Simulating from marginal structural model under violation of positivity assumption

Huang, Zhicong

Simulating from marginal structural model under violation of positivity assumption

Zhicong Huang

Daily supervisor: Dr. Marta Spreafico
Second supervisor: Prof.dr. Marta Fiocco

Defended on November 8, 2023

MASTER THESIS
STATISTICS AND DATA SCIENCE
UNIVERSITEIT LEIDEN

Universiteit
Leiden
The Netherlands

Abstract

Marginal structural models (MSMs) combined with the method of inverse probability of treatment weighting (IPTW) are frequently used to estimate the causal effects of treatment in longitudinal studies. The validity of the IPTW estimator relies heavily on the positivity assumption, that does not always hold in practice, especially in a longitudinal setting with time-varying treatments and confounding. The aim of this thesis is hence to investigate and assess the effect of structural positivity violations on the IPTW estimator in a longitudinal survival setting, where multiple treatments are sequentially assigned over time and time-dependent confounding is present.

Two simulation approaches to generate longitudinal data from a known survival MSM under the violation of the positivity assumption are introduced. Specifically, the simulation algorithms proposed by Havercroft and Didelez (2012) [1] and Keogh *et al.*(2021) [2] are extended to incorporate positivity violations. To investigate the effect of structural violations on the performance of the IPTW estimator, several simulation scenarios are considered by increasing the severity of the violation and varying the sample sizes. Two performance measures (i.e., bias and mean square error) are used to assess the violation effect on the accuracy of the IPTW estimator.

Results showed that as the degree of positivity violation intensifies, both the bias and mean square error of the IPTW estimate for causal parameters tend to increase. Increasing the sample size mitigates the bias and the mean square error of the IPTW estimator but does not eliminate the violation issues. Indeed, the bias of the IPTW estimator is composed of two parts: the bias arising from the structural positivity violation and the bias due to the finite sample bias. Only the second one can be mitigated by increasing sample size. Furthermore, the performance of the IPTW estimator using simulation approach II is less sensitive to varying sample size compared to the IPTW estimator of simulation approach I. This difference is attributed to the different treatment receiving mechanisms employed by the two simulation procedures.

In conclusion, this study suggests violating the positivity assumption has a detrimental impact on the accuracy of the IPTW estimator in a longitudinal survival framework. Therefore, it is imperative for analysts to systematically evaluate the presence of positivity violations at all time-points when conducting causal analyses using real data.

Contents

Chapter 1

Introduction

Relying on the counterfactual framework, several causal inference approaches for estimating causal effects in longitudinal studies have been developed. Due to its straightforward implementation, marginal structural models (MSM) estimated using inverse probability of treatment weighting method (IPTW) have become a standard tool for investigating the causal effect of a time-dependent treatment (or exposure) on the clinical outcome of interest in the presence of time-dependent confouders [3, 4]. This approach is based on three key assumptions, also named *identifiability assumptions*: positivity, consistency, and (conditional) exchangeability [4, 5, 6]. Specifically, positivity is satisfied when, for any combination of the covariates, there exists a non-null probability of receiving each possible treatment or being unexposed. However, in clinical research, it is common to encounter challenges in obtaining adequate data support to ensure positivity. This can pose a risk of violating the positivity assumption, leading to substantial bias and increased variance in the IPTW estimator [7]. In fact, the accuracy of the IPTW estimator relies heavily on this assumption [8] and, when it is violated, the IPTW estimator is undefined.

There exist two types of positivity violations. *Structural violations* of positivity correspond to situations in which a subject is prevented to receive a certain treatment, for example, if certain characteristics of this subject construct a contraindication to a particular treatment. On the other hand, *practical violations* of positivity represent situations in which the allocation of a certain treatment to a given subgroup of subjects is not observed by chance, even though it is theoretically possible for this subgroup to receive such treatment. Several simulation studies have been conducted to investigate the impact of positivity violation on IPTW estimator of casual effect in the point treatment setting. Neugebauer and van der Laan (2005) has shown that both practical and structural positivity violations can lead to substantial bias in the IPTW estimator in the point treatment setting [9]. Léger *et al.* (2022) illustrated that even a near-violation of the positivity assumption can impact the bias and precision of the IPTW estimator [10]. Petersen *et al.* (2012) discussed that the sparsity in the data due to positivity violations may increase bias with or without an increase in the variance of the IPTW estimator [7].

The main aim of this thesis is to extend these investigations into a longitudinal survival context, where multiple treatments are sequentially assigned over time and time-dependent confounding is present. Assessing the effect of positivity violations on IPTW estimator in a longitudinal setting is not straightforward. In practice, practical violations of positivity can readily occur in a setting with multiple time-points. This is due to the sequential form of the positivity assumption, which requires that the conditional probability of each potential treatment history remains positive. Consequently, it becomes easy to obtain a small conditional probability that spans multiple time-points. This affects the weights used in IPTW, which are determined by the product of time point-specific treatment probabilities given the past.

To evaluate the impact of structural positivity violations in longitudinal studies, it is essential to employ a well-defined simulation procedure. Havercroft and Didelez (2012) [1] and Keogh *et al.* (2021) [2] proposed two simulation algorithms to generate longitudinal data from a known MSM for survival outcomes in the presence of time-varying confounding. By carefully setting up positivity violations whithin these procedures, we can effectively assess the effect of structural positivity violations on the performance of the IPTW estimator in a longitudinal survival context. Our simulation study designs are hence extensions of the approaches by Havercroft and Didelez (2012) [1] and Keogh *et al.* (2021) [2] where structural violations of the positivity assumption are intentionally introduced into the simulation procedures. Several simulation scenarios, i.e., by increasing the severity of the violation and varying the sample sizes, are then investigated to examine the effect of these violations on the performance of the IPTW estimator performing multiple repetitions. Results for the various scenarios are evaluated using two metrics: the bias and the mean square error [11]. Furthermore, the various simulation settings and the special properties of MSM between the two simulation algorithms will be compared and discussed.

The remainder of this thesis is organised as follows. In Chapter 2, we introduce the basic notions and the mathematical notation of causal inference and marginal structural models. In Chapter 3, we illustrate the basic concepts of survival analysis and explain the related models in the simulation procedures. In Chapter 4, we introduce the idea of simulation study and provide a general explanation on the simulating mechanisms of the two simulation algorithms by Havercroft and Didelez (2012) [1] and Keogh *et al.* (2021) [2]. In Chapter 5, we describe the violation of positivity assumption setup and we extend the two simulation algorithms by including violations. Results for the two extended algorithms under different simulation scenarios are presented in Chapter 6. In Chapter 7, we concludes this thesis discussing the implications of our study and possible future extensions.

Chapter 2

Causal inference

The aim of this chapter is to introduce the basic notions and the mathematical notation of causal inference and marginal structural models. First, we discuss the concepts of causal effect, counterfactual outcomes, and identifiability assumptions. Then, we illustrate an important type of causal diagram called directed acyclic graphs and the concept of confounding. We end the chapter by introducing marginal structural model estimated by inverse probability of treatment weighting.

2.1 Causal effect

In the counterfactual framework [12], the *cause* usually refers to an action, an exposure or a state. The *effect* is then a consequence of the defined cause. The causal effect of an exposure A on an outcome Y is defined as a contrast between what would potentially happen under competing actions or states (i.e., exposed or unexposed).

To make the causal intuition amenable to mathematical and statistical analysis we now introduce some notation. Let us consider a dichotomous treatment variable A (for example, 1 if treated, 0 if untreated) and a dichotomous outcome variable Y (for example, 1 if death, 0 if survival). Let us define:

- $Y^{a=1}$ as the outcome variable that would have been observed under the treatment value $a = 1$;
- $Y^{a=0}$ as the outcome variable that would have been observed under the treatment value $a = 0$.

The variables $Y^{a=1}$ and $Y^{a=0}$ are referred to as *potential outcomes* or *counterfactual outcomes*. The term "potential" is to emphasize that, depending on the treatment that is received, only one of these two outcomes can actually be observed and the other remains *potential*. The term "counterfactual" is to emphasize that these outcomes represent situations that may not actually occur, which is the *counter-to-the-fact* situations. In fact, for each individual there are two possible ways that treatment A could be assigned

Table 2.1: Potential outcomes for a single subject.

A	Y	$Y^{a=0}$	$Y^{a=1}$
0	0	0	?
0	1	1	?
1	0	?	0
1	1	?	1

and two possible outcomes Y that can be observed. The resulting four combinations are shown in Table 2.1 along with the corresponding factual and counterfactual outcomes, where "?" represents the missing data for the potential outcome that is not observed. In the following we will refer to Y^a to indicate the counterfactual outcome under treatment $A = a$.

We can now provide a formal definition of the causal effect of a binary treatment A for a single subject.

Definition 2.1.1 *The binary treatment A has a causal effect on the outcome Y of an individual if*

$$Y^{a=1} \neq Y^{a=0}. \tag{2.1}$$

We notice that only one counterfactual outcome is observed for each individual. Due to the unobserved counterfactual, the *causal contrast* $Y^{a=1} - Y^{a=0}$ cannot be individually evaluated. Nevertheless, in general the interest is directed towards the average causal effect for a population rather than for a single individual.

Definition 2.1.2 *The binary treatment A has an average causal effect in a population if and only if*

$$E[Y^{a=1}] \neq E[Y^{a=0}]. \tag{2.2}$$

Since we are considering the case of dichotomous Y, condition (2.2) can be rewritten as:

$$P(Y^{a=1} = 1) \neq P(Y^{a=0} = 1). \tag{2.3}$$

In the following, we refer to $P(Y^a)$ as the counterfactual risk of experiencing event $Y = 1$ under treatment $A = a$.

2.1.1 Causation versus association

When considering the relationship between treatment and outcome, it is important to distinguish between association and causation. *Causation* implies that the treatment variable causes the direct effect in the outcome. For a binary treatment A and a dichotomous outcome Y, this means that condition (2.3) holds. Statistical *association* implies that knowing the value of the treatment variable provides information on the value of the outcome, but does not necessarily imply that the former causes the latter. We say that

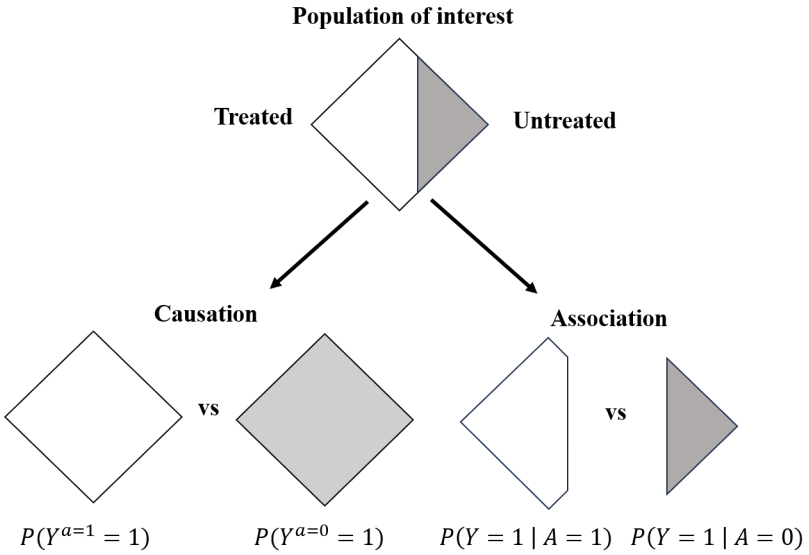

Figure 2.1: Comparison between causation and association [12] in a population of interest with binary treatment A and binary outcome Y.

treatment A and outcome Y are associated when $P(Y = 1|A = 1) \neq P(Y = 1|A = 0)$. Figure 2.1 depicts the causation-association difference. Causation corresponds to a contrast between the pseudo-group had all individuals been treated and the pseudo-group had all individuals been untreated, whereas association corresponds to a contrast between two disjoint subsets of the population: the observed treated group and the observed untreated group.

Randomized experiments are considered perfect for establishing causation. In a randomized experiment, study participants are randomly assigned to different treatment groups, which creates two (or more) comparable groups with similar characteristics, both observed and unobserved. As a result, causation coincides with association and effect measures can be consistently estimated despite the missing data. This is not true for observational studies: when drawing causal inferences from observational data, it is essential to ensure that causation coincides with association.

2.2 Identifiability assumptions

We need three assumptions to ensure that an observational study can be viewed as a randomized experiment: (conditional) exchangeability, consistency, and positivity. Under randomized experiment, the missing data occurs by chance. As a result, the measure effect can be consistently estimated despite the missing data. These three assumptions are referred to as *identifiability assumptions* [12].

Exchangeability means that the risk of death in the treated group would have been the same as the risk of death in the untreated group had individuals in the treated group received the treatment given to those in the untreated group. This implies that

$$P(Y^a = 1|A = 1) = P(Y^a = 1|A = 0) = P(Y^a = 1), \tag{2.4}$$

which is equivalent to

$$Y^a \perp\!\!\!\perp A \quad \forall a. \tag{2.5}$$

Consistency means that the observed outcome for every treated individual equals the outcome if individual receive the treatment, and that the observed outcome for every untreated individual equals the outcome if individual has remained untreated. In other words

$$P(Y^a = 1|A = a) = P(Y = 1|A = a), \tag{2.6}$$

which is equivalent to

$$Y^a = Y \quad \text{if } A = a. \tag{2.7}$$

By combining exchangeabability (2.5) and consistency (2.7) assumptions, it results in

$$P(Y^a = 1) = P(Y^a = 1|A = a) = P(Y = 1|A = a). \tag{2.8}$$

This means that under the exchangeability and consistency assumptions, the causal effect can be estimated via observed values.

We now consider the presence of a prognostic factor L, which was measured before treatment was assigned (for example, 1 if the individual was in critical condition, 0 otherwise). A randomized experiment where the randomization is conditional on a prognostic factor L is referred as as the *conditionally randomized experiment*. In a conditionally randomized experiment, randomization probabilities vary among different subsets of subjects based on the values of prognostic factor L. Within each subset, a marginally randomized experiment is assumed to be conducted. Thus, although conditional randomization does not guarantee marginal exchangeability (2.5), it guarantees conditional exchangeability.

Conditional exchangeability holds if and only if

$$P(Y^a = 1|A = 1, L = l) = P(Y^a = 1|A = 0, L = l) = P(Y^a = 1|L = l) \quad \forall l, \tag{2.9}$$

which is equivalent to

$$Y^a \perp\!\!\!\perp A|L \quad \forall a. \tag{2.10}$$

Similarly to (2.8), under conditional exchangeability (2.9), we can imply that

$$P(Y^a = 1|L = l) = P(Y = 1|A = a, L = l). \tag{2.11}$$

In marginal randomized experiments, the probabilities $P(A = 1)$ and $P(A = 0)$ are both positive by design. In conditional randomized experiments, the conditional probabilities

$P(A = 1|L = l)$ and $P(A = 0|L = l)$ are also positive by design for all levels of the variable L that are eligible for the study. Thus the positivity assumption can be defined as follows.

Positivity means the probability of receiving every value of treatment is greater than zero, i.e., positive. Positivity holds if and only if

$$P(A = a|L = l) > 0 \tag{2.12}$$

for all values l with $P(L = l) \neq 0$ in the population of interest.

In observational studies where stratification given by prognostic factor L is present, the causal effect cannot be computed in subsets $L = l$ in which there are only treated, or untreated, individuals. Mathematically this can be view expanding (2.11) as follows:

$$\begin{aligned} P(Y^a = 1|L = l) &= P(Y = 1|A = a, L = l) = \\ &= \frac{P(Y = 1, A = a, L = l)}{P(A = a, L = l)} = \frac{P(Y = 1, A = a, L = l)}{P(A = a \mid L = l)P(L = l)} \end{aligned} \tag{2.13}$$

that is undefined when $P(A = a|L = l) = 0$, i.e., if the positivity assumption is violated.

Provided that the three identifiability conditions hold, the marginal counterfactual risk is obtained in each strata $L = l$. Then, averaging across all strata defined by L yields the correct marginal counterfactual risk in the population

$$P(Y^a = 1) = \sum_l P(Y^a = 1|L = l)P(L = l). \tag{2.14}$$

2.3 Directed acyclic graphs

Causal effects, as explained in the preceding sections, can be effectively represented using graphs. Causal graphs provide a powerful visual tool for understanding causal relationships between variables in complex systems. Figure 2.2 shows an example of causal graphs, also known as directed acyclic graphs (DAGs) [13]: the vertices (nodes) represent variables and the edges represent direct causal effects. "Directed" means the edges imply a direction. For example, the arrow pointing from L to A indicates that L may cause A, but not the other way around. "Acyclic" means there are no cycles, which indicates that a variable cannot cause itself, either directly or through another variable.

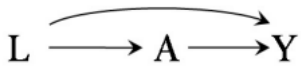

Figure 2.2: Directed Acyclic Graph (DAG).

A fundamental property of causal DAG is that, conditional on its direct causes, any variable on the DAG is independent of any other variable that is not its cause. This assumption, referred to as the causal Markov assumption, implies that in a causal DAG the common causes of any pair of variables in the graph must be also in the graph. This is mathematically equivalent to the Markov factorization:

$$P(L, A, Y) = P(L)P(A|L)P(Y|A, L).$$

DAGs offer a distinct advantage by providing a natural framework for simulating data, as they explicitly encode conditional independencies within the graph structure.

It is also important to note that in Figure 2.2, the association between treatment A and outcome Y arises from two types of sources: (i) the causal path $A \longrightarrow Y$ that represents the causal effect of treatment A on outcome Y, and (ii) the non-causal path $A \longleftarrow L \longrightarrow Y$ that links treatment A and outcome Y through their common cause L. This non-causal path, which has an arrow pointing into the treatment, is commonly referred to as "*backdoor path*".

2.4 Confounding

The primary limitation in conducting causal analysis through observational studies lies in the presence of confounding.

Definition 2.4.1 *Confounding is the bias that arises when treatment and outcome share common causes.*

In the presence of confounding factors, the association is not causation even if the study population is arbitrarily large. In general, a *confounding variable* or *confounder* is a variable correlated with both the dependent and independent variables. As in Figure 2.2, the treatment A and the outcome Y share a common cause L. If the common cause L did not exist, then the only path between treatment and outcome would be $A \longrightarrow Y$, and thus the entire association between A and Y would be due to the causal effect of A on Y. But the presence of the common cause L here creates an additional source of association between the treatment A and the outcome Y, which is referred to as confounding for the effect of A on Y, so association is not causation.

The traditional definition of confounder refers as a variable which meets the following three conditions: (1) it is associated with the treatment, (2) it is associated with the outcome conditional on the treatment, and (3) it does not lie on a causal pathway between treatment and outcome. Adjusting for confounders is hence crucial for causal inference.

We can also relate the confounding to exchangeability. If there is no common causes of treatment and outcome, it is a marginally randomized experiment in which marginal exchangeability holds and confounding is not expected. Therefore, marginal exchangeability is equivalent to no confounding by either measured or unmeasured covariates.

Furthermore, if there are common causes of both the treatment A and the outcome variables Y, but a subset L of measured variables (which are non-descendants of treatment A) is sufficient to block all backdoor paths, then conditioning on the covariates L will block all backdoor paths. Consequently, conditional exchangeability will hold. For this reason, the conditional exchangeability assumption is also called *no unmeasured confounding*.

2.5 Time-dependent treatment and confounding

We now consider a time-dependent dichotomous treatment variable A_k that may change at every time-point k of the follow-up, where $k = 0, 1, 2, ..., K$. Let $\bar{\boldsymbol{A}}_k = (A_0, A_1, ..., A_k)$ denote the treatment history as from time 0 to time k. In observational studies, decisions about treatment often depend on prognostic factors which may change over time. Let $\boldsymbol{L}_k$ be the vector of covariates at time k, whose history at time k is denote by $\bar{\boldsymbol{L}}_k = (\boldsymbol{L}_0, \boldsymbol{L}_1, ..., \boldsymbol{L}_k)$. In this time-dependent setting, it can be assumed that the treatment A_k at time k depends on the evolution of an time-dependent covariates $\bar{\boldsymbol{L}}_k$, and DAGs can again be used to represent the interconnections between treatment, prognostic factors and outcome variables.

For example, Figure 2.3 represents a randomized experiment in which treatment A_k at each time k depends on prior treatment A_{k-1} and measured covariate history $\bar{\boldsymbol{L}}_k$. At each time point k, treatment A_k shares unmeasured causes $\boldsymbol{U}_k$ with the outcome Y, which indicates the confounding. However, the backdoor path, for example, of A_0 is $A_0 \longleftarrow \boldsymbol{L}_0 \leftarrow \boldsymbol{U}_0 \longrightarrow Y$, which can be blocked by conditioning on $\boldsymbol{L}_0$, which is measured. Thus, if data on $\boldsymbol{L}_0$ is collected for all individuals, there would be no unmeasured confounding for the effect of A_0. We then say that $\boldsymbol{L}_0$ is the confounder for the effect of A_0.

If we consider all time points and we want to estimate the causal effects on the outcome Y of treatment strategies, then, at each time k, the covariate history $\bar{\boldsymbol{L}}_k$ and treatment history $\bar{\boldsymbol{A}}_{k-1}$ will be needed to block the backdoor paths between treatment A_k and the outcome Y to ensure no unmeasured confounding. We then say that the time-dependent covariates in $\boldsymbol{L}_k$ are *time-dependent confounders* for the effect of the time-dependent treatment $\bar{\boldsymbol{A}}$ on outcome Y at several times k in the study.

2.5.1 Sequential identifiability assumptions

In the presence of time-dependent treatment and confounding, the identifiability assumptions presented in Section 2.2 must be adapted to a sequential setting. Indeed, causal inference with time-dependent treatments requires adjusting for the time-dependent covariates $\bar{\boldsymbol{L}}_k$ to achieve the identifiability assumption at each time point k, i.e., sequential identifiability assumptions.

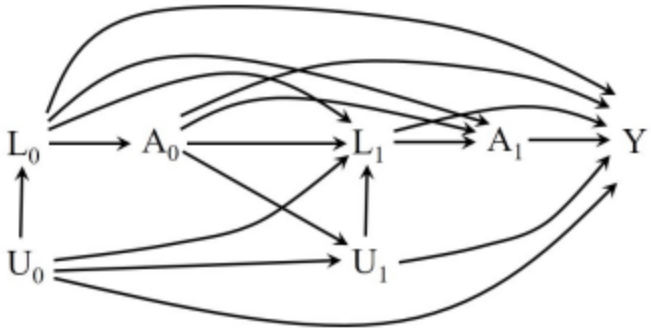

Figure 2.3: DAG for time-dependent treatment and time-dependent confounding.

Sequential conditional exchangeability means that, for any strategy g, the treated and the untreated at each time k are exchangeable for Y^g conditional on prior covariate history $\bar{\boldsymbol{L}}_k$ and any observed treatment history $\bar{\boldsymbol{A}}_{k-1} = g(\bar{\boldsymbol{A}}_{k-2}, \bar{\boldsymbol{L}}_{k-1})$ compatible with strategy g:

$$Y^g \perp\!\!\!\perp A_k | \bar{\boldsymbol{A}}_{k-1} = g(\bar{\boldsymbol{A}}_{k-2}, \bar{\boldsymbol{L}}_{k-1}), \bar{\boldsymbol{L}}_k \tag{2.15}$$

for all strategies g and $k = 0, 1, ..., K$.

In addition to sequential exchangeability, causal inference involving time-dependent treatments also requires a sequential version of consistency assumption and positivity assumption.

Sequential consistency requires that for any strategy g:

$$Y^g = Y \quad \text{if } A_k = g(\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k) \tag{2.16}$$

at each time k.

If for static strategies, the identification of effects of time-varying treatments on Y requires weaker consistency conditions:

$$Y^{\bar{a}} = Y \quad \text{if } \bar{\boldsymbol{A}} = \bar{\boldsymbol{a}} \tag{2.17}$$

Sequential positivity requires that

$$P(A_k = a_k | \bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{L}}_k = \bar{\boldsymbol{l}}_k) > 0 \tag{2.18}$$

for all $(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{l}}_k)$ with $P(\bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{L}}_k = \bar{\boldsymbol{l}}_k) \neq 0$ in the population of interest.

2.5.2 Causal effects of time-dependent treatment

Let $\bar{\boldsymbol{a}} = (a_0, a_1, \ldots, a_K)$ denote the complete treatment strategy, where element a_k is the treatment at time k $(k = 0, \ldots, K)$. In this situation, there are many possible causal effects for the time-dependent treatment, each of them is defined by a contrast of outcomes under two particular treatment strategies.

Definition 2.5.1 *The average causal effect for a time-dependent treatment is given by*

$$E[Y^{\bar{\boldsymbol{a}}_1}] - E[Y^{\bar{\boldsymbol{a}}_2}] \tag{2.19}$$

and compares complete treatment strategies $\bar{\boldsymbol{a}}_1$ *and* $\bar{\boldsymbol{a}}_2$*, with* $\bar{\boldsymbol{a}}_1 \neq \bar{\boldsymbol{a}}_2$*.*

In case of time-dependent binary treatment, elements a_k are 1 if the subject is treated at time k, or 0 if untreated. An average causal effect commonly considered for that case is the comparison of *always treated* strategy (i.e., $a_k = 1$ for all $k = 0, \ldots, K$) versus *never treated* one (i.e., $a_k = 0$ for all $k = 0, \ldots, K$):

$$P\left(Y^{\bar{\boldsymbol{a}}_1=(1,1,\ldots,1)}\right) - P\left(Y^{\bar{\boldsymbol{a}}_2=(0,0,\ldots,0)}\right). \tag{2.20}$$

2.6 Inverse Probability Treatment Weighting

Inverse Probability of Treatment Weighting (IPTW) is a statistical method used in causal inference to address confounding in observational studies. Under the identifiability assumptions, IPTW allows to create a hypothetical population in which every individual appears as a treated and as an untreated individual. This hypothetical population is known as the *pseudo-population*. By making the treated and untreated groups comparable, the effects of confounders $\boldsymbol{L}$ on treatment A are mitigated in the pseudo-population.

The pseudo-population is created by weighting each individual by the inverse of the probability of receiving the treatment level received conditional on confounders. These weights are computed as

$$W^A = \frac{1}{f(A|\boldsymbol{L})}, \tag{2.21}$$

where $f(A|\boldsymbol{L})$ is the conditional density of treatment A given covariates $\boldsymbol{L}$. These weights are called *unstabilized weights* and the pseudo-population thus created is twice as large as the original population under a dichotomous treatment.

Since the mean of W^A is expected to be 2, we introduce the *stabilized weights* whose mean is expected to be 1. The pseudo-population created by the stabilized weights has the same size as the original population, and maintains the same estimate of the causal effect as in the case of the unstabilized weights. Stabilized weights is defined as:

$$SW^A = \frac{f(A)}{f(A|\boldsymbol{L})}, \tag{2.22}$$

where $f(A)$ is the density of treatment A.

Under the identifiability assumptions, association is causation in both the pseudo-population and the original population. To create the pseudo-population, we need to calculate the stabilized weights for each subject i:

$$sw_i^A = \frac{f(A_i = a)}{f(A_i = a|\boldsymbol{L}_i = \boldsymbol{l}_i)}, \tag{2.23}$$

where $\boldsymbol{l}_i$ is the observed values of covariates $\boldsymbol{L}_i$. However, in the observational studies, both $f(A)$ and $f(A|\boldsymbol{L})$ are generally unknown. In case of binary treatment, we can fit a logistic model (or an equivalent binary regression model) to estimate $P(A_i = a|\boldsymbol{L}_i = \boldsymbol{l}_i)$ for each individual i:

$$\text{logit}P(A_i = 1|\boldsymbol{L}_i = \boldsymbol{l}) = \alpha_0 + \boldsymbol{\alpha}_1\boldsymbol{l}. \tag{2.24}$$

To estimate $P(A_i = a)$ we can use a non-parametric estimator based on the empirical data or fit a saturated logistic model for $P(A_i = 1)$ with only the intercept. Finally, to create the pseudo-population, the treated individuals with covariates $\boldsymbol{L}$ are weighted by $\hat{P}(A = 1)/\hat{P}(A = 1|\boldsymbol{L})$, and the untreated by $[1 - \hat{P}(A = 1)]/[1 - \hat{P}(A = 1|\boldsymbol{L})]$.

