

# Assessing causal effects in the presence of treatment switching through principal stratification

Alessandra Mattei<sup>\*1,2</sup>, Peng Ding<sup>†3</sup>, Veronica Ballerini<sup>‡1</sup>, and  
Fabrizia Mealli<sup>§1,4</sup>

<sup>1</sup>University of Florence

<sup>2</sup>Florence Center of Data Science

<sup>3</sup>University of California Berkeley

<sup>4</sup>European University Institute

## Abstract

Clinical trials often allow patients in the control arm to switch to the treatment arm if their physical conditions are worse than certain tolerance levels. For instance, treatment switching arises in the Concorde clinical trial, which aims to assess causal effects on the time-to-disease progression or death of immediate versus deferred treatment with zidovudine among patients with asymptomatic HIV infection. The Intention-To-Treat analysis does not measure the effect of the actual receipt of the treatment and ignores the information on treatment switching. Other existing methods reconstruct the outcome a patient would have had if they had not switched under strong assumptions. Departing from the literature, we re-define the problem of treatment switching using principal stratification and focus on causal effects for patients belonging to subpopulations defined by the switching behavior under control. We use a Bayesian approach to inference, taking into account that *(i)* switching happens in continuous time; *(ii)* switching time is not defined for patients who never switch in a particular experiment; and *(iii)* survival time and switching time are subject to censoring. We apply this framework to analyze synthetic data based on the Concorde study. Our data analysis reveals that immediate treatment with zidovudine increases survival time for never switcher and that treatment effects are highly heterogeneous across different types of patients defined by the switching behavior.

**Keywords:** Bayesian causal inference; Censoring; Competing risks; Noncompliance; Potential outcomes; Survival

---

\*alessandra.mattei@unifi.it

†pengdingpku@berkeley.edu

‡veronica.ballerini@unifi.it

§Fabrizia.Mealli@eui.eu

# 1 Introduction

Treatment switching is a post-randomization event that commonly occurs in clinical trials designed to assess the effect of a treatment on the incidence of a disease. There exist various types of treatment switching. During the follow-up period, the treatment may cause unwanted side effects for some patients, preventing them from continuing the treatment; such a kind of treatment switching is known as “treatment discontinuation”. For instance, in clinical trials where the control group is the standard of care, patients may be allowed to switch from the active to the control treatment if unbearable toxicity occurs under treatment. In clinical trials aiming to assess the causal effects of an active versus a placebo treatment plus the existing standard of care, patients may be allowed to discontinue both treatments while remaining on the standard of care. In other cases, a sudden disease worsening for some weaker patients forces physicians to allow them to switch to the treatment arm or take a non-trial treatment. In this work, we focus on clinical trials where patients in the treatment arm never switch to the control arm, but patients in the control arm can switch to the treatment arm if their physical conditions are worse than certain tolerance levels. This type of switching often happens in clinical trials for patients suffering from AIDS-related illnesses or particularly painful cancers in advanced stages (see, e.g., [Robins and Tsiatis, 1991](#); [Robins, 1994](#); [White et al., 1997, 1999](#); [Zeng et al., 2011](#); [Chen et al., 2013](#)). Such a type of switching also occurs in the Concorde Trial ([Concorde Coordinating Committee, 1994](#)), which we will use as a running example to illustrate the methodological framework we propose to deal with the problem of treatment switching. An additional example of treatment switching is the BREAK-3 Trial ([Hauschild et al., 2012](#)). The BREAK-3 Trial and the CheckMate 067 phase III trial, which is an example of treatment discontinuation ([Larkin et al., 2015](#)), will serve as additional case studies to describe our methodology at work, even though no data will be analyzed (see details in Section 6).

The Concorde Study is a randomized clinical trial that aims to assess the causal effects of immediate versus deferred treatment with an antiretroviral medication (zidovudine) on time-to-disease progression or death among symptom-free individuals infected with HIV. According to the trial protocol, patients assigned to the control group should not receive the active treatment until they progress to AIDS-related complex (ARC) or AIDS. However, physicians may judge it unethical to keep patients in the control arm if their physical conditions worsen considerably, e.g., if they experience persistently low CD4 cell counts even before the onset of ARC or symptoms of HIV.

Intention-to-treat (ITT) analysis compares groups formed by randomization regardless of the treatment actually received, ignoring the information on treatment switching in the control group. It is valid for measuring the effect of assignment but does not estimate the effect of the actual receipt of the treatment. In the Concorde study, an ITT analysis compares outcomes by the assignment to immediate versus deferred treatment with zidovudine, ignoring whether control patients stay on the control arm for the entire follow-up period (or, according

to the protocol, up to the onset of ARC or AIDS). However, we cannot ignore treatment switching if the focus is on assessing the effects of the treatment itself, that is, of receiving zidovudine immediately versus subsequently after the onset of ARC or AIDS.

Unfortunately, we cannot adjust for treatment switching by simply conditioning on its observed value because treatment switching is a non-randomized post-assignment variable. Imagine that immediate treatment with zidovudine increases every individual’s survival, but weaker patients, who are most at risk of death, would switch very early if assigned to control. A naive analysis that compares observed immediate versus observed deferred treatment with zidovudine may unfairly conclude that the first has no or little effect on survival. Web Appendix A reviews various existing methods to evaluate the effect of a treatment accounting for treatment switching. They focus on causal effects for the whole population under the assumption that each individual has an outcome that would have happened under assignment to treatment and an outcome that would have happened under assignment to control if that individual had not switched. The recent release of an Addendum to the E9 guideline on ‘Statistical principles in clinical trials’ by the ICH (ICH E9(R1) addendum) refers to this approach as a “hypothetical strategy” for dealing with inference on treatment effects in the presence of intercurrent events, such as treatment switching (ICH, 2019). In Web Appendix B, we also describe and discuss a semi-competing risks approach to the analysis of randomized studies with survival outcomes suffering from treatment switching. In the classical competing risks literature, controlled direct effects and total effects are usually the targets of inference. Controlled direct effects are hypothetical estimands as those usually considered in the treatment switching literature, comparing the time to the primary event (e.g., disease progression or death) under assignment to treatment versus control after somehow eliminating the competing event (e.g., switching). Total effects are also causal effects for the whole population. They are a type of ITT effect, namely the contrasts of the probabilities of experiencing the primary event before a time  $t$ . As total effects, they do not account for the mechanisms by which the treatment affects the occurrence of the primary event, e.g., through other (secondary) events like tolerance implying treatment switching.

We propose to re-define the problem of treatment switching using principal stratification (Frangakis and Rubin, 2002), which is also recognized in the ICH E9(R1) addendum (ICH, 2019) as a strategy to deal with intercurrent events. The novel causal estimands are the principal causal effects (PCEs) for subpopulations defined by the switching behavior under control. The key insight underlying our approach is that treatment switching can be viewed as a general form of noncompliance. Principal stratification plays an important role in the analysis of randomized studies with all-or-none noncompliance, where it classifies units into groups defined by compliance status. These studies usually focus on the causal effects for the principal stratum of compliers (Angrist et al., 1996). In clinical trials with treatment switching, classifying patients into subpopulations defined by the switching behavior is an extension of classifying units based on the compliance status (see Web Appendix C for details on the

connection to the noncompliance literature). To the best of our knowledge, no published studies before our study first published on ArXiv (Mattei et al., 2020) used principal stratification to deal with the problem of treatment switching. Principal stratification has been recently used to define the causal effects of treatment with semi-competing risks (Comment et al., 2019; Xu et al., 2022), and strong connections exist between our study and the existing studies. Nevertheless, some distinguishing features make our contribution unique. Comment et al. (2019) and Xu et al. (2022) focus on assessing the causal effects of treatment on non-terminal time-to-event outcomes and use principal stratification to account for the fact that the non-terminal endpoints are subject to truncation by death, that is, they are not well-defined after death. Specifically, their causal estimands are time-varying survivor causal effects for the principal strata of patients who would survive regardless of treatment assignment. Our causal estimands are PCEs on a terminal time-to-event outcome (i.e., time-to-disease progression or death), with principal strata defined by the switching behavior considering that the occurrence of the primary terminal endpoint precludes any future non-terminal switching event. Because switching is a non-terminal event that does not truncate death, here the causal effects are well-defined for each principal stratum. See Web Appendix B for an in-depth discussion of the similarities and differences between our framework and the existing frameworks in the presence of semi-competing risks.

The PCEs are *local* causal effects for patients who are homogeneous with respect to the switching behavior. Therefore, the PCEs provide information on treatment effect heterogeneity with respect to the switching behavior. In the Concorde trial, the principal stratum of non-switchers will be of particular interest. Non-switchers are patients who would never switch to the active treatment if assigned to the control. They take the treatment and control according to the protocol and thus can provide evidence on the causal effect of treatment versus control.

Treatment switching complicates causal inference. First, the switching of patients under control either never happens or happens in continuous time. Second, assumptions such as exclusion restrictions (Angrist et al., 1996), typically invoked in the noncompliance setting, are untenable in studies with treatment switching. Section 3 will discuss these issues in detail. We deal with inferential issues in the analysis of the Concorde trial using a flexible model-based Bayesian approach, which allows us to take into account that (i) switching happens in continuous time, generating a continuum of principal strata, (ii) switching time is not defined for patients who never switch in a particular experiment, and (iii) survival time and switching time are subject to censoring.

## 2 The Concorde trial

The Concorde trial is a double-blind, randomized clinical trial aimed to evaluate the effect of immediate/active versus deferred/control treatment with zidovudine in symptom-free individuals infected with HIV (Concorde Coordinating

Committee, 1994). Due to privacy constraints, we use a synthetic dataset produced by White et al. (2002), which closely mimics the Concorde trial. The data comprise  $n = 1000$  patients with asymptomatic HIV infection. Half the patients are randomized to immediate zidovudine, and the other half to deferred zidovudine. In principle, patients in the deferred arm should not receive zidovudine until they progress to AIDS-related complex (ARC) or AIDS. Nevertheless, some patients in the deferred arm are allowed to switch to the active treatment arm, starting zidovudine before the onset of ARC or symptoms of HIV on the basis of persistently low CD4 cell counts. The outcome is time-to-disease progression or all causes of death. The survival time and the switching time are subject to censoring. The trial lasts 3 years, with staggered entry over the first 1.5 years; therefore, the censoring time ranges from 1.5 to 3 years. The data do not include any pretreatment covariates.

For each patient  $i$ , let  $Z_i$  denote the treatment assignment:  $Z_i = 1$  for immediate zidovudine and  $Z_i = 0$  for deferred zidovudine. Let  $Y_i^{\text{obs}}$  and  $S_i^{\text{obs}}$  denote the survival time and switching time under the actual treatment assignment without censoring. Let  $C_i$  be the censoring time. Let  $\tilde{Y}_i^{\text{obs}} = \min\{Y_i^{\text{obs}}, C_i\}$  denote the censored time-to-disease progression or death. Because patients cannot switch from the treatment to control, for patients assigned to immediate zidovudine, we set the switching time to be  $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = \bar{\mathbb{S}}$ , where the symbol “ $\bar{\mathbb{S}}$ ” is a non-real value. Under control, patients could either experience the event of interest (disease progression or death) without switching or switch before progressing or dying; they can switch from the control to the treatment arm only before their time-to-disease progression or death under control. Therefore, for patients assigned to deferred treatment with zidovudine, we observe the censored switching time:

$$\tilde{S}_i^{\text{obs}} = \begin{cases} S_i^{\text{obs}} & \text{if } S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} \leq C_i, \\ C_i & \text{if } S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} > C_i, \\ C_i & \text{if } S_i^{\text{obs}} = \bar{\mathbb{S}}, \end{cases}$$

where we set  $S_i^{\text{obs}} = \bar{\mathbb{S}}$  for control patients who progress the disease/die without switching.

Table 1 presents some summary statistics. The upper panel in Table 1 provides some insights that immediate treatment with zidovudine increases survival time. However, this simple comparison between survival times under treatment and control cannot even be interpreted as the average causal effect of the assignment due to censoring. Figure 1 shows the Kaplan–Meier estimates of the survival functions. The one under treatment dominates the one under control. This suggests that being assigned to immediate treatment with zidovudine is beneficial, although the difference between the two survival curves is quite small, and the 95% confidence intervals overlap. The comparison between the survival curves provides a non-parametric estimate of the ITT effect. Nevertheless, to assess the effect of immediate versus deferred treatment with zidovudine, we cannot ignore information on the switching status.

Table 1: Synthetic Concorde data: Descriptive statistics

Variable	All ( $n = 1000$ )	$Z_i = 0$ ( $n = 500$ )	$Z_i = 1$ ( $n = 500$ )
Treatment assignment ( $Z_i$ )	0.5	0	1
Indicator for the switching time being censored or taking on a non-real value ( $\mathbb{I}\{(S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} > C_i) \text{ or } S_i^{\text{obs}} = \bar{S}\}$ )	–	0.62	–
Censored switching time ( $\tilde{S}_i^{\text{obs}}$ )	–	1.55	–
Censoring indicator for the survival time ( $\mathbb{I}\{Y_i^{\text{obs}} > C_i\}$ )	0.69	0.66	0.71
Censored survival time ( $\tilde{Y}_i^{\text{obs}}$ )	1.93	1.89	1.97

Variable	$Z_i = 0$		
	$Y_i^{\text{obs}} \leq C_i$ $\tilde{S}_i^{\text{obs}} = C_i$ ( $n = 119$ )	$S_i^{\text{obs}} \leq C_i$ ( $n = 189$ )	$\tilde{Y}_i^{\text{obs}} = C_i$ $\tilde{S}_i^{\text{obs}} = C_i$ ( $n = 192$ )
Indicator for the switching time taking on a non-real value ( $\mathbb{I}\{S_i^{\text{obs}} = \bar{S}\}$ )	1	0	–
Indicator for the switching time being censored or taking on a non-real value ( $\mathbb{I}\{(S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} > C_i) \text{ or } S_i^{\text{obs}} = \bar{S}\}$ )	1	0	1
Censored switching time ( $\tilde{S}_i^{\text{obs}}$ )	–	1.24	2.11*
Censoring indicator for the survival time ( $\mathbb{I}\{Y_i^{\text{obs}} > C_i\}$ )	0	0.74	1
Censored survival time ( $\tilde{Y}_i^{\text{obs}}$ )	1.16	2.14	2.11*

\*Average censoring time

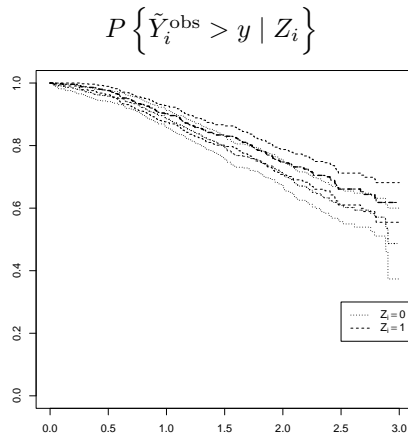


Figure 1: ITT analysis: Kaplan–Meier estimates of the survival functions with 95% confidence intervals. The solid line corresponds to the control, and the dashed line corresponds to the treatment.

### 3 Treatment switching with censoring

#### 3.1 Potential outcomes

The objective is to assess the causal effects of immediate versus deferred treatment with zidovudine on the time-to-event outcome,  $Y$  (e.g., survival time or time-to-disease progression). We use the potential outcomes to define causal effects and make the stable unit treatment value assumption. Let  $Y_i(z) \geq 0$  and  $C_i(z) \geq 0$  be the potential survival time and censoring time for patient  $i$  under treatment assignment  $z$  ( $z = 0, 1$ ). The survival time is subject to censoring. The trial starts and ends at specific calendar times, which determine a fixed duration of the study,  $c = 3$  years. Therefore, the censoring time depends on patients’ entry, which is staggered over time. Thus,  $C_i(z) \leq c$  represents the duration till the end of the study for patient  $i$  given treatment assignment  $z$ . Because the censoring time is determined by the date of entrance in the study (which varies with  $i$ ) and the date the study ended (which is determined a priori and is the same for all  $i$ ), we assume  $C_i(0) = C_i(1) = C_i$  for all  $i = 1 \dots n$ .

As the trial goes on, keeping the patients in the control arm is unethical if their physical conditions are worse than certain tolerance levels. Therefore, some patients might switch to the treatment arm even if they had been assigned to the control. Some trials permit patients in the treatment arm to switch to control if, e.g., they experience an adverse reaction to the treatment. Here, we focus on one-sided switching, as in the Concorde trial, where only patients in the control arm can switch to the treatment. Let  $S_i(z)$  be the potential switching time of patient  $i$  under treatment assignment  $z$ . The value of  $S_i(z)$  needs careful discussion. First, in the presence of one-sided switching behavior, patient  $i$ ’s

switching time is the potential switching time under control  $S_i(0)$ . Because patients in the treatment arm cannot switch, we define  $S_i(1) = \bar{\mathbb{S}}$ . Second, a patient  $i$  may not switch from the control to treatment no matter how long the follow-up is, implying  $S_i(0) = \bar{\mathbb{S}}$ . Third, a patient  $i$  can switch to the treatment arm only before their survival time, implying a natural constraint  $S_i(0) < Y_i(0)$ . The natural constraint implies that for patients who would die under control without switching to the active treatment, the switching time,  $S_i(0)$ , is censored by death, with the censoring event (death/survival time) defined by the potential outcome under control for the primary endpoint,  $Y_i(0)$ . For this type of patients, the switching time is not only not observed but also undefined; thus,  $S_i(0) = \bar{\mathbb{S}}$ . Fourth, the switching time is also subject to censoring.

### 3.2 Causal estimands

Causal effects are comparisons of the treatment and control potential outcomes for a common set of units. The average causal effect of treatment assignment equals

$$\text{ACE} = \mathbb{E}\{Y_i(1)\} - \mathbb{E}\{Y_i(0)\}. \quad (1)$$

When assessing whether the treatment can prolong the survival of patients, we are also interested in the distributional causal effect:

$$\text{DCE}(y) = P\{Y_i(1) > y\} - P\{Y_i(0) > y\}, \quad (y \in \mathbb{R}_+). \quad (2)$$

Ju and Geng (2010) noted that  $\text{ACE} = \int_0^{+\infty} \text{DCE}(y) dy$ . Although the average causal effect in (1) and the distributional effect in (2) measure well-defined ITT causal effects, they ignore the information on treatment switching in the control group.

We adopt principal stratification (Frangakis and Rubin, 2002) to define causal estimands *adjusted* for the treatment switching behavior. A principal stratification with respect to the switching behavior classifies patients into latent groups named principal strata, defined by the joint potential values of the switching time under control and under treatment,  $(S_i(0), S_i(1))$ . In the presence of one-sided switching from the control to the treatment arm,  $S_i(1) = \bar{\mathbb{S}}$  for all patients; thus, principal strata are defined by the potential outcome of the treatment switching time under control,  $S_i(0)$ , only. Frangakis and Rubin (2002) pointed out that  $S_i(0)$  acts as a pretreatment covariate unaffected by the treatment assignment. The variable  $S(0)$  is semi-continuous because switching either does not happen or happens in continuous time. Therefore, the basic principal stratification with respect to the treatment switching behavior consists of a continuum of principal strata. Each principal stratum comprises patients with the same value of the switching time:  $\{i : S_i(0) = s\}$ ,  $s \in \{\bar{\mathbb{S}}\} \cup \mathbb{R}_+$ . Throughout the paper, we refer to patients with  $S_i(0) = \bar{\mathbb{S}}$  as *non-switchers*, and to patients with a positive real value  $S_i(0) = s$ ,  $s \in \mathbb{R}_+$ , as *switchers*. Non-switchers are patients who experience disease progression or death without switching if assigned to control. Switchers belong to  $\cup_{s \in \mathbb{R}_+} \{i : S_i(0) = s\}$ , the union of the basic principal strata  $\{i : S_i(0) = s\}$ ,  $s \in \mathbb{R}_+$ . Hereafter we also refer to  $S_i(0)$



as the “switching status” of patient  $i$ : “non-switcher” is the switching status of a patient  $i$  with  $S_i(0) = \bar{S}$ , and “switcher (at some point in time)” is the switching status of a patient  $i$  with  $S_i(0) = s$ ,  $s \in \mathbb{R}_+$ .

The causal effects within principal strata are called principal causal effects (PCEs). For instance,

$$\text{ACE}(s) = \mathbb{E}\{Y_i(1) \mid S_i(0) = s\} - \mathbb{E}\{Y_i(0) \mid S_i(0) = s\} \quad (3)$$

is the principal average causal effect for  $s \in \{\bar{S}\} \cup \mathbb{R}_+$ , and

$$\text{DCE}(y \mid s) = P\{Y_i(1) > y \mid S_i(0) = s\} - P\{Y_i(0) > y \mid S_i(0) = s\} \quad (4)$$

is the principal distributional causal effect, for  $y \in \mathbb{R}_+$  and  $s \in \{\bar{S}\} \cup \mathbb{R}_+$ .

Because non-switchers would not switch to treatment if assigned to control, for them, the treatment received coincides with the treatment assigned. Thus, the PCEs for non-switchers are attributable to treatment received, that is,  $\text{ACE}(\bar{S})$  and  $\text{DCE}(y \mid \bar{S})$  can be interpreted as the effects of the treatment. They provide information on the causal effects of immediate versus deferred treatment with zidovudine for the subpopulation of patients who would never start zidovudine before the onset of ARC or AIDS if assigned to deferred zidovudine.

