of the complier in the study population. Subjects assigned to either active treatment or control groups are heterogeneous and are mixture of always takers, never takers, and compliers. The distribution of these subjects' survival times does not follow Weibull distribution any more. However, Weibull model assumes these heterogeneous subjects assigned to each arm homogeneous and impose a Weibull distribution upon them. The distributional parameters are chosen in such way that makes this Weibull distribution as close to the true mixture distribution as possible. Thus, 2SPS estimator can not be simplified into an expression consisting of distributional parameters from the complier only. Utilizing this fact, we derived the closed form expression of the probability limit of 2SPS estimator in terms of defined parameters of potential survival times for always taker, never taker, and complier. The difference between this expression of probability limit of 2SPS estimator and the true log causal hazard ratio is the asymptotic bias of 2SPS estimator. To derive the close form expression of asymptotic bias of 2SRI estimator, we can only derive the closed form expression of the asymptotic bias under the assumption that there is no always taker, although we can still evaluate the consistency of 2SRI

estimator without imposing no always taker assumption using simulation. The true model should include an interaction term between the treatment received and the residual term, thus 2SRI is a mis-specified model. Similarly, the second stage 2SRI Weibull model including treatment received indicator and residual term is equivalent to a model including treatment received indicator and treatment assignment indicator (IV). Such re-parameterization does not change the values of linear predictors of Weibull model. Utilizing this relationship between two Weibull models, we derived the closed form expression of asymptotic bias of 2SRI estimator. Through analytic formula and simulations, we showed that when we use two stage IV methods to estimate the causal hazard ratio among compliers, 2SPS estimator is less volatile and perform better than 2SRI estimator when the hazard is a decreasing function ($\alpha < 1$) but both 2SRI and 2SPS estimators have very large variability when the hazard is an increasing function ($\alpha > 1$). Another interesting finding is that the biases of 2SPS and 2SRI estimators are also associated with censoring distribution. The closed form expression of asymptotic bias for 2SPS and 2SRI are useful when we design a two arm trial with possible non-compliance to determine the magnitude of bias of using the two stage IV method in estimating the causal hazard ratio among compliers. In chapter 3, we studied the consistency of two stage IV methods from another perspective. that is, Are 2SPS and 2SRi biased in estimating conditional treatment effect given observed and unobserved covariates? The current two stage modeling framework are very restrictive and may not be suitable for clinical settings involving a single endogenous exposure variable and multiple unmeasured covariates. Under this new framework, we successfully demonstrated the consistency problems of 2SPS and 2SRI in estimating the conditional treatment effect in non-linear model context can be transformed into the omitted-variable bias problems in non-linear models. The latter question is a topic under extensive research. Using this result, we easily revealed that 2SPS and 2SRI are unbiased under the null hypothesis that there is no causal effect for Poisson, logistic, and Cox proportional hazard models. When the treatment effect is not null, the results are mixed. 2SPS and 2SRI are unbiased for Poisson model but the two methods are biased for logistic and Cox proportional hazard model. The biases are influenced by several factors: (1) larger magnitude of treatment effect is associated with larger biases; (2) larger magnitude of the effects of unobserved covariates is associated with larger biases; (3) The magnitude of coefficients of omitted error terms in 2SPS and 2SRI are associated with the biases. Specially for 2SRI, the coefficient of omitted error term can be represented by a dissimilarity metric $\sin(\theta)$. This sine dissimilarity metric measures the dis-

tance between two coefficient vectors of unmeasured covariates in both treatment and outcome models. As two coefficient vectors are more dissimilar, the bias becomes larger. (4) For time to event outcome, we found via simulation that 2 stage IV methods have smaller biases when the hazard is a decreasing function ($\alpha < 1$), and larger biases when the hazard is a increasing function ($\alpha > 1$). This finding is consistent with our results in chapter 2. Overall, 2SRI performs better than 2SPS when estimating the conditional treatment effect. This new two stage modeling framework and related techniques presented in chapter 3 have several immediate extensions. For example, we can use this new framework to assess the bias in estimating the conditional treatment effect when using propensity score adjusted non-linear regression models, rather than using covariates adjusted models. In chapter 4, we discussed the extension of a covariates adjusted non-parametric conditional log rank test by using propensity score as a summary score of measured covariates for observational studies. This extension relaxed the limitation of Heller and Venkatraman's conditional log rank test (Heller and Venkatraman, 2004) on the number of covariates that can be adjusted. Simulation shows that the propensity score adjusted conditional log rank test is satisfactory.

BIBLIOGRAPHY

Abadie, A (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics* 113, 231–263.

Angrist, J, Imbens, G, and Rubin, D (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.

Baiocchi, M, Cheng, J, and Small, D (2014). Instrumental variable methods for causal inference. *Statistics in Medicine* 33, 2297–2340.

Bekelman, J, Mitra, N, Handorf, E, Uzzo, R, Hahn, S, Polsky, D, and Armstrong, K (2015). Effectiveness of Androgen Deprivation Therapy and Radiotherapy for Older Men with Locally Advanced Prostate Cancer. *Journal of Clinical Oncology* 33, 716–722.