2.6.1 IPTW for time-dependent treatment

In the context of time-dependent treatment and confounders, the model cannot possibly be saturated and the stabilized weights could result in narrower confidence intervals than unstabilized weights. For this reason,stabilized weights are preferred usually here.

When treatment and confounders are time-dependent, the IPT weights for time-fixed treatment in Equation (2.22) need to be generalized. For a time-dependent treatment $\bar{\boldsymbol{A}} = (A_0, A_1, ..., A_K)$ and time-dependent confounders $\bar{\boldsymbol{L}} = (\boldsymbol{L}_0, \boldsymbol{L}_1, ..., \boldsymbol{L}_K)$ with times $k = 0, 1, ..., K$, the stabilized weight is given by:

$$SW^{\bar{\boldsymbol{A}}} = \prod_{k=0}^{K} \frac{f(A_k|\bar{\boldsymbol{A}}_{k-1})}{f(A_k|\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k)}, \tag{2.25}$$

where A_{-1} is defined to be 0.

When the identifiability conditions hold, these IPT weights create a pseudo-population ps where the mean of $Y^{\bar{\boldsymbol{a}}}$ is identical to that in the actual population. Thus the average causal effect $E[Y^{\bar{\boldsymbol{a}}_1}] - E[Y^{\bar{\boldsymbol{a}}_2}]$ in Equation (2.19) is equivalent to the average causal effect in the pseudo-population, i.e., $E_{ps}[Y^{\bar{\boldsymbol{a}}_1}] - E_{ps}[Y^{\bar{\boldsymbol{a}}_2}]$.

To estimate the stabilized weights in the case of binary treatment, similarly to before, we can fit a logistic regression model to estimate the conditional probability of a dichotomous treatment $P(A_k = 1|\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k)$ at each time k, and use it as the estimate of $f(A_k|\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k)$. The numerator of the weights can be estimated in a similar way.

2.7 Marginal structural models

Models for the marginal mean of a counterfactual outcome are referred to as *marginal structural models*. IPTW is commonly used in conjunction with Marginal Structural Models (MSMs) to estimate the causal effect of an exposure or treatment on an outcome of interest, while accounting for (time-dependent) confounders.

2.7.1 MSM with time-fixed treatment

A linear MSM for the mean outcome under the dichotomous treatment a can be expressed as:

$$E[Y^a] = \gamma_0 + \gamma_1 a. \tag{2.26}$$

In this model, parameter γ_1 is equal to $E[Y^{a=1}] - E[Y^{a=0}]$, which is the average causal effect of treatment A on outcome Y. To estimate γ_1, we can use IPTW to construct a pseudo-population (see Section 2.6), and then fit a regression model

$$E[Y|A] = \theta_0 + \theta_1 A \tag{2.27}$$

to the pseudo-population data to estimate θ_1. The estimate $\hat{\theta}_1$ of the associational parameter in the pseudo-population is also a consistent estimator of the causal effect γ_1 in the original population. In addition, the MSM in Equation (2.26) is saturated since the number of the unknown parameters (the counterfactual quantities) are the same on the right and on the left of the equations.

In case of dichotomous outcome Y, we can consider Equation (2.26) as a *marginal structural logistic model* like

$$\text{logit}\left[P(Y^a = 1)\right] = \gamma_0 + \gamma_1 a, \tag{2.28}$$

whose parameters can be estimated from the logistic regression model

$$\text{logit}\left[P(Y = 1|A)\right] = \theta_0 + \theta_1 A \tag{2.29}$$

fitted on the pseudo-population.

2.7.2 MSM with time-dependent treatment

When treatment is time-dependent, the linear MSM in Equation 2.26 can be rewritten considering the complete treatment strategy $\bar{\boldsymbol{a}}$ as follows:

$$E[Y^{\bar{\boldsymbol{a}}}] = \gamma_0 + g\left(\boldsymbol{\gamma}; \bar{\boldsymbol{a}}\right) \tag{2.30}$$

where $g(\cdot)$ is a function (to be specified) of the complete treatment strategy that combines information from the time-dependent treatment. For example, by hypothesizing that the effect of treatment history $\bar{\boldsymbol{a}}$ on the mean outcome increases linearly as a function of the cumulative treatment $\text{cum}(\bar{\boldsymbol{a}}) = \sum_{k=0}^{K} a_k$, MSM in (2.30) can be rewritten as

$$E[Y^{\bar{\boldsymbol{a}}}] = \gamma_0 + \gamma_1 \text{cum}(\bar{\boldsymbol{a}}) \tag{2.31}$$

for all strategy $\bar{\boldsymbol{a}}$. In this setting, the average causal effect $E[Y^{\bar{\boldsymbol{a}}}] - E[Y^{\bar{\boldsymbol{a}}=\bar{\boldsymbol{0}}}]$ is equal to $\gamma_1 \text{cum}(\bar{\boldsymbol{a}})$.

To estimate the parameters of MSM in Equation (2.31), as before, we can fit a regression model to the pseudo-population created by IPTW:

$$E[Y|\bar{\boldsymbol{A}}] = \theta_0 + \theta_1 \text{cum}(\bar{\boldsymbol{A}}), \quad \text{with} \quad \text{cum}(\bar{\boldsymbol{A}}) = \sum_{k=0}^{K} A_k, \tag{2.32}$$

where A_k is the treatment value at time k. Under the identifiability conditions, the estimate of θ_1 is consistent for the causal parameter γ_1.

In case of dichotomous outcome Y, we can consider Equation (2.31) as a *marginal structural logistic model* like

$$\text{logit}\left[P(Y^{\bar{\boldsymbol{a}}} = 1)\right] = \gamma_0 + \gamma_1 \text{cum}(\bar{\boldsymbol{a}}), \tag{2.33}$$

whose parameters can be estimated from the logistic regression model

$$\text{logit}\left[P(Y = 1|\bar{\boldsymbol{A}})\right] = \theta_0 + \theta_1 \text{cum}(\bar{\boldsymbol{A}}) \tag{2.34}$$

fitted on the pseudo-population.

2.8 Collapsibility

Collapsibility is a property of certain statistical estimators, including those used in MSMs, where the adjustment for other variables does not affect the estimated association of interest. In the case of collapsible estimators, the conditional and marginal associations are equal, and thus the marginal measure can be expressed as a weighted average of the conditional measures. In the context of MSMs, collapsibility refers to the property that the estimate of the causal effect of a treatment A on an outcome Y remains unchanged, regardless of whether additional covariates are adjusted for or not, as long as these additional covariates are not affected by treatment A.

Collapsibility. Let us consider a generalized linear model for the regression of outcome Y on treatment A and a set of covariates $\boldsymbol{L}$

$$g\left[E(Y|A = a, \boldsymbol{L} = \boldsymbol{l})\right] = \gamma_0 + \gamma_1 a + \boldsymbol{\gamma_2 l}, \tag{2.35}$$

and the corresponding regression model omitting $\boldsymbol{L}$

$$g\left[E(Y|A = a)\right] = \gamma_0^* + \gamma_1^* a. \tag{2.36}$$

Model (2.35) is said to be *collapsible* for γ_1 over $\boldsymbol{L}$ if

$$\gamma_1 = \gamma_1^*,$$

otherwise it is *non-collapsible*.

Collapsibility is a desirable property in causal inference as it simplifies the analysis and interpretation of results. It is particularly useful when dealing with time-dependent confounders and helps ensure that the estimated causal effects are robust to different modeling choices. In this thesis, we employ two simulation algorithms [1, 2] to simulate longitudinal data from desired MSMs, both relying on a conditional model specification. The proved collapsibility property of these model specifications ensures that the simulated data will be consistent with the properties of the desired MSMs.

Chapter 3

Survival analysis

This chapter will illustrate the basic concepts of survival analysis and discuss the related models. First, we introduce the concepts of censoring and truncation, and illustrate two important functions in survival analysis: the survival function and the hazard rate function. Then, we discuss the non-parametric and parametric (regression) estimation methods for survival function and hazard function. In the end, the marginal structural hazard models are described.

3.1 Survival analysis

Survival analysis (or *time-to-event* analysis) refers to a set of statistical methodologies for studying the duration of time elapsed from a so-called *origin event* until the occurrence of a well-defined event of interest, i.e., the so-called *survival time*. Many examples exists in many research fields. For example, when doing research in the survival time under cancer, the survival time can be defined as the time from the the diagnosis of cancer until the death of the patient. Other examples are: the time until a specific event (e.g., disease progression, relapse, or death) between the treatment and control groups in clinical trial; the time until unemployment, namely the job duration, in economics; the time until failure for mechanical system or electronic device in the field of engineering.

It has to be notice that survival time data usually come as a mixture of complete and incomplete observations. If individuals have experienced the event during period of observation, their data is complete observations. However, it is common that the event have not occurred during the period of observation for some subjects, and their real survival time will be unknown to us when we want to analyze the data. This phenomenon is called *censoring* and may arise in different ways.

3.1.1 Censoring and Truncation

Censoring occurs when we only have partial information about an individual's event time: we know that the event has happened after or before a specific time point, but

we do not know the exact event time. Some general reasons may lead to censoring: the individual did not experience the event during the study; the individual was lost to follow-up during the study; the individual dropped out of the study.

There are three types of censoring: right censoring, left censoring, and interval censoring.

Definition 3.1.1 *Right censoring occurs when an individual has not experienced the event of interest by the end of the study.*

Definition 3.1.2 *Left censoring occurs when an individual has already experienced the event of interest before entering into the study.*

Definition 3.1.3 *Interval censoring occurs when the exact event time of an individual is not known but is known to fall within a specific interval.*

Another common phenomenon in survival analysis is *truncation*. Truncation occurs when the information of individual can only be observed within a certain observational interval.

Definition 3.1.4 *Right truncation occurs when values of random variable can only be observable when they are smaller than a specific upper bound τ^R.*

Definition 3.1.5 *Left truncation occurs when values of random variable can only be observable when they are larger than a specific lower bound τ^L.*

In this thesis we will deal with right censoring. In particular we will consider the so-called *type I censoring* where the event is observed only if it occurs prior to a pre-specified time point, the censoring time.

3.1.2 Time-to-event outcome notation

We now introduce the notation for the time-to-event outcome considered in this thesis. For a specific individual under study, we assume that there is a lifetime X and a right censoring time C_r, and thus the exact lifetime X of the individual will be known if, and only if, X is less than or equal to C_r. We use $T = \min(X, C_r)$ to represent the survival time, i.e., if the lifetime X is observed within the study, the survival time T is equal to X, and to C_r if it is censored. In addition, we use δ to indicate whether the lifetime X corresponds to an event ($\delta = 1$) or is censored ($\delta = 0$). Hence, the time-to-event outcome can be conveniently represented by pairs of random variables (T, δ).

3.2 Basic functions

For time-to-event data, conducting standard analysis, such as ordinary linear regression, will not always work, since the censored survival time might underestimate the true but unknown time to event. The collection of statistical methods specifically applied for

such time-to-event outcome is called survival analysis. The survival time T is generally modelled by two related functions, the survival function $S(t)$ and the hazard rate function $\lambda(t)$.

The survival function $S(t)$ is used to describe the survival experience of a study cohort in practice.

Definition 3.2.1 *The survival function is defined as the probability of an individual surviving to time t, which can be formulated as:*

$$S(t) = P(T > t). \tag{3.1}$$

The survival function can also be expressed using the probability density function of survival time $f(t)$:

$$S(t) = \int_t^\infty f(x)\,dx = 1 - F(t), \tag{3.2}$$

where $F(t)$ is the cumulative probability distribution of survival time T. This implies

$$f(t) = -\frac{dS(t)}{dt}. \tag{3.3}$$

The hazard rate function $\lambda(t)$ is primarily used for specifying a model for survival analysis to describe the distribution of life time.

Definition 3.2.2 *The hazard rate function is defined as the chance an individual of age t experiences the event in the next instant of time, which can be formulated as:*

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}. \tag{3.4}$$

Definition 3.2.3 *The cumulative hazard function at time t corresponding to* (3.4) *is*

$$\Lambda(t) = \int_0^t \lambda(x)\,dx. \tag{3.5}$$

Additionally, the cumulative hazard function $\lambda(t)$ and the survival function $S(t)$ can be transformed to each other via a specific equation. It can be proved that the relationship between the survival function and the cumulative hazard function is

$$S(t) = \exp\{-\Lambda(t)\}, \tag{3.6}$$

which implies that

$$f(t) = -\frac{dS(t)}{dt} = \lambda(t) \cdot \exp\{-\Lambda(t)\} = \lambda(t) \cdot S(t). \tag{3.7}$$

This means hazard function share the same information as survival function, but from a different perspective.

3.3 Non-parametric estimation

This section will illustrate the procedure of how to estimate the survival function $S(t)$ and the cumulative hazard function $\Lambda(t)$ using the non-parametric approach.

The Kaplan-Meier estimator [14] is a non-parametric estimator that can be used to estimate the survival function from censored survival data. Sometimes, it is also called the product-limit estimator.

Definition 3.3.1 *Suppose that the events occur at D distinct times $t_1 < t_2 < \cdots < t_D$. At time t_i there are d_i events and Y_i is the number of individuals who are at risk at time t_i. The Kaplan-Meier estimator is defined as:*

$$\hat{S}(t) = \begin{cases} 1, & \text{if } \ t < t_1, \\ \prod_{t_i \leq t}[1 - \frac{d_i}{Y_i}], & \text{if } \ t \geq t_1, \end{cases} \tag{3.8}$$

where the quantity $\frac{d_i}{Y_i}$ provides an estimate of the probability that an individual who survives to just prior to time t_i experiences the event at time t_i.

By using the equation (3.6), the Kaplan-Meier estimator can also be used to estimate the cumulative hazard function $\Lambda(t)$.

To directly estimate the cumulative hazard function, the Nelson-Aalen estimator [15] can be used.

Definition 3.3.2 *The Nelson-Aalen estimator is defined as:*

$$\hat{\Lambda}(t) = \begin{cases} 0, & \text{if } \ t < t_1, \\ \sum_{t_i \leq t} \frac{d_i}{Y_i}, & \text{if } \ t \geq t_1. \end{cases} \tag{3.9}$$

3.4 Regression models for the hazard function

Regression models for the hazard function are usually used to describe the relationship of covariates to a survival or other censored outcome. In the following we will introduce the semi-parametric Cox proportional hazard model [16] and the Aalen's additive hazard model [17].

3.4.1 The Cox proportional hazard model

One of the most widely used models in survival analysis is the Cox proportional hazard regression model [16].

Definition 3.4.1 *Proportional hazards assumption states that the ratio of two hazards $HR = \lambda_2(t)/\lambda_1(t)$ does not depend on time t, but is constant over time.*

Under the proportional hazards assumption, the Cox model can be used to model the relationship of covariates to survival or other censored outcome. Suppose that $\boldsymbol{X}_i(t)$ is the vector of covariates for the i-th individual at time t which may affect the survival distribution of the survival time T. The Cox model assumes the hazard for individual i as

$$\lambda_i(t|\boldsymbol{X}_i(t)) = \lambda_0(t)\exp\{\boldsymbol{\beta}\cdot\boldsymbol{X}_i(t)\},$$

where $\lambda_0(\cdot)$ is the baseline hazard function and $\boldsymbol{\beta}$ is the vector of regression coefficients. The effect of the covariates is to act multiplicatively on the baseline hazard.

This model is a semi-parametric model, as the baseline hazard does not assume any shape or parametric form. It is also called the proportional hazards (PH) model as the hazard ratio for two subjects i and j with time-fixed covariate vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$

$$HR = \frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp\{\boldsymbol{X}_i\boldsymbol{\beta}\}}{\lambda_0(t)\exp\{\boldsymbol{X}_j\boldsymbol{\beta}\}} = \frac{\exp\{\boldsymbol{X}_i\boldsymbol{\beta}\}}{\exp\{\boldsymbol{X}_j\boldsymbol{\beta}\}} = \exp\{\boldsymbol{\beta}\cdot(\boldsymbol{X}_i-\boldsymbol{X}_j)\} \tag{3.10}$$

is constant over time, satisfying the definition.

The estimation of the coefficients β is based on maximizing the logarithm of the partial likelihood, given by:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{D}\frac{\exp\{\boldsymbol{X}_k\boldsymbol{\beta}\}}{\sum_{j\in R(t_k)}\exp\{\boldsymbol{X}_j\boldsymbol{\beta}\}}, \tag{3.11}$$

where $t_1 < t_2 < \cdots < t_D$ denote the ordered true event times, and $R(t_k)$ denotes the risk set at time t_k, namely the set of all individuals who are still under study at a time just prior to t_k.

3.4.2 The Aalen's additive hazard model

As an alternative to proportional hazard models, in 1989 Aalen proposed the additive hazard model [17]. For each subject i, let us consider a vector of p time-dependent covariates $\boldsymbol{X}_i(t) = (X_{i1}(t),\cdots,X_{ip}(t))$. The hazard in the additive framework is more directly similar to a linear model and can be expressed as:

$$\lambda(t|X_i) = \alpha_0(t) + \alpha_1(t)X_{i1}(t) + \cdots + \alpha_p(t)X_{ip}(t), \tag{3.12}$$

where $\alpha_0(t)$ is the baseline hazard corresponding to the hazard rate of an individual with all covariates identically equal to zero, and $\alpha_j(t)$ is the regression function related to the j-th covariate ($j = 1,\cdots,p$), which can vary over time. The key idea of Aalen's additive hazard model is that the hazard rate for an individual i with observed covariates $\boldsymbol{x}_i(t)$ at time t is assumed to be a linear combination of the covariates $x_{ij}(t)$, with $j = 1,\cdots,p$.

Parameter estimation is based on least-squares based technique. Instead of directly estimating $\alpha_j(t)$, which is difficult in practice, their cumulative version $A_j(t)$ are estimated

by least-squares technique at first, where

$$A_j(t) = \int_0^t \alpha_j(u)\, du, \quad j = 1, \ldots, p.$$

The crude estimates of $\alpha_j(t)$ are given by the slope of the estimate of $A_j(t)$, while the better estimates of $\alpha_j(t)$ can be derived by kernel-smoothing technique [18].

It should be notice that in the Cox model the hazard rate is necessarily non-negative, while the additive hazard model does not have this natural restriction. The Aalen hazard rate can occasionally be negative values, especially when the model fit is not very good. On the other hand, additive hazard model has some appealing properties compared to Cox model. One main advantage is that the parameters of additive hazard model are collapsible [19] (see Section 2.8), while hazard ratios are non-collapsible, implying that the Cox model does not have this property.

3.5 Marginal structural hazard models

Marginal structural hazards model are causal models for the marginal distribution of the counterfactual variables of survival time $T^{\bar{\boldsymbol{a}}}$, where $T^{\bar{\boldsymbol{a}}}$ represents the subject's time-to-event had the subject followed the treatment history $\bar{\boldsymbol{A}} = \bar{\boldsymbol{a}}$ from the start of follow-up.

Similarly to Section 2.6.1, to model for the marginal distribution of counterfactual variables $T^{\bar{a}}$ for survival time in the presence of time-dependent confounders $\bar{\boldsymbol{L}}_k$, we need to create a pseudo-population by IPTW using the stabilized weights

$$sw_i(t) = \prod_{k=0}^{\lfloor t \rfloor} \frac{P(A_k = a_{k,i} | \bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1,i})}{P(A_k = a_{k,i} | \bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1,i}, \bar{\boldsymbol{L}}_k = \bar{\boldsymbol{l}}_{k,i})}, \tag{3.13}$$

where $sw_i(t)$ represents the IPT-weight for subject i at time t, $\lfloor t \rfloor$ is the largest integer less than or equal to t, and A_{-1} is defined to be 0 [20]. In the pseudo-population thus created, the effects of time-dependent confounding are balanced, so association in hazard regression models is causation.

3.5.1 MSMs based on Cox PH model

Given treatment history $\bar{\boldsymbol{a}}$, we can define a marginal structural Cox proportional hazard model as:

$$\lambda_{T^{\bar{\boldsymbol{a}}}}(t) = \lambda_0(t) \exp\left\{ g\left(\tilde{\boldsymbol{\beta}}_A; \bar{\boldsymbol{a}}_{\lfloor t \rfloor} \right) \right\}, \tag{3.14}$$

where $\lambda_0(t)$ is the baseline hazard function, $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$ denotes treatment pattern up to the most recent visit prior to time t, $g(\cdot)$ is a function (to be specified) of treatment pattern $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$, and $\tilde{\boldsymbol{\beta}}_A$ is a vector of log hazard ratios.

For example, we can assume a Cox MSM where the hazard at time t is specified to depend only on the current level of treatment by setting $g\left(\tilde{\boldsymbol{\beta}}_A; \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right) = \tilde{\beta}_{A1} \cdot a(t)$. In that case, $\exp\{\tilde{\beta}_{A1}\}$ can be interpreted as the causal hazard ratio at any time t had all subjects been continuously exposed to treatment compared with the hazard rate at time t had all subjects remained unexposed.

3.5.2 MSMs based on Aalen's additive model

Given treatment history $\bar{\boldsymbol{a}}$, we can define a marginal structural Aalen's additive hazard model as:

$$\lambda_{T^{\bar{\boldsymbol{a}}}}(t) = \tilde{\alpha}_0(t) + g\left(\tilde{\boldsymbol{\alpha}}_A(t); \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right) \tag{3.15}$$

where $\tilde{\alpha}_0(t)$ is the baseline hazard at time t, $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$ denotes treatment pattern up to the most recent visit prior to time t, $g(\cdot)$ is a function (to be specified) of treatment pattern $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$, and $\tilde{\boldsymbol{\alpha}}_A$ is a vector of log hazard ratios whose values can vary over time.

Similar to the Cox MSM example, we can assume an Aalen MSM where the hazard at time t is specified to depend only on the current level of treatment. Another example is for the hazard at time t to depend on the history of treatment up to time t through the main effect terms for treatment at each visit, by using $g\left(\tilde{\boldsymbol{\alpha}}_A(t); \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right) = \sum_{j=0}^{t} \tilde{\alpha}_{Aj}(t) a_{t-j}$.

3.5.3 Counterfactual survival probability

The Cox MSM provides estimates of the log hazard ratios $\tilde{\boldsymbol{\beta}}_A$, and the Aalen MSM yields estimates of cumulative regression coefficients $\int_0^t \tilde{\boldsymbol{\alpha}}_A(s)\, ds$. However, both of these hazard-based estimands do not have a direct causal interpretation. To derive a causal estimand from the estimates obtained through MSM, we can utilize the concepts of counterfactual survival probability and marginal risk difference.

Definition 3.5.1 *The counterfactual survival probability at time t is defined as the survival probability at time t under the possibly counter-to-fact treatment $\bar{\boldsymbol{a}}$, i.e.,* $\Pr(T^{\bar{\boldsymbol{a}}} \geq t)$.

Specifically, the counterfactual survival probability at time t based on Cox MSM in (3.14) is given by

$$\begin{aligned} Pr(T^{\bar{\boldsymbol{a}}} \geq t) = \exp\Bigg(& - e^{g(\tilde{\boldsymbol{\beta}}_A; a_0)} \int_0^1 \lambda_0(s)\, ds - e^{g(\tilde{\boldsymbol{\beta}}_A; \bar{\boldsymbol{a}}_1)} \int_1^2 \lambda_0(s)\, ds \\ & - \cdots - e^{g(\tilde{\boldsymbol{\beta}}_A; \bar{\boldsymbol{a}}_{\lfloor t \rfloor})} \int_{\lfloor t \rfloor}^t \lambda_0(s)\, ds \Bigg), \end{aligned} \tag{3.16}$$

while the counterfactual survival probability at time t based on the Aalen MSM in (3.15) is given by

$$Pr(T^{\bar{\boldsymbol{a}}} \geq t) = \exp\Bigg(- \int_0^t \tilde{\alpha}_0(s)\, ds - \int_0^1 g(\tilde{\boldsymbol{\alpha}}_A(s); a_0)\, ds \\ - \int_1^2 g(\tilde{\boldsymbol{\alpha}}_A(s); \bar{\boldsymbol{a}}_1)\, ds - \cdots - \int_{\lfloor t \rfloor}^t g(\tilde{\boldsymbol{\alpha}}_A(s); \bar{\boldsymbol{a}}_{\lfloor t \rfloor})\, ds \Bigg). \tag{3.17}$$

Definition 3.5.2 *The marginal risk difference at time t is defined as the difference in counterfactual survival probabilities if all individuals had been treated up to that time compared to when none of the individuals have been treated up to that time, i.e.,*

$$Pr(T^{\bar{\boldsymbol{a}}=\mathbf{1}} \geq t) - Pr(T^{\bar{\boldsymbol{a}}=\mathbf{0}} \geq t).$$

c

Chapter 4

Simulating from marginal structural models

In this chapter, we will briefly recall the idea of simulation study, and present the simulation mechanisms of the two algorithms introduced to generate longitudinal data from known Marginal Structural Models (MSMs).

4.1 Introduction to simulation study

In statistical research, mathematical theory is the cornerstone of supporting and evaluating statistical methods addressing well-defined problems. However, in longitudinal studies, the complexity of statistical modelling often hinders the derivation of corresponding mathematical evidence to support the behaviour of the considered statistical method under certain assumptions. In such cases, a simulation study becomes an appropriate approach to evaluate the statistical methods. Simulation studies are useful when theoretical arguments alone are insufficient to determine the validity of the method in a real-life application or when violations of the assumptions underlying the available theory may affect the validity of the results [21].

The key idea behind a simulation study is to generate synthetic data sets with known properties to investigate how different methods perform when applied to them. By repeatedly sampling from a specified probability distribution and applying a certain method or function, a simulated distribution of desired estimator can be created, and its expected value can be used as the final estimate. Since the true value of the setting is known, the simulation study enables us to evaluate whether the research methods recover the known truth, and explore the behavior of the estimator under different conditions, such as varying sample sizes or violations of certain assumptions. For example, longitudinal data can be generated based on a specific DAG, where the treatment effect is pre-specified. Then, the performance of MSMs, estimated by Inverse Probability of Treatment Weighting (IPTW), can be evaluated to assess how accurately they estimate this known effect.

When performing simulation studies, it is hence crucial to assess how well the estimators perform. Two common performance measures used for this purpose are bias and mean square error [11].

Bias is a measure of the systematic difference between the expected value of an estimator and the true value of the effect. It is used to quantify whether the estimator targets the true value on average. A biased estimator may consistently overestimate or underestimate the true value, resulting in inaccurate results. In a simulation study with B repetitions, the bias of estimator $\hat{\theta}$ can be estimated as:

$$Bias = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b - \theta \tag{4.1}$$

Mean square error (MSE) is the sum of the squared bias and variance of estimator $\hat{\theta}$. It is a way to integrate both measures into one summary performance measure. A low MSE indicates a high prediction accuracy. In a simulation study with B repetitions, the mean square error of estimator $\hat{\theta}$ can be estimated as:

$$MSE = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b - \theta)^2. \tag{4.2}$$

In the following sections, we will illustrate two simulation algorithms designed to generate longitudinal data from a given MSM with defined causal parameters. These algorithms can be used to evaluate the effectiveness of the IPTW estimator in determining the causal effect of time-varying treatment in longitudinal studies where time-varying confounding is present.

4.2 Simulating from MSMs

Simulating longitudinal data from MSMs when the data-generating process exhibits time-dependent confounding is challenging, especially in the context of survival analysis using proportional hazards or pooled logistic regressions. The data generating process based on a special DAG relies on the conditional model specification. However, this may not be compatible with the desired MSM. Noncollapsibility (see Section 2.8) is one of the factors that can contribute to this issue. For example, the regression in Equation (2.35) can refer to the conditional model that we sample from, while the regression in Equation (2.36) can represent the desired MSM.

In survival analysis, the hazard functions or their discrete equivalent based on generalized linear models, such as pooled logistic regression model, are typically noncollapsible. This can result in conditional distributions used to draw the simulated data that are incompatible with the desired properties of the given marginal model. Consequently, it

is difficult to simulate data from MSMs in the context of survival outcomes. The two algorithms we present in the following Sections 4.3 and 4.4 address this issue by using different methods to overcome the noncollapsibility problem and successfully simulate longitudinal data from given MSMs.