The estimands  $\text{ACE}(y \mid s)$  and  $\text{DCE}(y \mid s)$  for  $s \in \mathbb{R}_+$  measure the average causal effect and the distributional causal effect for patients who would switch to the treatment arm at time  $s$  had they been assigned to the control arm. For  $y \in \mathbb{R}_+$  and  $s \in \mathbb{R}_+$ ,  $\text{DCE}(y \mid s)$  defines a two-dimensional surface on  $\mathbb{R}_+ \times \mathbb{R}_+$ . The natural constraint  $S_i(0) < Y_i(0)$  implies that  $P\{Y_i(0) > y \mid S_i(0) = s\} = 1$  for  $y < s$ , and thus the principal distributional causal effect reduces to  $\text{DCE}(y \mid s) = P\{Y_i(1) > y \mid S_i(0) = s\} - 1$  for  $y < s$ . If we further assume monotonicity of survival time with respect to treatment assignment:  $Y_i(1) \geq Y_i(0)$ , then  $Y_i(1) > S_i(0)$  and the principal distributional causal effect reduces to  $\text{DCE}(y \mid s) = 0$  for  $y < s$ . In this case, for a fixed value of  $S_i(0) = s$ ,  $s \in \mathbb{R}_+$ , the principal distributional causal effect curve is non-negative within the interval  $[s, c]$  as depicted by Figure 2.

Monotonicity states that the treatment does not shorten survival compared to the control. In the Concorde trial, monotonicity amounts to assuming that immediate zidovudine does not shorten time-to-disease progression or death compared to deferred zidovudine. This assumption cannot be directly validated and can be suspicious. Without monotonicity, the principal distributional causal effect  $\text{DCE}(y \mid s)$  is negative (or at most zero) by construction for  $y < s$ . A structural negative effect may lead to a misleading interpretation of the effectiveness of the treatment. Therefore, it is sensible to consider the conditional principal distributional causal effect for the subpopulation with  $Y_i(1) > S_i(0)$ :

$$\begin{aligned} \text{cDCE}(y \mid s) & \quad (5) \\ &= P\{Y_i(1) > y \mid Y_i(1) > S_i(0), S_i(0) = s\} - P\{Y_i(0) > y \mid Y_i(1) > S_i(0), S_i(0) = s\} \\ &= P\{Y_i(1) > y \mid Y_i(1) > s, S_i(0) = s\} - P\{Y_i(0) > y \mid Y_i(1) > s, S_i(0) = s\}, \end{aligned}$$

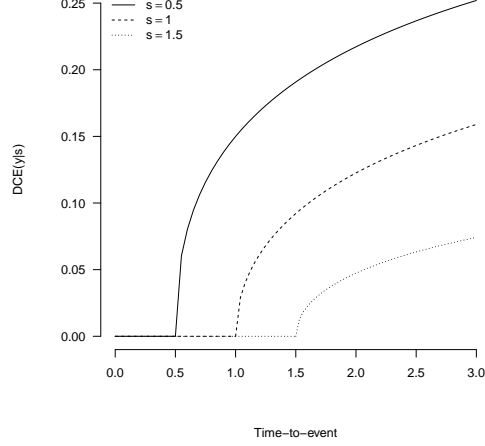


Figure 2: Examples of principal distributional causal effects under monotonicity ( $c = 3$ )

for  $y, s \in \mathbb{R}_+$ . For  $y \leq s$ ,  $\text{cDCE}(y | s) = 1 - 1 = 0$ . The estimand  $\text{cDCE}(y | s)$ ,  $s \in \mathbb{R}_+$  measures the distributional causal effect on the residual survival time from the switching time for patients who would switch to the treatment arm at time  $s$  had they been assigned to the control.

The principal causal effects in (3)–(5) are causal effects for groups of patients defined by the detailed value of  $S_i(0)$ , and thus they are basic PCEs (Frangakis and Rubin, 2002). Furthermore, we can define coarsened PCEs

$$\text{ACE}(\mathcal{A}) = \mathbb{E}\{Y_i(1) | S_i(0) \in \mathcal{A}\} - \mathbb{E}\{Y_i(0) | S_i(0) \in \mathcal{A}\}, \quad (6)$$

$$\text{DCE}(y | \mathcal{A}) = P\{Y_i(1) > y | S_i(0) \in \mathcal{A}\} - P\{Y_i(0) > y | S_i(0) \in \mathcal{A}\}, \quad (7)$$

$$\begin{aligned} \text{cDCE}(y | \mathcal{A}) &= P\{Y_i(1) > y | Y_i(1) > S_i(0), S_i(0) \in \mathcal{A}\} \\ &\quad - P\{Y_i(0) > y | Y_i(1) > S_i(0), S_i(0) \in \mathcal{A}\}, \end{aligned} \quad (8)$$

where  $\mathcal{A}$  is a subset of  $\mathbb{R}_+$ . The simplest example is  $\mathcal{A} = \mathbb{R}_+$ , which implies that the causal effects in (6)–(8) are the causal effects for the coarsened stratum of all switchers. They are the causal effects for patients who would switch earlier than or at and later than time  $s$  if assigned to control for  $\mathcal{A} = [0, s]$  and for  $\mathcal{A} = (s, +\infty)$ , respectively. Explicit formulae for these examples are shown in Web Appendix D.

In general, we can discretize the switching time into several disjoint intervals  $\mathbb{R}_+ = \mathcal{A}_1 \cup \dots \cup \mathcal{A}_K$ , and define  $\text{ACE}(\mathcal{A}_k)$ ,  $\text{DCE}(y | \mathcal{A}_k)$  and  $\text{cDCE}(y | \mathcal{A}_k)$  for  $k = 1, \dots, K$ . When  $K = 2$ ,  $\mathcal{A}_1 = [0, s]$  and  $\mathcal{A}_2 = (s, +\infty)$ , these coarsened PCEs reduce to causal effects for patients who would switch earlier and later than time  $s$  for  $k = 1$  and for  $k = 2$ , respectively. However, patients switching at

different times have different characteristics, and the basic PCEs conditioning on the potential switching time give detailed information on treatment effect heterogeneity.

Randomized clinical trials on survival outcomes with treatment switching have a similar structure to survival studies with semi-competing risks if we view the switching time and the time-to-disease progression or death as semi-competing events. The switching time for patients who would switch is a non-terminal competing event to the event of interest, and the event of interest (i.e., disease progression or death) is a terminal truncating event for the switching time. Our principal stratification approach offers an innovative approach to deal with semi-competing risks, with distinguishing features that make it crucially different from existing approaches, including the classical approach based on competing risks models. The classical semi-competing risks literature focuses on the causal effects for the whole population, namely, on total effects or (controlled) direct effects (Young et al., 2020). Recently, Stensrud and Dukes (2022) proposed to target separable effects in the presence of semi-competing risks (see Section 1 and Web-Appendix B). Our principal stratification analysis does not target causal effects for the whole population; it is a type of “sub-group” analysis with a focus on the PCEs, which are local causal effects for latent subpopulations of units defined by the switching behavior. The focus on local causal effects offers methodological and substantive advantages. The PCEs are defined without envisaging any hypothetical scenarios or hypothetical decomposition of the treatment. The PCEs may be of great interest because they naturally provide information on the heterogeneity of the treatment effect with respect to the switching behavior and on the effect of the treatment for the subpopulation of non-switchers.

### 3.3 Observed data

The potential outcome for the switching status under control,  $S_i(0)$ , and the potential outcomes for survival,  $Y_i(0)$  and  $Y_i(1)$ , are well-defined and a-priori observable for all patients in the sense that they could be observed if the patients were assigned to the corresponding treatment level (at least in the absence of censoring). A-posteriori, once the treatment has been assigned, for each patient, only the potential outcome corresponding to the treatment actually assigned is observed; the other potential outcome is missing. Specifically, for each patient  $i$ , we observe:  $\tilde{Y}_i^{\text{obs}} = \min\{Y_i^{\text{obs}}, C_i\}$ , where  $Y_i^{\text{obs}} = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$ ; and  $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = S_i(1) = \bar{S}$  under treatment and

$$\tilde{S}_i^{\text{obs}} = \begin{cases} S_i^{\text{obs}} = S_i(0) & \text{if } S_i^{\text{obs}} = S_i(0) \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} = S_i(0) \leq C_i, \\ C_i & \text{if } S_i^{\text{obs}} = S_i(0) \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} = S_i(0) > C_i, \\ C_i & \text{if } S_i^{\text{obs}} = S_i(0) = \bar{S} \end{cases}$$

under control.

In general, we do not observe the principal stratum of a patient for different reasons in the treatment and control arms. In the treatment arm, we observe

no treatment switching, and the potential outcome  $S_i(0)$  is missing. Therefore, a patient in the treatment arm may belong to any principal stratum defined by  $S_i(0)$ , and the treatment group results in an infinite mixture of principal strata. In the control arm, both the survival time and switching time are subject to censoring. We have the following cases.

- (a) The patient dies at time  $Y_i^{\text{obs}} \leq C_i$  and does not switch to zidovudine before the onset of ARC or AIDS, i.e.,  $\tilde{S}_i^{\text{obs}} = C_i$  and  $\tilde{Y}_i^{\text{obs}} = Y_i^{\text{obs}}$ . The natural constraint implies  $S_i^{\text{obs}} = S_i(0) = \bar{S}$ . Since we observe the time-to-disease progression or death without switching and the switching time is not well-defined after disease progression/death, this patient is a non-switcher belonging to the stratum  $\{i : S_i(0) = \bar{S}\}$ .
- (b) The patient switches to the treatment arm, starting zidovudine before the onset of ARC or AIDS, at time  $S_i^{\text{obs}}$  and dies at time  $Y_i^{\text{obs}}$  with  $S_i^{\text{obs}} < Y_i^{\text{obs}} \leq C_i$ , i.e.,  $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = S_i(0) = s \in \mathbb{R}_+$ , and  $\tilde{Y}_i^{\text{obs}} = Y_i^{\text{obs}} = Y_i(0)$ . This patient is a switcher belonging to the stratum  $\{i : S_i(0) = s\}$ .
- (c) The patient switches to the treatment arm, starting zidovudine before the onset of ARC or AIDS, at time  $S_i^{\text{obs}} \leq C_i$  but does not die before the end of the study, i.e.,  $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = S_i(0) = s \in \mathbb{R}_+$ , and  $\tilde{Y}_i^{\text{obs}} = C_i < Y_i^{\text{obs}}$ . This patient is a switcher belonging to the stratum  $\{i : S_i(0) = s\}$ .
- (d) The patient neither switches to zidovudine before the onset of ARC or AIDS nor dies before the end of the study (with  $S_i^{\text{obs}} \in \{\bar{S}\} \cup (C_i, +\infty)$  and  $Y_i^{\text{obs}} > C_i$ ), i.e.,  $\tilde{S}_i^{\text{obs}} = \tilde{Y}_i^{\text{obs}} = C_i$ . This patient may be a switcher with  $C_i < S_i^{\text{obs}} = S_i(0) < Y_i^{\text{obs}} = Y_i(0)$ , or a non-switcher with  $S_i^{\text{obs}} = S_i(0) = \bar{S}$  and  $Y_i^{\text{obs}} = Y_i(0) > C_i$ . This patient belongs to either the stratum  $\{i : S_i(0) = \bar{S}\}$  or the union of strata  $\cup_{s > C_i} \{i : S_i(0) = s\}$ .

Cases (a)–(c) have clear values of the switching time and survival time, at least hypothetically, so we directly observe the principal strata for these types of patients. Case (d) is less clear due to censoring because the principal stratum membership for patients with  $\tilde{S}_i^{\text{obs}} = \tilde{Y}_i^{\text{obs}} = C_i$  is missing. Table 2 shows the data pattern and latent principal strata associated with each observed group.

### 3.4 Identification issues under randomization

Although the synthetic data of the Concorde trial do not have any covariates, we discuss a general case with a  $K$ -dimensional vector of pretreatment variables  $X_i$  for each patient. We consider a completely randomized trial where the following assumption holds by design:

**Assumption 1.**  $P\{Z_i \mid S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i\}$ .

We assume that the censoring mechanism is independent of both the survival time and the switching time.

**Assumption 2.**  $P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i\} = P\{C_i\}$ .

Table 2: Observed data pattern and possible latent principal strata

$Z_i$	$\tilde{S}_i^{\text{obs}}$	$\tilde{Y}_i^{\text{obs}}$	Principal strata	Principal stratum label
0	$C_i$	$Y_i^{\text{obs}} \in [0, C_i]$	$\{i : S_i(0) = \bar{S}\}$	Non-switchers
0	$S_i^{\text{obs}} \leq C_i$	$Y_i^{\text{obs}} \in (S_i^{\text{obs}}, C_i]$	$\{i : S_i(0) = S_i^{\text{obs}}\}$ $(S_i^{\text{obs}} \in \mathbb{R}_+)$	Switchers at time $S_i^{\text{obs}} \in \mathbb{R}_+$
0	$S_i^{\text{obs}} \leq C_i$	$C_i$	$\{i : S_i(0) = S_i^{\text{obs}}\}$ $(S_i^{\text{obs}} \in \mathbb{R}_+)$	Switchers at time $S_i^{\text{obs}} \in \mathbb{R}_+$
0	$C_i$	$C_i$	$\{i : S_i(0) = \bar{S}\}$ or $\{i : S_i(0) = s \in (C_i, +\infty)\}$	Non-switchers or Switchers at some time $s > C_i$
1	$\bar{S}$	$Y_i^{\text{obs}} \in [0, C_i]$	$\{i : S_i(0) = \bar{S}\}$ or $\{S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers
1	$\bar{S}$	$C_i$	$\{i : S_i(0) = \bar{S}\}$ or $\{S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers

Assumption 2 implies that the distribution of the censoring times contains no information about the distributions of the potential survival and switching time. We can also extend the discussion under unconfounded treatment assignment  $P\{Z_i | S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i | X_i\}$ , and ignorability of the censoring mechanism conditional on observed pretreatment variables  $X_i$ ,  $P\{C_i | S_i(0), Y_i(0), Y_i(1), X_i\} = P\{C_i | X_i\}$ . The following discussion would be applicable within cells defined by  $X_i$ .

Randomization helps inference. It implies that the distribution of the switching behavior,  $S_i(0)$ , is the same in the treatment and control arms. Moreover, it allows us to express the distributional causal effects of the treatment assignment on the survival time in (2) by the distribution of the observed data:

$$\text{DCE}(y) = P\{Y_i^{\text{obs}} > y | Z_i = 1\} - P\{Y_i^{\text{obs}} > y | Z_i = 0\}.$$

Under ignorability of the censoring mechanism, we can estimate the survival functions  $P\{Y_i^{\text{obs}} > y | Z_i = z\}$  for  $y \in [0, c]$  by the empirical survival functions under treatment  $z$ ,  $z = 0, 1$ . Without imposing further assumptions, the data provide no information about the survival functions for  $y \in (c, +\infty)$ . Therefore, the identification of the average causal effect must rely on further (parametric) assumptions on  $Y$  because  $\text{ACE} = \int_0^c \text{DCE}(y)dy + \int_c^{+\infty} \text{DCE}(y)dy$  depends on the distributional causal effect within both the intervals  $[0, c]$  and  $(c, +\infty)$ .

Identifying the principal average and distributional causal effects is even more challenging. For instance, the distributional effect for the non-switchers,  $\text{DCE}(y | \bar{S})$ , is, in general, different from the prima facie distributional effect,

$$\text{FDCE}(y | \bar{S}) = P\{Y_i^{\text{obs}} > y | Z_i = 1\} - P\{Y_i^{\text{obs}} > y | Z_i = 0, S_i^{\text{obs}} = \bar{S}\},$$

i.e., the naive comparison between the patients that do not switch under treatment and control. The prima facie effect would differ from  $DCE(y | \bar{S})$ , even if no censored cases existed. Without censoring, randomization implies

$$P \{Y_i(0) > y | S_i(0) = \bar{S}\} = P \{Y_i^{\text{obs}} > y | Z_i = 0, S_i^{\text{obs}} = \bar{S}\},$$

and if we assume that switchers are less healthy people than non-switchers, then  $FDCE(y | \bar{S})$  is a lower bound for  $DCE(y | \bar{S})$ . More precisely, if  $P \{Y_i(1) > y | S_i(0) = \bar{S}\} \geq P \{Y_i(1) > y | S_i(0) \in \mathbb{R}_+\}$ , then

$$\begin{aligned} & P \{Y_i^{\text{obs}} > y | Z_i = 1\} \\ &= P \{Y_i(1) > y | S_i(0) = \bar{S}\} P \{S_i(0) = \bar{S}\} + P \{Y_i(1) > y | S_i(0) \in \mathbb{R}_+\} P \{S_i(0) \in \mathbb{R}_+\} \\ &\leq P \{Y_i(1) > y | S_i(0) = \bar{S}\}, \end{aligned}$$

which implies that  $FDCE(y | \bar{S}) \leq DCE(y | \bar{S})$ .

## 4 Bayesian Inference

In the Concorde trial, inference on the PCEs is particularly challenging due to the nature of the intermediate variable, which is a time-to-event outcome subject to censoring. Since we generally do not observe the principal stratum membership, we have to deal with a large amount of missing data, and the PCEs of interest are either not or only partially identified. We propose to use a flexible Bayesian parametric approach, which is often adopted in principal stratification analysis where inference involves techniques for incomplete data (see, e.g., [Mattei and Mealli, 2007](#); [Jin and Rubin, 2008, 2009](#); [Zigler and Belin, 2012](#); [Schwartz et al., 2011](#); [Kim et al., 2017](#)). Conceptually, the Bayesian approach does not require full identification. Bayesian inference is based on the posterior distribution of the parameters of interest, which is derived by updating a prior distribution via a likelihood, irrespective of whether the parameters are fully or partially identified, and it is always proper if the prior distribution is proper (e.g., [Lindley, 1972](#); [Ding and Li, 2018](#)). Nevertheless, in finite samples, posterior distributions of partially identified parameters may be *weak identifiable* in the sense that they may have a substantial region of flatness (e.g., [Imbens and Rubin, 1997](#); [Gustafson, 2010](#); [Schwartz et al., 2011](#)). Another appealing feature of the Bayesian approach is that it allows us to deal with all complications – missing data, truncation by death, and censoring – simultaneously in a natural way. Moreover, in Bayesian analysis, inferences are directly interpretable in probabilistic terms. The following subsections introduce and discuss a specific parametric model, and Web Appendix E details the description of our Bayesian principal stratification approach. Nevertheless, it is worth noting that alternative model specifications, possibly with a different parameterization, can be used, also depending on the specific substantive setting. The principal stratification method we propose is general and does not rely on any particular model.

## 4.1 Parametric assumptions

We adopt flexible parametric models for the switching status and survival times. We use the Weibull distribution to model the potential switching time and the potential survival times. The Weibull model has appealing features: its hazard and survival functions have a simple form, and it is flexible and easy to interpret. We can similarly consider alternative survival models, such as Burr models (e.g., Mealli and Pudney, 2003) or Bayesian semi-parametric or non-parametric models (e.g. Ibrahim et al., 2001; Schwartz et al., 2011; Kim et al., 2017). The Weibull model has two positive parameters  $\alpha$  and  $\xi$ . The parameter  $\alpha$  allows for different shapes of the hazard function. The hazard function monotonically decreases if  $\alpha < 1$ , is constant if  $\alpha = 1$ , and monotonically increases if  $\alpha > 1$ . We write the Weibull model in terms of the parameterization  $(\alpha, \log(\xi))$ .

First, we model  $S_i(0)$ . We assume that the binary indicator  $\mathbb{I}\{S_i(0) = \bar{S}\}$  follows a Bernoulli distribution with probability of success

$$\pi(x_i) = \frac{\exp(\eta_0 + x_i' \boldsymbol{\eta})}{1 + \exp(\eta_0 + x_i' \boldsymbol{\eta})}, \quad (\eta_0, \boldsymbol{\eta}) \in \mathbb{R}^{K+1}.$$

Conditionally on  $S_i(0)$  taking on real values, we assume that it follows a Weibull distribution:  $S_i(0) \mid S_i(0) \in \mathbb{R}_+, X_i \sim \text{Weibull}(\alpha_S, \log(\xi_S) = \beta_S + X_i' \boldsymbol{\eta}_S)$ ,  $\alpha_S > 0, \beta_S \in \mathbb{R}, \boldsymbol{\eta}_S \in \mathbb{R}^K$ .