Burgess, S (2013). Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Statistics in Medicine* 32, 4726–4747.

Cai, B, Small, D, and Ten Have, T (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Statistics in Medicine* 30, 1809–1824.

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

Gail, M, Wieand, S, and Piantadosi, S (1984). Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates. *Biometrika* 71, 431–444.

Gore, J, Litwin, M, Lai, J, Yano, E, Madison, R, Setodji, C, Adams, J, and Saigal, C (2010). Use of Radical Cystectomy for Patients with Invasive Bladder Cancer. *Journal of the National Cancer Institute* 102, 802–811.

Hadley, J, Yabroff, KR, Barrett, MJ, Penson, DF, Saigal, CS, and Potosky, AL (2010). Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. *Journal of the National Cancer Institute* 102, 1780–1793.

Heller, G and Venkatraman, ES (2004). A nonparametric test to compare survival distributions with covariate adjustment. *Journal of the Royal Statistical Society, Series B* 66, 719–733.

Hernán, M and Hernández-Díaz, S (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* 9, 48–55.

Hernán, M and Robins, J (2006a). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health* 60, 578–586.

Hernán, M and Robins, J (2006b). Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 29, 360–372.

Holland, P (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* 81, 945–960.

Imbens, G and Angrist, J (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, 467–475.

Kong, F and Slud, E (1997). Robust covariate-adjusted logrank tests. 84, 847–862.

Lin, DY and Wei, LJ (1989). The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association* 84, 1074–1078.

Lin, D, Psaty, B, and Kronmal, R (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 54, 948–963.

Lin, N, Logan, S, and Henley, W (2013). Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates. *Biometrics* 69, 850–860.

Martens, E, Pestman, W, Boer, A de, Belitser, S, and Klungel, O (2006). Instrumental variables: application and limitations. *Epidemiology* 17, 260–267.

Mitra, N and Heitjan, D (2007). Sensitivity of the hazard ratio to nonignorable treatment assignment in an observational study. *Statistics in Medicine* 26, 1398–1414.

Mortensen, E, Halm, E, Pugh, M, Copeland, L, Metersky, M, Fine, M, Johnson, C, Alvarez, C, Frei, C, Good, C, Restrepo, M, Downs, J, and Anzueto, A (2014). Association of azithromycin with mortality and cardiovascular events among older patients hospitalized with pneumonia. *Journal of the American Medical Association* 307, 1629–1635.

Nagelkerke, N, Fidler, V, Bernsen, R, and Borgdorff, M (2000). Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Statistics in Medicine* 19, 1849–1864.

Nallamothu, B and Hayward, R (2008). Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 118, 1294–1303.

Neuhaus, J and Jewell, N (1993). A Geometric Approach to Assess Bias Due to Omitted Covariates in Generalized Linear Models. *Biometrika* 80, 807–815.

Newhouse, J and McClellan, M (1998). Econometrics in outcomes research:the use of instrumental variables. *Annual Review Public Health* 19, 17–34.

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

Palmer, T, Thompson, J, Tobin, M, Sheehan, N, and Burton, P (2008). Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *International Journal of Epidemiology* 37, 1161–1168.

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

Rosenbaum, P and Rubin, D (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 41–55.

Rubin, D (1986). Statistics and causal inference–Which ifs have causal answers. *Journal of the American Statistical Association* 81, 961–962.

Rubin, D (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* 5, 472–480.

Rubin, D (2005). Causal Inference Using Potential Outcomes: Design, Modelling, Decisions. *Journal of the American Statistical Association* 100, 322–331.

Struthers, C and Kalbfleisch, J (1986). Misspecified proportional hazard models. *Biometrika* 73, 363–369.

Tan, H, Norton, E, Ye, Z, Hafez, K, Gore, J, and Miller, D (2012). Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *Journal of the American Medical Association* 307, 1629–1635.

Ten Have, T, M, J, and M., C (2003). Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* 22, 1255–1283.

Ten Have, T, Joffe, M, and Cary, M (2003). Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* 22, 1255–1283.

Terza, J, Basu, A, and Rathouz, P (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27, 531–543.

Wan, F, Small, D, Justin, B, and Mitra, N (2015). Bias in estimating the causal hazard ratio when using two-stage instrumental variable methods. *Statistics in Medicine*. DOI: 10.1002/sim.6470.

Wang, J (2012). Dose as instrumental variable in exposure-safety analysis using count models. *Journal of Biopharmaceutical Statistics* 22, 565–581.

Warde, P, Mason, M, Ding, K, Kirkbride, P, Brundage, M, Cowan, R, Gospodarowicz, M, and Sanders, K (2011). Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: a randomised, phase 3 trial. *Lancet* 378, 2104–2111.

Widmark, A, Klepp, O, Solberg, A, Damber, J, Angelsen, A, Fransson, P, Lund, J, Tasdemir, I, Hoyer, M, and Wiklund, F (2009). Endocrine treatment, with or without radiotherapy, in locally advanced prostate cancer (SPCG-7/SFUO-3): an open randomised phase III trial. *Lancet* 373, 301–308.

Zeger, S, Liang, K, and Albert, P (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060.