4.3 Simulation algorithm by Havercroft and Didelez (2012)

The first simulation algorithm considered in this thesis was proposed by Havercroft and Didelez (2012) [1]. Based on the data structure of the Swiss HIV Cohort Study [22], the authors developed an algorithm to simulate longitudinal data with a survival outcome from a given discrete-time hazard model. The algorithm is built on a specific DAG that exhibits time-dependent confounding in the data-generating process.

Emulating the Swiss HIV Cohort Study, this simulation algorithm generates the data of HIV patients to investigate the effect of highly active antiretroviral therapy (HAART) versus no treatment on survival, with the only measured time-dependent confounding by CD4 cell count. The data structure operates in discrete time, considering a follow-up with $t = 0, 1, \ldots, K$ time points.

Notation: The four processes included in the simulation process are as follows.

- $\{A_t\}$ corresponds to the binary treatment indicator process. $A_t = 1$ means the subject is on treatment at time t, otherwise ($A_t = 0$) the subject is not on treatment. Once treatment has started for a subject, it continues until failure or end of the follow-up period.
- $\{L_t\}$ represents the CD4 cell count in the Swiss HIV Cohort Study. L_t is the subject's CD4 count in $cells/\mu L$ at time t. The normal CD4 count for healthy adults and teens is between 500 $cells/\mu L$ to 1200 $cells/\mu L$, while a lower value indicates more severe illness.
- $\{Y_t\}$ is the binary survival outcome process, where $Y_t = 1$ indicates that death has occurred before time t, while $Y_t = 0$ otherwise.
- $\{U_t\}$ represents a latent general health process, with $U_t \in [0, 1]$. A value of U_t close to 0 indicates a poor health status at time t, while a value close to 1 indicates good health at time t.

In addition, k_c denotes the time interval for regular check-ups. Confounder L_t and treatment A_t are measured and updated at each check-up point (i.e., every k_c-th time points). On the other hand, U_t and Y_t are measured at each time point t, regardless of the check-up schedule.

Marginal structural model: The authors proposed a simulation procedure to generate longitudinal data from the following logistic MSM:

$$\begin{aligned}\lambda_t^{\bar{a}} &= logit^{-1}(\gamma_0 + \gamma_1[(1-a_t)t + a_t t^*] + \gamma_2 a_t + \gamma_3 a_t(t-t^*)) \\ &= logit^{-1}(\gamma_0 + \gamma_1 d_{1,t} + \gamma_2 a_t + \gamma_3 d_{3,t}),\end{aligned} \tag{4.3}$$

where a_t is the binary treatment at time t, t^* is the treatment initiation time, $d_{1,t} = \min\{t, t^*\}$ represents the time elapsed before treatment initiation, and $d_{3,t} = \max\{t - t^*, 0\}$ represents the time elapsed after treatment initiation. This model structure indicates that the hazard depends on a summary of the treatment history, rather than only on the current treatment.

Simulation procedure: For each subject $i = 1, \ldots, n$, the simulation procedure with K discrete time points and k_c check-up times is as follows.

- At baseline (time $t = 0$):

  1. Initialize a baseline value for latent general health variable from standard uniform distribution:
  $$U_{0,i} \sim Uniform[0, 1].$$
  2. To approximate the sample distribution of baseline CD4 counts in the Swiss HIV Cohort Study [22], generate a baseline CD4 cell count as a transformation of $U_{0,i}$ by the inverse cumulative distribution function of $\Gamma(k = 3, \theta = 154)$ distribution plus an error $\epsilon_{0,i} \sim \mathcal{N}(0, 20)$:
  $$L_{0,i} \leftarrow F^{-1}_{\Gamma(3,154)}(U_{0,i}) + \epsilon_{0,i}.$$
  3. Draw the binary treatment decision based on
  $$A_{0,i} \sim Bernoulli(P_{A_{0,i}}), \quad P_{A_{0,i}} = logit^{-1}(\theta_0 + \theta_2(L_{0,i} - 500)),$$
  where θ_0 and θ_2 are the parameters to be defined. Specifically, θ_2 should be negative since a lower value of CD4 cell count $L_{0,i}$ indicates more severe illness and thus the subject is more likely to receive the treatment. If $A_{0,i} = 1$, the treatment initiation time T_i^* is set as 0.
  4. Compute the hazard of the subject as
  $$\lambda_{0,i} = logit^{-1}(\gamma_0 + \gamma_2 A_{0,i}),$$
  where γ_0 and γ_2 correspond to the parameters in MSM (4.3) for time $t = 0$. If $\lambda_{0,i} \geq U_{0,i}$, we assume the death has occurred in the interval $t \in (0, 1]$ and set $Y_{1,i} \leftarrow 1$. Otherwise, the subject survived and set $Y_{1,i} \leftarrow 0$.

- If the subject is still alive ($Y_{t,i} = 0$), for each time point $t = 1, 2, 3 \cdots K$:

1. Draw $U_{t,i} \leftarrow \min\{1, \max\{0, U_{t-1,i} + \mathcal{N}(0, 0.05)\}\}$, representing the value of the latent general health status at time t that depends on $U_{t-1,i}$ and is constrained within $[0, 1]$.
2. If t is not a check-up time point, treatment and CD4 cell count are not updated: $A_{t,i} \leftarrow A_{t-1,i}$ and $L_{t,i} \leftarrow L_{t-1,i}$. Else,
   - if treatment has not started yet,

     $$L_{t,i} \leftarrow \max(0, L_{t-1,i} + \mathcal{N}(100(U_{t,i} - 2), 50)),$$

     where the Gaussian drift term implies that the closer $U_{t,i}$ is to 0 (i.e., bad general health condition), the stronger the negative drift in CD4. Then draw a binary treatment decision based on

     $$A_{t,i} \sim Bernoulli(P_{A_{t,i}}), \quad P_{A_{t,i}} = logit^{-1}(\theta_0 + \theta_1 t + \theta_2(L_{t,i} - 500)). \quad (4.4)$$

   - Otherwise, if treatment has started at previous check-up, the subject remains on treatment so $A_{t,i} \leftarrow 1$, and

     $$L_{t,i} \leftarrow \max(0, L_{t-1,i} + 150 + \mathcal{N}(100(U_{t,i} - 2), 50))$$

     where the addition of 150 indicates the positive effect of starting a treatment on CD4.

   If treatment starts at this check-up, set treatment start time $T_i^* \leftarrow t$.
3. Compute the hazard of subject at time t based on

   $$\lambda_{t,i} = logit^{-1}(\gamma_0 + \gamma_1[(1 - A_{t,i})t + A_{t,i}T_i^*] + \gamma_2 A_{t,i} + \gamma_3 A_{t,i}(t - T_i^*)), \quad (4.5)$$

   where γ_0, γ_1, γ_2 and γ_3 correspond to the parameters in MSM (4.3).
4. Compute the survival probability up to time $t + 1$:

   $$S_i(t) = \prod_{j=0}^{t} (1 - \lambda_{j,i})$$

   If $1 - S_i(t) \geq U_{0,i}$, we assume the death has occurred in the interval $t \in (t, t+1]$ and set $Y_{t+1,i} \leftarrow 1$. Otherwise, the subject survived and set $Y_{t+1,i} \leftarrow 0$.

The complete pseudocode of the simulation process is reported in Algorithm 1. Specifically, in the original paper [1], the authors considered the following input parameters for their simulation studies:

- $K = 40$: follow-up occurs over 40 time points;
- $k_c = 5$: check-up every 5^{th} time points;

Algorithm 1: Simulation algorithm by Havercroft and Didelez (2012) [1].

Input:
K: number of time points during follow-up
k_c: check-up times
n: number of subjects
$(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$: parameters of MSM hazard in Equation (4.3)
$(\theta_0, \theta_1, \theta_2)$: parameters of the conditional distributions of treatment in Equation (4.4)
Result: The longitudinal dataset of n subjects

Simulation process:
for *subject* $i \leftarrow 1$ ***to*** n **do**
  $U_{0,i} \sim Uniform[0,1]$ *(simulate baseline latent health status)*
  $\epsilon_{0,i} \sim \mathcal{N}(0, 20)$
  $L_{0,i} \leftarrow F^{-1}_{\Gamma(3,154)}(U_{0,i}) + \epsilon_{0,i}$ *(simulate baseline CD4 cell count)*
  Draw the baseline binary treatment decision based on Equation (4.4) *at time* $t = 0$
  $A_{0,i} \sim Bernoulli(logit^{-1}(\theta_0 + \theta_2(L_{0,i} - 500)))$
  if $A_{0,i} = 1$ **then**
    $T_i^* \leftarrow 0$ *(indicates treatment starts at time point 0)*
  end
  Compute the hazard based on Equation (4.5) *at time* $t = 0$
  $\lambda_{0,i} = logit^{-1}(\gamma_0 + \gamma_2 A_{0,i})$
  if $\lambda_{0,i} \geq U_{0,i}$ **then**
    $Y_{1,i} \leftarrow 1$ *(subject is dead)*
  else
    $Y_{1,i} \leftarrow 0$ *(subject is still alive)*
  end
  for $t \leftarrow 1, \ldots, K$ **do**
    If subject still alive, continue sampling the data
    while $Y_{t,i} = 0$ **do**
      $\Delta_{t,i} \sim \mathcal{N}(0, 0.05)$
      $U_{t,i} \leftarrow \min\{1, \max\{0, U_{t-1,i} + \Delta_{t,i}\}\}$
      if t *mod* $k_c \neq 0$ **then**
        If not at a check-up time point, $L_{t,i}$ *and* $A_{t,i}$ *remain the previous values*
        $L_{t,i} \leftarrow L_{t-1,i}$
        $A_{t,i} \leftarrow A_{t-1,i}$
      else
        If treatment has started at the last check-up, CD4 count will have a shift of 150
        $\epsilon_{t,i} \sim \mathcal{N}(100(U_{t,i} - 2), 50)$
        $L_{t,i} \leftarrow \max\{0, L_{t-1,i} + 150 A_{t-k_c,i}(1 - A_{t-k_c-1,i}) + \epsilon_{t,i}\}$
        if $A_{t-1,i} = 0$ **then**
          Draw a binary treatment decision based on Equation (4.4)
          $A_{t,i} \sim Bernoulli(logit^{-1}(\theta_0 + \theta_1 t + \theta_2(L_{t,i} - 500)))$
        else
          $A_{t,i} \leftarrow 1$ *(subject remains on treatment once started)*
        end
        if $A_{t,i} = 1$ *and* $A_{t-k_c,i} = 0$ **then**
          $T_i^* \leftarrow t$ *(indicates treatment starts at time point* t*)*
        end
      end
      Compute the hazard to generate the survival outcome based on Equation (4.5)
      $\lambda_{t,i} \leftarrow logit^{-1}(\gamma_0 + \gamma_1[(1 - A_{t,i})t + A_{t,i}T_i^*] + \gamma_2 A_{t,i} + \gamma_3 A_{t,i}(t - T_i^*))$
      if $1 - \Pi_{j=0}^{t}(1 - \lambda_{j,i}) \geq U_{0,i}$ **then**
        $Y_{t+1,i} = 1$ *(subject is dead)*
      else
        $Y_{t+1,i} = 0$ *(subject is still alive)*
      end
    end
  end
end

- The parameters for the conditional distributions of treatment in Equation (4.4) are

$$(\theta_0, \theta_1, \theta_2) = (-0.405, 0.0205, -0.00405)$$

to ensure the treatment assignment probabilities not close to 0 or 1 and calibrate the logistic function such that $P(A_0 = 1|L_0 = 500) = 0.4$, $P(A_0 = 1|L_0 = 400) = 0.5$ and $P(A_{10} = 1|L_{10} = 500) = 0.45$;

- The parameters of MSM in Equation (4.3) are

$$(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (-3, 0.05, -1.5, 0.1)$$

to address the nontrivial scenario where initiating treatment results in a one-time decrease in $\lambda_t^{\bar{a}}$ and also increases the rate at which $\lambda_t^{\bar{a}}$ increases over time.

Figure 4.1 displays the DAGs representing the data-generating process for two specific cases where the follow-up occurs over different time points K (panel a: $K = 1$; panel b: $K = 2$) with check-ups at each time ($k_c = 1$). In particular, the DAG in panel (b) extends the one in panel (a) by including an additional follow-up time point. Note that L_t depends on L_{t-1}, A_{t-1} and A_{t-2} in the case of $k_c = 1$. Following such pattern, the DAG expands as K increases.

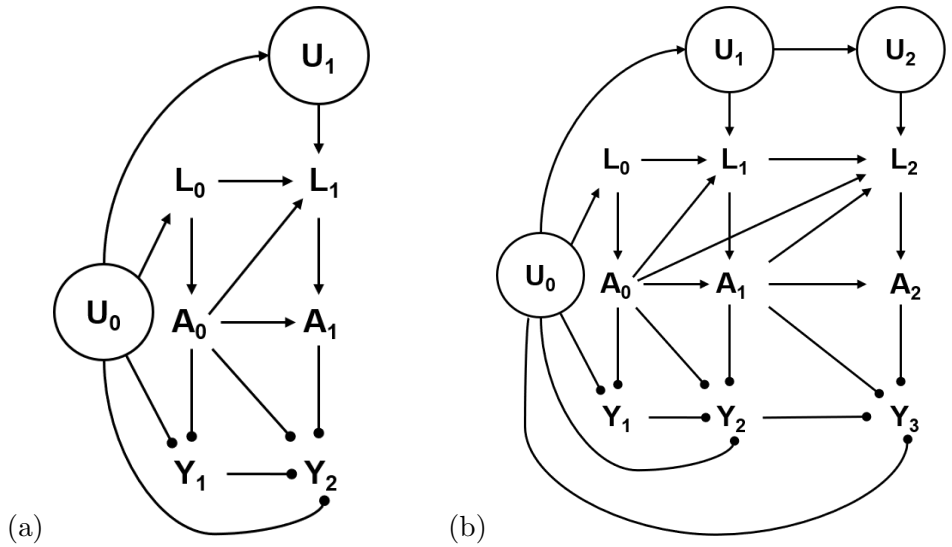

Figure 4.1: Directed Acyclic Graphs (DAGs) representing the simulation algorithm proposed by Havercroft and Didelez (2012) [1] in two cases: (a) $K = 1$, $k_c = 1$; (b) $K = 2$, $k_c = 1$. Dot arrowheads represent deterministic relationships and circled nodes represent latent variables.

This algorithm efficiently samples survival data from the MSM given in Equation (4.3). The key mechanism lies in how the treatment history $\bar{\boldsymbol{A}}$ and survival outcome history $\bar{\boldsymbol{Y}}$

depend on $U_{0,i}$, while CD4 cell count $L_{t,i}$ sufficiently adjusts for confounding. $U_{0,i}$ acts as a probability threshold against the survival function, calculated from the conditional hazard, to generate the survival outcome history $\bar{\boldsymbol{Y}}$. Under this mechanism, it can be proved that the conditional distributions that we relies upon to simulate the longitudinal data are compatible with the desired properties of the marginal model, enabling us to generate samples from the desired MSM. This simulation procedure can also be considered as one way of matching the left-hand side with the right-hand side in the following equation:

$$P(Y^a) = \sum_u P(Y|U = u, A = a)P(U = u), \tag{4.6}$$

which implies that the conditional model specification here is collapsible.

The main contribution of this algorithm lies in its ability to address the noncollapsibility of the pooled logistic regression model. This noncollapsibility makes it challenging to confirm whether the data generated from this conditional model specification is compatible with the desired marginal model. However, by using $U_{0,i}$ as a probability threshold to determine the survival outcome, the generating procedure in Algorithm 1 can be proved to simulate from the desired marginal model, and thus the conditional model specification is collapsible.

4.4 Simulation algorithm by Keogh *et al.* (2021)

The second simulation algorithm considered in this thesis has been proposed by Keogh *et al.* (2021) [2]. The authors proved that if longitudinal data are simulated from a conditional additive hazard model, then the corresponding marginal hazard model is also additive and hence can be correctly specified. The algorithm simulates longitudinal data with a event time outcome from a given conditional hazard model, specifically an Aalen additive hazard model. It is built on a specific DAG that exhibits time-dependent confounding in the data-generating process with event times generated in continuous time.

Compared to Algorithm 1, this simulation algorithm generates longitudinal data in a general setting, with a single time-dependent continuous confounding representing a biomarker. Besides, event (death) times are generated in continuous time. The data structures considers a follow-up with $t = 0, 1, \ldots, 4$ time points, with an administrative censoring at time 5.

Notation: The three processes included in the simulation process are as follows.

- $\{A_t\}$ corresponds to the binary treatment indicator process. $A_t = 1$ means the subject is on treatment at time t, otherwise ($A_t = 0$) the subject is not on treatment. Unlike simulation algorithm 1, there is no requirement for subject to maintain the treatment once they initiate it.

- $\{L_t\}$ denotes a biomarker value process at time t. In this case, a high value of the biomarker L_t is associated with higher propensity to receive the treatment and higher hazard. The biomarker value tends to rise over time, but it can be reduced through treatment.
- $\{Y_t\}$ is the binary survival outcome process, where $Y_t = 1$ indicates that death has occurred before time t, while $Y_t = 0$ otherwise.

In addition, U is an unmeasured continuous variable representing a subject frailty term. It does not vary over the simulation process. L_t , A_t and Y_t are updated at each time point t, which corresponds to the case of $k_c = 1$ in Algorithm 1.

Marginal structural model: Keogh *et al.* (2021) proved that if the conditional hazard model is additive, the corresponding MSM is also additive. Furthermore, even if the conditional hazard model depends only on the current level of treatment, the MSM depends on the whole treatment history. Based on a conditional additive hazard model of this form:

$$\lambda(t|\bar{\boldsymbol{A}}_{\lfloor t \rfloor}, \bar{\boldsymbol{L}}_{\lfloor t \rfloor}, U) = \alpha_0 + \alpha_A A_{\lfloor t \rfloor} + \alpha_L L_{\lfloor t \rfloor} + \alpha_U U, \tag{4.7}$$

where $\bar{\boldsymbol{A}}_{\lfloor t \rfloor}$ and $\bar{\boldsymbol{L}}_{\lfloor t \rfloor}$ denote the treatment and time-dependent covariate history up to the most recent visit prior to time t (i.e., $\lfloor t \rfloor$ is the largest integer less than t), their simulation procedure generate data from the following Aalen's additive MSM:

$$\lambda_{T^{\bar{\boldsymbol{a}}}}(t) = \tilde{\alpha}_0(t) + \sum_{j=0}^{\lfloor t \rfloor} \tilde{\alpha}_{Aj}(t) a_{\lfloor t \rfloor - j}, \tag{4.8}$$

where $T^{\bar{\boldsymbol{a}}}$ denotes the counterfactual event time had an individual followed treatment regime $\bar{\boldsymbol{a}}$ from visit 0 onwards, $\tilde{\alpha}_0(t)$ corresponds to the baseline hazard and $\tilde{\alpha}_{Aj}(t)$ are the risk coefficients correspond to the treatment with a j-time-point lag.

Simulation procedure: For each subject $i = 1, \ldots, n$, the simulation procedure with $K + 1$ as administrative censoring time is as follows.

- At baseline (time $t = 0$):
  1. Initialize the individual frailty term from a normal distribution with mean 0 and standard deviation 0.1:
  $$U_i \sim \mathcal{N}(0, 0.1).$$
  2. Initialize the biomarker value from a normal distribution with mean U_i and standard deviation 1:
  $$L_{0,i} \sim \mathcal{N}(U_i, 1).$$
  3. Draw a binary treatment decision based on
  $$A_{0,i} \sim Bernoulli(P_{A_{0,i}}), \quad P_{A_{0,i}} = logit^{-1}(\gamma_0 + \gamma_L L_{0,i}). \tag{4.9}$$

4. Compute the hazard of the subject, i.e., $\lambda_i(t = 0|A_{0,i}, L_{0,i}, U_i)$, based on Equation (4.7), where α_0, α_A, α_L and α_U are the parameters to be defined.
5. Generate event time in the period of $(0, 1)$ as follows:
   - Generate $V_{0,i} \sim Uniform(0, 1)$.
   - Calculate $\Delta T_{0,i} = \frac{-log(V_{0,i})}{\lambda_i(t=0|A_{0,i}, L_{0,i}, U_i)}$.
   - If $\Delta T_{0,i} < 1$, we assume the death has occurred in the interval $t \in (0, 1)$, then the event time is set to be $\tilde{T}_i \leftarrow \Delta T_{0,i}$ and survival outcome is set to be $Y_{1,i} \leftarrow 1$. Else, the subject remains at risk at time $t = 1$ and set $Y_{1,i} \leftarrow 0$.

- If the subject is still alive ($Y_{t,i} = 0$), for each time point $t = 1, 2, 3 \cdots K$:

  1. Update the biomarker value:

  $$L_{t,i} \sim \mathcal{N}(0.8L_{t-1,i} - A_{t-1,i} + 0.1t + U_i, 1).$$

  2. Draw a binary treatment decision based on

  $$A_{t,i} \sim Bernoulli(P_{A_{t,i}}), \quad P_{A_{t,i}} = logit^{-1}(\gamma_0 + \gamma_A A_{t-1,i} + \gamma_L L_{t,i})). \quad (4.10)$$

  3. Compute the hazard of subject at time t, i.e., $\lambda_i(t|A_{t,i}, L_{t,i}, U_i)$, based on Equation (4.7).
  4. Generate event time in the period of $[t, t + 1)$ as follows:
     - Generate $V_{t,i} \sim Uniform(0, 1)$.
     - Calculate $\Delta T_{t,i} = \frac{-\log(V_{t,i})}{\lambda_i(t|A_{t,i}, L_{t,i}, U_i)}$.
     - If $\Delta T_{t,i} < 1$, we assume the death has occurred in the interval $t \in [t, t+1)$, then the event time is set to be $\tilde{T}_i \leftarrow \Delta T_{t,i} + t$ and survival outcome is set to be $Y_{t+1,i} \leftarrow 1$. Else, the subject remains at risk at time $t + 1$ and set $Y_{t+1,i} \leftarrow 0$.

- If the subject does not have an event time generated in the period $(0, K + 1)$, the subject is administratively censored at time $t = K + 1$.

The complete pseudocode of the simulation process is reported in Algorithm 2. Specifically, in the original paper [2], the authors considered the following input parameters for their simulation studies:

- $K = 4$: longitudinal data are generated at 5 visits $t = 0, \cdots, 4$ and administrative censoring is at time 5;

- The parameters for the conditional distributions of treatment in Equation (4.10) are

$$(\gamma_0, \gamma_A, \gamma_L) = (-2, 1, 0.5);$$

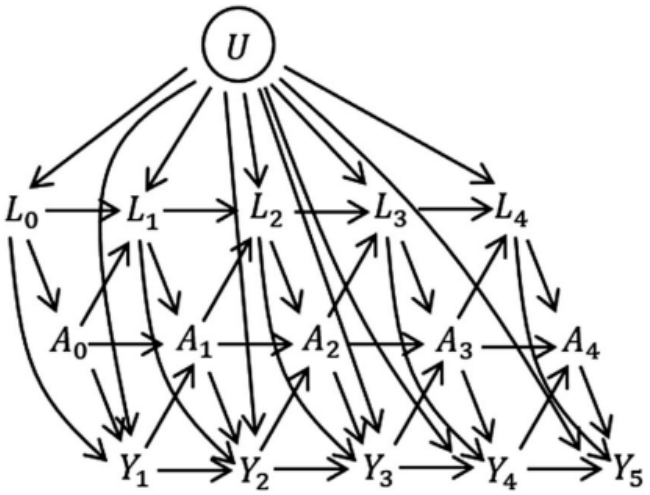

Figure 4.2: Directed Acyclic Graphs (DAGs) representing the simulation algorithm proposed by Keogh *et al.* (2021) illustrated for a discrete-time setting where $Y_t = I(T > t)$ in the case of $K = 5$ visits.

- The parameters of conditional hazard model in Equation (4.7) are

  $$(\alpha_0, \alpha_A, \alpha_L, \alpha_U) = (0.7, -0.2, 0.05, 0.05)$$

  to ensure the probability of obtaining a negative hazard is negligible under this simulation setting.

Figure 4.2 displays the data-generating process for the case with 5 visits. Please note that this DAG is presented in the discrete-time setting where $Y_t = I(\tilde{T} > t)$, with $\tilde{T}$ representing the event time and t being the visit time. However, in the actual generating process, it will be generalized to a continuous-time setting, meaning that the event times are generated in continuous time, and the corresponding outcome $Y_t = I(\tilde{T} > t)$ can be observed.

This simulation algorithm provides an effective approach to evaluate causal inference methods in a broad simulation scenario. Building on the work from Martinussen and Vansteelandt (2013) [19], which established the relationship between conditional and marginal additive hazard models in the point-treatment setting, the authors extended the results to the longitudinal setting with time-dependent confounding. They proved that using conditional Aalen additive hazard models for simulation also yields an additive form for the MSM, allowing for correct specification and fitting using standard software. In contrast, a MSM derived from a conditional Cox model is no longer a Cox model and lacks a closed-form expression in general.

This algorithm relying on the conditional Aalen additive hazard models offers significant advantages. It enables the simulation of longitudinal data with time-dependent

Algorithm 2: Simulation algorithm by Keogh *et al.* (2021) [2].

Input:
K: number of time points during follow-up
n: number of subjects
$(\alpha_0, \alpha_A, \alpha_L, \alpha_U)$: parameters of conditional hazard model in Equation (4.7)
$(\gamma_0, \gamma_A, \gamma_L)$: parameters for the conditional distributions of treatment in Equation (4.10)
Result: The longitudinal dataset of n subjects

Simulation process:
for *subject* $i \leftarrow 1$ ***to*** n **do**
  simulate individual frailty
  $U_i \sim \mathcal{N}(0, 0.1)$
  initialize a value for biomarker
  $L_{0,i} \sim \mathcal{N}(U_i, 1)$
  draw treatment decision based on Equation (4.9)
  $A_{0,i} \sim Bernoulli(logit^{-1}(\gamma_0 + \gamma_L L_{0,i}))$
  compute conditional hazard based on Equation (4.7)
  $\lambda_i(t = 0|A_{0,i}, L_{0,i}, U_i) = \alpha_0 + \alpha_A A_{0,i} + \alpha_L L_{0,i} + \alpha_U U_i$
  $V_{0,i} \sim U(0, 1)$
  $\Delta T_{0,i} = -\log(V_{0,i})/\lambda_i(t = 0|A_{0,i}, L_{0i}, U_i)$
  if $\Delta T_{0,i} < 1$ **then**
    $\tilde{T}_i \leftarrow \Delta T_{0,i}$
    $Y_{1,t} = 1$ *(indicates subject i died within the period* $0 < t < 1$*, with event time as* $\tilde{T}_i$*)*
  else
    $Y_{1,t} = 0$ *(indicates subject i remains at risk at next visit time* $(t = 1)$*)*
  end
  for $t \leftarrow 1 \dots K$ **do**
    if no event occurred, continue sampling the data
    if $\tilde{T}_i = null$ **then**
      $L_{t,i} \sim \mathcal{N}(0.8L_{t-1,i} - A_{t-1,i} + 0.1t + U_i, 1)$
      draw treatment decision based on Equation (4.10)
      $A_{t,i} \sim Bernoulli(logit^{-1}(\gamma_0 + \gamma_A A_{t-1,i} + \gamma_L L_{t,i}))$
      compute conditional hazard based on Equation (4.7)
      $\lambda_i(t|\bar{A}_{t,i}, \bar{L}_{t,i}, U_i) = \alpha_0 + \alpha_A A_{t,i} + \alpha_L L_{t,i} + \alpha_U U_i$
      $V_{t,i} \sim U(0, 1)$
      $\Delta T_{t,i} = -\log(V_{t,i})/\lambda_i(t|\bar{A}_{s,i}, \bar{L}_{s,i}, U_i)$
      if $\Delta T_{t,i} < 1$ **then**
        $\tilde{T}_i \leftarrow t + \Delta T_{t,i}$
        $Y_{t+1,i} = 1$*(subject i died within the period* $[t, t + 1)$*, with event time as* $\tilde{T}_i$*)*
      else
        $Y_{t+1,i} = 0$ *(subject i remains at risk at next visit time* $(t + 1)$*)*
      end
    end
  end
  if still no event occurred, subject i is administratively censored at time $t = K + 1$
  if $\tilde{T}_i = null$ **then**
    $\tilde{T}_i = K + 1$
  end
end

confounding, where the MSM specifying the marginal hazard is known. This also allows us to evaluate the performance of IPTW estimator under different scenarios. It is worthy to note that, compared to previous algorithms for simulating longitudinal data corresponding to specified Cox MSMs [1, 23, 24], this algorithm based on additive hazard models does not require restrictive assumptions about the longitudinal relationships between variables or the distributions of variables, and can be applied in more general simulation settings.