Second, we model  $Y_i(0) \mid S_i(0), X_i$ . Conditionally on  $S_i(0) = \bar{S}$ , we model  $Y_i(0) \mid S_i(0) = \bar{S}, X_i$  as a Weibull distribution with parameters  $(\bar{\alpha}_Y, \log(\bar{\xi}_0) = \bar{\beta}_Y + X_i' \bar{\boldsymbol{\eta}}_Y)$ ,  $\bar{\alpha}_Y > 0, \bar{\beta}_Y \in \mathbb{R}$  and  $\bar{\boldsymbol{\eta}}_Y \in \mathbb{R}^K$ . Conditionally on  $S_i(0) = s \in \mathbb{R}_+$ , we model  $Y_i(0)$  as a location shifted Weibull distribution:

$$Y_i(0) \mid S_i(0) = s, X_i \sim s + \text{Weibull}(\alpha_Y, \log(\xi_0) = \beta_Y + \lambda_0 \log(s) + X_i' \boldsymbol{\eta}_Y)$$

with  $\alpha_Y > 0, \beta_Y, \lambda_0 \in \mathbb{R}, \boldsymbol{\eta}_Y \in \mathbb{R}^K$ . This location shift parameterization reflects the constraint  $Y_i(0) > S_i(0)$  for switchers.

Third, we model  $Y_i(1) \mid S_i(0), Y_i(0), X_i$ . Conditionally on  $S_i(0) = \bar{S}$ , we model  $Y_i(1)$  for non-switchers as a location shifted Weibull distribution:

$$Y_i(1) - \kappa Y_i(0) \mid S_i(0) = \bar{S}, Y_i(0), X_i \sim \text{Weibull}(\bar{\nu}_Y, \log(\bar{\xi}_1) = \bar{\gamma}_Y + X_i' \bar{\boldsymbol{\zeta}}),$$

with  $\kappa \in [0, 1], \bar{\nu}_Y > 0, \bar{\gamma}_Y \in \mathbb{R}, \bar{\boldsymbol{\zeta}} \in \mathbb{R}^K$ . Conditionally on  $S_i(0) = s \in \mathbb{R}_+$ , we model  $Y_i(1)$  as a location shifted Weibull distribution:

$$Y_i(1) - \kappa Y_i(0) \mid S_i(0) = s, Y_i(0), X_i \sim \text{Weibull}(\nu_Y, \log(\xi_1) = \gamma_Y + \lambda_1 \log(s) + X_i' \boldsymbol{\zeta})$$

with  $\kappa \in [0, 1], \nu_Y > 0, \gamma_Y, \lambda_1 \in \mathbb{R}, \boldsymbol{\zeta} \in \mathbb{R}^K$ .

In Web Appendix F, we explicitly show the probability density functions, the survivor functions, and the hazard functions corresponding to these model assumptions. The entire parameter vector is  $\boldsymbol{\theta} = [(\eta_0, \boldsymbol{\eta}), (\alpha_S, \beta_S, \boldsymbol{\eta}_S), (\bar{\alpha}_Y, \bar{\beta}_Y, \bar{\boldsymbol{\eta}}_Y), (\alpha_Y, \beta_Y, \lambda_0, \boldsymbol{\eta}_Y), (\bar{\nu}_Y, \bar{\gamma}_Y, \bar{\boldsymbol{\zeta}}_Y), (\nu_Y, \gamma_Y, \lambda_1, \boldsymbol{\zeta}_Y), \kappa]$ .

## 4.2 Identification of some model parameters

The parameters  $\lambda_1$  and  $\kappa$  deserve some discussion. The parameter  $\kappa$  characterizes the dependence between  $Y_i(1)$  and  $Y_i(0)$  given  $\{S_i(0), X_i\}$ . The observed data provide little information on  $\kappa$  because we can observe only one of the potential survival times for each patient. We can view  $\kappa$  as a sensitivity parameter: when  $\kappa = 0$ , the potential survival times,  $Y_i(1)$  and  $Y_i(0)$  are conditionally independent; when  $\kappa = 1$ , monotonicity  $Y_i(1) \geq Y_i(0)$  holds, which describes a type of perfect dependence structure. In practice, we suggest conducting sensitivity analysis by varying  $\kappa$  within the range  $[0, 1]$ .

The parameter  $\lambda_1$  describes the association between  $Y_i(1)$  and  $S_i(0)$  given  $Y_i(0)$  for switchers. Because  $S_i(0)$  is never observed for treated patients, the observed data provide no direct information about the partial association between  $Y_i(1)$  and  $S_i(0)$  given  $Y_i(0)$ . This lack of information may affect the causal analysis, leading to imprecise inference on the causal estimands of interest.

We propose to deal with this identifiability issue by introducing parametric assumptions that allow us to leverage better the information we have in the observed data, borrowing information on  $\lambda_1$  from other parameters and the modeling structure. We assume equality of the association parameters  $\lambda_0$  and  $\lambda_1$ :  $\lambda \equiv \lambda_0 = \lambda_1$ , so that a common parameter,  $\lambda$ , is used to describe the association between  $Y_i(1)$  and  $S_i(0)$  given  $Y_i(0)$  and between  $Y_i(0)$  and  $S_i(0)$ . Because  $Y_i(0)$  and  $S_i(0)$  are jointly observed for some control patients, we have some direct information on the association between  $Y_i(0)$  and  $S_i(0)$ , and thus on the parameter  $\lambda$ . It is worth further highlighting that Bayesian principal stratification analysis does not require the assumption of equality of the association parameters  $\lambda_0$  and  $\lambda_1$ . Nevertheless, this parametric assumption may help sharpen inference, leading to more informative and firm causal conclusions, unless results are to some extent sensitive to it.

Under the parametric assumption that  $\lambda \equiv \lambda_0 = \lambda_1$  the entire parameter vector is  $\boldsymbol{\theta} = [(\eta_0, \boldsymbol{\eta}), (\alpha_S, \beta_S, \boldsymbol{\eta}_S), (\bar{\alpha}_Y, \bar{\beta}_Y, \bar{\boldsymbol{\eta}}_Y), (\alpha_Y, \beta_Y, \boldsymbol{\eta}_Y), (\bar{\nu}_Y, \bar{\gamma}_Y, \bar{\boldsymbol{\zeta}}_Y), (\nu_Y, \gamma_Y, \boldsymbol{\zeta}_Y), \lambda, \kappa]$ .

It is worth noting that some values of the parameters  $(\lambda, \kappa)$  correspond to invoking specific structural assumptions. For instance, under our model specification, if  $\kappa = 1$  and  $\lambda < 0$ , then  $Y_i(1) \approx Y_i(0)$  for each patient  $i$  with  $S_i(0) \in \mathbb{R}_+$  and  $S_i(0) \approx 0$ . In fact, if  $\kappa = 1$  and  $\lambda < 0$ , then

$$\lim_{s \rightarrow 0} G_{Y(1)}(y | s, y_0, x_i) = \lim_{s \rightarrow 0} P\{Y_i(1) - y_0 > y | S_i(0) = s, Y_i(0) = y_0, X_i = x_i\} = 0$$

for each  $y, y_0 \in \mathbb{R}_+$ , and thus  $Y_i(1) \approx Y_i(0)$  with probability one. This is a type of “exclusion restriction,” which assumes that the assignment has no or little effect on the survival outcome for switchers if they would immediately switch to the treatment arm had they been assigned to the control arm.

The association between  $Y_i(1)$  and  $S_i(0)$  given  $Y_i(0)$  for switchers and the dependence between  $Y_i(1)$  and  $Y_i(0)$  given  $S_i(0)$ , conditional on covariates, are not identifiable nonparametrically. The little information contained in the data on these associations affects inference on the parameters  $(\lambda, \kappa)$ . Under our parametric assumptions,  $(\lambda, \kappa)$  enter the observed data likelihood and thus enter the



Bayesian posterior inference. Therefore, they are at least partially identified and could be parametrically identified depending on the modeling assumptions (Gustafson, 2010). Information on these parameters is implicitly embedded in the model for the joint potential outcomes  $Y_i(0)$  and  $Y_i(1)$  conditional on  $\{S_i(0), X_i\}$ . Such a model provides the structure to recover the relationship between the observed and missing potential outcomes. We factorize the joint conditional distribution of  $P\{Y_i(0), Y_i(1) | S_i(0), X_i\}$  into the product of  $P\{Y_i(0) | S_i(0), X_i\}$  and  $P\{Y_i(1) | Y_i(0), S_i(0), X_i\}$ . The model for  $P\{Y_i(0) | S_i(0), X_i\}$  characterizes the relationship between the survival outcome and the switching status under control. The model for  $P\{Y_i(1) | Y_i(0), S_i(0), X_i\}$  provides the structure to recover the relationship among the potential survival outcomes and the switching status under control. The data-augmentation algorithm, detailed in Web Appendix H for the Concorde study, further provides intuition about how the observed data and model specification together allow for drawing information on the missing potential outcomes and thus on the parameters  $(\lambda, \kappa)$ . We draw the missing switching status for control patients from a distribution that depends on  $\tilde{S}_i^{\text{obs}}$  through the distribution of  $\tilde{Y}_i^{\text{obs}}$ . Then, we draw the missing switching status and the missing survival time under control for treated patients from a joint distribution that depends on the distribution of  $\tilde{Y}_i^{\text{obs}}$ .

Given the possible sensitivity of the prior specifications for  $(\lambda, \kappa)$ , we will conduct various sensitivity checks in the data analysis.

### 4.3 Prior distribution, posterior distribution and sensitivity checks

We assume that the parameters are a priori independent. We propose to use Normal prior distributions for the parameters of the logistic regression model for the mixing probability,  $\pi(X_i)$ , Gamma prior distributions for the shape parameters of the Weibull distributions, and Normal prior distributions for the other parameters of the Weibull distributions. Finally, we use a Dirac delta prior for the sensitivity parameter  $\kappa$  concentrated at a pre-fixed value  $\kappa_0 \in [0, 1]$ , which is essentially the same as fixing  $\kappa$  at  $\kappa_0$  a priori. See Web Appendix G for details.

The observed-data likelihood has a complex form involving infinite mixtures because we do not observe the switching status under treatment and only partially observe the switching status under control due to censoring. Therefore, it is extremely complicated to infer the causal estimands of interest based on the observed-data likelihood directly. We use the data augmentation algorithm to derive the complete-data posterior, which is easy to deal with because it does not involve any mixture distributions. We can compute the causal estimands as byproducts, and therefore, we can simulate their posterior distributions.

We investigate the sensitivity of the results with respect to the prior specification for  $\lambda$  using both more informative priors (e.g., Normal priors with a smaller variance) as well as a less informative prior (e.g., a uniform prior distribution). We also investigate the sensitivity of the results with respect to the parametric assumption of equality of the association parameters  $\lambda_0$  and  $\lambda_1$ .

We conduct the main analysis fixing  $\kappa = 0$ , that is, assuming that  $Y_i(1)$  and  $Y_i(0)$  are conditionally independent given  $S_i(0)$ . We then assess the sensitivity of the conclusions to different assumptions on  $\kappa$  by examining how the posterior distributions of the causal estimands change with respect to different  $\kappa_0$  within the range  $[0, 1]$ .

## 5 Bayesian causal inference in the Concorde Trial

### 5.1 Bayesian ITT analysis

We first conduct a Bayesian model-based ITT analysis, which compares survival times by assignment, ignoring the switching status (see the estimands in (1) and (2)). This analysis aims to further highlight our substantive contribution to the analysis of clinical trials suffering from treatment switching.

We assume that  $Y_i(0)$  and  $Y_i(1)$  marginally follow Weibull distributions, with parameters  $(\alpha_Y, \beta_Y)$ , and  $(\nu_Y, \gamma_Y)$ , respectively, where  $\alpha_Y, \nu_Y > 0$ , and  $\beta_Y, \gamma_Y \in \mathbb{R}$ . We conduct Bayesian inference using Gamma prior distributions with shape parameter 0.01 and scale parameter 100, and thus with mean 1 and variance 100, for  $\alpha_Y$  and  $\nu_Y$ , and Normal prior distributions with zero mean and variance 10 000 for  $\beta_Y$  and  $\gamma_Y$ . The posterior median of the average causal effect is approximately 0.42, with a relatively wide 95% posterior credible interval,  $(-0.51, 1.42)$ , which covers zero. Although the posterior probability that this effect is positive is relatively high (approximately 0.83), there is little evidence that being assigned to immediate treatment with zidovudine increases the average survival time. Similarly, the estimated distributional causal effects are positive and increase monotonically over time. Still, there is little difference between survival curves, with the 95% posterior credible intervals always covering zero except for durations between 1.65 and 2.10, where the lower bound of the point-wise credible intervals is very close to zero though. See Figure 3 showing the posterior medians and 95% posterior credible intervals of the distributional causal effects. Thus, there is evidence that immediate treatment with zidovudine extends life in individuals infected with HIV, but the estimated effects are small and statistically negligible. Nevertheless, it is sensible to expect that the causal effects are heterogeneous across non-switchers and switchers or, more generally, across principal strata, making the ITT analysis an inadequate summary of the evidence in the data for the efficacy of the treatment.

To overcome the limitations of the ITT analysis, we conduct Bayesian inference on the PCEs using the framework and the parametric assumptions introduced in Section 4.

### 5.2 Bayesian principal stratification analysis

As a starting point, we assume  $\kappa = 0$ , i.e.,  $Y_i(0)$  and  $Y_i(1)$  are independent given  $S_i(0)$ . We simulate the posterior distributions of the causal estimands of interest using three independent chains from different starting values. We run each chain

$$\text{DCE}(y) = P\{Y(1) > y\} - P\{Y(0) > y\}$$

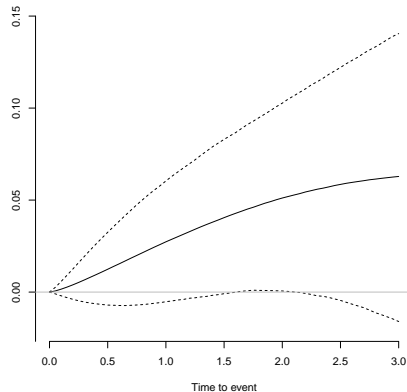


Figure 3: Bayesian ITT analysis using Weibull models. The solid line corresponds to the posterior median, and the dashed lines correspond to the 95% posterior credible interval.

for 125 000 iterations, discarding the first 25 000 iterations and saving every 20th iteration. The Markov chains mix well. We combine the three chains and use the remaining 15 000 iterations to draw inferences. See Web Appendices H and I for details on the model and prior specification, the posterior distribution of the model parameters, the MCMC algorithm, and convergence checks.

Based on the posterior medians, on average, immediate treatment with zidovudine increases survival time for non-switchers by 2.66 years, from 2.05 under deferred treatment with zidovudine to 4.76 years under immediate treatment with zidovudine. The 95% posterior credible interval (0.71, 7.73) only comprises positive values. In Figure 4, the posterior medians of the distributional causal effects for non-switchers,  $\text{DCE}(y | \bar{\mathbb{S}})$ , are positive and increase over time from 0 to 0.33 (approximately 4 months). The posterior credible intervals include only positive values except for survival times less than  $y = 0.60$  (about 7 months), where the lower bound is very close to zero though. Thus, there is evidence that immediate versus deferred treatment with zidovudine increases survival time for non-switchers.

The interpretation of the results for switchers deserves some care. For switchers,  $Y_i(1)$  is the survival value if they were assigned and actually exposed to the active treatment. The potential outcome under assignment to control,  $Y_i(0)$ , is the value of survival if switchers were initially assigned to the control treatment, exposed to the control treatment up to the time of switching, e.g.,  $s, s \in \mathbb{R}_+$ , and then exposed to the active treatment from the time of switching,  $s$ , onward. Therefore, the PCEs for switchers at time  $s$  compare the potential outcome that would have happened if they had been initially assigned to treatment and the

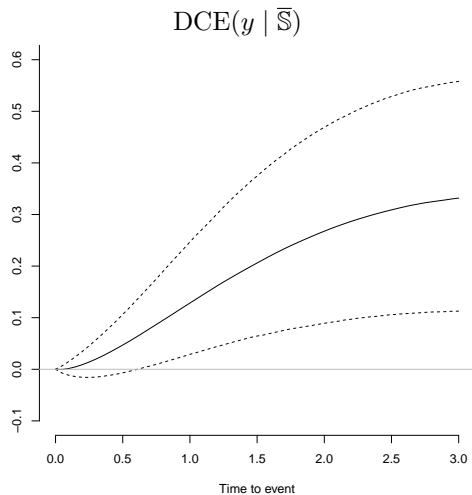


Figure 4: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of the distributional causal effects for non-switchers

potential outcome that would have happened if they had been initially assigned to control and received control treatment up to time  $s$ , and active treatment from  $s$  onward.

The average causal effects for switchers are very small and statistically negligible, irrespective of the time to switching. See Figure 5(a). Therefore, the assignment to immediate treatment with zidovudine does not affect the average survival time of patients who would have switched to zidovudine before the onset of ARC or symptoms of HIV had they been assigned to deferred treatment with zidovudine. We can interpret these results as evidence that for switchers starting to take the active treatment before the onset of ARC or symptoms of HIV is beneficial, in the sense that their survival is the same as if they had received the active treatment from the time of assignment.

We focus on the conditional distributional causal effects for switchers and relegate the results on the (unconditional) distributional causal effects to Web Appendix I. Figure 5(b) shows that the conditional distributional causal effects are always positive and show a trend increase throughout the years irrespective of the time to switching. Nevertheless, the longer the time to switching, the smaller the effects. Therefore, the distributional causal effects for switchers are highly heterogeneous with respect to the switching time. This seems plausible scientifically. For example, patients switch later because their CD4 cell counts remain sufficiently high for longer. Therefore, early switchers comprise sicker patients, and spending even a short time under control may harm them. Un-

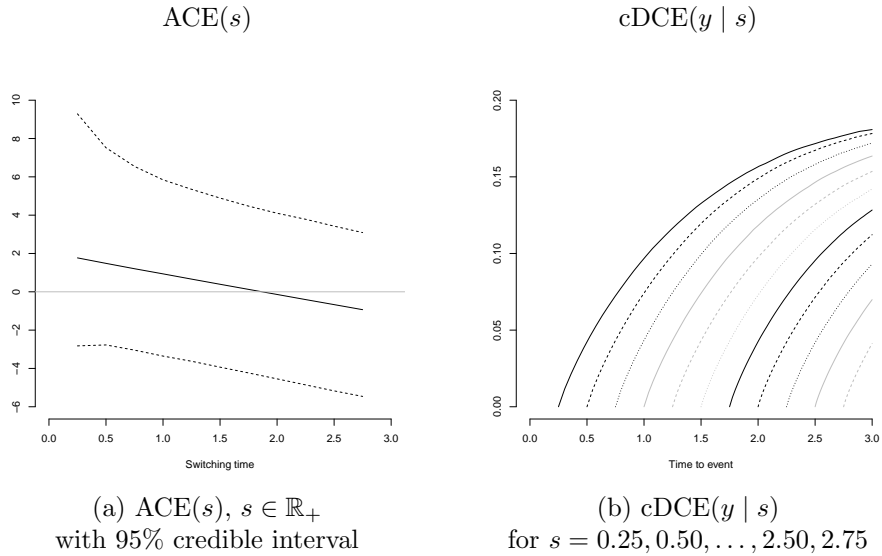


Figure 5: Posterior medians of the PCEs for switchers

der this mechanism, the benefits of immediate versus deferred treatment with zidovudine will be bigger for early switchers, i.e., the conditional distributional causal effects for early switchers will be larger than those for late switchers. Most of the posterior credible intervals for the conditional distributional causal effects only include positive values, except for patients who would switch later than 2.75 years had they been assigned to deferred treatment with zidovudine. Thus, taking the active treatment from the beginning rather than later increases the survival time for switchers. The posterior credible intervals for the conditional distributional causal effects are not shown to make Figure 5(b) easy to read.

### 5.3 Sensitivity Analyses and Model Checking

Previous results are obtained by fixing  $\kappa = 0$  and using a weakly informative prior distribution for  $\lambda$ , namely,  $N(0, 10^4)$ . We assess the sensitivity of the results to  $\kappa$ , the partial association between  $Y_i(1)$  and  $Y_i(0)$  given the switching status,  $S_i(0)$ , and to the prior specification for  $\lambda$ , by investigating how the posterior distributions of causal estimands change under different  $\kappa$  values and different prior specifications for  $\lambda$ . Results appear robust to prior specifications for  $\lambda$ ; different prior distributions for  $\lambda$  change the posterior distribution of the causal estimands only slightly. We find some sensitivity of inferences on the causal estimands to  $\kappa$ , especially for switchers.

We also investigate the robustness of the results with respect to the para-

metric assumption imposing prior equality of the association parameters  $\lambda_0$  and  $\lambda_1$ . Relaxing the parametric assumption  $\lambda_0 = \lambda_1$  only slightly changes the results for switchers by leading to posterior distributions of the causal effects for switchers with a larger posterior variability. The increased uncertainty in the causal estimands for switchers makes it more difficult to draw firm causal conclusions for them, especially for early switchers.

Our analysis is based on parametric modeling assumptions and weakly informative priors. It is important to conduct model checking. We use posterior predictive  $p$ -values to evaluate parametric assumptions. We find no evidence against the model. We relegate details on sensitivity analyses and model checks to Web Appendix I for brevity.

## 6 Principal stratum strategy at work

We have used the Concorde trial as an illustrative case study to introduce, describe, and discuss our methodology’s key concepts and provide useful insights into the interpretation of the results. However, the principal stratification approach we propose is general, defining an innovative methodological framework for the analysis of randomized clinical trials with time-to-event primary outcomes suffering from problems of treatment switching or treatment discontinuation, which may lead to important contributions from a substantial perspective. To better convey the great power of our methodology in answering substantial questions, in this Section, we briefly revisit two randomized controlled oncology trials involving patients with metastatic melanoma: the BREAK-3 Trial (Hauschild et al., 2012; Latimer et al., 2015) and the CheckMate 067 phase III trial (Larkin et al., 2015).

The BREAK-3 Trial is a multicenter, open-label, phase 3 randomized controlled clinical trial conducted between December 23, 2010, and September 1, 2011, where patients with previously untreated metastatic melanoma (BRAF V600E mutation-positive melanoma) are randomly assigned to receive either dabrafenib, a selective BRAF inhibitor, or dacarbazine, a chemotherapy medication (Hauschild et al., 2012). The primary endpoint is progression-free survival. In the BREAK-3 study, the trial protocol allows patients to switch from dacarbazine to dabrafenib at disease progression. Patients who permanently stop taking dacarbazine for any reason other than the progression of the disease are not eligible for crossover. The BREAK-3 Trial has been previously analyzed using an intention-to-treat approach (Hauschild et al., 2012) and a hypothetical approach (Latimer et al., 2015). Our principal stratum strategy offers an appealing alternative to assess the efficacy of the treatment accounting for the problem of treatment switching. The BREAK-3 Trial presents features very similar to the Concorde Trial. Both studies are randomized controlled trials with one-sided treatment switching, where patients are allowed to switch from the control to the active treatment during the follow-up period if their physical conditions worsen above certain tolerance levels. In both studies, the outcome of interest is a time-to-event outcome, and the time to switch is censored by

death, with the censoring event defined by the potential outcome under control for the primary endpoint. Therefore, the methodological setup we have introduced for the Concorde Trial can be used for the BREAK-3 Trial, with the principal causal effects defined as local causal effects for the subpopulation of non-switchers, patients who would not switch from dacarbazine to dabrafenib if assigned to dacarbazine, and for the subpopulations of switchers, patients who would switch from dacarbazine to dabrafenib if assigned to dacarbazine at some time point (before death). The average and distributional causal effects for non-switchers provide information on the efficacy of dabrafenib versus dacarbazine. The average and distributional causal effects for switchers provide information on the heterogeneity of treatment effects with respect to the switching time. Physicians may also be interested in investigating the heterogeneity of the treatment effects across non-switchers and switchers by comparing the principal causal effects for non-switchers and switchers.

The CheckMate 067 phase III trial is a multicenter, double-blinded, randomized trial where patients with metastatic melanoma are randomly assigned to receive a combination of nivolumab and ipilimumab, ipilimumab monotherapy, or nivolumab monotherapy. Patients randomized to the combination therapy receive ipilimumab combined with nivolumab once every 3 weeks for four doses, followed by nivolumab alone every 2 weeks. Patients randomized to the monotherapy receive either ipilimumab or nivolumab combined with placebo once every 3 weeks for four doses, followed by placebo alone every 2 weeks. An outcome of interest is progression-free survival, defined as the time between the date of randomization and the first date of documented progression, as determined by the Investigator, or death due to any cause, whichever occurs first. Per protocol, patients are treated until progression or unacceptable toxicity and are allowed to discontinue the assigned treatment in the presence of Adverse Events (AEs). Although patients participating in the CheckMate 067 study may discontinue both the combination therapy and the monotherapy due to AEs, discontinuation of the combination therapy is particularly interesting (Schadendorf et al., 2017). Thus, we revisit the CheckMate 067 study focusing on one-sided discontinuation of the combination therapy. We can deal with the problem of treatment discontinuation using the proposed principal stratification framework. Let  $S_i(z)$  denote the discontinuation time under treatment assignment  $z$ , with  $z = 1$  for the combination therapy and  $z = 0$  for the monotherapy. Under one-sided discontinuation of the combination therapy, the discontinuation time under monotherapy,  $S_i(0)$ , is not defined, so we set it to the non-real value  $\bar{S}$ , and define principal strata by the discontinuation behavior under the active treatment (the combination therapy),  $S_i(1)$ . Specifically, we can cross-classify patients into the following principal strata: (i) the principal stratum of patients who would never discontinue the combination therapy if assigned to it; and (ii) the principal strata of patients who would discontinue the combination therapy at some point in time if assigned to it. For the former type of patients, the discontinuation time under treatment is not defined:  $S_i(1) = \bar{S}$ ; for the latter,  $S_i(1) \in \mathbb{R}_+$ . Similar to the Concorde Trial and the BREAK-3 Trial, which suffer from one-sided treatment switching, the key features of the

CheckMate 067 phase III trial with one-sided treatment discontinuation are as follows: (a) Since discontinuation never happens or happens in continuous time, there is a continuum of principal strata with patients who would discontinue the combination therapy; (b) time-to-discontinuation is censored by death, with the censoring event defined by the potential outcome under the combination therapy for the primary endpoint, time-to-disease progression or death; (c) both time to discontinuation under treatment and time-to-disease progression or death are subject to censoring due to the end of follow-up. Patients who would discontinue the combination therapy if assigned to it probably have prognosis factors good enough to experience the progression-free survival event not immediately but are at the same time fragile from suffering treatment discontinuation due to adverse events before disease progression. Our principal stratification framework may provide useful information to physicians. The principal causal effects for patients who would discontinue the combination therapy if assigned to it can be interpreted as evidence of whether patients who would discontinue the combination therapy due to AE still benefit from it; the principal causal effects for patients who would not discontinue the combination therapy can be interpreted as the “pure effect” of the combination therapy versus the monotherapy because these patients are essentially compliers who would take the treatment assigned and would not discontinue. Moreover, comparing the principal causal effects provides information on the heterogeneity of the treatment effects with respect to the discontinuation behavior. We can investigate the heterogeneity of the treatment effects between patients who would not discontinue the combination therapy and patients who would discontinue the combination therapy. We can also study the differences in treatment effects across subsets of patients who would discontinue the combination therapy.

In some clinical trials, especially oncology trials, the sample size may be relatively small, and thus, the number of patients who switch/discontinue may be minimal. The Bayesian principal stratification approach we propose, not relying on asymptotic approximations, is a natural mode of inference in small samples by conveying and correctly quantifying uncertainty due to small sample sizes and a large amount of missingness. For instance, in clinical trials suffering from treatment switching and involving a small number of patients, a small proportion of switchers will increase uncertainty in the PCEs for switchers, but inference on the PCEs for the never-switchers will be more precise, and because never-switchers will represent a larger portion of the population, the PCE for them will be even more meaningful and easier to generalize.

## 7 Discussion

In this work, we have proposed to use principal stratification to assess the causal effects in the Concorde trial with one-sided treatment switching. The *principal causal effects* (PCEs) allow for treatment comparisons with proper adjustment for the post-treatment switching behavior. The PCEs provide valuable information on treatment effect heterogeneity across different types of patients:



non-switchers and switchers, and switchers at different time points. In particular, the PCEs for non-switchers provide information on the “pure effect” of the treatment because they are essentially compliers who would take the treatment assigned and would not switch.

Although we focus on a specific setting – clinical trials with one-sided switching from the control arm to the treatment arm – we can extend our approach to general cases. For example, we can extend our principal stratification framework to analyze: (a) multi-arm clinical trials where patients in one treatment group may switch to another active treatment group; (b) clinical trials where the control group is standard of care and patients are allowed to switch from the active treatment to the control treatment if unbearable toxicity occurs (a type of treatment discontinuation; see, e.g., [Lipkovich et al., 2020](#)); (c) clinical trials where treatment switching or treatment discontinuation is two-sided so patients can switch or discontinue both treatments. Regarding point (c), consider, for instance, a clinical trial aiming to assess the causal effects of an active versus a placebo oncological treatment on time-to-disease progression or death. Suppose all patients are exposed to the existing standard of care and are allowed to discontinue both treatments while remaining on the standard of care; the study suffers from two-sided treatment discontinuation. Let  $S_i(z)$  denote the potential outcome for the discontinuation behavior under assignment to treatment  $z$ ;  $z = 0$  for placebo or control, and  $z = 1$  for treatment. The joint potential discontinuation behavior under control and under treatment defines the principal stratum to which a patient belongs. Specifically, patients can be cross-classified into the following principal strata: (i) the principal stratum of patients who would discontinue neither the placebo nor the active treatment, irrespective of their assignment. For this type of patients, the time to discontinuation is not defined under either treatment status. Using the notation we introduce in Section 3, we can set the discontinuation time for patients who would not discontinue either treatment to the non-real value  $\bar{S}$ , so  $S_i(0) = S_i(1) = \bar{S}$  for the principal stratum of patients who would discontinue neither the placebo nor the active treatment, irrespective of their assignment; (ii) the principal strata of patients who would discontinue the placebo treatment but would not discontinue the active treatment. For this type of patients  $S_i(0) = \bar{S}$  and  $S_i(1) \in \mathbb{R}_+$ ; (iii) the principal strata of patients who would discontinue the active treatment but would not discontinue the placebo treatment. For this type of patients  $S_i(0) \in \mathbb{R}_+$  and  $S_i(1) = \bar{S}$ ; (iv) the principal strata of patients who would discontinue both the placebo and the active treatment, irrespective of their assignment. For this type of patients,  $S_i(0) \in \mathbb{R}_+$  and  $S_i(1) \in \mathbb{R}_+$ . The principal average and distributional causal effects can be defined as causal effects for the subpopulations of patients with the same discontinuation behavior under treatment and under control (the principal strata), as we have defined the principal average and distributional causal effects for non-switchers and switchers. The data structure is similar to that in the Concorde trial. For patients assigned to treatment  $z$  who do not discontinue before experiencing either disease progression or death, the discontinuation time under treatment  $z$  is undefined; the discontinuation time under treatment  $z$  is censored by death with the censoring

event defined by the potential outcome under treatment  $z$  for the primary endpoint. Moreover, both the survival and the discontinuation times are subject to censoring due to the end of follow-up.

Background covariate information is valuable in various ways. First, if pretreatment variables enter the treatment assignment mechanism, such as in stratified randomized experiments, analyses must be conditional on them. Second, in completely randomized experiments, although pretreatment covariates do not enter the treatment assignment mechanism, they can make parametric assumptions more plausible. Moreover, they can improve the prediction of the missing potential outcomes, leading to more precise inferences. Third, relevant information could also be obtained in the principal stratification analysis by looking at the distribution of baseline characteristics within each principal stratum. Characterizing the latent subgroups of patients in terms of their background characteristics can provide insights into the type of patients for which the treatment is more effective. Therefore, covariates might help explain the heterogeneity of the effects across principal strata defined by the switching status.

In clinical trials involving duration outcomes, censoring may be due to other events such as dropout and loss to follow-up. We have assumed the ignorability of the censoring mechanism, which implies that the censoring mechanism is independent of the survival potential outcomes and the switching time. A valuable topic for future research is to relax the assumption of an ignorable censoring mechanism addressing the problem of treatment switching with non-ignorable random censoring. An appealing approach to deal with non-ignorable random censoring is to extend the principal stratification analysis we present here to multiple intermediate variables, the switching status and the censoring time, considering alternative sets of assumptions on the censoring mechanism and investigating the sensibility of the results with respect to them.

## Acknowledgments

The authors thank Kaifeng Lu for the precious comments and suggestions.

## References

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). “Identification of causal effects using instrumental variables.” *Journal of the American Statistical Association*, 91: 444–455. [3](#), [4](#), [38](#)
- Barnard, J., Frangakis, C. E., Hill, J. L., and Rubin, D. B. (2003). “Principal Stratification Approach to Broken Randomized Experiments.” *Journal of the American Statistical Association*, 98: 299–323. [72](#)
- Bartolucci, F. and Grilli, L. (2011). “Modeling Partial Compliance Through Copulas in a Principal Stratification Framework.” *Journal of the American Statistical Association*, 106(494): 469–479. [38](#), [39](#)

- Chen, Q., Zeng, D., Ibrahim, J. G., Akacha, M., and Schmidli, H. (2013). “Estimating time-varying effects for overdispersed recurrent events data with treatment switching.” *Biometrika*, 100(2): 339–354. 2, 34
- Comment, L., Mealli, F., Haneuse, S., and Zigler, C. (2019). “Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks.” *arXiv preprint arXiv:1902.09304*. 4, 37
- Concorde Coordinating Committee (1994). “Concorde: MRC / ANRS randomised double blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection.” *Lancet*, 343: 871–881. 2, 4
- De Finetti, B. (1937). “La prévision: ses lois logiques, ses sources subjectives.” *Annales de l’institut Henri Poincaré*, 7: 1–68. 40
- Ding, P. and Li, F. (2018). “Causal inference: A missing data perspective.” *Statistical Science*, 33(2): 214–237. 14
- Ding, P. and Lu, J. (2017). “Principal stratification analysis using principal scores.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3): 757–777. 37
- Feller, A., Grindal, T., Miratrix, L., and Page, L. C. (2016). “Compared to what? Variation in the impacts of early childhood education by alternative care type.” *The Annals of Applied Statistics*, 10(3): 1245–1285. 37
- Fisher, L. D. and Lin, D. Y. (1999). “Time-dependent covariates in the Cox proportional-hazards regression model.” *Annual Review of Public Health*, 20: 145–157. 33
- Forastiere, L., Mealli, F., and Miratrix, L. (2018). “Posterior Predictive  $p$  - Values with Fisher Randomization Tests in Noncompliance Settings: Test Statistics vs Discrepancy Measures.” *Bayesian Analysis*. 67
- Frangakis, C. E. and Rubin, D. B. (1999). “Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes.” *Biometrika*, 86(2): 365–379. 38
- (2002). “Principal stratification in causal inference.” *Biometrics*, 58: 191–199. 3, 8, 10, 32
- Gelman, A. E., Meng, X.-L., and Stern, H. S. (1996). “Posterior predictive assessment of model fitness via realized discrepancies (with discussion).” *Statistica Sinica*, 6: 733–807. 67
- Gelman, A. E. and Rubin, D. R. (1992). “Inference from Iterative Simulation Using Multiple Sequences (with discussion).” *Statistica Science*, 7: 457–472. 54

- Geskus, R. B. (2016). *Data analysis with competing risks and intermediate states*. CRC Press Boca Raton. 35
- Gilbert, P. B. and Hudgens, M. G. (2008). “Evaluating candidate principal surrogate endpoints.” *Biometrics*, 64: 1146–1154. 39
- Gustafson, P. (2010). “Bayesian inference for partially identified models.” *International Journal of Biostatistics*, 2. 14, 17
- Guttman, I. (1967). “The use of the concept of a future observation in goodness-of-fit problems.” *Journal of the Royal Statistical Society B*, 29(1): 83–100. 67
- Hauschild, A., Grob, J.-J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., Rutkowski, P., Blank, C. U., Miller Jr, W. H., Kaempgen, E., et al. (2012). “Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial.” *The Lancet*, 380(9839): 358–365. 2, 22
- Hernán, M. and Robins, J. (2006). “Instruments for causal inference: An epidemiologist’s dream?” 17: 360–372. 32
- Hernan, M. A., Brumback, B., and Robins, J. M. (2000). “Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.” *Epidemiology*, 1: 561–570. 33
- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer. 15
- ICH (2019). “Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, E9(R1).” URL [https://database.ich.org/sites/default/files/E9-R1\\_Step4\\_Guideline\\_2019\\_1203.pdf](https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf) 3
- Imbens, G. W. and Rubin, D. B. (1997). “Bayesian inference for causal effects in randomized experiments with noncompliance.” *The Annals of Statistics*, 25: 305–327. 14
- Jin, H. and Rubin, D. B. (2008). “Principal stratification for causal inference with extended partial compliance.” *Journal of the American Statistical Association*, 103(481): 101–111. 14, 38, 39
- (2009). “Public schools versus private schools: Causal inference with partial compliance.” *Journal of Educational and Behavioral Statistics*, 34(1): 24–45. 14, 39
- Jo, B. and Stuart, E. A. (2009). “On the use of propensity scores in principal causal effect estimation.” *Statistics in Medicine*, 28: 2857–2875. 37

- Ju, C. and Geng, Z. (2010). “Criteria for surrogate end points based on causal distributions.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72: 129–142. [8](#)
- Kim, C., Daniels, M. J., Hogan, J. W., Choirat, C., and Zigler, C. M. (2017). “Bayesian methods for multiple mediators: Principal stratification and causal mediation analysis of power plant emission controls.” *Working Paper*. [14](#), [15](#), [38](#), [39](#)
- Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., et al. (2015). “Combined nivolumab and ipilimumab or monotherapy in untreated melanoma.” *New England journal of medicine*, 373(1): 23–34. [2](#), [22](#)
- Latimer, N. R., Abrams, K. R., Amonkar, M. M., Stapelkamp, C., and Swann, R. S. (2015). “Adjusting for the confounding effects of treatment switching—the BREAK-3 trial: dabrafenib versus dacarbazine.” *The Oncologist*, 20(7): 798–805. [22](#)
- Lindley, D. V. (1972). *Bayesian Statistics: A review*. SIAM. [14](#)
- Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020). “Causal inference and estimands in clinical trials.” *Statistics in Biopharmaceutical Research*, 12(1): 54–67. [25](#)
- Ma, Y., Roy, J., and Marcus, B. (2011). “Causal models for randomized trials with two active treatments and continuous compliance.” *Statistics in Medicine*, 30(19): 2349–2362. [38](#), [39](#)
- Mattei, A., Forastiere, L., and Mealli, F. (2023). “Assessing Principal Causal Effects Using Principal Score Methods.” In Zubizarreta, J., Stuart, E. A., Small, D. S., and R., R. P. (eds.), *Handbook of Matching and Weighting Adjustments for Causal Inference*, chapter 17, 313–348. Chapman and Hall/CRC. [37](#)
- Mattei, A. and Mealli, F. (2007). “Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination.” *Biometrics*, 63(2): 437–446. [14](#)
- Mattei, A., Mealli, F., and Ding, P. (2020). “Assessing causal effects in the presence of treatment switching through principal stratification.” *arXiv preprint arXiv:2002.11989v1*. [4](#)
- Mealli, F. and Pudney, S. (2003). *Applying heterogeneous transition models in labour economics: the role of youth training in labour market transitions*, Chapter 16. Wiley. [15](#)
- Meng, X. L. (1994). “Posterior predictive p-values.” *Annals of Statistics*, 22: 1142–1160. [67](#)

- Morden, J. P., Lambert, N., Paul C. and Latimer, Abrams, K. R., and Wailoo, A. J. (2011). “Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study.” *Medical Research Methodology*, 1(4). 32
- Nevo, D. and Gorfine, M. (2022). “Causal inference for semi-competing risks data.” *Biostatistics*, 23(4): 1115–1132. 37, 38
- Pearl, J. (2001). “Direct and indirect effects.” In Breese, J. S. and Koller, D. (eds.), *17th Conference on Uncertainty in Artificial Intelligence*, 411–420. Morgan Kaufmann. 35
- Robins, J. M. (1989). *The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies*, 113–159. Wiley. 32
- (1994). “Correcting for non-compliance in randomized trials using structural nested mean models.” *Communications in Statistics-Theory and methods*, 23(8): 2379–2412. 2
- Robins, J. M. and Finkelstein, D. M. (2000). “Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests.” *Biometrics*, 56(3): 779–788. 33
- Robins, J. M. and Greenland, S. (1992). “Identifiability and exchangeability for direct and indirect effects.” *Epidemiology*, 143–155. 35
- Robins, J. M. and Tsiatis, A. A. (1991). “Correcting for non-compliance in randomized trials using rank preserving structural failure time models.” *Communications in Statistics-Theory and Methods*, 20(8): 2609–2631. 2, 33
- Rubin, D. B. (1984). “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.” *The Annals of Statistics*, 12(4): 1151–1172. 67
- Schadendorf, D., Wolchok, J. D., Hodi, F. S., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.-J., Cowey, C. L., Lao, C. D., Chesney, J., et al. (2017). “Efficacy and safety outcomes in patients with advanced melanoma who discontinued treatment with nivolumab and ipilimumab because of adverse events: a pooled analysis of randomized phase II and III trials.” *Journal of Clinical Oncology*, 35(34): 3807. 23
- Schwartz, S., Li, F., and Mealli, F. (2011). “A Bayesian semiparametric approach to intermediate variables in causal inference.” *Journal of the American Statistical Association*, 31(10): 949–962. 14, 15, 38, 39
- Shao, J., Chang, M., and Chow, S.-C. (2005). “Statistical inference for cancer trials with treatment switching.” *Statistics in Medicine*, 24(12): 1783–1790. 34

- Stensrud, M. J. and Dukes, O. (2022). “Translating questions to estimands in randomized clinical trials with intercurrent events.” *Statistics in Medicine*, 11, 36
- Therneau, T., Grambsch, P., and Fleming, T. (1990). “Martingale based residuals for survival models.” *Biometrika*, 77: 147–160. 70
- Varadhan, R., Xue, Q.-L., and Bandeen-Roche, K. (2014). “Semicompeting risks in aging research: methods, issues and needs.” *Lifetime data analysis*, 20(4): 538–562. 36
- Walker, A. S., White, I. R., and Babiker, A. G. (2004). “Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment.” *Statistics in Medicine*, 23(4): 571–590. 34
- White, I. R. (2006). “Estimating treatment effects in randomized trials with treatment switching.” *Statistics in Medicine*, 25(9): 1619–1622. 34
- White, I. R., Babiker, A. G., Walker, S., and Darbyshire, J. H. (1999). “Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial.” *Statistics in Medicine*, 18(19): 2617–2634. 2, 33
- White, I. R., Walker, S., and Babiker, A. (2002). “strbee: Randomization-based efficacy estimator.” *Stata Journal*, 2(2): 140–150. 5
- White, I. R., Walker, S., Babiker, A. G., and Darbyshire, J. H. (1997). “Impact of treatment changes on the interpretation of the Concorde trial.” *AIDS*, 11(8): 999–1006. 2
- Xu, Y., Scharfstein, D., Müller, P., and Daniels, M. (2022). “A Bayesian non-parametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks.” *Biostatistics*, 23(1): 34–49. 4, 37
- Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., and Hernán, M. A. (2020). “A causal framework for classical statistical estimands in failure-time settings with competing events.” *Statistics in Medicine*, 39(8): 1199–1236. 11, 35
- Zeng, D., Chen, Q., Chen, M.-H., Ibrahim, J. G., and Groups, A. R. (2011). “Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study.” *Biometrika*, 99(1): 167–184. 2, 34
- Zhang, J. and Chen, C. (2016). “Correcting treatment effect for treatment switching in randomized oncology trials with a modified iterative parametric estimation method.” *Statistics in Medicine*, 35(21): 3690–3703. 34
- Zigler, C. M. and Belin, T. R. (2012). “A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint.” *Biometrics*, 68(3): 922–932. 14, 38, 39

Web appendix for  
“Assessing causal effects in the presence of  
treatment switching through principal  
stratification”

## A Methods for treatment switching: A review

In causal inference, various methods have been proposed to evaluate the effect of a treatment accounting for treatment switching. To the best of our knowledge, all the existing methods generally focus on causal effects for the whole population, which are defined under the assumption that, for each individual, both the outcome that would have happened under assignment to treatment and the outcome that would have happened under assignment to control exist, if that individual had not switched. Unfortunately, for switchers, the outcome that would have happened if they had not switched does not exist conceptually in the data; it is an a-priori counterfactual (Frangakis and Rubin, 2002). The data contain no or little information on these a-priori counterfactual outcomes for switchers; thus, assumptions that allow one to extrapolate from the observed data information on them are required.