However, using additive hazard models also has some drawbacks. It can not inherently restrict the hazard to be non-negative, which may lead to calculated survival probabilities greater than one. A feasible solution proposed by Keogh *et al.* (2021) is to carefully select a set of proper values for the parameters in the hazard model and the parameters determining the distribution of observed covariate L_t. This ensure that the probability of obtaining a negative hazard is small and negligible, mitigating the issue of unrealistic survival probabilities.

4.5 Comparison of the two simulation algorithms

The two simulation algorithms presented in Sections 4.3 and 4.4 both demonstrate good performances in simulating longitudinal data with time-dependent confounding from the desired MSMs. As summarized in Table 4.1, the algorithms differs in several aspects, including the key mechanism of the data-generating processes, the DAG structures, the used conditional hazard models, and their flexibility in simulation scenarios.

The main difference between these two algorithms lies in the different structure of their DAGs, which indicates distinct dependency setting and, consequently, different simulating procedures. Although the settings in both DAGs are similar, a key difference is that the DAG of Keogh *et al.* (2021) includes the direct arrow from L_t to Y_{t+1}, which is omitted in the DAG of Havercroft and Didelez (2012). This makes the setting of Keogh *et al.* (2021) more realistic than the other. In addition, as shown in Figure 4.1, the algorithm proposed by Havercroft and Didelez (2012) defines U_t as a time-dependent variable and specifies the dependence of survival outcome Y_t on only U_0. Conversely, for the DAG in Figure 4.2, U is constant over time and consistently affects the survival outcome Y_t across time.

Moreover, the mechanisms of data-generating process simulating from the desired MSM differ. By using $U_{0,i}$ as a probability threshold to determine the survival outcome, the algorithm by Havercroft and Didelez (2012) successfully simulates longitudinal data with time-dependent confounding from the desired MSM incorporating a well-defined conditional pooled logistic hazard model. On the other hand, taking advantage of the attractive property of Aalen's additive hazard model, where the derived MSM remains in additive form and can be correctly specified, the data-generating process presented by Keogh *et al.*(2021) simulates longitudinal data with time-dependent confounding from

Table 4.1: Comparison between the simulation algorithms by Havercroft and Didelez (2012) and Keogh *et al.* (2021).

	Havercroft and Didelez (2012)	**Keogh *et al.* (2021)**
DAG structure	Omit the direct arrow from L_t to Y_{t+1}	Include the direct arrow from L_t to Y_{t+1}
Conditional hazard model	Pooled logistic regression model (hazard within $(0, 1)$)	Aalen's additive hazard model (hazard could be negative)
Key mechanism	U_0 as a probability threshold in determining survival outcome	MSM derived from Aalen's additive hazard model is also additive and can be correctly specified
Simulation scenarios	Limited	Versatile

a known MSM specifying the marginal hazard under a conditional model simulation setting.

According to the generalization of algorithms, the algorithm by Havercroft and Didelez (2012) aims to generate data that closely match the Swiss HIV Cohort Study [22]. Achieving this requires restrictive assumptions on some variable distributions, limiting its generalizability to other scenarios. In contrast, the algorithm by Keogh *et al.* (2021) does not have these restrictions, and is expected to have better generalization properties, making it more versatile in simulating data across various settings.

Finally, it is worth noting that the additive hazard model can produce negative hazard values, whereas the pooled logistic regression model restricts the hazard within the range of $(0, 1)$. This characteristic of the models should be taken into account while simulating data and interpreting the results.

Chapter 5

Simulating from MSMs under positivity violation

In this chapter, we will provide an introduction to the violations of positivity assumption, and describe the approaches used to introduce the positivity violations into the two simulation Algorithms 1 and 2 presented in Chapter 4. Finally we illustrate the various simulation scenarios employed to investigate positivity violations.

5.1 Violations of the positivity assumption

When estimating the causal effects of joint treatments over time based on the longitudinal data with time-dependent confounding, directly fitting the standard regression model to the data is not appropriate due to the presence of confounding. Instead, the most commonly used estimation approach is to employ marginal structural models (MSMs), estimated using Inverse Probability of Treatment Weighting (IPTW).

As explained in Section 2.6, IPTW enables us to reweight individuals using time-dependent weights as in Equation (2.22), to account for the confounding and estimate the causal effects accurately. By reweighting the data, IPTW aims to balance the treatment and control groups, effectively creating a pseudo-population in which the distribution of covariates is independent of treatment assignment. This balancing leads to the unbiased estimation of the treatment effect by removing the bias introduced by the confounding variables. However, the use of this method to estimate the causal effect requires three key identifiability assumptions introduced in Section 2.2 (or in Section 2.5.1 for the time-dependent sequential setting): consistency, conditional exchangeability and positivity.

Specifically, the (sequential) positivity assumption for time-dependent treatment and confounder in Equation (2.18), i.e.,

$$P(A_t = a_t | \bar{\boldsymbol{A}}_{t-1} = \bar{\boldsymbol{a}}_{t-1}, \bar{\boldsymbol{L}}_t = \bar{\boldsymbol{l}}_t) > 0,$$

states that, for any combination of values of the observed covariates and the previous

treatment history, there should be a non-zero probability of receiving any treatment level at any given time point. In other words, all potential treatment levels must be feasible for all individuals at each time point, given their observed characteristics and prior treatment history. Therefore, violating the positivity assumption for certain subsets of the population means that those subsets have a zero probability of receiving a specific treatment. In the sequential setting, the stabilized weights used in IPTW are given by:

$$sw(t) = \prod_{k=0}^{\lfloor t \rfloor} \frac{P(A_k = a_k | \bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1})}{P(A_k = a_k | \bar{\boldsymbol{A}}_{k-1} = \bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{L}}_k = \bar{\boldsymbol{l}}_k)}. \tag{5.1}$$

Therefore, if certain subsets have a zero probability of receiving a specific sequence of treatments (i.e., individuals with certain covariate history to receive a given exposure history of interest are impossible to be observed), the denominator in stabilized weight (5.1) will become zero and the thus the weights according to the subsets will be undefined.

5.1.1 Types of positivity violations

There are two forms of positivity violation: structural violation and practical violation [7]. *Structural violation* occurs when it is impossible for a certain subject to receive certain treatment. For example, certain patient characteristics may act as contraindication for specific treatment. In clinical trials, various indicators are used to assess and monitor the health condition of patients. These indicators serve as valuable measures to evaluate the efficacy and safety of treatments being tested. If the indicator values of a patient surpass or fall below a specific threshold, it may indicate a deteriorating health condition. In such cases, it might be necessary to withhold certain specific treatments to avoid potential harm to the patients. On the other hand, *practical* or *random violation* refers to the situation where the assignment to specific treatment for patients in a certain subgroup is theoretically possible but is not observed in the data due to randomness. Increasing sample size is likely to ameliorate this problem.

In the following sections, our focus is on the structural positivity violations occurring when there is a threshold τ above or below which (depending on the role of the confounder) no subject in specific subgroups is unexposed to the treatment.

5.2 Structural positivity violation by thresholding

We now intentionally introduce the violation of the positivity assumption as the form of structural violation into the simulation procedures presented in Sections 4.3 and 4.4, aiming to investigate the impact of such violation on the performance and robustness of the approach of IPTW. By deliberately creating scenarios where the treatment assignment probabilities approach zero or one, we seek to gain insights into the potential biases and limitations that the IPTW method may encounter in real-world situations. Through these intentional violations, we can explore the sensitivity of the IPTW approach to structural positivity violations and understand the potential consequences of

such violations in practical applications.

Specifically, we extend the Algorithms 1 and 2 proposed by Havercroft and Didelez (2012) (Section 4.3) and by Keogh *et al.* (2021) (Section 4.4), respectively, by deterministically assigning the treatment for specific subgroups of subjects with a confounder above or below (depending on the role of the confounder) a given threshold value τ.

5.2.1 Simulation approach I

Simulation approach I extends the one proposed by Havercroft and Didelez (2012) (see Section 4.3; Algorithm 1) by introducing structural positivity violation by thresholding as we explain in the following.

According to the DAG in Figure 4.1, the data-generating process of Algorithm 1 ensures that the measured covariate L_t is sufficient to adjust for confounding, and treatment decision A_t depends on its previous treatment A_{t-1} and L_t. In this process of emulating the Swiss HIV Cohort Study [22], L_t represents the subject's CD4 count in $cells/\mu L$ at time t. Briefly, CD4 count is a blood test that measures the number of CD4 cells in a sample of the subject's blood [25]. Healthy adults and teens typically have a normal CD4 count ranging from 500 $cells/\mu L$ to 1200 $cells/\mu L$. It is mostly used to assess the health of the immune system in patients infected with human immunodeficiency virus (HIV). HIV attacks and destroys CD4 cells, and a low CD4 count indicates the patient's immune system may be weakened, making them susceptible to develop serious infections from viruses, bacteria, or fungi that typically do not cause problems in healthy individuals. Hence, a low value of CD4 count L_t implies a more severe illness in the subject. In this scenario, it is reasonable to start treatment for the subjects when their measured CD4 count L_t falls below a certain threshold, provided they have not started it yet (subjects remains on treatment once initiated).

The idea of imposing the violation of positivity assumption in this data-generating process is to control a threshold τ for the CD4 count L_t. Below this threshold, subjects are always exposed to treatment. Since this disease scenario requires that once a subject started treatment, they remain on treatment until failure or end of follow-up, the structural positivity violation by thresholding can be specified as follows: if the subject is not yet on treatment at time t (i.e., $A_{t-1} = 0$), there should be a structural zero probability that a subject will remain unexposed when his or her CD4 count L_t falls below or equals τ, i.e.,

$$\begin{aligned} &P(A_t = 1 | L_t \leq \tau, A_{t-1} = 0) = 1, \text{ and} \\ &P(A_t = 0 | L_t \leq \tau, A_{t-1} = 0) = 0. \end{aligned} \tag{5.2}$$

This means that we force the subject to start the treatment when their health condition is poor (i.e., low CD4 count).

The complete pseudocode for simulation approach I under the violation of positivity assumption in Equation (5.2) is reported in Algorithm 3. The pseudocode in purple highlights the parts where structural positivity violations by thresholding τ is introduced. In other words, below the threshold exposure is assigned deterministically and above the threshold exposure is assigned stochastically.

5.2.2 Simulation approach II

Simulation approach II extends the one proposed by Keogh *et al.* (2021) (see Section 4.4; Algorithm 2) by introducing structural positivity violation by thresholding as we explain in the following.

According to the DAG shown in Figure 4.2, the simulation process proposed in Algorithm 2 ensures that the measured covariate L_t is sufficient to adjust for confounding, and treatment decision A_t depends on its previous treatment A_{t-1} and L_t. In this scenario, the time-dependent confounder L_t represents a biomarker, and higher values of this biomarker are associated with both a higher hazard and a higher propensity to receive the treatment.This implies that a high value of the L_t indicates a poor state of the patient's health and can be considered a compelling reason to initiate the treatment. In contrast to the previous case, the strustural positivity violation by thresholding is imposed in an opposite manner here.

In this simulation process, instead of controlling a threshold for CD4 count below which subjects are always exposed to treatment, we now control a threshold τ for the biomarker L_t, above which the subjects are exposed to treatment. In addition, the simulation process by Keogh *et al.* (2021) does not require that the subject should remain on treatment until failure or end of follow-up once the subject has started treatment. Therefore, the violation rule in this second scenario is different. Regardless of the subject's previous treatment status at time t, there should be a structural zero probability that a subject will be unexposed when his or her biomarker value L_t is greater than or equal to τ, i.e., for all treatment a

$$\begin{aligned} &P(A_t = 1 | L_t \geq \tau, A_{t-1} = a) = 1, \text{ and} \\ &P(A_t = 0 | L_t \geq \tau, A_{t-1} = a) = 0. \end{aligned} \tag{5.3}$$

This means that we will force the subject to receive the treatment at time t when their health condition is poor (i.e., high biomarker value L_t).

The complete pseudocode for simulation approach II under the violation of positivity assumption in Equation (5.3) is reported in Algorithm 4. The pseudocode in purple highlights the parts where structural positivity violations by thresholding τ is introduced. In other words, above the threshold exposure at time t is assigned deterministically, and below the threshold exposure at time t is assigned stochastically.

Algorithm 3: Simulation approach I with structural positivity violation by thresholding τ in Equation (5.2) (shown in purple).

Input:
K: number of time points during follow-up
k_c: check-up times
n: number of subjects
τ: the threshold of CD4 cell count L_t to determine the treatment decision mechanism
$(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$: parameters of MSM hazard in Equation (4.3)
$(\theta_0, \theta_1, \theta_2)$: parameters of the conditional distributions of treatment in Equation (4.4)
Result: The longitudinal dataset of n subjects

Simulation process:
for *subject* $i \leftarrow 1$ ***to*** n **do**
  $U_{0,i} \sim Uniform(0,1)$
  $\epsilon_{0,i} \sim \mathcal{N}(0, 20)$
  simulate baseline CD4 cell count
  $L_{0,i} \leftarrow F^{-1}_{\Gamma(3,154)}(U_{0,i}) + \epsilon_{0,i}$
  **Force assignment to treatment for CD4 cell count below τ or draw treatment decision based on Equation (4.4)*
  if $L_{0,i} \leq \tau$ **then**
    $A_{0,i} = 1$
  else
    $A_{0,i} \sim Bernoulli(logit^{-1}(\theta_0 + \theta_2(L_{0,i} - 500)))$
  end
  if $A_{0,i} = 1$ **then**
    $T_i^* \leftarrow 0$ *(indicates treatment starts at time point 0)*
  end
  calculate the hazard based on Equation (4.5)
  $\lambda_{0,i} = logit^{-1}(\gamma_0 + \gamma_2 A_{0,i})$
  if $\lambda_{0,i} \geq U_{0,i}$ **then**
    $Y_{1,i} \leftarrow 1$ *(subject is dead)*
  else
    $Y_{1,i} \leftarrow 0$ *(subject is still alive)*
  end
  for $t \leftarrow 1, \ldots, K$ **do**
    if no death occurred, continue sampling the data
    while $Y_{t,i} = 0$ **do**
      $\Delta_{t,i} \sim \mathcal{N}(0, 0.05)$
      $U_{t,i} \leftarrow \min\{1, \max\{0, U_{t-1,i} + \Delta_{t,i}\}\}$
      if t *mod* $k_c \neq 0$ **then**
        if not at a check-up time point, $L_{t,i}$ and $A_{t,i}$ remain the previous values
        $L_{t,i} \leftarrow L_{t-1,i}$
        $A_{t,i} \leftarrow A_{t-1,i}$
      else
        if treatment starts at the last check-up, CD4 count will have a shift of 150
        $\epsilon_{t,i} \sim \mathcal{N}(100(U_{t,i} - 2), 50)$
        $L_{t,i} \leftarrow \max\{0, L_{t-1,i} + 150 A_{t-k_c,i}(1 - A_{t-k_c-1,i}) + \epsilon_{t,i}\}$
        if $A_{t-1,i} = 0$ **then**
          **Force assignment to treatment for CD4 cell count below τ or draw treatment decision based on Equation (4.4)*
          if $L_{t,i} \leq \tau$ **then**
            $A_{t,i} = 1$
          else
            $A_{t,i} \sim Bernoulli(logit^{-1}(\theta_0 + \theta_1 t + \theta_2(L_{t,i} - 500)))$
          end
        else
          $A_{t,i} \leftarrow 1$ *(subject remains on treatment once initiated)*
        end
        if $A_{t,i} = 1$ *and* $A_{t-k_c,i} = 0$ **then**
          $T_i^* \leftarrow t$ *(indicates treatment starts at time point t)*
        end
      end
      compute the hazard to generate the survival outcome based on Equation (4.5)
      $\lambda_{t,i} \leftarrow logit^{-1}(\gamma_0 + \gamma_1[(1 - A_{t,i})t + A_{t,i}T_i^*] + \gamma_2 A_{t,i} + \gamma_3 A_{t,i}(t - T_i^*))$
      if $1 - \Pi_{\tau=0}^{t}(1 - \lambda_{\tau,i}) \geq U_{0,i}$ **then**
        $Y_{t+1,i} = 1$ *(subject is dead)*
      else
        $Y_{t+1,i} = 0$ *(subject is still alive)*
      end
    end
  end
end

Algorithm 4: Simulation approach II with structural positivity violation by thresholding τ in Equation (5.3) (shown in purple).

Input:
K: number of time points during follow-up
n: number of subjects
τ: the threshold of biomarker value L_t to determine the treatment decision mechanism
$(\alpha_0, \alpha_A, \alpha_L, \alpha_U)$: parameters of conditional hazard model in Equation (4.7)
$(\gamma_0, \gamma_A, \gamma_L)$: parameters for the conditional distributions of treatment in Equation (4.10)
Result: The longitudinal dataset of n subjects

Simulation process:
for *subject* $i \leftarrow 1$ ***to*** n **do**
  simulate individual frailty
  $U_i \sim \mathcal{N}(0, 0.1)$
  initialize a value for biomarker
  $L_{0,i} \sim \mathcal{N}(U_i, 1)$
  **Force assignment to treatment for biomarker value above τ or draw treatment decision based on Equation (4.9)*
  if $L_{0,i} \geq \tau$ **then**
    $A_{0,i} = 1$
  else
    $A_{0,i} \sim Bernoulli(logit^{-1}(\gamma_0 + \gamma_L L_{0,i}))$
  end
  compute conditional hazard based on Equation (4.7)
  $\lambda_i(t = 0|A_{0,i}, L_{0i}, U_i) = \alpha_0 + \alpha_A A_{0,i} + \alpha_L L_{0,i} + \alpha_U * U_i$
  $V_{0,i} \sim U(0, 1)$
  $\Delta T_{0,i} = \frac{-log(V_{0,i})}{\lambda_i(t=0|A_{0,i}, L_{0i}, U_i)}$
  if $\Delta T_{0,i} < 1$ **then**
    $\tilde{T}_i \leftarrow \Delta T_{0,i}$
    $Y_{1,t} = 1$ *(indicates subject i died within the period $0 < t < 1$, with event time as $\tilde{T}_i$)*
  else
    $Y_{1,t} = 0$ *(indicates subject i remains at risk at next visit time $(t = 1)$)*
  end
  for $t \leftarrow 1 \dots K$ **do**
    if no event occurred, continue sampling the data
    if $\tilde{T}_i = null$ **then**
      $L_{t,i} \sim \mathcal{N}(0.8L_{t-1,i} - A_{t-1,i} + 0.1t + U_i, 1)$
      **Force assignment to treatment for biomarker value above τ or draw treatment decision based on Equation (4.10)*
      if $L_{t,i} \geq \tau$ **then**
        $A_{t,i} = 1$
      else
        $A_{t,i} \sim Bernoulli(logit^{-1}(\gamma_0 + \gamma_A A_{t-1,i} + \gamma_L L_{t,i}))$
      end
      compute conditional hazard based on Equation (4.7)
      $\lambda_i(t|\bar{A}_{t,i}, \bar{L}_{t,i}, U_i) = \alpha_0 + \alpha_A A_{t,i} + \alpha_L L_{t,i} + \alpha_U U_i$
      $V_{t,i} \sim U(0, 1)$
      $\Delta T_{t,i} = \frac{-\log(V_{t,i})}{\lambda_i(t|\bar{A}_{s,i}, \bar{L}_{s,i}, U_i)}$
      if $\Delta T_{t,i} < 1$ **then**
        $\tilde{T}_i \leftarrow t + \Delta T_{t,i}$
        $Y_{t+1,i} = 1$*(indicates subject i died within the period $[t, t + 1)$, with event time as $\tilde{T}_i$)*
      else
        $Y_{t+1,i} = 0$ *(indicates subject i remains at risk at next visit time $(t + 1)$)*
      end
    end
  end
  if still no event occurred, subject i is administratively censored at time $t = K + 1$
  if $\tilde{T}_i = null$ **then**
    $\tilde{T}_i = K + 1$
  end
end

5.3 Simulation scenarios

Algorithms 3 and 4 enable data simulation in a time-dependent survival context in situations where the positivity assumption is structurally violated. In this section, we present an overview of the various scenarios used to investigate positivity violations. Specifically, we explore various scenarios by controlling different values of thresholds τ to introduce more or less extreme violations of positivity. Additionally, we examine the impact of different sample sizes n on the results. Detailed results for each scenario are presented in Chapter 6.

5.3.1 Scenarios for simulation approach I

In Algorithm 3, the parameter τ controls the threshold of CD4 cell count below which the subject is deterministically assigned to treatment following Equation (5.2). The normal CD4 count for healthy adults and teens is between 500 $cells/\mu L$ to 1200 $cells/\mu L$, while a CD4 count of below 500 $cells/\mu L$ is considered as a low CD4 count. To explore the performance of the IPTW estimator under different scenarios with varying scales of positivity violation, we conduct simulations using different values of the threshold τ, namely

$$\tau = 500, 400, 300, 200, 100.$$

We simulate longitudinal data based on each threshold to explore how the IPTW estimator behaves in different scenarios with varying degrees of positivity violation. Specifically, a threshold value $\tau = 100$ can be considered as a mild case of positivity violations, whereas $\tau = 500$ corresponds to a more severe violation, as it indicates a larger proportion of the sampling population being impacted by this structural positivity violation.

Furthermore, we repeat the experiment using different sample sizes:

$$n = 100, 200, 300, 500, 1000,$$

to assess the impact of sample size on the performance of the IPTW estimator in the presence of positivity violation. By exploring the estimator's behavior under different sample sizes, we gain insights into how the precision and accuracy of the estimator are affected in scenarios with varying degrees of positivity violation.

IPTW model

To obtain the IPTW estimator, the MSM in Equation (4.3) (Section 4.3) is fitted to each simulated dataset using stabilized weights as described in Equation (5.1). As mentioned in [1], the weight components at time k (i.e., numerator and denominator) are estimated by employing logistic regression models for the probability of treatment initiation at time k. In particular, following the approach introduced by [26], the numerator is modelled as

$$logit(P(A_k = 1|\bar{\boldsymbol{A}}_{k-1})) = \theta_0 + \theta_1 k, \tag{5.4}$$

and the denominator as

$$logit(P(A_k = 1|\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k)) = \theta_0 + \theta_1 k + \theta_2(L_k - 500), \tag{5.5}$$

under the condition that treatment has not started yet (i.e., $\bar{\boldsymbol{A}}_{k-1} = \mathbf{0}$) and event has not yet occurred (i.e., $Y_k = 0$).

5.3.2 Scenarios for simulation approach II

In Algorithm 4, the parameter τ controls the threshold of the biomarker value above which the subject is deterministically assigned to treatment following Equation (5.3). Since the biomarker value L used in this context does not have a direct interpretation in the real world, the choice of the threshold τ relies on the distribution of L in this simulation procedure. By simulating L using Algorithm 2 with 1000 subjects, we obtained a distribution of L as shown in left panel of Figure 5.1. The according QQ plot shown in right panel supports that L approximately follows a normal distribution.

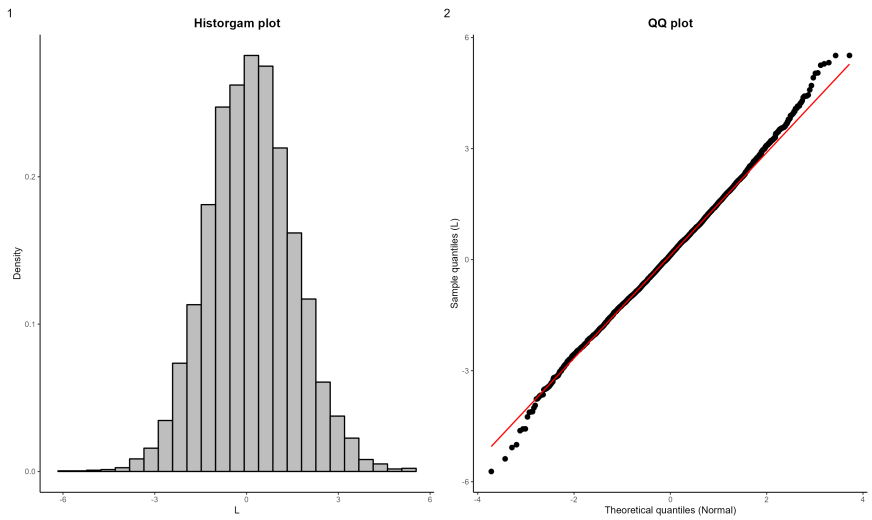

Figure 5.1: Left panel: histogram of L sampled by Algorithm 2 with 1000 subjects. Right panel: QQ-plot of L sampled by Algorithm 2 with 1000 subjects.

To set the thresholds τ, we use the approximated quantiles of the simulated L of order $\{0.4, 0.55, 0.65, 0.75, 0.85\}$, resulting in respective values of

$$\tau = -0.3, 0.3, 0.6, 1.0, 1.5.$$

These thresholds are then used to simulate the longitudinal data according to the specified violation rule in Equation (5.3). This allows us to explore how the IPTW estimator behaves in different scenarios with varying degrees of positivity violation. Specifically, a value $\tau = -0.3$ can be considered as a more extreme case of positivity violations,

whereas $\tau = 1.5$ corresponds to a mild violation, as it indicates a smaller proportion of the sampling population being impacted by the structural positivity violation.

Furthermore, we repeat the experiment using different sample sizes:

$$n = 100, 200, 300, 500, 1000,$$

to assess the impact of sample size on the performance of IPTW estimator under different scenarios of positivity violation.

IPTW model

To obtain the IPTW estimator, the MSM in Equation (4.8) (Section 4.4) is fitted to each simulated data set using stabilized weights as described in Equation (5.1). As specified in the original paper [2], the weight components at time k are estimated using logistic regression models for the probability of being on treatment at time k. In particular, numerator is modelled as

$$logit(P(A_k = 1|\bar{\boldsymbol{A}}_{k-1})) = \gamma_0 + \gamma_A A_{k-1}, \tag{5.6}$$

and denominator as

$$logit(P(A_k = 1|\bar{\boldsymbol{A}}_{k-1}, \bar{\boldsymbol{L}}_k)) = \gamma_0 + \gamma_A A_{k-1} + \gamma_L L_k, \tag{5.7}$$

under the condition that event has not occurred yet (i.e., $T \geq k$).

Chapter 6

Results

In this chapter, we will provide a detailed explanation of the simulation results, focusing on bias, MSE and counterfactual survival probabilities. These results will be presented in two contexts: (1) under the violation scenarios outlined in Section 5.3.1, using data generated through the simulation approach I in Algorithm 3, and (2) under the violation scenarios outlined in Section 5.3.2, using data generated thorugh the simulation approach II as in Algorithm 4.

Statistical analyses have been performed in the R software environment [27]. The code to reproduce the simulation studies is reported here: `https://github.com/MillarHuang/Thesis-project-simulation-.git`.

6.1 Simulation studies

The objective is to assess the effectiveness of Inverse Probability of Treatment Weighting (IPTW) under the different scenarios presented in Section 5.3. For each simulation approach, we conducted $B = 100$ replications encompassing various combinations of sample sizes and threshold values. As performance measures, we focused on bias in Equation (4.1) and mean square error (MSE) in Equation (4.2) introduced in Section 4.1.

Bias serves as a measure for assessing the average deviation of the IPTW estimator from the true value. It provides insights into whether the estimator consistently overestimates or underestimates the actual true value. On the other hand, MSE serves as an indicator of the precision of the IPTW estimator. It quantifies the overall accuracy of the estimator by combining the squared bias and the variance of the estimator. Thus, MSE acts as a comprehensive performance measure that summarizes both bias and variance.