It is worth noting that the problem of introducing assumptions that allow one to extrapolate from the observed data information on quantities that do not exist in the data for some units also arises in randomized experiments with non-compliance when the focus is on causal effects for the whole population. In these experiments, the Instrumental Variable (IV) assumptions (exogeneity of the instrument, existence of an association between the instrument and the treatment, exclusion restrictions, and monotonicity) are sufficient to identify average causal effects for the subpopulation of compliers. Still, they are not sufficient to identify the average effect of the treatment for the whole population; in addition to the IV assumptions (with or without monotonicity), we need to introduce additional assumptions that allow one to infer the overall average effect, i.e., assumptions on a-priori counterfactual outcomes for never-takers and always-takers. In the literature, alternative sets of assumptions have been considered, including the relatively strong assumption of identical treatment effect for all units and the weaker homogeneity assumption of no additive effect modification across levels of the instrument within the treated and the untreated (Robins, 1989; Hernán and Robins, 2006).

In the treatment switching literature, naive methods include excluding patients who switch, censoring patients who switch, and using the treatment as a time-varying covariate in a regression model. See Morden et al. (2011) for a review. Excluding switchers results in a comparison of all patients who receive the treatment to patients who are assigned to the control and do not switch. This analysis compares groups that are not formed by randomization and, therefore,



it may produce heavily biased results unless the switching behavior is completely at random. Censoring the survival time at the switch relies on the assumption that the switching status is ignorable, i.e., the prognosis of patients who switch is equal to that of patients who do not switch. This assumption is untenable in studies with non-ignorable switching behavior. An alternative approach considers the treatment as a time-varying covariate and includes a time-varying indicator for the treatment received in a (Cox proportional hazards) model. It is difficult to interpret the regression coefficients in these models (Fisher and Lin, 1999), especially their relationships with causal effects of interest. Moreover, this model-based approach compares groups that are not formed by treatment assignment, and thus it loses the benefits of randomization and can bias the estimates.

More sophisticated approaches address treatment switching by reconstructing the outcome a patient would have had if they had not switched. These are inverse probability of censoring weighting (IPCW) methods, marginal structural models, and rank-preserving structural failure time (RPSFT) models. The IPCW approach censors the switchers at the time point of switching and weights the subjects inversely proportional to their probability to switch (Robins and Finkelstein, 2000). Marginal structural models impose structure on potential outcomes that would have been observed under different treatment histories (Hernan et al., 2000). A key assumption underlying these approaches is that the switching status is independent of the switch-free outcomes conditional on the observed covariates. The plausibility of this assumption rests on the information contained in the covariates. It is worth noting that the IPCW approach is generally not applicable if no pretreatment variable is available, as in our synthetic study. Moreover, when the covariates are strong predictors of the switching behavior, the estimated switching probability will be close to zero or one for some patients, and the weights can be large. As a result, in such settings, IPCW estimators can be sensitive to minor changes in the specification of the model for the probability of switching.

The RPSFT model relates the observed survival time for each individual to the time-to-event that would have been observed for that individual if s/he had never received the treatment. The RPSFT model is rank preserving in the sense that, given any two patients,  $i$  and  $i'$ , if patient  $i$  survives longer than patient  $i'$  under a treatment regime, then  $i$  survives longer than  $i'$  under another treatment regime. This approach, initially proposed by Robins and Tsiatis (1991) and further developed by White et al. (1999), explicitly assumes that the time-varying treatment received status is the actual intervention and the random treatment assignment acts as an instrumental variable. Since the instrumental variable is binary, Robins and Tsiatis (1991) and White et al. (1999) require a model linking the potential outcome and the observed outcome by a scalar parameter. This scalar parameter is a real value by which the treatment would extend each patient's baseline lifetime, regardless of when the patient eventually switches. Therefore, the scalar parameter is the causal effect of interest, which is assumed to be the same for all patients regardless of the switching time. The assumption that the treatment effect is constant allows one

to extrapolate treatment effects across different subpopulations of patients (i.e., from non-switchers to switchers, irrespective of their switching time). Along this line, Walker et al. (2004) and Zhang and Chen (2016) further imposed additional parametric assumptions.

Other researchers have focused on modeling the observed data using parametric or semiparametric approaches (Zeng et al., 2011; Chen et al., 2013). However, they usually rely on strong assumptions, like that there exists no relation between a patient’s prognosis and switching behavior. Clearly, the switching status of the patients in the control group contains important post-treatment information, which is useful to characterize treatment effect heterogeneity. Shao et al. (2005) realize this problem and propose a model incorporating the “switching effect.” However, as pointed by White (2006), in their likelihood-based inference, Shao et al. (2005) again assume independence of the switching time and the survival time.

## B Competing risks models, Survivor PCEs and PCEs with respect to treatment switching: A comparison

Randomized clinical trials on survival outcomes with treatment switching have a similar structure to survival studies with semi-competing risks, where typically two (or multiple) events happen over time, and at least one of the events is an absorbing/terminal event, that is, it prevents from observing the other event(s). Therefore, it is worthwhile to clarify similarities and differences between the principal stratification approach we propose and both approaches based on classical semi-competing risks models and other methods based on principal stratification analysis with semi-competing risks.

We first summarize the distinguishing features of the principal stratification framework we propose (see the main text for details). In our setting, the outcome of primary interest is time-to-disease progression or death, and the switching status is an intercurrent event. Two distinguishing features characterize our setting: (i) it is biologically possible that a patient could either progress/die without switching or switch before progressing/dying and (ii) patients can switch from the control to the treatment arm if assigned to control only before their time-to-disease progression or death under control. Therefore the switching status and the time-to-disease progression or death can be viewed as semi-competing events: the switching time for patients who would switch is a non-terminal competing event to the event of interest; the event of interest (i.e., disease progression or death) is instead a terminal truncating event for the switching time. Since the time to switching is not well defined after disease progression or death, the switching time is “censored by death” with the censoring event defined by the primary endpoint. Thus, assessing the effect of treatment on a terminal event, such as time-to-disease progression or death, in the presence of treatment switching requires accounting for the fact that switching is a

non-terminal semi-competing event that is not well-defined after progression or death.

We focus on describing and addressing these complications in our study using a Bayesian approach with the principal stratification framework. We first introduce principal causal effects (PCEs) for subpopulations of patients defined by their switching behavior under the control treatment, that is, for non-switchers and switchers. It is worth noting that, in our setting, causal effects of treatment on time-to-disease progression or death are well-defined for all types of patients defined by the switching status (non-switchers and switchers at some point in time); only causal effects for non-switchers are interpretable as the “pure” effects of the treatment though. Then, we adopt a Bayesian parametric approach to inference, specifying parametric models for the switching status, the switching time for switchers, and the joint distribution of the potential outcomes for the time-to-event primary endpoint conditional on the switching status.

## Principal stratification versus semi-competing risks approaches

Two key aspects make the principal stratification framework we propose crucially different from classical approaches based on competing risks models: (i) the target causal estimands, and (ii) the use of the observed data to draw inferences on the causal effects of interest.

The causal estimands we focus on, namely PCEs for non-switchers and switchers by time to switching, are particular to our principal stratification approach.

The classical semi-competing risks literature focuses on statistical estimands generally defined as contrasts of risks, without using a formal framework for characterizing causal effects and their identifying conditions. Recently, [Young et al. \(2020\)](#) clarified that total effects or (controlled) direct effects ([Robins and Greenland, 1992](#); [Pearl, 2001](#)) of the treatment on the event of interest are usually the targets in the classical competing risks literature. Generally, controlled direct effects are related to the marginal cumulative incidence or net risk, and total effects are related to the concept of sub-distribution function, cause-specific cumulative incidence function, or crude risk ([Geskus, 2016](#)).

(Controlled) direct effects measure treatment effects on the event of interest not mediated through the competing event. They are causal effects for the whole population, defined under the assumption that there exists, for each individual, the outcome that would have happened under assignment to a treatment if the competing events had been somehow eliminated, assuming the existence of a-priori counterfactuals. In clinical trials with one-sided treatment switching, controlled direct effects compare the outcome that would have happened under assignment to the active treatment and the outcome that would have happened under assignment to control if that individual had not switched. Controlled direct effects are the hypothetical estimands usually targeted by the literature on treatment switching, where the focus is on causal effects for the whole population in the hypothetical scenario that the competing event (i.e., switching) can be somehow eliminated for all patients (see [Web-Appendix A](#) for a review of the

literature on methods for treatment switching).

Total effects of the treatment on the event of interest are defined as a comparison of the joint distribution of the time to the event of interest (e.g., disease progression or death) and the time to the competing event (e.g., switching time) under treatment versus control. Therefore they are a type of ITT effect; thus, they do not account for the mechanisms by which the treatment affects the occurrence of the primary event, e.g., through other (secondary) events like tolerance implying treatment switching.

Recently, [Stensrud and Dukes \(2022\)](#) proposed to target separable effects in the presence of semi-competing risks, under the assumption that the treatment can be, at least conceptually, separated into components such that each component affects a different competing event.

In a principal stratification analysis, potential outcomes are defined as a function of the initial treatment assignment only, and no a-priori counterfactual is required. In principle, a principal stratification analysis is like a “sub-group” analysis, where groups are defined by a latent variable (the principal stratum membership). The focus is not on the causal effects for the whole population but on the principal causal effects, which are local causal effects for the principal strata. Although we cannot observe the principal stratum membership for any patient, principal strata exist in the data; we know that each patient belongs to a principal stratum, which can be viewed as an intrinsic latent characteristic of each patient. Principal causal effects are sensible and may be of great interest in randomized clinical trials with treatment switching. They provide information on the heterogeneity of the treatment effect across principal strata, that is, with respect to the switching behavior, and on the ‘pure’ effect of the treatment for the subpopulation of non-switchers. In the principal stratification framework, we can also naturally deal with the problem that the switching time under control is not well-defined for patients who would experience disease progression or death under control without switching from the control to the treatment arm.

Another critical difference between a principal stratification analysis and an analysis based on models for semi-competing risks concerns the use of the observed data to draw inferences on the causal effects of interest.

Suppose that the censoring mechanism is ignorable. In principal stratification analysis, the observed-data likelihood involves infinite mixtures (see [Web-Appendix E](#)). Models typically used in the classical competing risk literature for analyzing semi-competing data can be divided into two broad classes: models for the distribution of the observable data, which usually target total effects, and models for the distribution of latent failure times, which usually target controlled direct effects (see [Varadhan et al., 2014](#), for a review). Models for the distribution of the observable data, which include cause-specific models and sub-distribution functions, only consider the time and type of the first event that occurs to an individual, ignoring the information available after the non-terminal event. Models for latent failure times attempt to model the joint distribution of the time to the non-terminal event and the time to the terminal event or the marginal distributions of the time to the non-terminal event and the time to the terminal event under the assumption that the time to the

non-terminal event without the terminal event is well defined for all subjects.

## Principal stratification with semi-competing risks

Principal stratification analysis has been previously proposed for evaluating causal effects of treatment with semi-competing risks (Comment et al., 2019; Xu et al., 2022), and there exist strong connections between our study and the existing studies, although distinguishing features make our contribution unique.

Comment et al. (2019) and Xu et al. (2022) focus on assessing the causal effects of treatment on non-terminal time-to-event outcomes – hospital readmission and disease progression, respectively – accounting for the fact that readmission and disease progression are subject to truncation by death; since patients could die without experiencing hospital readmission/disease progression, assessing the effect on hospital readmission/disease progression requires to take into account that readmission and disease progression are not well defined after death. In these types of studies, death is not the primary endpoint, but it is a terminal event that precludes the occurrence of the primary non-terminal time-to-event outcome.

Comment et al. (2019) and Xu et al. (2022) propose to handle the problem of truncation by death with principal stratification, defining principal strata by the pair of potential death times, to account for the fact that causal effects on the primary outcome are well defined only for principal strata of patients who would not die regardless of treatment assignment. They introduce new survivor causal effects for patients who would survive regardless of treatment assignment that explicitly account for the time-to-event nature of the non-terminal outcome. A nice methodological contribution of these papers is that survivor causal effects are defined over time rather than at a single point in time so that they can also investigate how the proportion of patients who would survive regardless of treatment assignment evolves. Recently, Nevo and Gorfine (2022) used the potential outcome approach with principal stratification in time-to-event studies with two semi-competing risks, where the focus is on assessing causal effects on both event times. Their key insight is to define principal stratification with respect to the order of the two events under both treatment and control.

In principal stratification analysis, inference is usually conducted by factorizing the joint distribution of all the potential outcomes for the intercurrent outcome and the primary outcome into the product of the marginal distribution of the principal stratum membership (the joint distribution of the potential outcomes for the intercurrent outcome) and the conditional distribution of the potential outcomes for the primary outcome given the principal stratum membership. We use this approach in our study. Xu et al. (2022) also use this approach, developing a Bayesian non-parametric model under a principal ignorability assumption (Jo and Stuart, 2009; Ding and Lu, 2017; Feller et al., 2016; Mattei et al., 2023). Comment et al. (2019) propose an alternative factorization of the joint distribution of all the potential outcomes. This is factorized into the product of the two joint distributions of potential intercurrent outcome and potential main outcome under treatment and under control. Inference is

then conducted using a Bayesian parametric approach under a conditional independence assumption between potential outcomes under treatment and under control, given covariates and an individual-level latent trait. A similar frailty-based approach with parametric assumptions is proposed by [Nevo and Gorfine \(2022\)](#).

## C Connection to the noncompliance literature

Treatment switching is a general form of the noncompliance problem. Consider the case where the switching of the patients under the control arm either occurs within a short period or never happens, i.e.,  $S_i(0) \in \{\bar{S}\} \cup [0, \epsilon]$ , with  $\epsilon > 0$  being a number smaller than any survival or censoring time. Some patients immediately switch to the treatment arm after the treatment assignment. In this case, treatment switching is equivalent to the so-called “all-or-none compliance problem” ([Angrist et al., 1996](#); [Frangakis and Rubin, 1999](#)). Non-switchers, i.e., those units such that  $S_i(0) = \bar{S}$ , and switchers, for whom  $S_i(0) \in [0, \epsilon]$ , correspond to compliers and an always-takers, respectively. Therefore,  $\text{DCE}(y \mid \bar{S})$  is the distributional effect for non-switchers or compliers, and  $\text{DCE}(y \mid [0, \epsilon])$  is the distributional effect for switchers or always-takers.

Since  $\epsilon$  is small, it is reasonable to assume that the treatment assignment affects only the outcomes of compliers but not those of always-takers. This is the exclusion restriction assumption ([Angrist et al., 1996](#)), meaning  $\text{DCE}(y \mid [0, \epsilon]) = 0$  for all  $y$ . Therefore, the compliers’ distributional effect can be identified by

$$\text{DCE}(y \mid \bar{S}) = \frac{P\{Y_i^{\text{obs}} > y \mid Z_i = 1\} - P\{Y_i^{\text{obs}} > y \mid Z_i = 0\}}{P\{S_i^{\text{obs}} = \bar{S} \mid Z_i = 0\}},$$

i.e., the ratio of the distributional effect on the outcome divided by the proportion of non-switchers.

### Connection to partial noncompliance and dose-response relationship

The switching status is a semi-continuous post-treatment variable, with a binary component that classifies patients into non-switchers and switchers, and a non-negative continuous component that classifies switchers according to their switching time. In this subsection, we focus on the switchers, a coarsened principal stratum defined by the union of uncountable sets.

Recently, assessing principal causal effects in the presence of continuous intermediate variables and infinitely many principal strata have received increasing attention ([Jin and Rubin, 2008](#); [Bartolucci and Grilli, 2011](#); [Ma et al., 2011](#); [Schwartz et al., 2011](#); [Zigler and Belin, 2012](#); [Kim et al., 2017](#)). Interest may lie either in principal causal effects for specific unions of principal strata (such as the average and distributional principal causal effects in (6)–(8) in the main

text) or in entire dose-response functions or surfaces describing how the causal effect on the outcome varies as a function of the basic principal strata membership. In our setting, the dose is the time to switching for switchers. In the main text, the average principal causal effect in (3) defines a dose-response function, and the distributional causal effects in (4) and (5) define dose-response surfaces. They describe how causal effects on survival time vary as functions of the dose. They are similar to the “causal effect predictiveness surfaces” in the literature on surrogate endpoints (e.g., [Gilbert and Hudgens, 2008](#); [Zigler and Belin, 2012](#)).

Our setting is related to randomized experiments with partial compliance ([Jin and Rubin, 2008](#); [Ma et al., 2011](#)). In particular, a monotonicity assumption holds by design because no patient in the treatment group can switch to control. The switching status,  $S_i(0)$ , can be viewed as the level (time) of control received by patient  $i$  if assigned to control, and (3), (4), and (5) in the main text are causal effects on survival time for patients who would comply with the assignment to the control arm for a specific amount of time,  $s$ , had they been assigned to the control arm. Similarly, the coarsened principal causal effects in (6)–(8) in the main text can be interpreted as causal effects in specific compliance regions ([Ma et al., 2011](#)). The principal causal effects are generally not identifiable with continuous intermediate variables. Flexible parametric (e.g., [Jin and Rubin, 2008, 2009](#); [Ma et al., 2011](#); [Zigler and Belin, 2012](#)) and semi-parametric models (e.g., [Schwartz et al., 2011](#); [Bartolucci and Grilli, 2011](#); [Kim et al., 2017](#)), possibly coupled with structural assumptions, have been developed to face the identification and estimation issues.

## D Coarsened principal causal effects

We define coarsened principal causal effects compared to Equations (6), (7), and (8) in the main text.

The simplest example is  $\mathcal{A} = \mathbb{R}_+$  and the causal effects for all switchers are

$$\begin{aligned} \text{ACE}(\mathbb{R}_+) &= \mathbb{E}[Y_i(1) \mid S_i(0) \in \mathbb{R}_+] - \mathbb{E}[Y_i(0) \mid S_i(0) \in \mathbb{R}_+], \\ \text{DCE}(y \mid \mathbb{R}_+) &= P\{Y_i(1) > y \mid S_i(0) \in \mathbb{R}_+\} - P\{Y_i(0) > y \mid S_i(0) \in \mathbb{R}_+\}, \\ \text{cDCE}(y \mid \mathbb{R}_+) &= P\{Y_i(1) > y \mid Y_i(1) > S_i(0), S_i(0) \in \mathbb{R}_+\} \\ &\quad - P\{Y_i(0) > y \mid Y_i(1) > S_i(0), S_i(0) \in \mathbb{R}_+\}. \end{aligned}$$

If  $\mathcal{A} = [0, s]$ , then the causal effects for units that switch earlier than or at time  $s$  are

$$\begin{aligned} \text{ACE}([0, s]) &= \mathbb{E}[Y_i(1) \mid S_i(0) \leq s] - \mathbb{E}[Y_i(0) \mid S_i(0) \leq s], \\ \text{DCE}(y \mid [0, s]) &= P\{Y_i(1) > y \mid S_i(0) \leq s\} - P\{Y_i(0) > y \mid S_i(0) \leq s\}, \\ \text{cDCE}(y \mid [0, s]) &= P\{Y_i(1) > y \mid Y_i(1) > S_i(0), S_i(0) \leq s\} \\ &\quad - P\{Y_i(0) > y \mid Y_i(1) > S_i(0), S_i(0) \leq s\}. \end{aligned}$$

If  $\mathcal{A} = (s, +\infty)$ , then the causal effects for units that switch later than time  $s$  are

$$\begin{aligned} \text{ACE}((s, +\infty)) &= \mathbb{E}[Y_i(1) \mid S_i(0) > s] - \mathbb{E}[Y_i(0) \mid S_i(0) > s], \\ \text{DCE}(y \mid (s, +\infty)) &= P\{Y_i(1) > y \mid S_i(0) > s\} - P\{Y_i(0) > y \mid S_i(0) > s\}, \\ \text{cDCE}(y \mid (s, +\infty)) &= P\{Y_i(1) > y \mid Y_i(1) > S_i(0), S_i(0) > s\} \\ &\quad - P\{Y_i(0) > y \mid Y_i(1) > S_i(0), S_i(0) > s\}. \end{aligned}$$

## E Bayesian Inference

Let  $\mathbf{Z}$ ,  $\mathbf{C}$ ,  $\mathbf{S}(0)$ ,  $\mathbf{Y}(0)$ , and  $\mathbf{Y}(1)$  be  $n$ -vectors with  $i$ th elements equal to  $Z_i$ ,  $C_i$ ,  $S_i(0)$ ,  $Y_i(0)$ , and  $Y_i(1)$ , respectively. Let  $\mathbf{X}$  be a  $n \times K$  matrix with  $i$ -th row equal to  $X_i$ . Under Assumption 1, the joint probability (density) function of these random variables is

$$P\{\mathbf{Z}, \mathbf{C}, \mathbf{S}(0), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\} = P\{\mathbf{C}, \mathbf{S}(0), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\} P\{\mathbf{Z}\}.$$

This allows us to ignore the model of  $P\{\mathbf{Z}\}$ .

We assume that  $P\{\mathbf{C}, \mathbf{S}(0), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\}$  is unit-exchangeable. By appealing to de Finetti's theorem (De Finetti, 1937), there exists an unknown parameter vector  $\boldsymbol{\theta}$  with prior distribution  $P(\boldsymbol{\theta})$  such that

$$\begin{aligned} &P\{\mathbf{C}, \mathbf{S}(0), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\} \\ &= \int \prod_{i=1}^n P\{C_i, S_i(0), Y_i(0), Y_i(1), X_i \mid \boldsymbol{\theta}\} P(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int \prod_{i=1}^n P\{X_i \mid \boldsymbol{\theta}\} P\{S_i(0) \mid X_i; \boldsymbol{\theta}\} P\{Y_i(0) \mid S_i(0), X_i; \boldsymbol{\theta}\} \\ &\quad P\{Y_i(1) \mid Y_i(0), S_i(0), X_i; \boldsymbol{\theta}\} P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i; \boldsymbol{\theta}\} P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

We condition on the observed distribution of covariates and assume that the parameters of the distribution of covariates are a priori independent of the other parameters. Then we do not need to model  $P\{X_i \mid \boldsymbol{\theta}\}$ . Under Assumption 2,

$$P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i; \boldsymbol{\theta}\} = P\{C_i \mid \boldsymbol{\theta}\}$$

Assuming that the parameters of the censoring mechanism are a priori independent of the other parameters, we can then ignore the model of  $P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i; \boldsymbol{\theta}\}$ . Therefore, Bayesian inference for principal stratification involves two sets of models: one for the principal strata defined by the switching status,  $S_i(0)$ , given the covariates,  $X_i$ , and the other for the distribution of potential survival times  $Y_i(0)$  and  $Y_i(1)$  conditional on the switching status and covariates,  $X_i$ .

First, we postulate a two-part model for  $S_i(0)$ . Let  $\pi(x_i) = P\{S_i(0) = \bar{S} \mid X_i = x_i; \boldsymbol{\theta}\}$  be the probability of being a non-switcher, and let  $f_{S(0)}(\cdot \mid x_i) = f_{S(0)}(\cdot \mid S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta})$  and  $G_{S(0)}(\cdot \mid x_i) = P\{S_i(0) > \cdot \mid S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta}\}$



denote the probability density function and the survival function of the switching time for switchers (that is, given that  $S_i(0)$  does not take on value  $\bar{S}$ ). Since we focus on time-to-event variables, it is helpful to introduce the notation for hazard functions. Let  $h_{S(0)}(\cdot | x_i) = h_{S(0)}(\cdot | S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta})$  be the hazard function of  $S_i(0)$  for switchers, which satisfies  $f_{S(0)}(\cdot | x_i) = h_{S(0)}(\cdot | x_i) \times G_{S(0)}(\cdot | x_i)$ . Second, we specify a model for the joint conditional distribution of  $Y_i(0)$  and  $Y_i(1)$  given  $S_i(0)$  and  $X_i$  by factorizing it as the product of the conditional distribution of  $Y_i(0)$  given  $S_i(0)$  and  $X_i$ , and the conditional distribution of  $Y_i(1)$  given  $Y_i(0)$ ,  $S_i(0)$  and  $X_i$ . Table A.1 shows the notation for the probability density functions, hazard functions, and the survival functions of the potential survival times  $Y_i(0)$  and  $Y_i(1)$ .

Let  $\mathbf{D}^{\text{obs}} = [\mathbf{Z}, \mathbf{C}, \tilde{\mathbf{S}}^{\text{obs}}, \mathbb{I}\{\mathbf{S}^{\text{obs}} \leq \mathbf{C}\}, \tilde{\mathbf{Y}}^{\text{obs}}, \mathbb{I}\{\mathbf{Y}^{\text{obs}} \leq \mathbf{C}\}]$  be an  $n \times 6$  matrix, with  $i$ th row equal to  $D_i^{\text{obs}} = [Z_i, C_i, \tilde{S}_i^{\text{obs}}, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}, \tilde{Y}_i^{\text{obs}}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}]$ . The complete-data contain the observed data  $\mathbf{X}$  and  $\mathbf{D}^{\text{obs}}$ , as well as the vector of switching statuses  $\mathbf{S}^*(0)$  with the  $i$ th element  $S_i^*(0) = (1 - Z_i)[\tilde{S}_i^{\text{obs}} \mathbb{I}\{S_i(0) \in \mathbb{R}_+\} + \bar{S} \mathbb{I}\{S_i(0) = \bar{S}\}] + Z_i S_i(0)$  and the vector of survival times under control  $\mathbf{Y}^*(0)$  with the  $i$ th element  $Y_i^*(0) = (1 - Z_i)\tilde{Y}_i^{\text{obs}} + Z_i Y_i(0)$ . We can then write the observed data likelihood function in terms of the observed data as:

$$\begin{aligned}
\mathcal{L}\{\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}^{\text{obs}}\} = & \prod_{i: Z_i=0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}=0, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=1} \pi(X_i) f_{Y(0)}^{\bar{S}}(Y_i^{\text{obs}} | X_i) \\
\times & \prod_{i: Z_i=0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}=1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=0} [1 - \pi(X_i)] f_{S(0)}(S_i^{\text{obs}} | X_i) \cdot G_{Y(0)}(C_i | S_i^{\text{obs}}, X_i) \\
\times & \prod_{i: Z_i=0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}=1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=1} [1 - \pi(X_i)] f_{S(0)}(S_i^{\text{obs}} | X_i) f_{Y(0)}(Y_i^{\text{obs}} | S_i^{\text{obs}}, X_i) \\
\times & \prod_{i: Z_i=0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}=0, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=0} \pi(X_i) G_{Y(0)}^{\bar{S}}(C_i | X_i) + [1 - \pi(X_i)] G_{S(0)}(C_i | X_i) \cdot 1 \\
\times & \prod_{i: Z_i=1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=1} \left[ \pi(X_i) \int_{\mathbb{R}_+} f_{Y(1)}^{\bar{S}}(Y_i^{\text{obs}} | Y_i(0) = y_0, X_i) f_{Y(0)}^{\bar{S}}(y_0 | X_i) dy_0 + \right. \\
& \left. [1 - \pi(X_i)] \int_{\mathbb{R}_+} \int_s^{+\infty} f_{Y(1)}(Y_i^{\text{obs}} | S_i(0) = s, Y_i(0) = y_0, X_i) f_{Y(0)}(y_0 | S_i(0) = s, X_i) f_{S(0)}(s | X_i) dy_0 ds \right] \\
\times & \prod_{i: Z_i=1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}=0} \left[ \pi(X_i) \int_{\mathbb{R}_+} G_{Y(1)}^{\bar{S}}(C_i | Y_i(0) = y_0, X_i) f_{Y(0)}^{\bar{S}}(y_0 | X_i) dy_0 + \right. \\
& \left. [1 - \pi(X_i)] \int_{\mathbb{R}_+} \int_s^{+\infty} G_{Y(1)}(C_i | S_i(0) = s, Y_i(0) = y_0, X_i) f_{Y(0)}(y_0 | S_i(0) = s, X_i) f_{S(0)}(s | X_i) dy_0 ds \right]
\end{aligned}$$

Table A.1: Probability density functions, hazard functions, and survival functions of the potential survival times conditional on the switching status under non-informative type one censoring.

---

Variable	Probability density function, hazard function and survival function
$S_i(0) \mid S_i(0) \in \mathbb{R}_+, X_i = x_i$	
$f_{S(0)}(\cdot \mid x_i)$	$= f_{S(0)}(\cdot \mid S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta})$
$h_{S(0)}(\cdot \mid x_i)$	$= h_{S(0)}(\cdot \mid S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta})$
$G_{S(0)}(\cdot \mid x_i)$	$= P\{S_i(0) > \cdot \mid S_i(0) \in \mathbb{R}_+, X_i = x_i; \boldsymbol{\theta}\}$
$Y_i(0) \mid S_i(0) = \bar{S}, X_i = x_i$	
$f_{Y(0)}^{\bar{S}}(\cdot \mid x_i)$	$= f_{Y(0)}(\cdot \mid S_i(0) = \bar{S}, X_i = x_i; \boldsymbol{\theta})$
$h_{Y(0)}^{\bar{S}}(\cdot \mid x_i)$	$= h_{Y(0)}(\cdot \mid S_i(0) = \bar{S}, X_i = x_i; \boldsymbol{\theta})$
$G_{Y(0)}^{\bar{S}}(\cdot \mid x_i)$	$= P\{Y_i(0) > \cdot \mid S_i(0) = \bar{S}, X_i = x_i; \boldsymbol{\theta}\}$
$Y_i(0) \mid S_i(0) = s, X_i = x_i \quad (s \in \mathbb{R}_+)$	
$f_{Y(0)}(\cdot \mid s, x_i)$	$= f_{Y(0)}(\cdot \mid S_i(0) = s, X_i = x_i; \boldsymbol{\theta})$
$h_{Y(0)}(\cdot \mid s, x_i)$	$= h_{Y(0)}(\cdot \mid S_i(0) = s, X_i = x_i; \boldsymbol{\theta})$
$G_{Y(0)}(\cdot \mid s, x_i)$	$= P\{Y_i(0) > \cdot \mid S_i(0) = s, X_i = x_i; \boldsymbol{\theta}\}$
$Y_i(1) \mid S_i(0) = \bar{S}, Y_i(0) = y_0, X_i = x_i$	
$f_{Y(1)}^{\bar{S}}(\cdot \mid y_0, x_i)$	$= f_{Y(1)}(\cdot \mid S_i(0) = \bar{S}, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta})$
$h_{Y(1)}^{\bar{S}}(\cdot \mid y_0, x_i)$	$= h_{Y(1)}(\cdot \mid S_i(0) = \bar{S}, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta})$
$G_{Y(1)}^{\bar{S}}(\cdot \mid y_0, x_i)$	$= P\{Y_i(1) > \cdot \mid S_i(0) = \bar{S}, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta}\}$
$Y_i(1) \mid S_i(0) = s, Y_i(0) = y_0, X_i = x_i \quad (s \in \mathbb{R}_+)$	
$f_{Y(1)}(\cdot \mid s, y_0, x_i)$	$= f_{Y(1)}(\cdot \mid S_i(0) = s, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta})$
$h_{Y(1)}(\cdot \mid s, y_0, x_i)$	$= h_{Y(1)}(\cdot \mid S_i(0) = s, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta})$
$G_{Y(1)}(\cdot \mid s, y_0, x_i)$	$= P\{Y_i(1) > \cdot \mid S_i(0) = s, Y_i(0) = y_0, X_i = x_i; \boldsymbol{\theta}\}$

---

The posterior distribution of  $\boldsymbol{\theta}$  based on the complete data is

$$\begin{aligned}
& P\{\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{D}^{\text{obs}}, \mathbf{S}^*(0), \mathbf{Y}^*(0)\} \propto P(\boldsymbol{\theta}) \\
& \times \prod_{i: Z_i=0, S_i^*(0)=\bar{S}} \pi(X_i) f_{Y(0)}^{\bar{S}}(Y_i^{\text{obs}} \mid X_i)^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} G_{Y(0)}^{\bar{S}}(C_i \mid X_i)^{\mathbb{I}\{Y_i^{\text{obs}} > C_i\}} \\
& \times \prod_{i: Z_i=0, S_i^*(0) \in \mathbb{R}_+} [1 - \pi(X_i)] G_{S(0)}(C_i \mid X_i)^{\mathbb{I}\{S_i^{\text{obs}} > C_i\}} \\
& \quad \left[ f_{S(0)}(S_i^{\text{obs}} \mid X_i) f_{Y(0)}(Y_i^{\text{obs}} \mid S_i^{\text{obs}}, X_i)^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} G_{Y(0)}(C_i \mid S_i^{\text{obs}}, X_i)^{\mathbb{I}\{Y_i^{\text{obs}} > C_i\}} \right]^{\mathbb{I}\{S_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i: Z_i=1, S_i^*(0)=\bar{S}} \pi(X_i) f_{Y(0)}^{\bar{S}}(Y_i^*(0) \mid X_i) f_{Y(1)}^{\bar{S}}(Y_i^{\text{obs}} \mid Y_i^*(0), X_i)^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} G_{Y(1)}^{\bar{S}}(C_i \mid Y_i^*(0), X_i)^{\mathbb{I}\{Y_i^{\text{obs}} > C_i\}} \\
& \times \prod_{i: Z_i=1, S_i^*(0) \in \mathbb{R}_+} [1 - \pi(X_i)] f_{S(0)}(S_i^*(0) \mid X_i) f_{Y(0)}(Y_i^*(0) \mid S_i^*(0), X_i) \\
& \quad f_{Y(1)}(Y_i^{\text{obs}} \mid S_i^*(0), Y_i^*(0), X_i)^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} G_{Y(1)}(C_i \mid S_i^*(0), Y_i^*(0), X_i)^{\mathbb{I}\{Y_i^{\text{obs}} > C_i\}}.
\end{aligned}$$

## F Parametric assumptions

### Weibull distribution

A Weibull random variable  $T$  with parameters  $(\alpha, \eta)$  has pdf

$$f_T(t) = \begin{cases} \alpha \eta t^{\alpha-1} \exp\{-\eta t^\alpha\} & \text{for } t > 0, \alpha > 0, \eta > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The survivor function, the hazard function, and the cumulative hazard function of  $T$  are

$$G_T(t) = \exp\{-\eta t^\alpha\}, \quad h_T(t) = \alpha \eta t^{\alpha-1}, \quad H_T(t) = \int_0^t h(u) du = \eta t^\alpha.$$

Under the parameterization  $\beta = \log(\eta)$ , we have

$$\begin{aligned}
& f_T(t) = \alpha t^{\alpha-1} \exp\{\beta - e^\beta t^\alpha\}, \\
& G_T(t) = \exp\{-e^\beta t^\alpha\}, \quad h_T(t; \alpha, \beta) = \alpha t^{\alpha-1} e^\beta \quad H_T(t) = e^\beta t^\alpha.
\end{aligned}$$

### Sub-model for $\mathbb{I}\{S_i(0) = \bar{S}\}$

$\mathbb{I}\{S_i(0) = \bar{S}\} \sim \text{Bernoulli}(\pi(x_i))$  with

$$\pi(x_i) = \frac{\exp(\eta_0 + x_i' \boldsymbol{\eta})}{1 + \exp(\eta_0 + x_i' \boldsymbol{\eta})}, \quad (\eta_0, \boldsymbol{\eta}) \in \mathbb{R}^{K+1}.$$

### Sub-model for $S_i(0)$ for switchers

$S_i(0) \mid S_i(0) \in \mathbb{R}_+, X_i \sim \text{Weibull}(\alpha_S, \beta_S + X_i' \boldsymbol{\eta}_S), \alpha_S > 0, \beta_S \in \mathbb{R}, \boldsymbol{\eta}_S \in \mathbb{R}^K:$

$$\begin{aligned} f_{S(0)}(s \mid x_i) &= \alpha_S s^{\alpha_S - 1} \exp\{[\beta_S + x_i' \boldsymbol{\eta}_S] - e^{\beta_S + x_i' \boldsymbol{\eta}_S} s^{\alpha_S}\}, \\ h_{S(0)}(s \mid x_i) &= \alpha_S s^{\alpha_S - 1} \exp\{\beta_S + x_i' \boldsymbol{\eta}_S\}, \quad G_{S(0)}(s \mid x_i) = \exp\{-e^{\beta_S + x_i' \boldsymbol{\eta}_S} s^{\alpha_S}\}. \end{aligned}$$

### Sub-model for $Y_i(0)$ for non-switchers

$Y_i(0) \mid S_i(0) = \bar{S}, X_i \sim \text{Weibull}(\bar{\alpha}_Y, \bar{\beta}_Y + X_i' \bar{\boldsymbol{\eta}}_Y), \bar{\alpha}_Y > 0, \bar{\beta}_Y \in \mathbb{R} \text{ and } \bar{\boldsymbol{\eta}}_Y \in \mathbb{R}^K:$

$$\begin{aligned} f_{Y(0)}^{\bar{S}}(y \mid x_i) &= \bar{\alpha}_Y y^{\bar{\alpha}_Y - 1} \exp\{[\bar{\beta}_Y + x_i' \bar{\boldsymbol{\eta}}_Y] - e^{\bar{\beta}_Y + x_i' \bar{\boldsymbol{\eta}}_Y} y^{\bar{\alpha}_Y}\}, \\ h_{Y(0)}^{\bar{S}}(y \mid x_i) &= \bar{\alpha}_Y y^{\bar{\alpha}_Y - 1} \exp\{\bar{\beta}_Y + x_i' \bar{\boldsymbol{\eta}}_Y\}, \quad G_{Y(0)}^{\bar{S}}(y \mid x_i) = \exp\{-e^{\bar{\beta}_Y + x_i' \bar{\boldsymbol{\eta}}_Y} y^{\bar{\alpha}_Y}\}. \end{aligned}$$

### Sub-model for $Y_i(0)$ for switchers

$Y_i(0) \mid S_i(0) = s, s \in \mathbb{R}_+, X_i \sim s + \text{Weibull}(\alpha_Y, \beta_Y + \lambda_0 \log(s) + X_i' \boldsymbol{\eta}_Y), \alpha_Y > 0, \beta_Y, \lambda_0 \in \mathbb{R}, \boldsymbol{\eta}_Y \in \mathbb{R}^K:$

$$\begin{aligned} f_{Y(0)}(y \mid s, x_i) &= \alpha_Y (y - s)^{\alpha_Y - 1} \exp\{[\beta_Y + \lambda_0 \log(s) + x_i' \boldsymbol{\eta}_Y] - e^{\beta_Y + \lambda_0 \log(s) + x_i' \boldsymbol{\eta}_Y} (y - s)^{\alpha_Y}\}, \\ h_{Y(0)}(y \mid s, x_i) &= \alpha_Y (y - s)^{\alpha_Y - 1} e^{\beta_Y + \lambda_0 \log(s) + x_i' \boldsymbol{\eta}_Y}, \\ G_{Y(0)}(y \mid s, x_i) &= \exp\{-e^{\beta_Y + \lambda_0 \log(s) + x_i' \boldsymbol{\eta}_Y} (y - s)^{\alpha_Y}\}. \end{aligned}$$

### Sub-model for $Y_i(1)$ for non-switchers

$Y_i(1) \mid S_i(0) = \bar{S}, Y_i(0), X_i \sim \kappa Y_i(0) + \text{Weibull}(\bar{\nu}_Y, \bar{\gamma}_Y + X_i' \bar{\boldsymbol{\zeta}}), \kappa \in [0, 1], \bar{\nu}_Y > 0, \bar{\gamma}_Y \in \mathbb{R}, \bar{\boldsymbol{\zeta}} \in \mathbb{R}^K:$

$$\begin{aligned} f_{Y(1)}^{\bar{S}}(y \mid y_0, x_i) &= \bar{\nu}_Y (y - \kappa y_0)^{\bar{\nu}_Y - 1} \exp\{[\bar{\gamma}_Y + x_i' \bar{\boldsymbol{\zeta}}] - e^{\bar{\gamma}_Y + x_i' \bar{\boldsymbol{\zeta}}} (y - \kappa y_0)^{\bar{\nu}_Y}\}, \\ h_{Y(1)}^{\bar{S}}(y \mid y_0, x_i) &= \bar{\nu}_Y (y - \kappa y_0)^{\bar{\nu}_Y - 1} e^{\bar{\gamma}_Y + x_i' \bar{\boldsymbol{\zeta}}}, \\ G_{Y(1)}^{\bar{S}}(y \mid y_0, x_i) &= \exp\{-e^{\bar{\gamma}_Y + x_i' \bar{\boldsymbol{\zeta}}} (y - \kappa y_0)^{\bar{\nu}_Y}\}. \end{aligned}$$