6.2 Results of simulation approach I

Below, we present the results of our simulation study based on simulation approach I aimed at assessing the efficacy of the IPTW-estimator in estimating MSMs within a time-dependent framework when the positivity assumption is imposed. The Algorithm 3 discussed in Section 5.2.1 is employed to generate data, enabling us to investigate the scenarios detailed in Section 5.3.1. Moreover, we adopted the input parameter values from the original paper by Havercroft and Didelez (2012), as outlined in Section 4.3.

6.2.1 Bias and MSE

The logistic MSM in Equation (4.3) used in simulation approach I presents four parameters to be estimated: γ_0 that is the baseline intercept, γ_1 related to the time elapsed before treatment initiation, γ_2 related to the current treatment value, and γ_3 related to the time elapsed after treatment initiation. Therefore, we focus our attention on γ_2 because it is the parameter most directly related to the effect of treatment.

Table 6.1 shows the average bias for γ_2 estimates under the simulation scenarios presented in Section 5.3.1. It can be observed that for a given sample size n, as the threshold τ increases (i.e., indicating a more severe violation of the positivity assumption), the IPTW estimate of γ_2 demonstrates an increasing positive bias tendency. This suggests that the treatment effect becomes less effective when the positivity assumption is more severely violated. A plausible explanation for this observation is that as τ increases, individuals with CD4 counts below τ are compelled to initiate the treatment. Consequently, more patients with lower CD4 counts, which typically correspond to poorer health conditions, are included in the treatment group. This particular group of patients is more likely to experience shorter survival times compared to the others. Therefore, including such patients in the treatment group due to positivity violation can lead to an underestimation of the treatment's effectiveness. Furthermore, it is noteworthy that as the sample size n increases, the bias of the IPTW estimator grows at a slower rate with increasing violation (i.e., larger values of τ). This behavior aligns with expectations, as larger sample sizes can partially mitigate the bias to some extent by minimizing finite sample bias.

Table 6.1: Bias of γ_2 estimates produced by IPTW method under different scenarios.

	Threshold τ				
Sample size n	100	200	300	400	500
100	0.011	0.781	7.245	8.977	9.097
200	0.069	0.786	3.933	5.480	7.708
300	0.094	0.759	3.106	4.519	6.463
500	0.085	0.787	2.670	4.175	5.369
1000	0.127	0.848	2.564	3.260	3.428

Table 6.2: MSE of γ_2 estimates produced by IPTW method under different scenarios.

	Threshold τ				
Sample size n	100	200	300	400	500
100	0.164	0.837	93.516	116.440	127.152
200	0.067	0.741	28.987	50.351	86.313
300	0.065	0.650	13.102	32.340	64.741
500	0.039	0.656	7.639	24.922	46.244
1000	0.031	0.755	6.745	11.699	16.393

Table 6.2 reports the MSE for γ_2 estimates under the simulation scenarios presented in Section 5.3.1. MSE shows a similar trend as bias: under a certain sample size, the MSE of IPTW estimate of γ_2 increases as the degree of violation intensifies (i.e., higher τ). Moreover, as the sample size increases, the rate at which the MSE of the IPTW estimate of γ_2 grows in response to the degree of violation decreases. This suggests that in scenarios where the positivity assumption is violated, the precision of the IPTW estimator could be compromised. However, increasing the sample size has the potential to mitigate the detrimental effect caused by the violation.

The overall insight of bias and MSE for all causal parameters under different scenarios are given in Figure 6.1 and Figure 6.2, respectively. In both figures, each panel refers to a different MSM parameter $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ with different lines/colours representing different sample sizes $n \in \{100, 200, 300, 500, 1000\}$. The general pattern observed in absolute bias and MSE of γ_0, γ_1, and γ_3 aligns with that of γ_2. However, positivity violations have a more significant impact on the magnitude of bias/MSE for γ_2, i.e., the parameter most directly related to the effect of exposure, and the intercept γ_0.

6.2.2 Counterfactual survival curves for *always* and *never* treated

Based on the observation in Section 6.2.1, it is evident that the violation of positivity introduces bias into the estimated causal parameters. Consequently, this bias can impact the counterfactual survival curves computed using these estimated causal parameters. To assess the influence of positivity violations on the counterfactual survival curves and enable a more meaningful comparison with the results in Section 6.3.2 later on, we proceed to estimate the counterfactual survival probabilities for the following two treatment regimes:

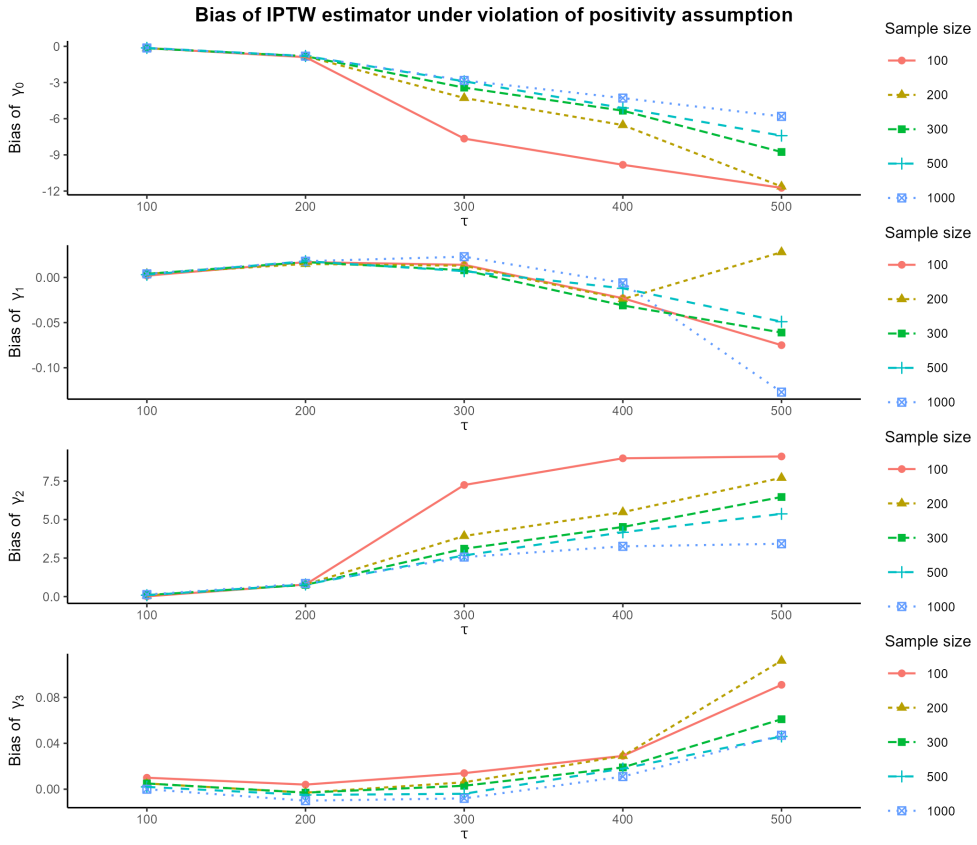

Figure 6.1: Bias of IPTW estimator under different scenarios presented in Section 5.2.1 using data generated from simulation approach I in Algorithm 3. Each panel refers to a different MSM parameter $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Different lines and colours refer to sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

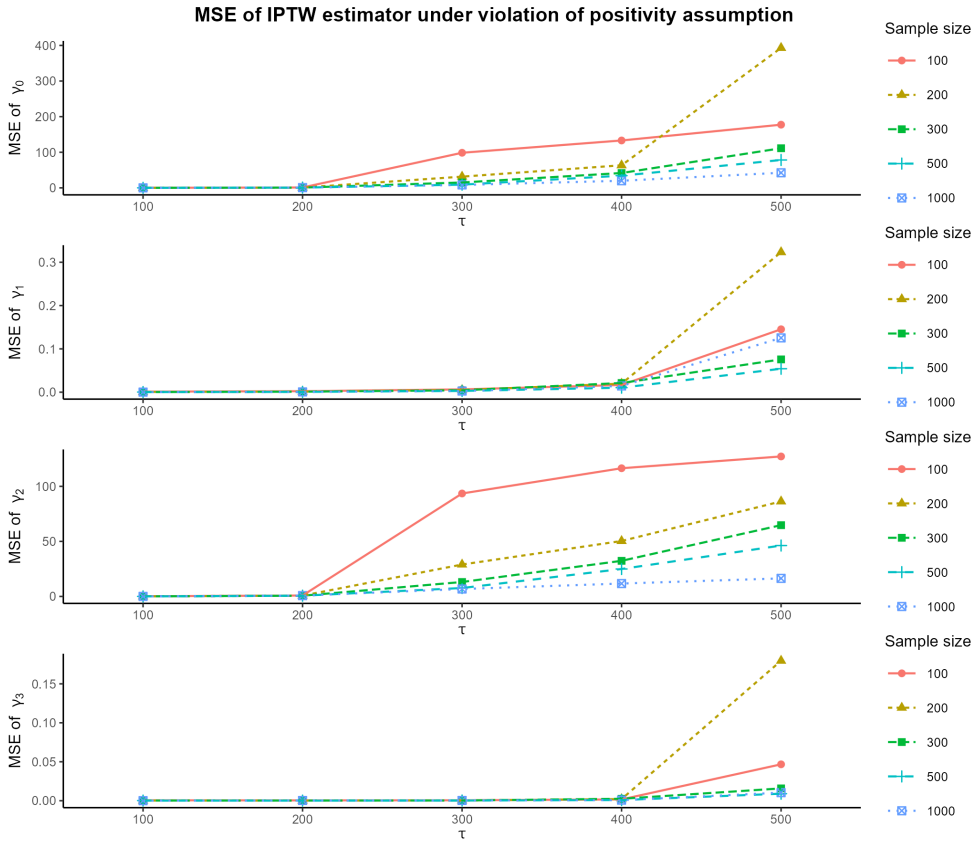

Figure 6.2: MSE of IPTW estimator under different scenarios presented in Section 5.3.1 using data generated from simulation approach I in Algorithm 3. Each panel refers to a different MSM parameter $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

- *always treated* with $\bar{\boldsymbol{a}} = (1, 1, \ldots, 1) = \mathbf{1}$:

$$\begin{aligned} Pr(T^{\bar{\boldsymbol{a}}=\mathbf{1}} \geq t) &= \prod_{x \leq t} [1 - \lambda(x)] = \prod_{x \leq t} \left[1 - \left(logit^{-1}(\gamma_0 + \gamma_1 d_{1,x} + \gamma_2 a_x + \gamma_3 d_{3,x})\right)\right] \\ &= \prod_{x \leq t} [1 - \left(logit^{-1}(\gamma_0 + \gamma_2 + \gamma_3 x)\right)], \; t = 1, \ldots, 40; \end{aligned} \tag{6.1}$$

- *never treated* with $\bar{\boldsymbol{a}} = (0, 0, \ldots, 0) = \mathbf{0}$:

$$\begin{aligned} Pr(T^{\bar{\boldsymbol{a}}=\mathbf{0}} \geq t) &= \prod_{x \leq t} [1 - \lambda(x)] \\ &= \prod_{x \leq t} \left[1 - \left(logit^{-1}(\gamma_0 + \gamma_1 d_{1,x} + \gamma_2 a_x + \gamma_3 d_{3,x})\right)\right] \\ &= \prod_{x \leq t} [1 - \left(logit^{-1}(\gamma_0 + \gamma_1 x)\right)], \; t = 1, \ldots, 40; \end{aligned} \tag{6.2}$$

Figure 6.3 displays the IPTW mean estimates of the counterfactual survival probabilities of *always treated* (solid lines) and *never treated* (dashed lines) under the different scenarios presented in Section 5.3.1. Each panel refers to a different sample size $n \in \{100, 200, 300, 500, 1000\}$ with different lines/colours representing various thresholds $\tau \in \{100, 200, 300, 400, 500\}$. The black lines represent the IPTW mean estimates of counterfactual survival probabilities in the absence of positivity violations. It is evident that as the extent of positivity violation increases (i.e., higher threshold τ), the estimated counterfactual survival curves deviate further from the true counterfactual survival curves for both treatment regimes. However, it is worth noting that the counterfactual survival curves of *never treated* show a greater deviation compared to those of *always treated*. Furthermore, in the presence of severe positivity violation (i.e., $\tau = 300, 400, 500$), the rate of decline in survival probabilities over time is more gradual for the *never treated* group compared to the *always treated* group. This observation can be attributed to the parameter settings detailed in Section 4.3 and the performance characteristics of the IPTW estimator in the presence of positivity violations as shown in Figure 6.1. Based on Equations (6.1) and (6.2), the estimated counterfactual survival probabilities of *always treated* primarily depend on the values of γ_2 and γ_3 as time progresses, while the estimated counterfactual survival probabilities of *never treated* mainly depends on the values of γ_1 over time. As depicted in Figure 6.1, when the positivity violation become more severe, the estimated γ_1 demonstrates a greater negative bias, whereas the estimated γ_2 and γ_3 both exhibit an greater positive bias. Consequently, it's expected that the survival probabilities of *never treated* will exhibit a slower rate of decline over time compared to those of *always treated*. Furthermore, with larger sample sizes, the survival curves become more distinct across varying thresholds τ, showing a more evident impact of positivity violations on the estimated survival probabilities.

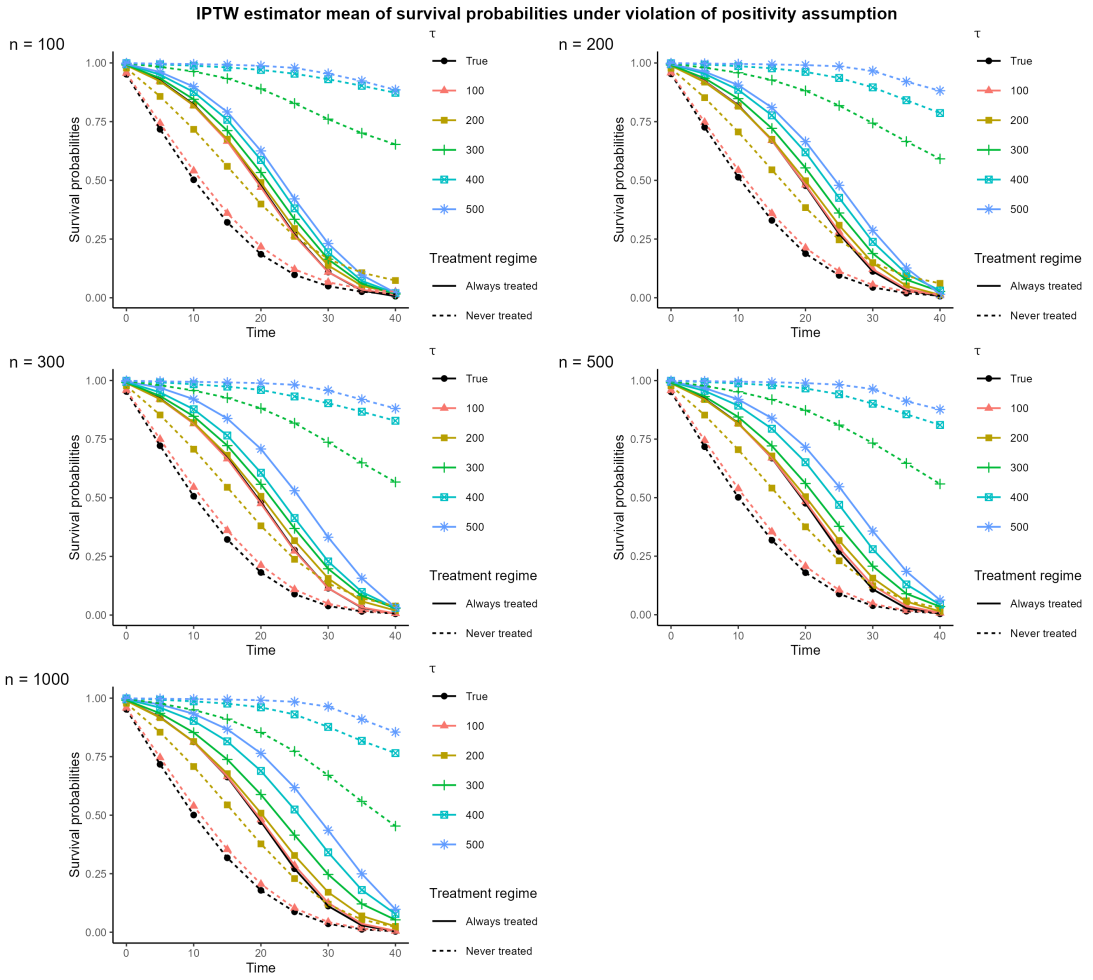

Figure 6.3: IPTW mean estimates of the counterfactual survival probabilities of *always treated* (solid lines) and *never treated* (dashed lines) under the different scenarios presented in Section 5.3.1 using data generated from simulation approach I in Algorithm 3. Each panel refers to a different sample size $n \in \{100, 200, 300, 500, 1000\}$. Different lines and colours refer to different thresholds $\tau \in \{100, 200, 300, 400, 500\}$.

6.3 Results of simulation approach II

Below, we present the results of our simulation study based on simulation approach II aimed at assessing the efficacy of the IPTW-estimator in estimating MSMs within a time-dependent framework when the positivity assumption is imposed. The Algorithm 4 discussed in Section 5.2.2 is employed to generate data, enabling us to investigate the scenarios detailed in Section 5.3.2. Moreover, we adopted the input parameter values from the original paper by Keogh *et al.* (2021), as outlined in Section 4.4.

6.3.1 Bias and MSE

Contrary to proportional hazard model or logistic hazard model where the risk coefficients are constant, the Aalen's additive hazard MSM in Equation (4.8) presents risk coefficients that are functions of time. This indicates that the effect of a covariate may vary over time. In this case, the estimands of interest are the cumulative coefficients

$$C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds \quad \text{and} \quad C_{Aj}(t) = \int_0^t \tilde{\alpha}_{Aj}(s)ds \text{ with } j = 0, 1, 2, 3, 4.$$

In particular, the longitudinal data are generated for a total of 5 visits with administrative censoring at time 5. Therefore, the cumulative coefficients are estimated at time points $t = 1, 2, 3, 4, 5$ to evaluate the performances. In addition, interpreting these cumulative coefficients is difficult, and to achieve a direct causal interpretation they have to be translated into a more interpretable causal estimand, such as the counterfactual survival probability (see Section 6.3.2 below). To be comparable with γ_2 in the MSM of simulation approach I, the cumulative estimand of main interest here is

$$C_{A0}(t) = \int_0^t \tilde{\alpha}_{A0}(s)ds,$$

where $\tilde{\alpha}_{A0}(t)$ represents the coefficient related to the current treatment at time t.

The results of $\widehat{C}_{A0}(t)$ from the simulation study are shown in Table 6.3 and Figure 6.4 (where each panel refers to a different time-point t with different lines/colours representing the various sample sizes n). In contrast to the previous simulation study, the violation of the positivity assumption here arises from enforcing treatment on subjects whose biomarker value L_t exceeds the threshold τ. The smaller the threshold τ is, the more severe is the positivity violation. Based on these results, several observations can be made. First, for a given sample size n and time t, the magnitude of bias in the IPTW estimate of $C_{A0}(t)$ tends to slightly increase as the threshold τ decreases, due to the more severe violation. This trend becomes more pronounced for later time points t, as $C_{A0}(t)$ represents cumulative coefficients that accumulate bias information over the entire time interval $[0, t]$. Moreover, with increasing sample size n, the IPTW estimate of $C_{A0}(t)$ generally exhibits reduced fluctuation across different τ values, particularly for a later time point t.

As shown in Table 6.4 and Figure 6.5 (where each panel refers to a time-points t with different lines/colours representing the various sample sizes n), the MSE of the IPTW estimate for $C_{A0}(t)$ tends to slightly increase as the degree of violation intensifies (i.e., with a smaller value of τ) and as time elapses. This trend becomes more pronounced with a larger sample size. Furthermore, as the sample size increases, the MSE of the IPTW estimate for $C_{A0}(t)$ decreases. These results suggest that when the positivity assumption is more severely violated, the IPTW estimators exhibit higher bias. However, increasing the sample size has the potential to reduce the variance of the IPTW estimator and mitigate the scale of bias associated with the violation effect.

Table 6.3: Bias of $C_{A0}(t)$ estimates produced by IPTW method under the different scenarios presented in Section 5.3.2 using data generated from the simulation approach II in Algorithm 4.

	Threshold τ				
Time t	-0.3	0.3	0.6	1	1.5
$n = 100$					
1	0.034	0.022	-0.038	-0.048	-0.028
2	0.040	0.120	0.037	-0.111	-0.033
3	0.039	0.230	-0.002	-0.037	-0.078
4	0.010	0.245	0.020	0.035	-0.021
5	0.234	0.377	0.177	0.222	0.175
$n = 200$					
1	0.048	0.073	0.051	0.040	-0.021
2	0.026	0.087	0.018	0.041	-0.092
3	0.038	0.112	0.066	0.050	-0.116
4	0.104	0.125	0.183	0.118	-0.023
5	0.271	0.235	0.260	0.235	0.130
$n = 300$					
1	0.041	0.015	-0.039	-0.001	-0.029
2	0.074	0.149	-0.062	-0.004	-0.066
3	0.131	0.145	-0.061	-0.006	-0.024
4	0.185	0.235	-0.041	0.057	0.023
5	0.179	0.246	-0.040	0.217	0.069
$n = 500$					
1	0.063	-0.044	-0.048	0.027	-0.022
2	-0.011	-0.054	-0.053	-0.049	-0.038
3	0.033	-0.054	0.016	-0.028	-0.050
4	0.066	0.043	-0.027	-0.051	-0.013
5	0.127	0.198	-0.037	0.017	0.094
$n = 1000$					
1	-0.060	-0.037	-0.008	-0.042	0.011
2	-0.084	-0.043	-0.005	-0.086	-0.011
3	-0.098	-0.076	-0.045	-0.068	-0.010
4	-0.026	-0.221	-0.036	-0.030	0.006
5	0.091	-0.166	0.129	0.012	0.110

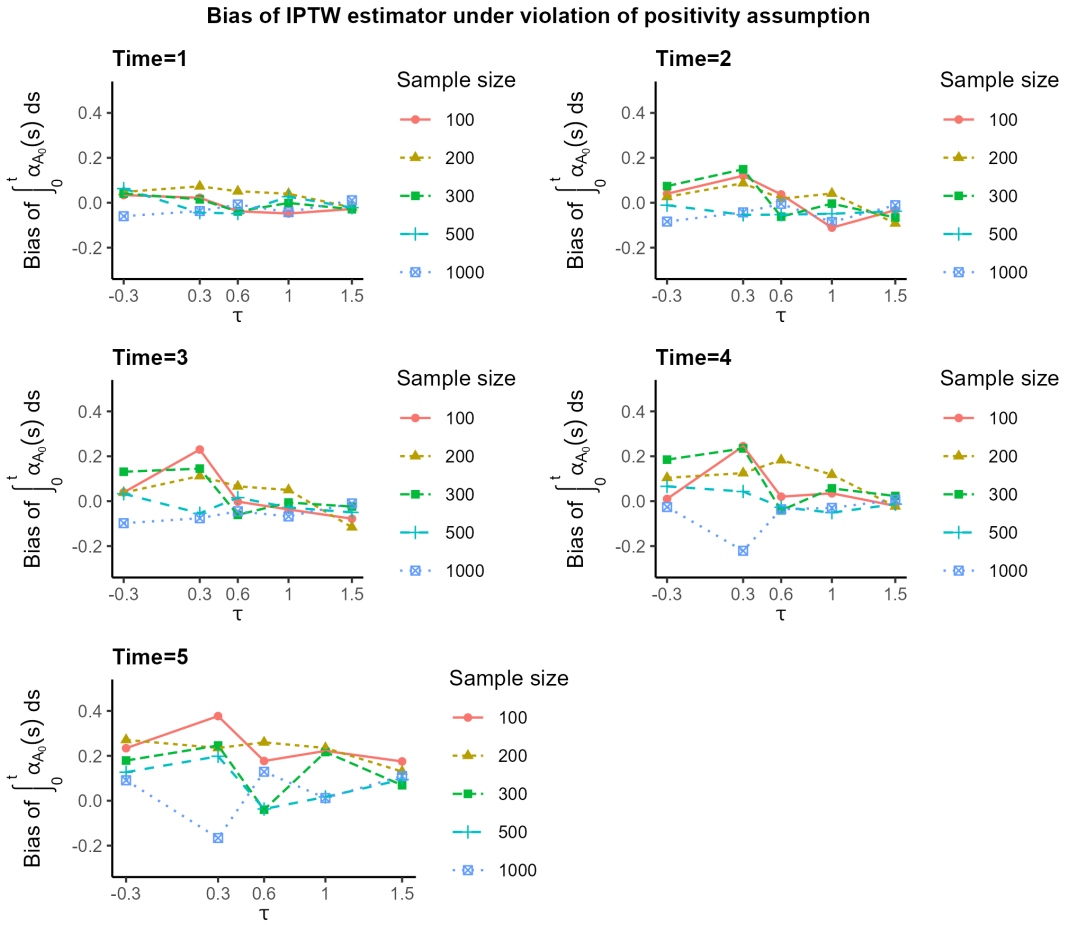

Figure 6.4: Bias of IPTW estimator of $C_{A0}(t)$ under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different time-point $t \in \{1, 2, 3, 4, 5\}$. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

Figures displaying the bias and MSE for all the cumulative coefficients $C_0(t)$ and $C_{Aj}(t)$ (for $j = 0, \ldots, 4$ and $t = 1, \ldots, 5$) under the different scenarios are provided in Appendixes A.1.1 and A.1.2, respectively. Overall, the IPTW estimates of cumulative coefficients exhibit a similar pattern to that of $C_{A0}(t)$ in terms of MSE. However, the pattern observed in the bias varies across different cumulative coefficients. This variability could be attributed to the randomness inherent in the generated data and the specific property of the Aalen's additive hazard model.

Table 6.4: MSE of $C_{A0}(t)$ estimates produced by IPTW method under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4.

	Threshold τ				
Time t	-0.3	0.3	0.6	1	1.5
$n = 100$					
1	0.114	0.164	0.106	0.120	0.101
2	0.453	0.428	0.372	0.271	0.268
3	0.830	0.888	0.616	0.546	0.664
4	1.287	1.242	1.167	0.962	1.310
5	1.857	1.671	1.823	1.312	1.581
$n = 200$					
1	0.090	0.143	0.136	0.111	0.063
2	0.321	0.346	0.352	0.340	0.189
3	0.804	0.546	0.703	0.526	0.400
4	1.227	0.989	1.096	0.899	0.916
5	1.941	1.364	1.755	1.483	1.185
$n = 300$					
1	0.116	0.100	0.076	0.078	0.038
2	0.357	0.359	0.253	0.199	0.152
3	0.645	0.716	0.466	0.430	0.327
4	0.872	0.997	0.693	0.676	0.532
5	1.185	1.306	1.183	1.184	1.067
$n = 500$					
1	0.100	0.085	0.084	0.062	0.027
2	0.319	0.206	0.260	0.175	0.090
3	0.614	0.562	0.608	0.323	0.208
4	1.091	0.781	0.862	0.568	0.417
5	1.667	1.258	1.226	0.852	0.689
$n = 1000$					
1	0.067	0.069	0.045	0.033	0.017
2	0.338	0.218	0.214	0.101	0.047
3	0.634	0.449	0.459	0.293	0.138
4	0.914	0.873	0.726	0.548	0.336
5	1.360	0.956	0.985	0.797	0.563

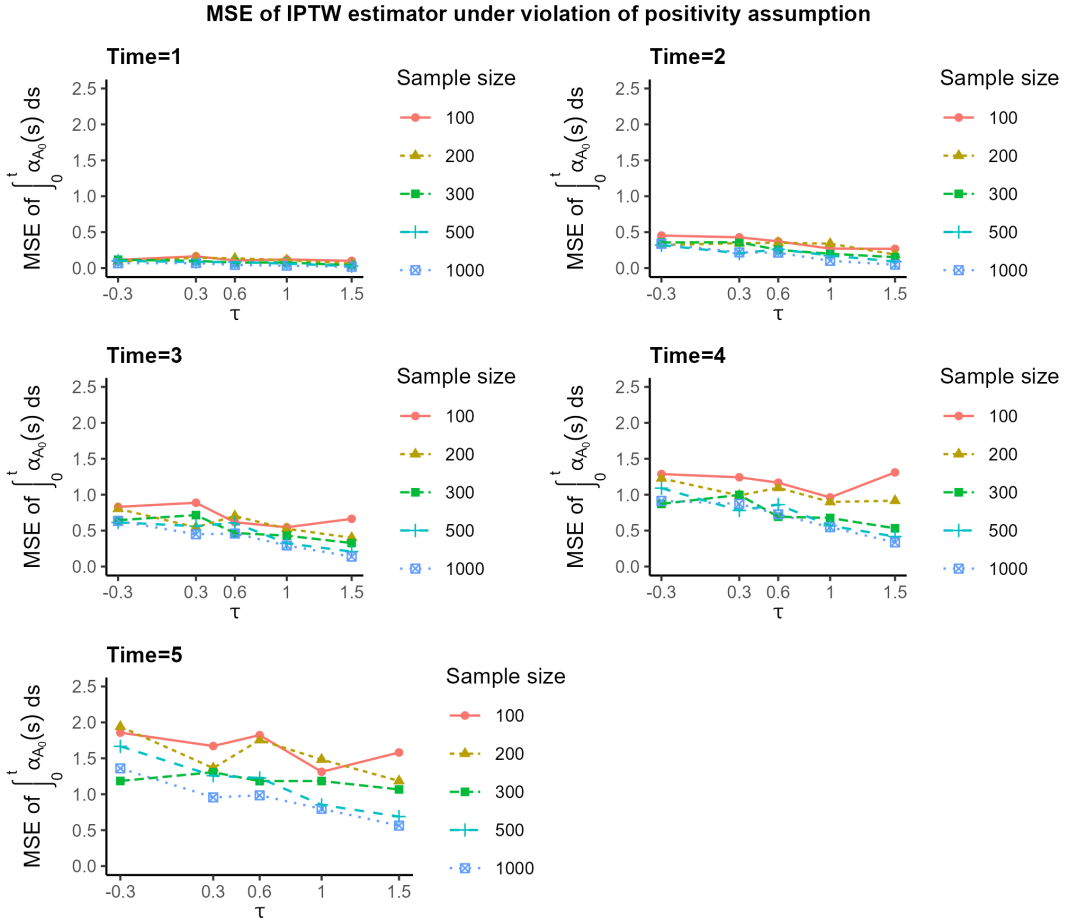

Figure 6.5: MSE of IPTW estimator of $\int_0^t \tilde{\alpha}_{A0}(s)ds$ under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different time-point $t \in \{1, 2, 3, 4, 5\}$. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

6.3.2 Counterfactual survival curves for *always* and *never* treated

As mentioned before, interpreting the cumulative coefficients $C_0(t)$ and $C_{Aj}(t)$ is challenging and to achieve a more direct causal interpretation they need to be translated into more interpretable causal estimands. To this purpose, we consider the counterfactual survival probabilities in Equation (3.17) (see Section 3.5.3) for two treatment regimes:

- *always treated* with $\bar{\boldsymbol{a}} = (1, 1, \ldots, 1) = \mathbf{1}$:

$$Pr(T^{\bar{\boldsymbol{a}}=\mathbf{1}} \geq t) = \exp\{-\Lambda(t)\} \tag{6.3}$$
$$= \exp\left\{ -\int_0^t \left(\tilde{\alpha}_0(x) + \sum_{j=0}^{\lfloor x \rfloor} \tilde{\alpha}_{Aj}(x) \right) dx \right\}, t = 1, \ldots, 5;$$

- *never treated* with $\bar{\boldsymbol{a}} = (0, 0, \ldots, 0) = \mathbf{0}$:

$$Pr(T^{\bar{\boldsymbol{a}}=\mathbf{0}} \geq t) = \exp\{-\Lambda(t)\} = \exp\left\{ -\int_0^t \tilde{\alpha}_0(x)\, dx \right\}, t = 1, \ldots, 5. \tag{6.4}$$

Estimates of Equation (6.3) and (6.4) can then be used to obtain the marginal risk difference between these two regimes at time t, i.e., $Pr(T^{\bar{\boldsymbol{a}}=\mathbf{1}} \geq t) - Pr(T^{\bar{\boldsymbol{a}}=\mathbf{0}} \geq t)$.

Figure 6.6 shows the bias of the IPTW estimates of counterfactual survival probabilities for both regimes under the different scenarios. Each panel refers to a different time-point t with different lines/colours representing the various sample sizes n. It is evident that the IPTW estimate of survival probabilities for the *always treated* regime (solid lines) tends to exhibit a larger bias compared to that of the *never treated* regime (dashed lines). This can be attributed to the inclusion of both the current treatment and all the past treatments in the hazard of the *always treated* regime, thereby introducing additional bias and variability to the estimated survival probabilities. Moreover, the bias in the estimated survival probabilities is approximately inversely proportional to the sample size (i.e., proportional to $1/n$). Consequently, as the sample size increases, the IPTW estimates of survival probabilities for either regime generally exhibits a reduction in bias.

Figure 6.7 illustrates the MSE of the IPTW estimates of counterfactual survival probabilities for both treatment regimes under the different scenarios. Each panel refers to a different time-point t with different lines/colours representing the various sample sizes n. It can be observed that when the sample size is small (e.g., $n = 100$), the MSE can present a significant magnitude. This is primarily due to the fact that the MSE is the sum of the variance and the square of the bias. As mentioned earlier, the bias of the IPTW estimate of survival probabilities is approximately inversely proportional to the sample size ($1/n$). Therefore, when the sample size decreases, the bias can increase, resulting in a even higher MSE.

Figure 6.8 displays the IPTW mean estimates of counterfactual survival probabilities for both regimes under the different scenarios. First, it can be noted that for small sample sizes, the mean estimate of survival probabilities for the *always treated* regime can present an unusual behavior: the survival curve can increase at certain time points. This phenomenon arises as a result of the characteristic of the Aalen additive model, which does not impose constraints on the estimated hazards to fall within the range of 0 and 1. When the sample size is small, the random sample effect can lead to negative

estimates of the hazard, which in turn results in an upward trend in the estimated survival probabilities over time.

Upon closer examinations of the estimated hazard components under the two treatment regimes, we found that they exhibit distinct behaviors. For the survival probabilities under the *never treated* regime, it solely relies on the baseline hazard $\tilde{\alpha}_0(t)$. The estimated cumulative coefficient $\widehat{C}_0(t)$ demonstrates an approximately monotonic increase over time with minimal fluctuations (see Appendix A.1.3, Figure A.11). This explains why the survival probabilities under the *never treated* regime consistently decrease monotonically over time without any unusual upward trends. However, when considering the survival probabilities under the *always treated* regime, which includes the effects of both the current treatment and the lag treatments, the situation is different. As shown in the previous section, the estimated cumulative coefficient associated with the current treatment, i.e., $\widehat{C}_{A0}(t)$, exhibits a decreasing trend over time with relatively high fluctuations. On the other hand, the estimated cumulative coefficients of the lag treatments, i.e., $\widehat{C}_{Aj}(t)$ where $j = 1, 2, 3, 4$, show diverse trends over time and generally exhibit considerable fluctuations (see Appendix A.1.3, Figures A.12 to A.16). Consequently, under a small sample size n, the higher variance associated with the estimated cumulative coefficients $\widehat{C}_{At}(t)$ lead the estimated cumulative hazard $\hat{\Lambda}(t)$ of the *always treated* regime to decrease at certain time points and thus cause unusual survival probabilities over time. This specifically occurs when the estimated hazard for *always treated*, i.e.,

$$\hat{\lambda}_{T^{\bar{a}=\mathbf{1}}}(t) = \hat{\tilde{\alpha}}_0(t) + \sum_{j=0}^{\lfloor t \rfloor} \hat{\tilde{\alpha}}_{Aj}(t),$$

is negative.

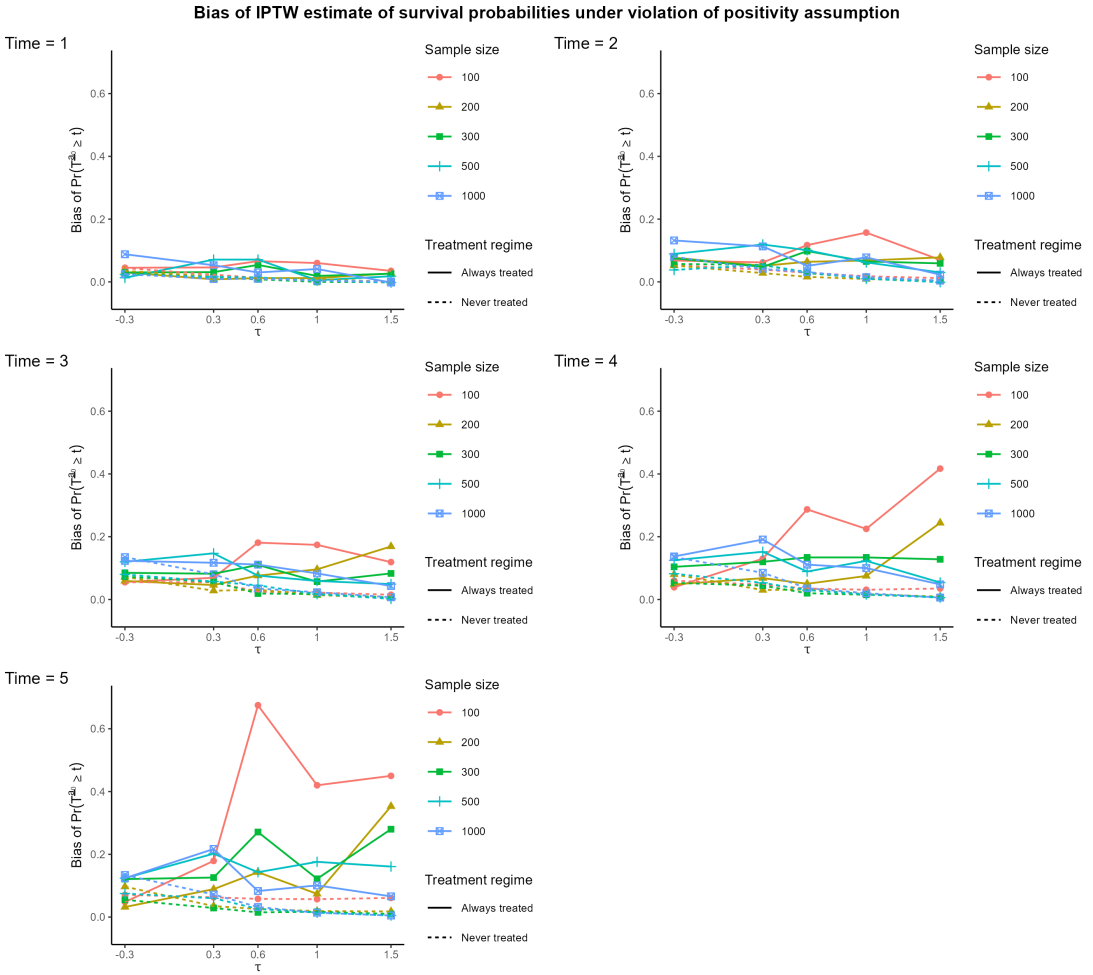

Figure 6.6: Bias of the IPTW estimates of counterfactual survival probabilities of *always treated* (solid lines) and *never treated* (dashed lines) under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different time-point $t \in \{1, 2, 3, 4, 5\}$. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

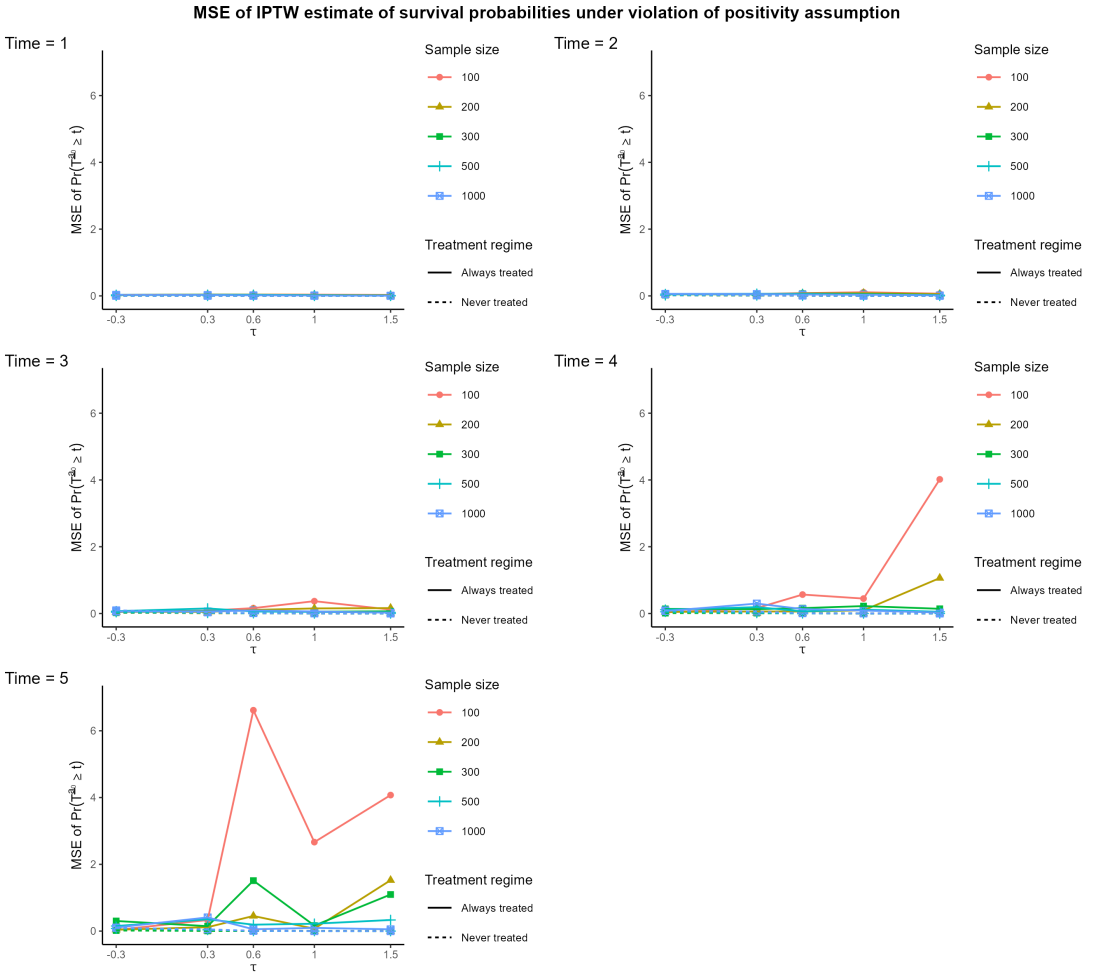

Figure 6.7: MSE of the IPTW estimates of counterfactual survival probabilities of *always treated* (solid lines) and *never treated* (dashed lines)under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different time-point $t \in \{1, 2, 3, 4, 5\}$. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

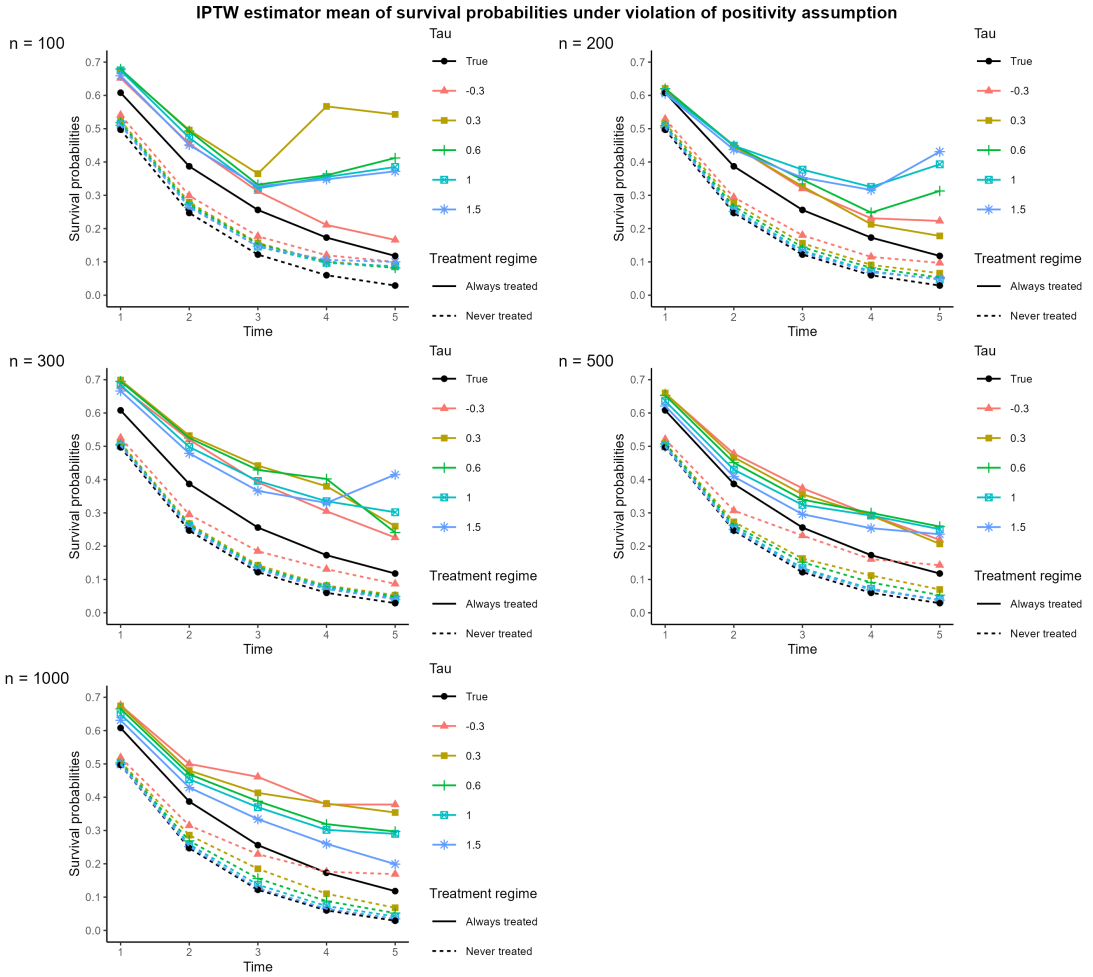

Figure 6.8: The IPTW mean estimates of counterfactual survival probabilities of *always treated* (solid lines) and *never treated* (dashed lines)under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different sample size $n \in \{100, 200, 300, 500, 1000\}$. Different lines and colours refer to various thresholds $\tau \in \{-0.3, 0.3, 0.6, 1, 1.5\}$.

Chapter 7

Discussion

The objective of this thesis was to explore the impact of structurally violating the positivity assumption on the performance of the inverse probability of treatment weighting (IPTW) estimator in longitudinal studies for survival outcomes. We proposed two simulation algorithms specifically designed to simulate longitudinal data exhibiting time-dependent confounding from given marginal structural models (MSMs) formulated using logistic regression and Aalen's additive regression, i.e., Algorithms 3 and 4 respectively. The investigation was conducted under different scenarios to provide a comprehensive analysis and the IPTW-estimator performances for both mechanisms were assessed using bias and mean square error (MSE).

Findings from the simulation studies revealed a consistent trend: as the degree of violation of the positivity assumption becomes more severe, both the bias and the MSE of the IPTW estimate for causal parameters tend to increase. This suggested that violating the positivity assumption has a detrimental impact on the accuracy of the IPTW estimator. Furthermore, the results showed that with larger sample sizes the IPTW estimates demonstrate a smaller scale of bias and MSE in general, and potentially exhibit a slower growth rate in terms of bias with increasing violation (Figure 6.1). This is due to the fact that increasing the sample size mitigated the bias due to the finite sample bias. However, the bias resulting from the structural positivity violation still remained.

Previous causal inference research has emphasized the need for an adequate exposure variability within confounder strata [28, 29] as the violation of positivity assumption can result in substantial bias, with or without a corresponding increase in variance, regardless of the used causal effect estimator [7]. In fact, the consistency of the IPTW estimator relies heavily on the positivity assumption. When the positivity assumption is violated, the IPTW estimator is undefined. This occurs because the denominator of the inverse probability weights reaches zero for some subjects. In that case, the observed bias is mainly due to the positivity violation, with some additional bias due to the finite sample bias [8]. Under the sequential version of the positivity assumption, the conditional probability of each possible treatment history (i.e., the product of multiple time

point-specific treatment probabilities given the past) is required to be positive regardless of covariate history. However, achieving a small value for this probability becomes more likely, especially when multiple time points are considered. Moreover, in the case of a small sample size, certain combinations of treatment and covariate history have fewer or even no observations in the given finite sample. Consequently, the weights of those rarely observed combinations become extreme, resulting in a substantial bias and high variability.

Differences in the results between the two simulation approaches can be due to the different treatment decision mechanisms. Simulation approach I requires that once a subject has started treatment, the subject will remain on treatment until failure or end of follow-up. This indicates a narrower range of possible combinations of treatment history and covariate history, making the IPTW estimator more sensitive to sample size as the violation severity increases. Under more severe violations, fewer combinations of treatment and covariate history can be observed, and a limited sample size can exacerbate this effect. Since the number of possible combinations is small, even a few combination missing in the observed data has an significant impact on the estimated effect, introducing substantial bias and inflating the variance. On the other hand, simulation approach II does not have any requirement on treatment decision mechanism, indicating a larger range of possible combinations of treatment and covariate history. Consequently, the bias and precision of its IPTW estimator under positivity violation is less sensitive to sample size compared to the estimator in the first setting.

Furthermore, it is worthwhile to note other distinctions between these approaches and the potential limitations they have in practice. The simulation approach I generates longitudinal data from a given discrete-time MSM overcoming the non-collapsibility issue of the pooled logistic regression model by omitting the direct arrow from L_t to Y_{t+1} in the direct acyclyc graph (DAG) in Figure 4.1. However, this omission is likely to be unrealistic in practice. On the contrary, the simulation approach II avoids this issue (see DAG in Figure 4.2) and highlights the benefits of using Aalen's additive hazard model in causal inference research, thanks to its collapsibility property. Moreover, simulation approach II is applicable in a more general scenario, while simulation approach I aims to generate data that closely match that of the Swiss HIV Cohort Study [22]. Nonetheless, the linear form of the Aalen's additive hazard model does not restrict the hazard to be non-negative, sometimes resulting in unrealistic survival probabilities. To overcome this issue, model parameter values need to be carefully selected so that the chance of obtaining a negative hazard can be negligible.

Previous research on the effect of positivity violations on IPTW estimator of causal effect has been limited to a point treatment setting. Wang *et al.* (2005) [8] and Petersen *et al.* (2012) [7] demonstrated that data sparsity can increase both the bias and variance of a causal effect estimator and that IPTW estimator is particularly sensitive to this issue. This thesis extended their investigations into a longitudinal survival setting with

time-varying treatments and covariates. Nonetheless, how to improve the identifiability of parameters in the presence of positivity violations in the longitudinal setting still remains a field to be explored, providing a challenge for future research. On the contrary, several approaches have been proposed to estimate causal effects in the presence of positivity violations in a point treatment context. For example, trimming-related methods – including propensity score-based trimming [30, 31], matching [32, 33], Bayesian additive regression trees-based trimming [34, 35] – have been proposed to address the issue by identifying and removing from the data the subgroup of subjects that violates positivity, and drawing inference on the remaining population. These methods, however, limit the subsample used to make inference to only those subjects for whom the positivity assumption is valid, shifting the inference target to the population reflected by such subsample. As an alternative, the so-called weighting scheme methods, which alter the covariate distribution by making certain characteristics more prominent, have been proposed [36, 37, 38]. However, these methods can also present the potential issue of shifting the target of inference. Therefore, the use of trimming and weighting approaches must be limited to situations of structral positivity violation. On the other hand, when there is a practical posivity violation, certain combination of treatment and covariate categories are not observed in the sample by chance and the treatment effect of these subjects must be taken into account in order to preserve the original target of the inference. In such a case, extrapolation methods are recommended, which consist of extrapolating trends in the region of overlap to the subjects in non-overlap region using their observed variables [39, 40]. However these methods suffer from the additional uncertainty resulting from the extrapolation. All these methods should be further investigated into a longitudinal setting.

In summary, this thesis investigated and assessed how positivity violations impact the IPTW estimator in two longitudinal survival settings with time-varying treatment and covariate. While this study proposed specific simulation procedures and revealed the detrimental effect of violations on the estimates, it suggests a need for a more comprehensive framework to simulate longitudinal data from a flexible MSM to gain deeper insights into this problem. In real practice, positivity violations are commonly observed in clinical data with many covariates, especially in contexts involving multiple sequentially assigned treatments over numerous time points. This often results in extreme weights in the IPTW approach, leading to substantial bias in the causal estimate. Therefore, it is imperative for analysts to systematically evaluate the presence of positivity violations at all time-points when conducting causal analyses using real data.

Bibliography

[1] W.G. Havercroft and V. Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in medicine*, 31(30):4190–4206, 2012.

[2] R.H. Keogh, S.R. Seaman, J.M. Gran, and S. Vansteelandt. Simulating longitudinal data from marginal structural models using the additive hazard model. *Biometrical Journal*, 63(7):1526–1541, 2021.

[3] J.M. Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer, 2000.

[4] J.M. Robins, M.A. Hernán, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*:550–560, 2000.

[5] T.J. VanderWeele. Concerning the consistency assumption in causal inference. *Epidemiology*, 20(6):880–883, 2009.

[6] R.M. Daniel, S.N. Cousens, B.L. De Stavola, M.G. Kenward, and J.A.C. Sterne. Methods for dealing with time-dependent confounding. *Statistics in medicine*, 32(9):1584–1618, 2013.

[7] M.L. Petersen, K.E. Porter, S. Gruber, Y. Wang, and M.J. Van Der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, 2012.

[8] Y. Wang, M.L. Petersen, D. Bangsberg, and M.J. van der Laan. Diagnosing bias in the inverse probability of treatment weighted estimator resulting from violation of experimental treatment assignment, 2006.

[9] R. Neugebauer and M. van der Laan. Why prefer double robust estimators in causal inference? *Journal of statistical planning and inference*, 129(1-2):405–426, 2005.

[10] M. Léger, A. Chatton, F. Le Borgne, R. Pirracchio, S. Lasocki, and Y. Foucher. Causal inference in case of near-violation of positivity: comparison of methods. *Biometrical Journal*, 64(8):1389–1403, 2022.

[11] T.P. Morris, I.R. White, and M.J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in medicine*, 38(11):2074–2102, 2019.

[12] M.A. Hernán and J.M. Robins. *Causal Inference: What If*. Boca Raton: Chapman Hall/CRC, 2020.

[13] J.C. Digitale, J.N. Martin, and M.M. Glymour. Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142:264–267, 2022.

[14] E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[15] O. Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*:701–726, 1978.

[16] D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[17] O. Aalen. A linear regression model for the analysis of life times. *Statistics in medicine*, 8(8):907–925, 1989.

[18] J.P. Klein and M.L. Moeschberger. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

[19] T. Martinussen and S. Vansteelandt. On collapsibility and confounding bias in cox and aalen regression models. *Lifetime data analysis*, 19(3):279–296, 2013.

[20] M.A. Hernán, B. Brumback, and J.M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*:561–570, 2000.

[21] A.L. Boulesteix, R.H.H. Groenwold, M. Abrahamowicz, H. Binder, M. Briel, R. Hornung, T.P. Morris, J. Rahnenführer, and W. Sauerbrei. Introduction to statistical simulations in health research. *BMJ open*, 10(12):e039921, 2020.

[22] J.A.C Sterne, M.A Hernán, B. Ledergerber, K. Tilling, R. Weber, P. Sendi, M. Rickenbach, J.M. Robins, and M. Egger. Long-term effectiveness of potent antiretroviral therapy in preventing aids and death: a prospective cohort study. *The Lancet*, 366(9483):378–384, 2005.