### Sub-model for $Y_i(1)$ for switchers

$Y_i(1) \mid S_i(0) = s, s \in \mathbb{R}_+, Y_i(0), X_i \sim \kappa Y_i(0) + \text{Weibull}(\nu_Y, \gamma_Y + \lambda \log(s) + X_i' \boldsymbol{\zeta}), \kappa \in [0, 1], \nu_Y > 0, \gamma_Y, \lambda_1 \in \mathbb{R}, \boldsymbol{\zeta} \in \mathbb{R}^K:$

$$\begin{aligned} f_{Y(1)}(y \mid s, y_0, x_i) &= \nu_Y (y - \kappa y_0)^{\nu_Y - 1} \exp\{[\gamma_Y + \lambda_1 \log(s) + x_i' \boldsymbol{\zeta}] - e^{\gamma_Y + \lambda_1 \log(s) + x_i' \boldsymbol{\zeta}} (y - \kappa y_0)^{\nu_Y}\}, \\ h_{Y(1)}(y \mid s, y_0, x_i) &= \nu_Y (y - \kappa y_0)^{\nu_Y - 1} e^{\gamma_Y + \lambda_1 \log(s) + x_i' \boldsymbol{\zeta}}, \\ G_{Y(1)}(y \mid s, y_0, x_i) &= \exp\{-e^{\gamma_Y + \lambda_1 \log(s) + x_i' \boldsymbol{\zeta}} (y - \kappa y_0)^{\nu_Y}\}. \end{aligned}$$

## G Prior distributions

Under the model specification introduced in the previous Section, we propose to use Normal prior distributions for the parameters of the logistic regression model for the mixing probability  $\pi(X_i)$ :  $(\eta_0, \boldsymbol{\eta}) \sim \mathbf{N}(\boldsymbol{\mu}_\eta, \sigma_\eta^2 I_{K+1})$ , where  $I_r$  is the  $r \times r$  identity matrix. We use Gamma prior distributions for the shape parameters of the Weibull distributions:  $\alpha_S \sim \text{Gamma}(a_S, b_S)$ ,  $\bar{\alpha}_Y \sim \text{Gamma}(\bar{a}_Y, \bar{b}_Y)$ ,  $\alpha_Y \sim \text{Gamma}(a_Y, b_Y)$ ,  $\bar{\nu}_Y \sim \text{Gamma}(\bar{d}_Y, \bar{s}_Y)$ , and  $\nu_Y \sim \text{Gamma}(d_Y, s_Y)$ . Finally, we use Normal prior distributions for the other parameters of the Weibull distributions:  $\beta_S \sim \mathbf{N}(\mu_{\beta_S}, \sigma_{\beta_S}^2)$ ,  $\boldsymbol{\eta}_S \sim \mathbf{N}(\boldsymbol{\mu}_{\eta_S}, \sigma_{\eta_S}^2 I_K)$ ;  $\bar{\beta}_Y \sim \mathbf{N}(\mu_{\bar{\beta}_Y}, \sigma_{\bar{\beta}_Y}^2)$ ,  $\bar{\boldsymbol{\eta}}_Y \sim \mathbf{N}(\boldsymbol{\mu}_{\bar{\eta}_Y}, \sigma_{\bar{\eta}_Y}^2 I_K)$ ;  $\beta_Y \sim \mathbf{N}(\mu_{\beta_Y}, \sigma_{\beta_Y}^2)$ ,  $\boldsymbol{\eta}_Y \sim \mathbf{N}(\boldsymbol{\mu}_{\eta_Y}, \sigma_{\eta_Y}^2 I_K)$ ;  $\bar{\gamma}_Y \sim \mathbf{N}(\mu_{\bar{\gamma}_Y}, \sigma_{\bar{\gamma}_Y}^2)$ ,  $\bar{\boldsymbol{\zeta}}_Y \sim \mathbf{N}(\mu_{\bar{\zeta}_Y}, \sigma_{\bar{\zeta}_Y}^2 I_K)$ ;  $\gamma_Y \sim \mathbf{N}(\mu_{\gamma_Y}, \sigma_{\gamma_Y}^2)$ ,  $\boldsymbol{\zeta}_Y \sim \mathbf{N}(\mu_{\zeta_Y}, \sigma_{\zeta_Y}^2 I_K)$ ; and  $\lambda \sim \mathbf{N}(\mu_\lambda, \sigma_\lambda^2)$ .

## H Application: Model and Computational Details

### Parametric Assumptions

#### Sub-model for the Switching Behavior.

$\pi = \mathbb{E}[\mathbb{I}\{S_i(0) = \bar{S}\}] = P(S_i(0) = \bar{S})$  and  $S_i(0) \mid S_i(0) \in \mathbb{R}_+ \sim \text{Weibull}(\alpha_S, \beta_S)$ ,  $\alpha_S > 0$ ,  $\beta_S \in \mathbb{R}$ :

$$f_{S(0)}(s) = \alpha_S s^{\alpha_S - 1} \exp\{\beta_S - e^{\beta_S} s^{\alpha_S}\}$$

$$h_{S(0)}(s) = \alpha_S s^{\alpha_S - 1} e^{\beta_S} \quad G_{S(0)}(s) = \exp\{-e^{\beta_S} s^{\alpha_S}\}$$

#### Sub-model for $Y_i(0) \mid S_i(0)$ .

$Y_i(0) \mid S_i(0) = \bar{S} \sim \text{Weibull}(\bar{\alpha}_Y, \bar{\beta}_Y)$ ,  $\bar{\alpha}_Y > 0$ ,  $\bar{\beta}_Y \in \mathbb{R}$ :

$$f_{Y(0)}^{\bar{S}}(y) = \bar{\alpha}_Y y^{\bar{\alpha}_Y - 1} \exp\{\bar{\beta}_Y - e^{\bar{\beta}_Y} y^{\bar{\alpha}_Y}\}$$

$$h_{Y(0)}^{\bar{S}}(y) = \bar{\alpha}_Y y^{\bar{\alpha}_Y - 1} e^{\bar{\beta}_Y} \quad G_{Y(0)}^{\bar{S}}(y) = \exp\{-e^{\bar{\beta}_Y} y^{\bar{\alpha}_Y}\}$$

and  $Y_i(0) \mid S_i(0) = s, s \in \mathbb{R}_+ \sim s + \text{Weibull}(\alpha_Y, \beta_Y + \lambda \log(s))$ ,  $\alpha_Y > 0$ ,  $\beta_Y \in \mathbb{R}$ ,  $\lambda \in \mathbb{R}$ :

$$f_{Y(0)}(y \mid s) = \alpha_Y (y - s)^{\alpha_Y - 1} \exp\{[\beta_Y + \lambda \log(s)] - e^{\beta_Y + \lambda \log(s)} (y - s)^{\alpha_Y}\}$$

$$h_{Y(0)}(y \mid s) = \alpha_Y (y - s)^{\alpha_Y - 1} e^{\beta_Y + \lambda \log(s)} \quad G_{Y(0)}(y \mid s) = \exp\{-e^{\beta_Y + \lambda \log(s)} (y - s)^{\alpha_Y}\}$$

**Sub-model for  $Y_i(1) \mid Y_i(0), S_i(0)$ .**

$Y_i(1) \mid Y_i(0), S_i(0) = \bar{\mathbb{S}} \sim \kappa Y_i(0) + \text{Weibull}(\bar{\nu}_Y, \bar{\gamma}_Y) \quad \bar{\nu}_Y > 0, \bar{\gamma}_Y \in \mathbb{R}$ :

$$f_{Y(1)}^{\bar{\mathbb{S}}}(y \mid y_0) = \bar{\nu}_Y (y - \kappa y_0)^{\bar{\nu}_Y - 1} \exp\{\bar{\gamma}_Y - e^{\bar{\gamma}_Y} (y - \kappa y_0)^{\bar{\nu}_Y}\}$$

$$h_{Y(1)}^{\bar{\mathbb{S}}}(y \mid y_0) = \bar{\nu}_Y (y - \kappa y_0)^{\bar{\nu}_Y - 1} e^{\bar{\gamma}_Y} \quad G_{Y(1)}^{\bar{\mathbb{S}}}(y \mid y_0) = \exp\{-e^{\bar{\gamma}_Y} (y - \kappa y_0)^{\bar{\nu}_Y}\}$$

and  $Y_i(1) \mid Y_i(0), S_i(0) = s, s \in \mathbb{R}_+ \sim \kappa Y_i(0) + \text{Weibull}(\nu_Y, \gamma_Y + \lambda \log(s))$ ,  
 $\nu_Y > 0, \gamma_Y \in \mathbb{R}, \lambda \in \mathbb{R}$ :

$$f_{Y(1)}(y \mid s, y_0) = \nu_Y (y - \kappa y_0)^{\nu_Y - 1} \exp\{[\gamma_Y + \lambda \log(s)] - e^{\gamma_Y + \lambda \log(s)} (y - \kappa y_0)^{\nu_Y}\}$$

$$h_{Y(1)}(y \mid s, y_0) = \nu_Y (y - \kappa y_0)^{\nu_Y - 1} e^{\gamma_Y + \lambda \log(s)} \quad G_{Y(1)}(y \mid s, y_0) = \exp\{-e^{\gamma_Y + \lambda \log(s)} (y - \kappa y_0)^{\nu_Y}\}$$

Therefore, the entire parameter vector is  $\boldsymbol{\theta} = [\pi, (\alpha_S, \beta_S), (\bar{\alpha}_Y, \bar{\beta}_Y), (\alpha_Y, \beta_Y), (\bar{\nu}_Y, \bar{\gamma}_Y), (\nu_Y, \gamma_Y), \lambda, \kappa]$ .

## Prior distributions

Parameters are assumed to be a priori independent, with the following prior distributions. We use a conjugate Beta prior distribution for the mixing probability  $\pi \sim \text{Beta}(a, b)$ :

$$\pi \sim \text{Beta}(a, b) : p(\pi) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \pi^{a-1} (1-\pi)^{b-1}$$

with  $a = b = 1$ . Therefore the full conditional distribution of  $\pi$  is Beta with parameters  $a + \sum_{i=1}^n \mathbb{I}\{S_i(0) = \bar{\mathbb{S}}\}$  and  $b + \sum_{i=1}^n \mathbb{I}\{S_i(0) \in \mathbb{R}_+\}$ .

We use Gamma priors for the shape parameters of the Weibull distributions,  $\alpha_S, \bar{\alpha}_Y, \alpha_Y, \bar{\nu}_Y$  and  $\nu_Y$ ,

$$\alpha_S \sim \text{Gamma}(a_S, b_S) : p(\alpha_S) = \frac{1}{(b_S)^{a_S} \Gamma(a_S)} \alpha_S^{a_S - 1} e^{-\alpha_S / b_S}$$

with  $a_S = 0.1$  and  $b_S = 10$ ,

$$\begin{aligned} \bar{\alpha}_Y &\sim \text{Gamma}(\bar{a}_Y, \bar{b}_Y) & \alpha_Y &\sim \text{Gamma}(a_Y, b_Y) \\ \bar{\nu}_Y &\sim \text{Gamma}(\bar{d}_Y, \bar{s}_Y) & \nu_Y &\sim \text{Gamma}(d_Y, s_Y) \end{aligned}$$

with  $\bar{a}_Y = a_Y = 0.1, \bar{b}_Y = b_Y = 10$ , and  $\bar{d}_Y = d_Y = 100, \bar{s}_Y = s_Y = 0.01$ .

We use Normal priors for  $\beta_S, \bar{\beta}_Y, \beta_Y, \bar{\gamma}_Y, \gamma_Y$  and  $\lambda_Y$ :

$$\beta_S \sim \text{N}(\mu_S, \sigma_S^2) : p(\beta_S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left\{-\frac{1}{2\sigma_S^2} (\beta_S - \mu_S)^2\right\}$$

$$\begin{aligned} \bar{\beta}_Y &\sim N(\bar{\mu}_Y, \bar{\sigma}_Y^2) & \beta_Y &\sim N(\mu_Y, \sigma_Y^2) & \bar{\gamma}_Y &\sim N(\bar{m}_Y, \bar{\tau}_Y^2) & \gamma_Y &\sim N(m_Y, \tau_Y^2) \\ & & & & \lambda &\sim N(\mu_\lambda, \sigma_\lambda^2) \end{aligned}$$

with  $\mu_S = \bar{\mu}_Y = \mu_Y = \bar{m}_Y = m_Y = \mu_\lambda = 0$  and  $\sigma_S^2 = \bar{\sigma}_Y^2 = \sigma_Y^2 = \sigma_\lambda^2 = 10^4$ , and  $\bar{\tau}_Y^2 = \tau_Y^2 = 0.25$ .

It is worth noting that we use more informative prior distributions for the parameters  $\bar{\nu}_Y, \bar{\gamma}_Y, \nu_Y$  and  $\gamma_Y$  to deal with the difficulty of untying the mixture of switchers and non-switchers under treatment. Since we never observe the switching behavior for units assigned to the active treatment, there is no unique way to disentangle the mixture of switchers and non-switchers under treatment, and thus we can end up with unrealistic draws for those parameters. The availability of covariates might, at least partially, address this issue, helping to better disentangle the mixture.

### Complete data posterior distribution

Let  $D_i^{\text{obs}} = [Z_i, C_i, \tilde{S}_i^{\text{obs}}, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\}, \tilde{Y}_i^{\text{obs}}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}]$  denote the observed data for unit  $i$  and let  $\mathbf{D}^{\text{obs}} = [\mathbf{Z}, \mathbf{C}, \tilde{\mathbf{S}}^{\text{obs}}, \mathbb{I}\{\mathbf{S}^{\text{obs}} \leq \mathbf{C}\}, \tilde{\mathbf{Y}}^{\text{obs}}, \mathbb{I}\{\mathbf{Y}^{\text{obs}} \leq \mathbf{C}\}]$  be the matrix stacking observations for all units. For  $\kappa = \kappa_0$ , with  $\kappa_0 \in (0, 1]$ , the complete data (w.r.t. the switching status and the survival time under control) posterior distribution for the parameter vector  $\boldsymbol{\theta} = [\pi, (\alpha_S, \beta_S), (\bar{\alpha}_Y, \bar{\beta}_Y), (\alpha_Y, \beta_Y), (\bar{\nu}_Y, \bar{\gamma}_Y), (\nu_Y, \gamma_Y), \lambda, \kappa = \kappa_0]$ , is

$$\begin{aligned}
& P\left\{\boldsymbol{\theta} \mid \mathbf{D}^{\text{obs}}, \mathbf{S}^*(0), \mathbf{Y}^*(0)\right\} \propto \\
& P\{\pi\}P\{\alpha_S\}P\{\beta_S\}P\{\bar{\alpha}_Y\}P\{\bar{\beta}_Y\}P\{\alpha_Y\}P\{\beta_Y\}P\{\bar{\nu}_Y\}P\{\bar{\gamma}_Y\}P\{\nu_Y\}P\{\gamma_Y\}P\{\lambda\}\delta_{\kappa_0}(\kappa) \\
& \times \prod_{i:Z_i=0, S_i^*(0)=\bar{S}} \pi \left[ \bar{\alpha}_Y (Y_i^{\text{obs}})^{\bar{\alpha}_Y-1} \exp\{\bar{\beta}_Y - e^{\bar{\beta}_Y} (Y_i^{\text{obs}})^{\bar{\alpha}_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \exp\{-e^{\bar{\beta}_Y} C_i^{\bar{\alpha}_Y}\}^{1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i:Z_i=0, S_i^*(0) \in \mathbb{R}_+} (1-\pi) \left\{ \alpha_S (S_i^{\text{obs}})^{\alpha_S-1} \exp\{\beta_S - e^{\beta_S} (S_i^{\text{obs}})^{\alpha_S}\} \right. \\
& \quad \left[ \alpha_Y (Y_i^{\text{obs}} - S_i^{\text{obs}})^{\alpha_Y-1} \exp\{[\beta_Y + \lambda \log(S_i^{\text{obs}})] - e^{\beta_Y + \lambda \log(S_i^{\text{obs}})} (Y_i^{\text{obs}} - S_i^{\text{obs}})^{\alpha_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \quad \left. \left[ \exp\{-e^{\beta_Y + \lambda \log(S_i^{\text{obs}})} (C_i - S_i^{\text{obs}})^{\alpha_Y}\} \right]^{1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \right\}^{\mathbb{I}\{S_i^{\text{obs}} \leq C_i\}} \left\{ \exp\{-e^{\beta_S} (C_i)^{\alpha_S}\} \right\}^{1-\mathbb{I}\{S_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i:Z_i=1, S_i^*(0)=\bar{S}} \pi \bar{\alpha}_Y (Y_i^*(0))^{\bar{\alpha}_Y-1} \exp\{\bar{\beta}_Y - e^{\bar{\beta}_Y} (Y_i^*(0))^{\bar{\alpha}_Y}\} \\
& \quad \left[ \bar{\nu}_Y (Y_i^{\text{obs}} - \kappa Y_i^*(0))^{\bar{\nu}_Y-1} \exp\{\bar{\gamma}_Y - e^{\bar{\gamma}_Y} (Y_i^{\text{obs}} - \kappa Y_i^*(0))^{\bar{\nu}_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \quad \exp\{-e^{\bar{\gamma}_Y} (C_i - \kappa Y_i^*(0))^{\bar{\nu}_Y}\}^{(1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\})\mathbb{I}\{Y_i^*(0) \leq C_i/\kappa\}} \mathbb{1}^{(1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\})\mathbb{I}\{Y_i^*(0) > C_i/\kappa\}} \\
& \times \prod_{i:Z_i=1, S_i(0) \in \mathbb{R}_+} (1-\pi) \alpha_S S_i^*(0)^{\alpha_S-1} \exp\{\beta_S - e^{\beta_S} S_i^*(0)^{\alpha_S}\} \\
& \quad \alpha_Y (Y_i^*(0) - S_i^*(0))^{\alpha_Y-1} \exp\{[\beta_Y + \lambda \log(S_i^*(0))] - e^{\beta_Y + \lambda \log(S_i^*(0))} (Y_i^*(0) - S_i^*(0))^{\alpha_Y}\} \\
& \quad \left[ \exp\{-e^{\gamma_Y + \lambda \log(S_i^*(0))} (C_i - \kappa Y_i^*(0))^{\nu_Y}\} \right]^{(1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\})\mathbb{I}\{Y_i^*(0) \leq C_i/\kappa\}} \mathbb{1}^{(1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\})\mathbb{I}\{Y_i^*(0) > C_i/\kappa\}} \\
& \quad \left[ \nu_Y (Y_i^{\text{obs}} - \kappa Y_i^*(0))^{\nu_Y-1} \exp\{[\gamma_Y + \lambda \log(S_i^*(0))] - e^{\gamma_Y + \lambda \log(S_i^*(0))} (Y_i^{\text{obs}} - \kappa Y_i^*(0))^{\nu_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}}
\end{aligned}$$

For  $\kappa = 0$ , the complete (switching status) data posterior distribution for the parameter vector  $\boldsymbol{\theta} = [\pi, (\alpha_S, \beta_S), (\bar{\alpha}_Y, \bar{\beta}_Y), (\alpha_Y, \beta_Y), (\bar{\nu}_Y, \bar{\gamma}_Y), (\nu_Y, \gamma_Y), \lambda]$ , is



$$\begin{aligned}
& P\left\{\boldsymbol{\theta} \mid \mathbf{D}^{\text{obs}}, \mathbf{S}^*(0), \mathbf{Y}^*(0)\right\} \propto \\
& P\{\pi\}P\{\alpha_S\}P\{\beta_S\}P\{\bar{\alpha}_Y\}P\{\bar{\beta}_Y\}P\{\alpha_Y\}P\{\beta_Y\}P\{\bar{\nu}_Y\}P\{\bar{\gamma}_Y\}P\{\nu_Y\}P\{\gamma_Y\}P\{\lambda\}\delta_{\kappa_0}(\kappa) \\
& \times \prod_{i:Z_i=0, S_i^*(0)=\bar{S}} \pi \left[ \bar{\alpha}_Y (Y_i^{\text{obs}})^{\bar{\alpha}_Y-1} \exp\{\bar{\beta}_Y - e^{\bar{\beta}_Y} (Y_i^{\text{obs}})^{\bar{\alpha}_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \exp\{-e^{\bar{\beta}_Y} C_i^{\bar{\alpha}_Y}\}^{1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i:Z_i=0, S_i^*(0) \in \mathbb{R}_+} (1-\pi) \left\{ \alpha_S (S_i^{\text{obs}})^{\alpha_S-1} \exp\{\beta_S - e^{\beta_S} (S_i^{\text{obs}})^{\alpha_S}\} \right. \\
& \quad \left[ \alpha_Y (Y_i^{\text{obs}} - S_i^{\text{obs}})^{\alpha_Y-1} \exp\{[\beta_Y + \lambda \log(S_i^{\text{obs}})] - e^{\beta_Y + \lambda \log(S_i^{\text{obs}})} (Y_i^{\text{obs}} - S_i^{\text{obs}})^{\alpha_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \quad \left. \left[ \exp\{-e^{\beta_Y + \lambda \log(S_i^{\text{obs}})} (C_i - S_i^{\text{obs}})^{\alpha_Y}\} \right]^{1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \right\}^{\mathbb{I}\{S_i^{\text{obs}} \leq C_i\}} \left\{ \exp\{-e^{\beta_S} (C_i)^{\alpha_S}\} \right\}^{1-\mathbb{I}\{S_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i:Z_i=1, S_i^*(0)=\bar{S}} \pi \left[ \bar{\nu}_Y (Y_i^{\text{obs}})^{\bar{\nu}_Y-1} \exp\{\bar{\gamma}_Y - e^{\bar{\gamma}_Y} (Y_i^{\text{obs}})^{\bar{\nu}_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \exp\{-e^{\bar{\gamma}_Y} (C_i)^{\bar{\nu}_Y}\}^{1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}} \\
& \times \prod_{i:Z_i=1, S_i(0) \in \mathbb{R}_+} (1-\pi) \alpha_S S_i^*(0)^{\alpha_S-1} \exp\{\beta_S - e^{\beta_S} S_i^*(0)^{\alpha_S}\} \\
& \quad \left[ \exp\{-e^{\gamma_Y + \lambda \log(S_i^*(0))} (C_i)^{\nu_Y}\} \right]^{(1-\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\})} \\
& \quad \left[ \nu_Y (Y_i^{\text{obs}})^{\nu_Y-1} \exp\{[\gamma_Y + \lambda \log(S_i^*(0))] - e^{\gamma_Y + \lambda \log(S_i^*(0))} (Y_i^{\text{obs}})^{\nu_Y}\} \right]^{\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\}}
\end{aligned}$$

## Details of Calculations

Note that if  $\kappa = 0$ , we only need to impute the missing switching status by drawing from its conditional distribution given  $(\mathbf{D}^{\text{obs}}, \boldsymbol{\theta})$ ; we do not need to impute  $Y_i(0)$  for treated units.