[23] J.G. Young, M.A. Hernán, S. Picciotto, and J.M. Robins. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime data analysis*, 16:71–84, 2010.

[24] J.G. Young, Tchetgen T., and Eric J. Simulation from a known cox msm using standard parametric models for the g-formula. *Statistics in medicine*, 33(6):1001–1014, 2014.

[25] R. Li, D. Duffee, and M.F. Gbadamosi-Akindele. Cd4 count, 2017.

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

[27] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: `https://www.R-project.org/`.

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

[29] J.M. Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association*, volume 1999, pages 6–10. Indianapolis, IN, 2000.

[30] R.K. Crump, V.J. Hotz, G. Imbens, and O. Mitnik. Moving the goalposts: addressing limited overlap in the estimation of average treatment effects by changing the estimand, 2006.

[31] T. Stürmer, K.J. Rothman, J. Avorn, and R.J. Glynn. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. *American journal of epidemiology*, 172(7):843–854, 2010.

[32] P.R. Rosenbaum. Optimal matching of an optimally chosen subset in observational studies. *Journal of Computational and Graphical Statistics*, 21(1):57–71, 2012.

[33] G. Visconti and J. Zubizarreta. Handling limited overlap in observational studies with cardinality matching. *Observational Studies*, 4(1):217–249, 2018.

[34] J. Hill and Y.S. Su. Assessing lack of common support in causal inference using bayesian nonparametrics: implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *The Annals of Applied Statistics*:1386–1420, 2013.

[35] L. Hu, C. Gu, M. Lopez, J. Ji, and J. Wisnivesky. Estimation of causal effects of multiple treatments in observational studies with a binary outcome. *Statistical methods in medical research*, 29(11):3218–3234, 2020.

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

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

[38] F. Li and F. Li. Propensity score weighting for causal inference with multiple treatments, 2019.

[39] R.C. Nethery, F. Mealli, and F. Dominici. Estimating population average causal effects in the presence of non-overlap: the effect of natural gas compressor station exposure on cancer mortality. *The annals of applied statistics*, 13(2):1242, 2019.

[40] A.Y. Zhu, N. Mitra, and J. Roy. Addressing positivity violations in causal effect estimation using gaussian process priors. *Statistics in Medicine*, 42(1):33–51, 2023.

Appendix

A.1 Appendix Figures to Section 6.3

This appendix contains the plots of bias, MSE and mean estimate of all the cumulative coefficients including $C_0(t)$ and $C_{Aj}(t)$ (for $j = 0, \ldots, 4$ and $t = 1, \ldots, 5$) under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4 .

A.1.1 Bias for all cumulative coefficients under different scenarios

Appendix A.1.1 contains the plots of bias for all the estimated cumulative coefficients under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. In particular

- Figure A.1 shows the bias of estimated cumulative coefficients at time point $t = 1$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.2 shows the bias of estimated cumulative coefficients at time point $t = 2$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.3 shows the bias of estimated cumulative coefficients at time point $t = 3$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.4 shows the bias of estimated cumulative coefficients at time point $t = 4$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.5 shows the bias of estimated cumulative coefficients at time point $t = 5$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

The pattern observed in the bias varies across different cumulative coefficients. This variability could be attributed to the randomness inherent in the generated data and the specific property of the Aalen's additive hazard model.

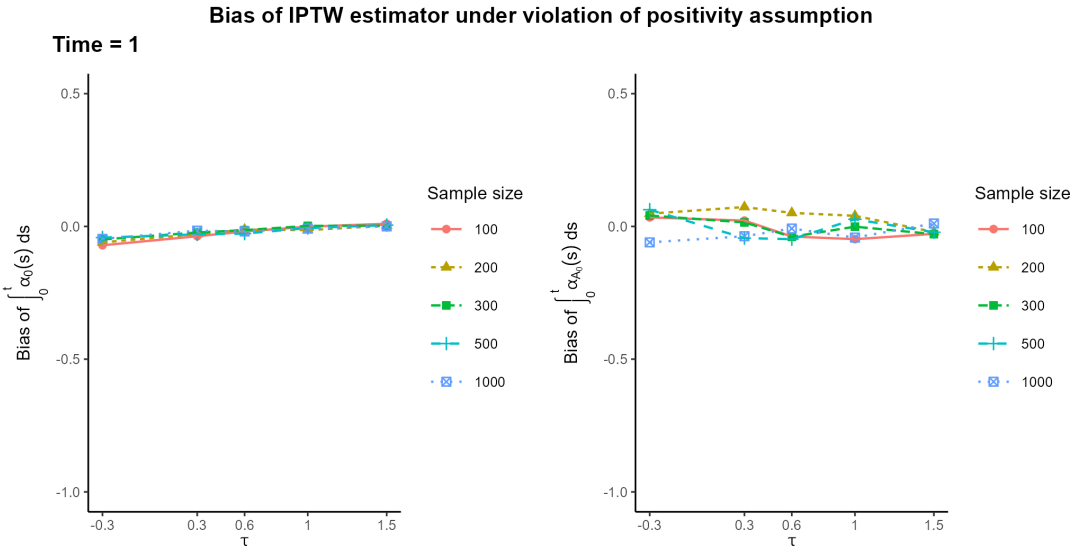

Figure A.1: Bias of IPTW estimator of cumulative coefficients at time point $t = 1$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

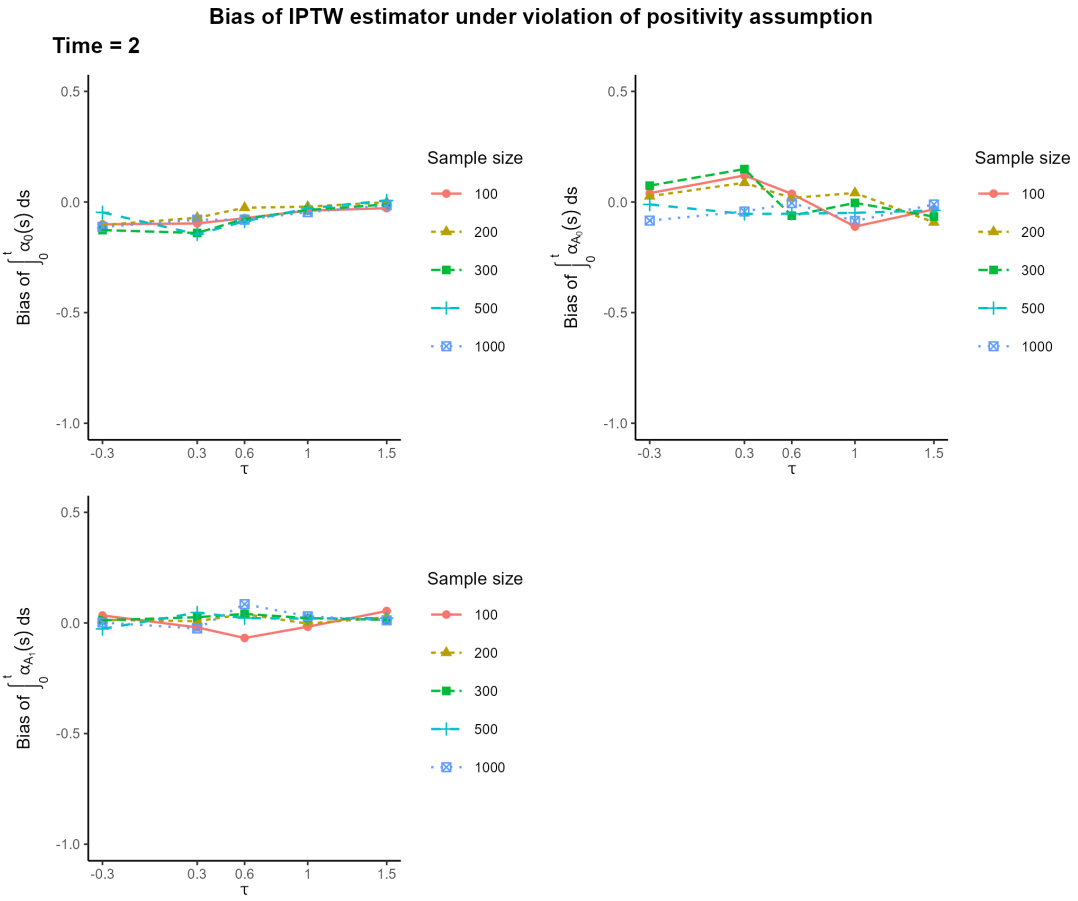

Figure A.2: Bias of IPTW estimator of cumulative coefficients at time point $t = 2$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

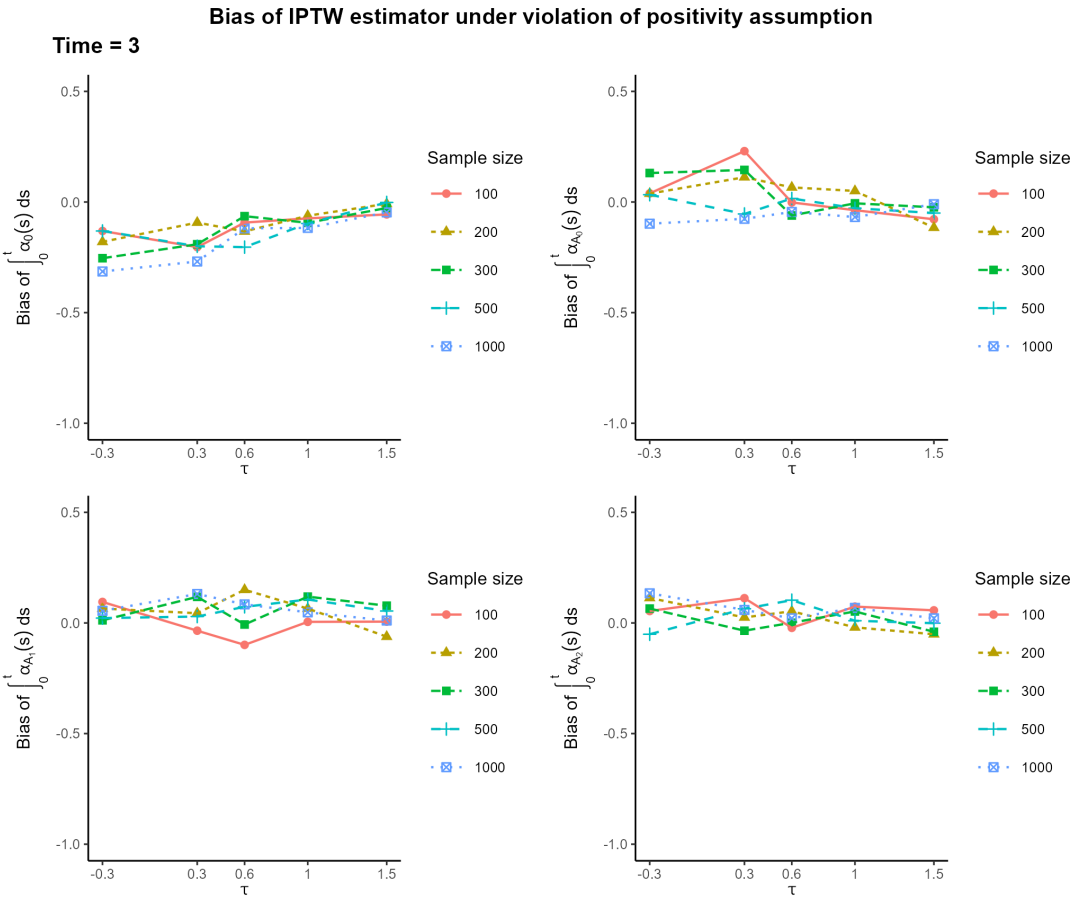

Figure A.3: Bias of IPTW estimator of cumulative coefficients at time point $t = 3$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

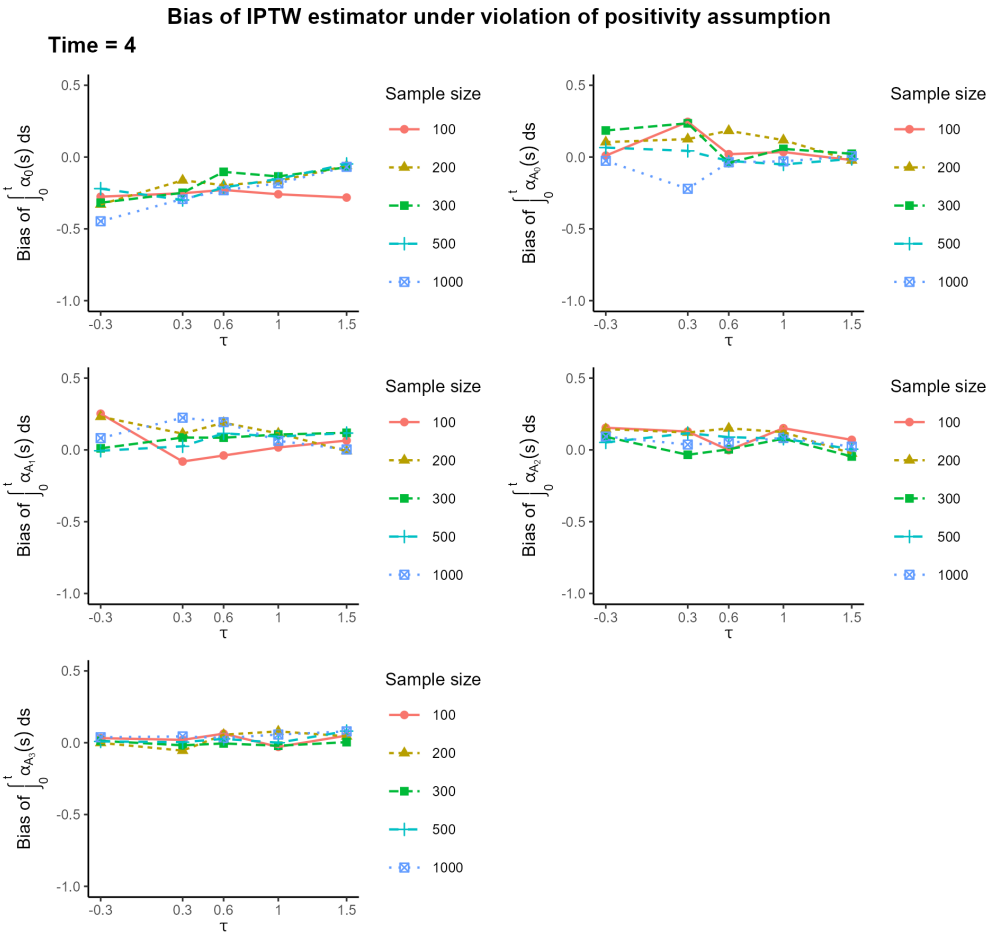

Figure A.4: Bias of IPTW estimator of cumulative coefficients at time point $t = 4$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

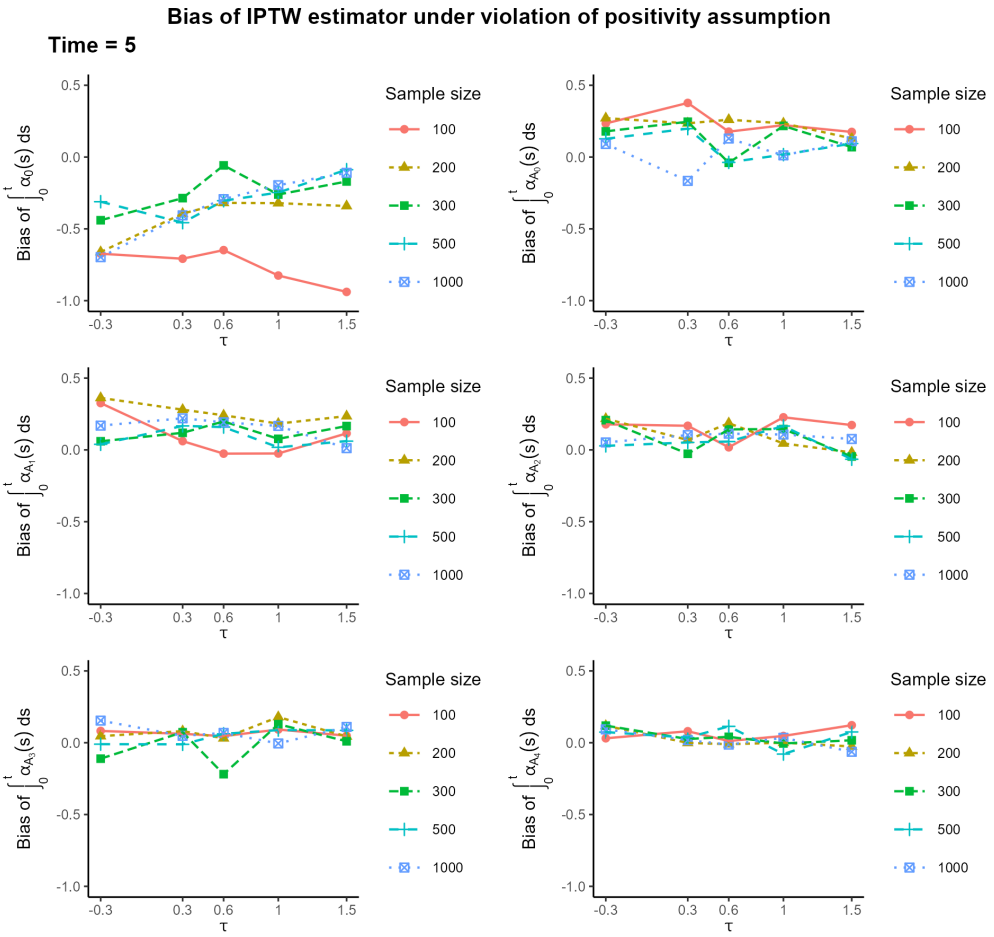

Figure A.5: Bias of IPTW estimator of cumulative coefficients at time point $t = 5$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

A.1.2 MSE for all cumulative coefficients under different scenarios

Appendix A.1.2 contains the plots of MSE for all the estimated cumulative coefficients under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. In particular

- Figure A.6 shows the MSE of estimated cumulative coefficients at time point $t = 1$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.7 shows the MSE of estimated cumulative coefficients at time point $t = 2$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.8 shows the MSE of estimated cumulative coefficients at time point $t = 3$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.9 shows the MSE of estimated cumulative coefficients at time point $t = 4$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.
- Figure A.10 shows the MSE of estimated cumulative coefficients at time point $t = 5$. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

In general, the MSE of cumulative coefficients $C_{Aj}(t)$ $(j = 0, \ldots, 4)$ exhibit a similar pattern to that of $C_{A0}(t)$.

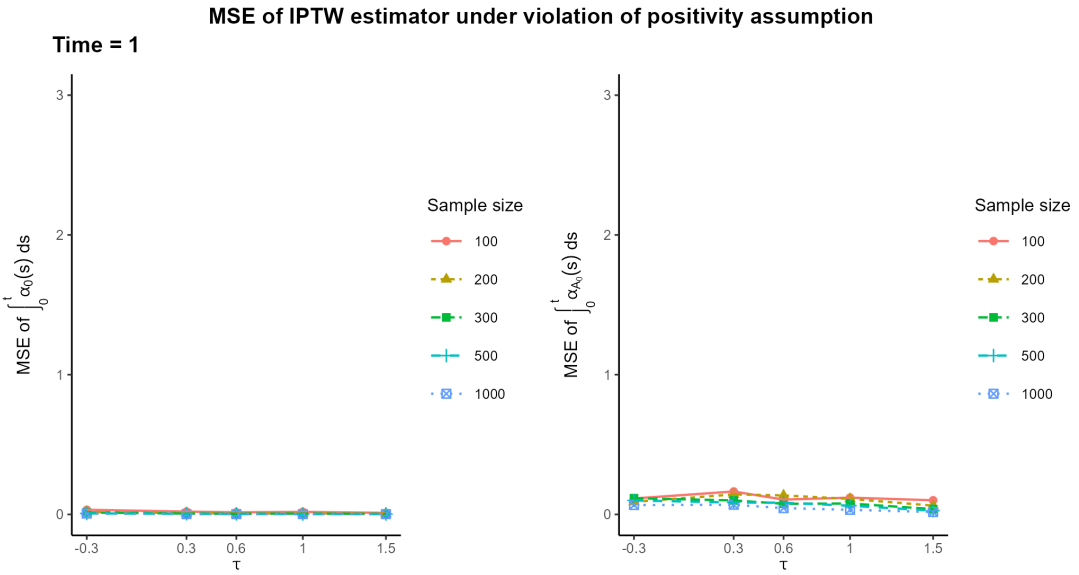

Figure A.6: MSE of IPTW estimator of cumulative coefficients at time point $t = 1$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

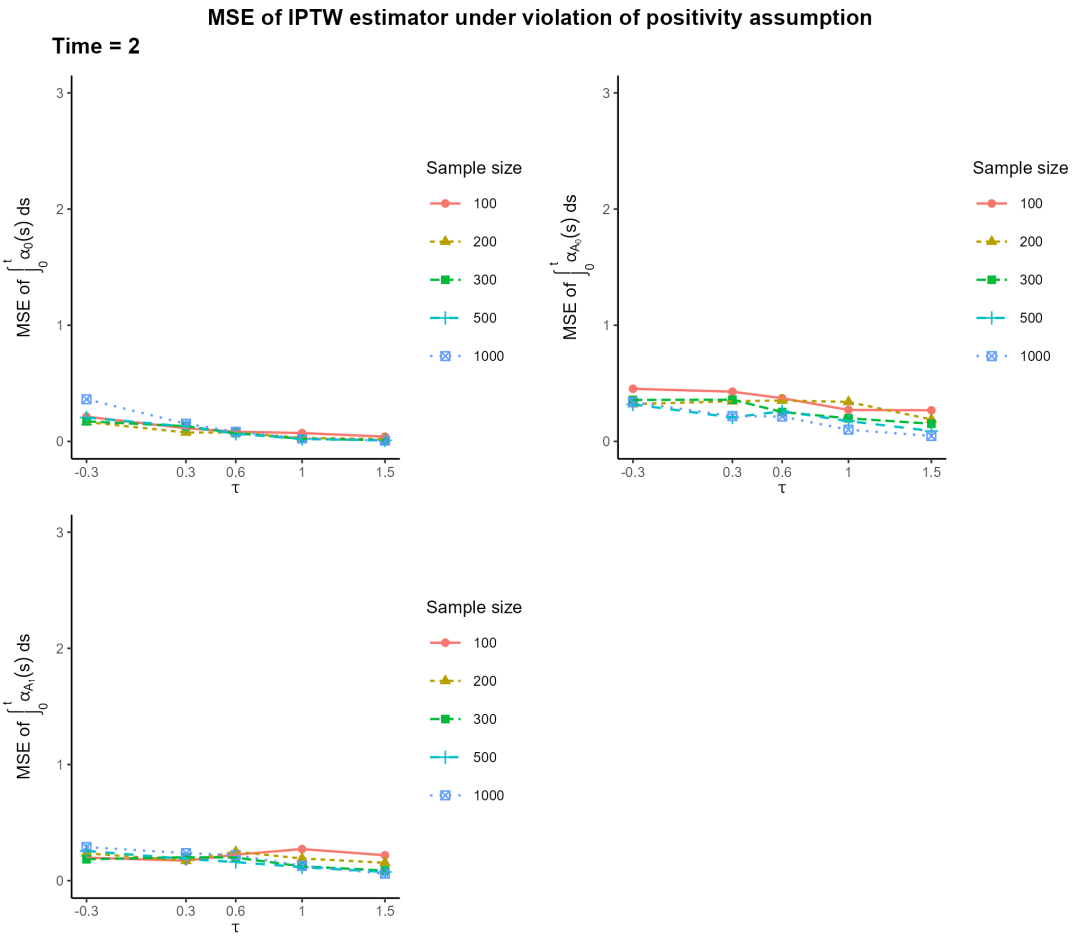

Figure A.7: MSE of IPTW estimator of cumulative coefficients at time point $t = 2$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

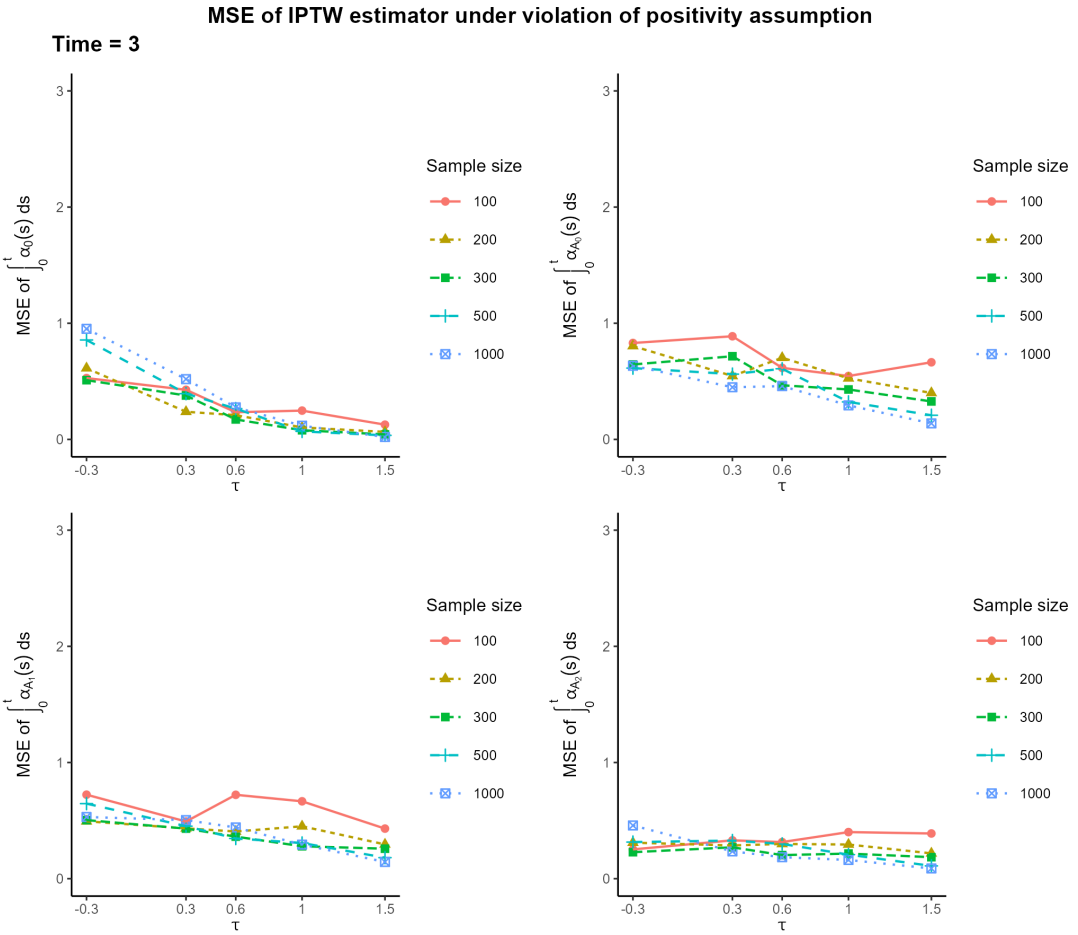

Figure A.8: MSE of IPTW estimator of cumulative coefficients at time point $t = 3$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

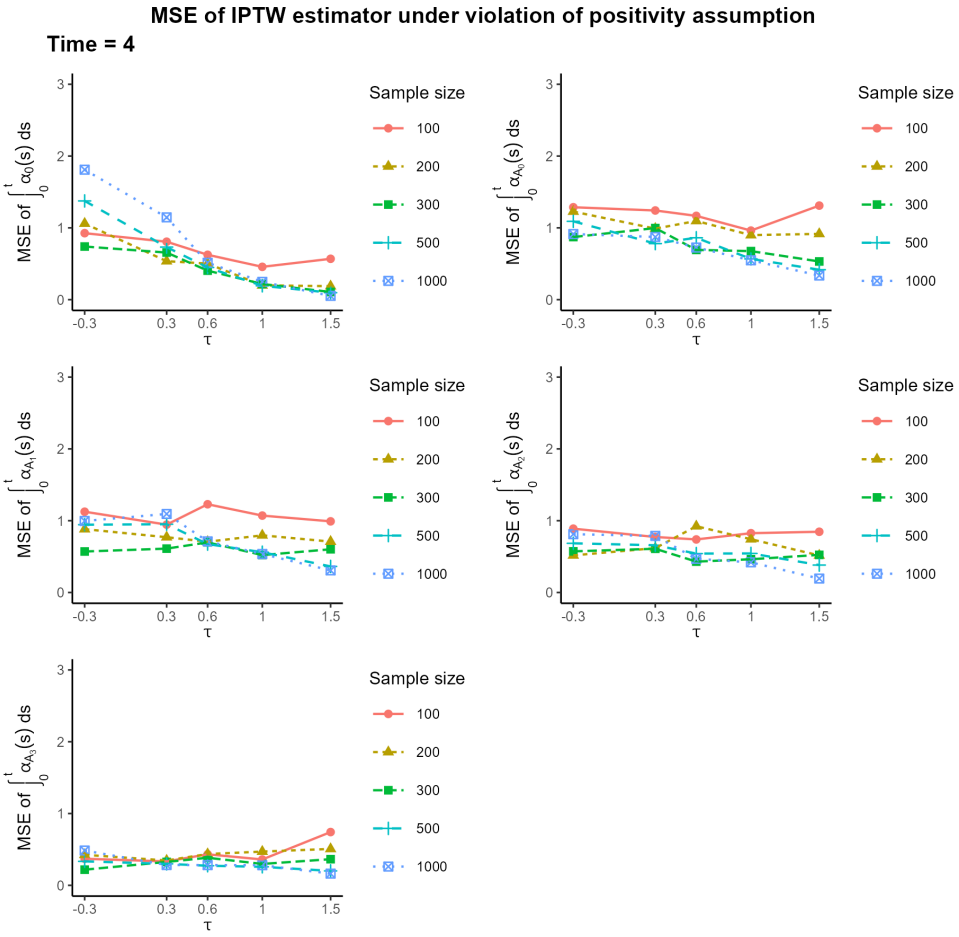

Figure A.9: MSE of IPTW estimator of cumulative coefficients at time point $t = 4$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

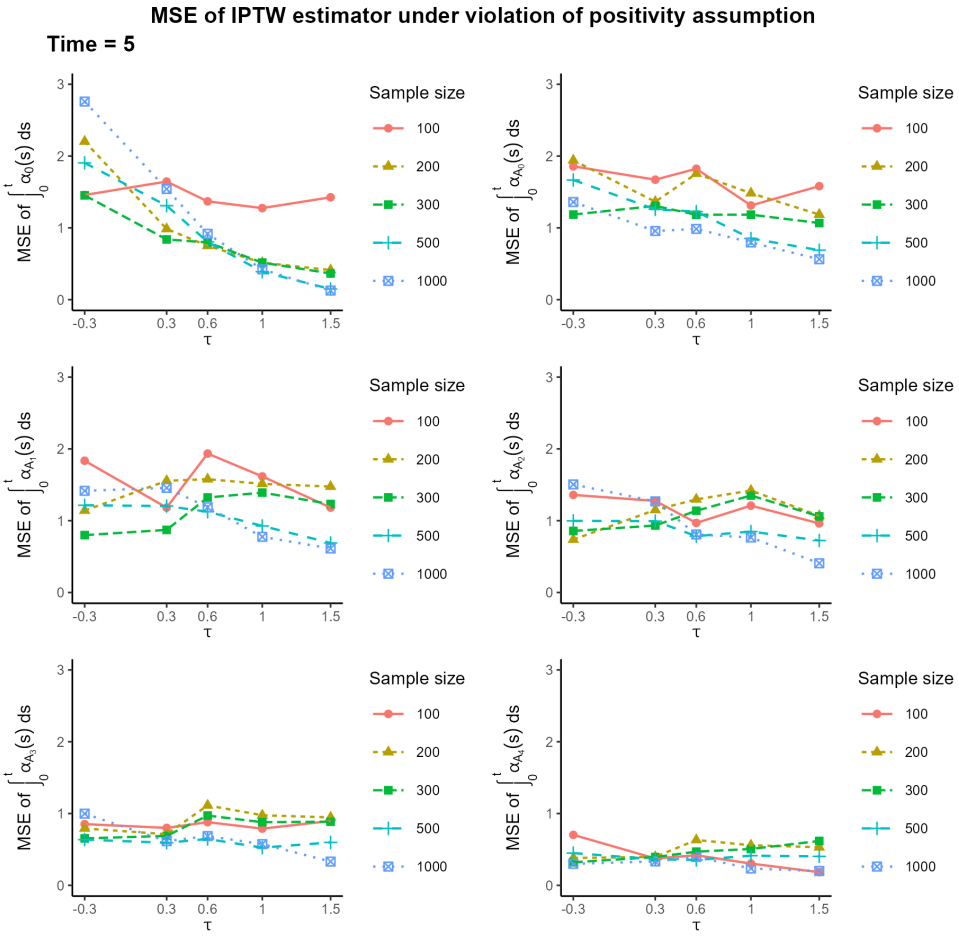

Figure A.10: MSE of IPTW estimator of cumulative coefficients at time point $t = 5$ and the different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. Each panel refers to a different accumulative coefficient. Different lines and colours refer to different sample sizes $n \in \{100, 200, 300, 500, 1000\}$.

A.1.3 IPTW estimators for all cumulative coefficients under different scenarios

Appendix A.1.3 contains the plots of all the IPTW estimated cumulative coefficients under the different scenarios presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4. In particular

- Figure A.11 shows the IPTW estimated cumulative coefficient of $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$ over time under the different scenarios.
- Figure A.12 shows the IPTW estimated cumulative coefficient of $C_{A0}(t) = \int_0^t \tilde{\alpha}_{A0}(s)ds$ over time under the different scenarios.
- Figure A.13 shows the IPTW estimated cumulative coefficient of $C_{A1}(t) = \int_0^t \tilde{\alpha}_{A1}(s)ds$ over time under the different scenarios.
- Figure A.14 shows the IPTW estimated cumulative coefficient of $C_{A2}(t) = \int_0^t \tilde{\alpha}_{A2}(s)ds$ over time under the different scenarios.
- Figure A.15 shows the IPTW estimated cumulative coefficient of $C_{A3}(t) = \int_0^t \tilde{\alpha}_{A3}(s)ds$ over time under the different scenarios.
- Figure A.16 shows the IPTW estimated cumulative coefficient of $C_{A4}(t) = \int_0^t \tilde{\alpha}_{A4}(s)ds$ over time under the different scenarios.

In general, the cumulative coefficient of baseline hazard $C_0(t)$ exhibits an approximately monotonic increasing trend over time, while the cumulative coefficient of current treatment $C_{A0}(t)$ exhibits a decreasing trend over time and the cumulative coefficients of the lag treatments, i.e., $C_{Aj}(t)$ where $j = 1, 2, 3, 4$, show diverse trends over time and exhibit considerable fluctuations.

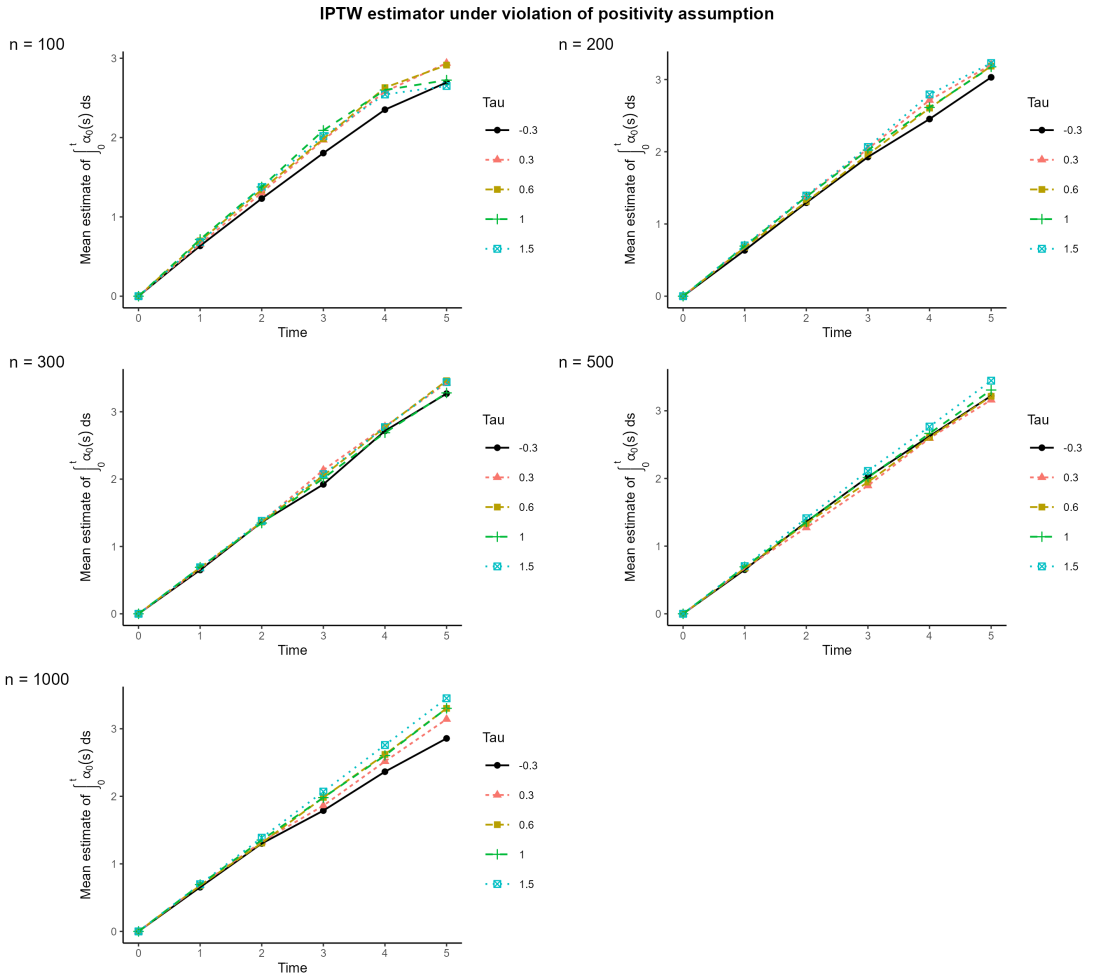

Figure A.11: IPTW estimator of cumulative coefficient $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4

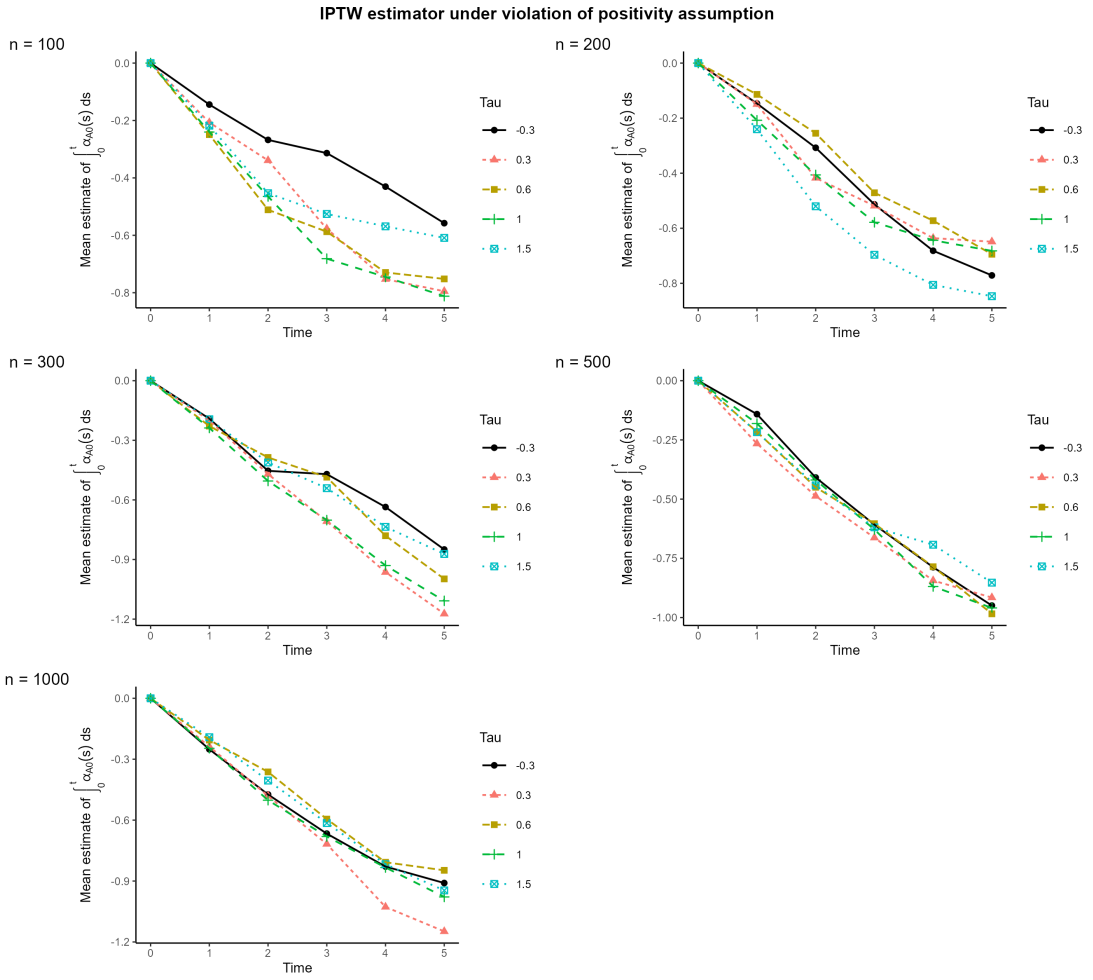

Figure A.12: IPTW estimator of cumulative coefficient $C_{A0}(t) = \int_0^t \tilde{\alpha}_{A0}(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4

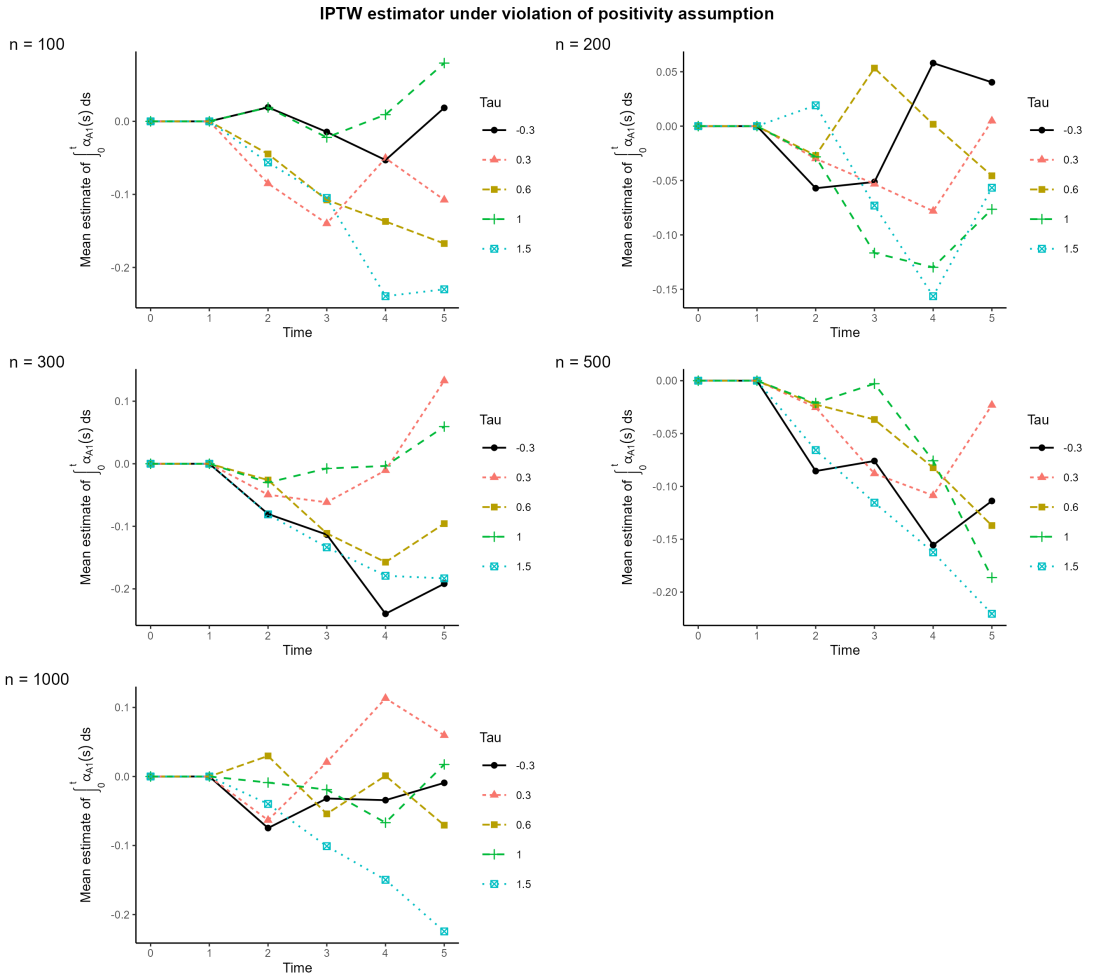

Figure A.13: IPTW estimator of cumulative coefficient $C_{A1}(t) = \int_0^t \tilde{\alpha}_{A1}(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4

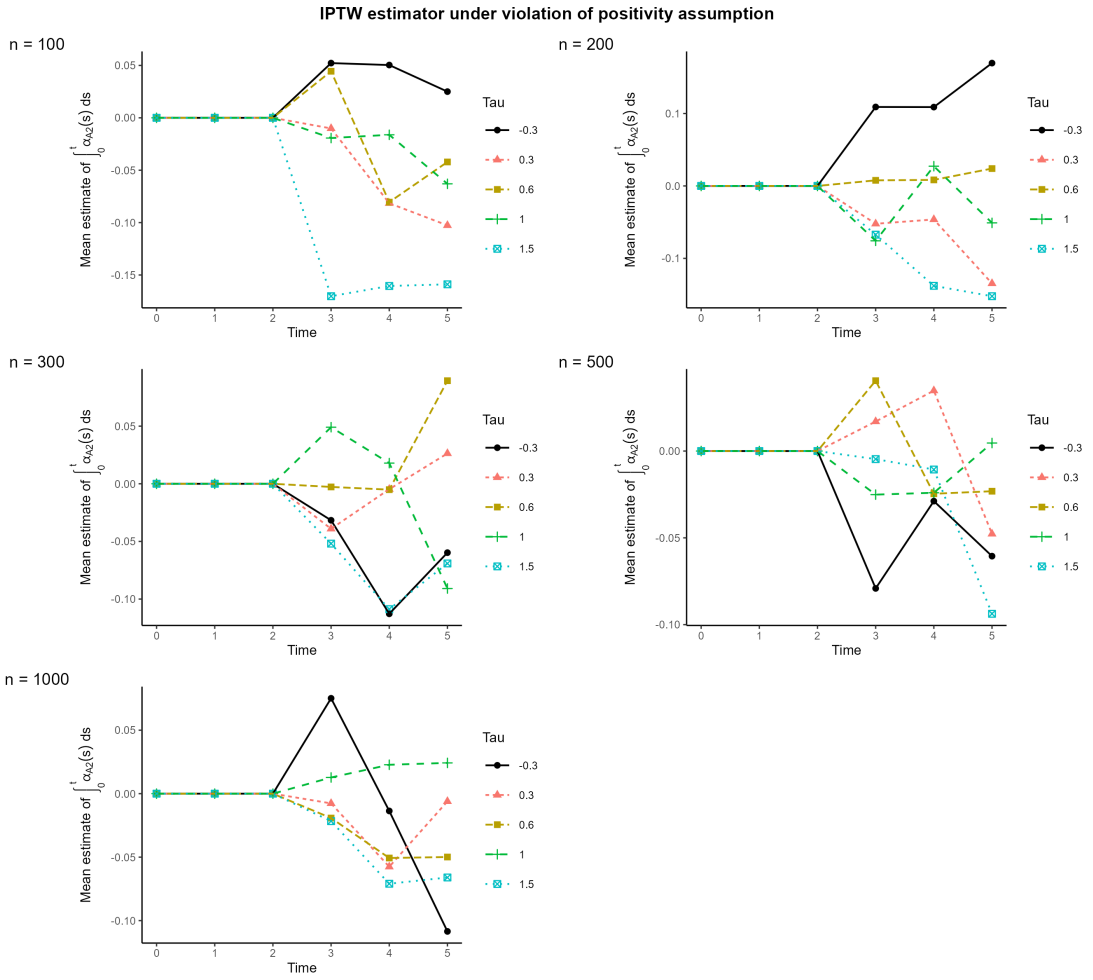

Figure A.14: IPTW estimator of cumulative coefficient $C_{A2}(t) = \int_0^t \tilde{\alpha}_{A2}(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4

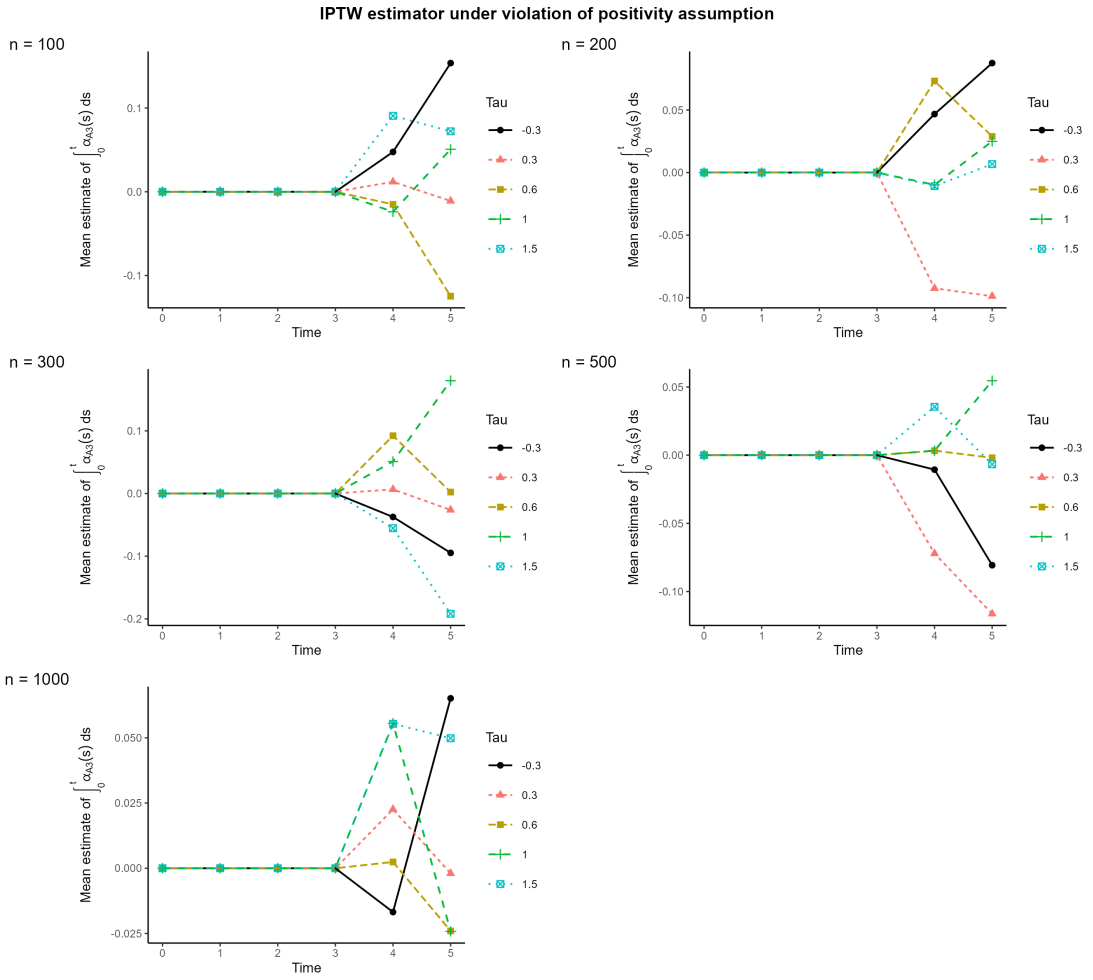

Figure A.15: IPTW estimator of cumulative coefficient $C_{A3}(t) = \int_0^t \tilde{\alpha}_{A3}(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4

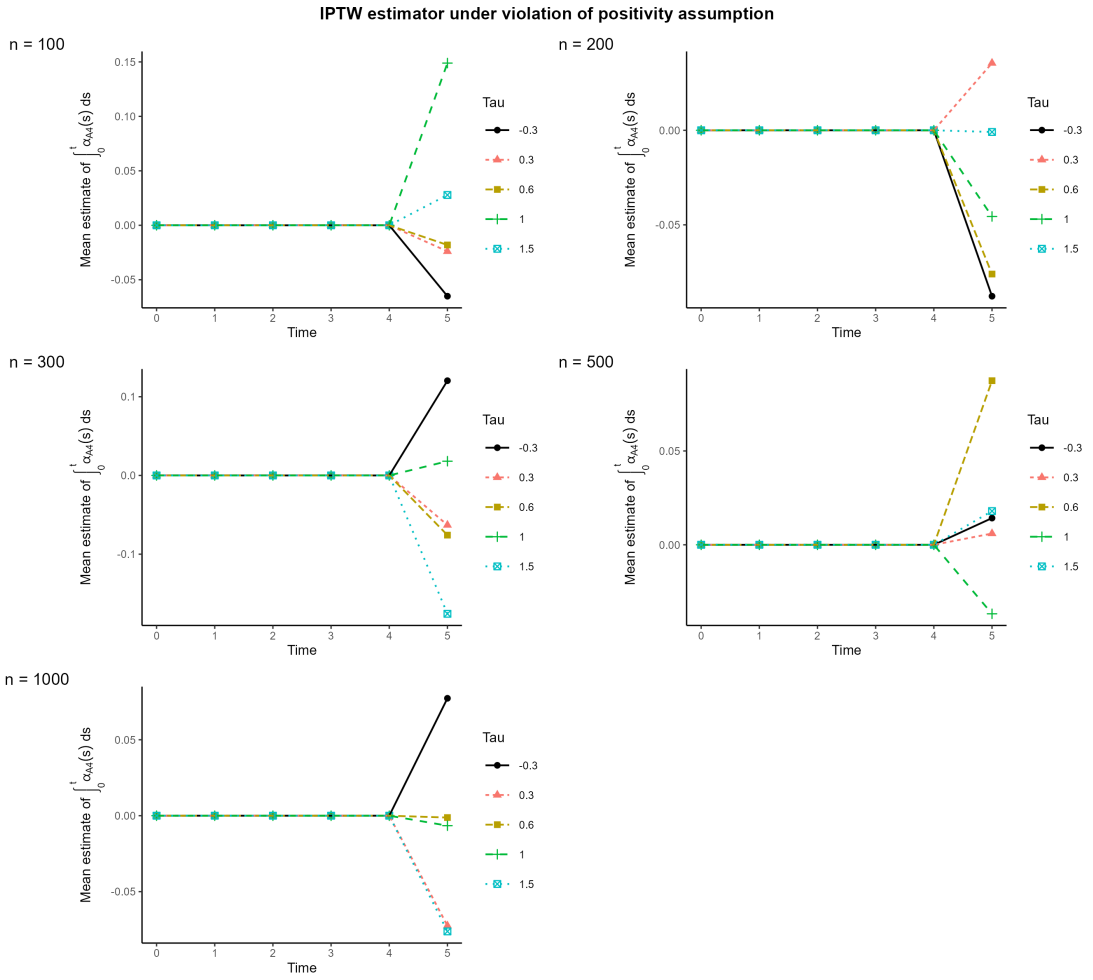

Figure A.16: IPTW estimator of cumulative coefficient $C_{A4}(t) = \int_0^t \tilde{\alpha}_{A4}(s)ds$ under different scenarios of positivity violations presented in Section 5.3.2 using data generated from simulation approach II in Algorithm 4