The random variables  $S_i^*(0)$  and  $Y_i^*(0)$  are independent across units  $i = 1, \dots, n$  given  $(\mathbf{D}^{\text{obs}}, \boldsymbol{\theta})$ ; therefore, sampling from the distributions of  $(\mathbf{S}^*(0) \mid \mathbf{D}^{\text{obs}}, \boldsymbol{\theta})$  (for  $\kappa = 0$ ) and  $(\mathbf{S}^*(0), \mathbf{Y}^*(0) \mid \mathbf{D}^{\text{obs}}, \boldsymbol{\theta})$  (for  $\kappa \in (0, 1]$ ) for data augmentation only involves independent drawing from  $(S_i^*(0) \mid D_i^{\text{obs}}, \boldsymbol{\theta})$  and  $(S_i^*(0), Y_i^*(0) \mid D_i^{\text{obs}}, \boldsymbol{\theta})$ .

### Details of Calculations: $\kappa = 0$ .

Let  $(\boldsymbol{\theta}, \mathbf{S}^*(0))$  denote the current state of the chain, with

$$\boldsymbol{\theta} = [\pi, (\alpha_S, \beta_S), (\bar{\alpha}_Y, \bar{\beta}_Y), (\alpha_Y, \beta_Y), (\bar{\nu}_Y, \bar{\gamma}_Y), (\nu_Y, \gamma_Y), \lambda, \kappa = 0].$$

1. Given the parameter  $\boldsymbol{\theta}$  and observed data,  $\mathbf{D}^{\text{obs}}$ , draw the missing data  $S_i^*(0)$

– For control patients, we have

$$S_i^*(0) = S_i(0) = \begin{cases} \bar{S} & \text{if } Z_i = 0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 1 \\ \tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} & \text{if } Z_i = 0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} \in \{0, 1\}, \end{cases}$$

For control patients with  $\mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0$  and  $\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0$ , we have

$$\begin{aligned} \pi_{NS} &\equiv \\ P\left(S_i^*(0) = \bar{S} \mid \boldsymbol{\theta}, Z_i = 0, C_i, \tilde{S}_i^{\text{obs}}, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0, \tilde{Y}_i^{\text{obs}}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0\right) &= \\ \frac{\pi G_{Y(0)}^{\bar{S}}(C_i)}{\pi G_{Y(0)}^{\bar{S}}(C_i) + (1 - \pi) G_{S_i(0)}(C_i)} \cdot 1 & \end{aligned}$$

Therefore, control patients with  $\mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0$  and  $\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0$  are classified as non-switchers ( $S_i^*(0) = \bar{S}$ ) with probability  $\pi_{NS}$  and as switchers with censored switching time ( $S_i^*(0) = \tilde{S}_i^{\text{obs}} = C_i$ ) with probability  $1 - \pi_{NS}$ .

- For treated patients, we never observe  $S_i(0)$ . We use Metropolis-Hasting steps to draw  $S_i(0)$  according to  $P(S_i(0) \mid \boldsymbol{\theta}, D_i^{\text{obs}})$ . We draw candidate values  $S_i^{\text{cand}}(0)$  from a semi-continuous distribution: We first draw  $n_1$  values from a Bernoulli distribution with probability  $\pi$  setting  $S_i^{\text{cand}}(0) = \bar{S}$  for treated units for which we obtain a success (a positive value). For treated units for which we obtain a failure, a missing value of  $S_i^{\text{cand}}(0)$  is then drawn from the Weibull distribution with parameters  $\alpha_S$  and  $\beta_S$ : Weibull( $\alpha_S, \beta_S$ ). For each  $i$  with  $Z_i = 1$ , we accept  $S_i^{\text{cand}}(0)$ , setting  $S_i^*(0) = S_i^{\text{cand}}(0)$ , with probability  $p_i = \min\{p_{S_i(0)}, 1\}$ , with

$$p_{S_i(0)} = \begin{cases} r_i & \text{if } S_i^*(0) = \bar{S}, S_i^{\text{cand}}(0) = \bar{S} \\ r_i \cdot \frac{\pi}{(1 - \pi) f_{S(0)}(S_i^{\text{cand}}(0))} & \text{if } S_i^*(0) = \bar{S}, S_i^{\text{cand}}(0) \in \mathbb{R}_+ \\ r_i \cdot \frac{(1 - \pi) f_{S(0)}(S_i^*(0))}{\pi} & \text{if } S_i^*(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) = \bar{S} \\ r_i \cdot \frac{f_{S(0)}(S_i^*(0))}{f_{S(0)}(S_i^{\text{cand}}(0))} & \text{if } S_i^*(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) \in \mathbb{R}_+ \end{cases}$$

where  $f_{S(0)}(\cdot)$  is the density of the proposal Weibull distribution, Weibull( $\alpha_S, \beta_S$ ), and

$$r_i = \frac{P\{S_i^{\text{cand}}(0) \mid \boldsymbol{\theta}, D_i^{\text{obs}}\}}{P\{S_i^*(0) \mid \boldsymbol{\theta}, D_i^{\text{obs}}\}}.$$

2. Given the imputed complete data,

$$\boldsymbol{D} = \left[ \boldsymbol{Z}, \boldsymbol{C}, \tilde{\boldsymbol{S}}^{\text{obs}}, \mathbb{I}\{\boldsymbol{S}^{\text{obs}} \leq \boldsymbol{C}\}, \tilde{\boldsymbol{Y}}^{\text{obs}}, \mathbb{I}\{\boldsymbol{Y}^{\text{obs}} \leq \boldsymbol{C}\}, \boldsymbol{S}^*(0) \right],$$

we then draw the following sub-vectors of  $\boldsymbol{\theta}$  in sequence, conditional on all others:  $\pi$ ,  $\alpha_S$ ,  $\beta_S$ ,  $\bar{\alpha}_Y$ ,  $\bar{\beta}_Y$ ,  $\alpha_Y$ ,  $\beta_Y$ ,  $\bar{\nu}_Y$ ,  $\bar{\gamma}_Y$ ,  $\nu_Y$ ,  $\gamma_Y$ ,  $\lambda$ . We draw  $\pi$  directly from its full conditional distribution, a Beta distribution with parameters  $a + \sum_{i=1}^n \mathbb{I}\{S_i^*(0) = \bar{\mathbb{S}}\}$  and  $b + \sum_{i=1}^n \mathbb{I}\{S_i^*(0) \in \mathbb{R}_+\}$ . We cannot draw directly from the appropriate conditional distributions for the other model parameters, but we use Metropolis–Hasting steps for drawing from their full-conditional distributions. For instance, to draw  $\alpha_S$ , we draw a candidate value  $\alpha_S^{\text{cand}}$  from a density  $g(\alpha_S \mid \boldsymbol{\theta})$ . The candidate draw is accepted with probability

$$p_{\alpha_S} = \min \left\{ \frac{P\{[\boldsymbol{\theta} \setminus \alpha_S], \alpha_S^{\text{cand}} \mid \mathbf{D}\} g(\alpha_S \mid [\boldsymbol{\theta} \setminus \alpha_S], \alpha_S^{\text{cand}})}{P\{[\boldsymbol{\theta} \setminus \alpha_S], \alpha_S \mid \mathbf{D}\} g(\alpha_S^{\text{cand}} \mid [\boldsymbol{\theta} \setminus \alpha_S], \alpha_S)}, 1 \right\}$$

For the candidate densities, we use Gamma densities for the parameters  $\alpha_S$ ,  $\bar{\alpha}_Y$ ,  $\alpha_Y$ ,  $\bar{\nu}_Y$ , and  $\nu_Y$ , and Normal densities for the parameters  $\beta_S$ ,  $\bar{\beta}_Y$ ,  $\beta_Y$ ,  $\bar{\gamma}_Y$ ,  $\gamma_Y$ , and  $\lambda$ , centered at the current values of the parameters. The scaling factors were chosen based on preliminary runs of the chains.

**Details of Calculations:**  $\kappa \in (0, 1]$ .

Let  $(\boldsymbol{\theta}, \mathbf{S}^*(0), \mathbf{Y}^*(0))$  denote the current state of the chain, with

$$\boldsymbol{\theta} = [\pi, (\alpha_S, \beta_S), (\bar{\alpha}_Y, \bar{\beta}_Y), (\alpha_Y, \beta_Y), (\bar{\nu}_Y, \bar{\gamma}_Y), (\nu_Y, \gamma_Y), \lambda, \kappa = \kappa_0] \quad \kappa_0 \in (0, 1].$$

1. Given the parameter  $\boldsymbol{\theta}$ , observed data,  $\mathbf{D}^{\text{obs}}$ , and  $\mathbf{S}^*(0)$ , draw the missing data  $Y_i^*(0)$ 
  - For control patients, we set  $Y_i^*(0) = \tilde{Y}_i^{\text{obs}}$
  - For treated patients, we never observe  $Y_i(0)$ . We use Metropolis–Hasting steps to draw  $Y_i(0)$  according to  $P(Y_i(0) \mid \boldsymbol{\theta}, S_i(0), D_i^{\text{obs}})$ . We draw candidate values  $Y_i^{\text{cand}}(0)$  from Weibull distributions: (a) For treated patients with  $S_i^*(0) = \bar{\mathbb{S}}$ , we draw  $Y_i^{\text{cand}}(0)$  from a Weibull distribution with parameters  $(\bar{\alpha}_Y, \bar{\beta}_Y)$ ; and (b) for treated patients with  $S_i^*(0) \in \mathbb{R}_+$ , we draw  $Y_i^{\text{cand}}(0)$  from the following location shifted Weibull distribution:  $S_i^*(0) + \text{Weibull}(\alpha_Y, \beta_Y + \lambda \log(S_i^*(0)))$ . For each  $i$  with  $Z_i = 1$ , we accept  $Y_i^{\text{cand}}(0)$ , setting  $Y_i^*(0) = Y_i^{\text{cand}}(0)$ ,

with probability  $p_i = \min\{p_{Y_i(0)}, 1\}$ , with

$$p_{Y_i(0)} = \begin{cases} r_i \cdot \frac{f_{\bar{Y}(0)}^{\bar{S}}(Y_i^*(0))}{f_{\bar{Y}(0)}^{\bar{S}}(Y_i^{\text{cand}}(0))} & \text{if } S_i^*(0) = \bar{S}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0 \\ r_i \cdot \frac{f_{\bar{Y}(0)}^{\bar{S}}(Y_i^*(0))}{f_{\bar{Y}(0)}^{\bar{S}}(Y_i^{\text{cand}}(0))} & \text{if } S_i^*(0) = \bar{S}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 1, Y_i^{\text{cand}}(0) \leq Y_i^{\text{obs}}/\kappa \\ r_i \cdot \frac{f_{Y(0)}(Y_i(0))}{f_{Y(0)}(Y_i^{\text{cand}}(0))} & \text{if } S_i^*(0) \in \mathbb{R}_+, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0 \\ r_i \cdot \frac{f_{Y(0)}(Y_i^*(0))}{f_{Y(0)}(Y_i^{\text{cand}}(0))} & \text{if } S_i^*(0) \in \mathbb{R}_+, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 1, Y_i^{\text{cand}}(0) \leq Y_i^{\text{obs}}/\kappa \\ 0 & \text{if } \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 1, Y_i^{\text{cand}}(0) > Y_i^{\text{obs}}/\kappa \end{cases}$$

where  $f_{\bar{Y}(0)}^{\bar{S}}(\cdot)$  and  $f_{Y(0)}(\cdot)$  are the densities of the proposal Weibull distributions, and

$$r_i = \frac{P\{Y_i^{\text{cand}}(0) \mid \boldsymbol{\theta}, D_i^{\text{obs}}, S_i^*(0)\}}{P\{Y_i^*(0) \mid \boldsymbol{\theta}, D_i^{\text{obs}}, S_i^*(0)\}}.$$

Note that we do not set  $p_{Y_i(0)} = 0$  for  $\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0$  and  $Y_i^{\text{cand}}(0) > C_i/\kappa$ , because, in principle, the survival outcome under control,  $Y_i(0)$ , can be greater than  $C_i/\kappa$ : For some units, we can have  $Y_i(1)/\kappa \geq Y_i(0) > C_i/\kappa$ . For this type of units, the probability that  $Y_i(1) > C_i$  is one.

2. Given the parameter  $\boldsymbol{\theta}$ , the observed data,  $\mathbf{D}^{\text{obs}}$ , and  $\mathbf{Y}^*(0)$  draw, the missing data  $S_i^*(0)$ .

– For control patients, we have

$$S_i^*(0) = S_i(0) = \begin{cases} \bar{S} & \text{if } Z_i = 0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 1 \\ \tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} & \text{if } Z_i = 0, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 1, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} \in \{0, 1\}. \end{cases}$$

For control patients with  $\mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0$  and  $\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0$ , we have

$$\pi_{NS} \equiv P\left(S_i^*(0) = \bar{S} \mid \boldsymbol{\theta}, Z_i = 0, C_i, \tilde{S}_i^{\text{obs}}, \mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0, \tilde{Y}_i^{\text{obs}}, \mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0\right) = \frac{\pi G_{\bar{Y}(0)}^{\bar{S}}(C_i)}{\pi G_{\bar{Y}(0)}^{\bar{S}}(C_i) + (1 - \pi) G_{S_i(0)}(C_i) \cdot 1}$$

Therefore, control patients with  $\mathbb{I}\{S_i^{\text{obs}} \leq C_i\} = 0$  and  $\mathbb{I}\{Y_i^{\text{obs}} \leq C_i\} = 0$  are classified as non-switchers ( $S_i^*(0) = \bar{S}$ ) with probability  $\pi_{NS}$  and as switchers with censored switching time ( $S_i^*(0) = \tilde{S}_i^{\text{obs}} = C_i$ ) with probability  $1 - \pi_{NS}$ .

- For treated patients, we never observe  $S_i^*(0) = S_i(0)$ . We use Metropolis-Hasting steps to draw  $S_i(0)$  according to  $P(S_i(0) | \boldsymbol{\theta}, D_i^{\text{obs}})$ . We draw candidate values  $S_i^{\text{cand}}(0)$  from a semi-continuous distribution: We first draw  $n_1$  values from a Bernoulli distribution with probability  $\pi$  setting  $S_i^{\text{cand}}(0) = \bar{S}$  for treated units for which we obtain a success (a positive value). For treated units for which we obtain a failure, a missing value of  $S_i^{\text{cand}}(0)$  is then drawn from the Weibull distribution with parameters  $\alpha_S$  and  $\beta_S$ : Weibull( $\alpha_S, \beta_S$ ). For each  $i$  with  $Z_i = 1$ , we accept  $S_i^{\text{cand}}(0)$ , setting  $S_i^*(0) = S_i^{\text{cand}}(0)$ , with probability  $p_i = \min\{p_{S_i(0)}, 1\}$ , with

$$p_{S_i(0)} = \begin{cases} r_i & \text{if } S_i^*(0) = \bar{S}, S_i^{\text{cand}}(0) = \bar{S} \\ r_i \cdot \frac{\pi}{(1-\pi)f_{S(0)}(S_i^{\text{cand}}(0))} & \text{if } S_i^*(0) = \bar{S}, S_i^{\text{cand}}(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) \leq Y_i^*(0) \\ r_i \cdot \frac{(1-\pi)f_{S(0)}(S_i^*(0))}{\pi} & \text{if } S_i^*(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) = \bar{S} \\ r_i \cdot \frac{f_{S(0)}(S_i^*(0))}{f_{S(0)}(S_i^{\text{cand}}(0))} & \text{if } S_i^*(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) \leq Y_i^*(0) \\ 0 & \text{if } S_i^{\text{cand}}(0) \in \mathbb{R}_+, S_i^{\text{cand}}(0) > Y_i^*(0) \end{cases}$$

where  $f_{S(0)}(\cdot)$  is the density of the proposal Weibull distribution, Weibull( $\alpha_S, \beta_S$ ), and

$$r_i = \frac{P\{S_i^{\text{cand}}(0), | \boldsymbol{\theta}, D_i^{\text{obs}}, Y_i^*(0)\}}{P\{S_i^*(0), | \boldsymbol{\theta}, D_i^{\text{obs}}, Y_i^*(0)\}}.$$

3. Given the imputed complete data,

$$\boldsymbol{D} = \left[ \boldsymbol{Z}, \boldsymbol{C}, \tilde{\boldsymbol{S}}^{\text{obs}}, \mathbb{I}\{\boldsymbol{S}^{\text{obs}} \leq \boldsymbol{C}\}, \tilde{\boldsymbol{Y}}^{\text{obs}}, \mathbb{I}\{\boldsymbol{Y}^{\text{obs}} \leq \boldsymbol{C}\}, \boldsymbol{S}^*(0), \boldsymbol{Y}^*(0) \right],$$

we then draw for the following sub-vectors of  $\boldsymbol{\theta}$  in sequence, conditional on all others:  $\pi, \alpha_S, \beta_S, \bar{\alpha}_Y, \bar{\beta}_Y, \alpha_Y, \beta_Y, \bar{\nu}_Y, \bar{\gamma}_Y, \nu_Y, \gamma_Y, \lambda$ , using the procedure described in step 2. in Section “Details of Calculations:  $\kappa = 0$ .”

Table A.2: Summary statistics of the posterior distributions

Parameter	<i>Mean</i>	<i>sd</i>	<i>Percentiles</i>					$\hat{R}$
			2.5%	25%	50%	75%	97.5%	
$\pi$	0.38	0.06	0.28	0.34	0.38	0.42	0.50	1.000
$\alpha_Y$	1.56	0.11	1.35	1.48	1.55	1.63	1.79	1.000
$\beta_S$	-1.28	0.15	-1.55	-1.38	-1.28	-1.18	-0.98	1.001
$\bar{\alpha}_Y$	1.37	0.13	1.13	1.28	1.37	1.46	1.64	1.000
$\bar{\beta}_Y$	-1.09	0.21	-1.50	-1.24	-1.09	-0.95	-0.69	1.001
$\alpha_Y$	0.93	0.12	0.71	0.85	0.93	1.01	1.18	1.000
$\beta_Y$	-1.21	0.15	-1.51	-1.30	-1.20	-1.10	-0.93	1.000
$\bar{\nu}_Y$	1.12	0.10	0.92	1.05	1.12	1.19	1.33	1.000
$\bar{\gamma}_Y$	-1.79	0.27	-2.33	-1.97	-1.79	-1.61	-1.29	1.000
$\nu_Y$	1.16	0.09	0.97	1.09	1.16	1.22	1.35	1.000
$\gamma_Y$	-2.10	0.21	-2.54	-2.24	-2.09	-1.95	-1.70	1.000
$\lambda$	0.10	0.17	-0.21	-0.01	0.10	0.21	0.43	1.001

## I Application: Additional Results

### Convergence Checks

We use the potential scale-reduction statistic (Gelman and Rubin, 1992) to assess convergence of the MCMC algorithm; the potential scale reduction statistic takes on values around 1 for all the model parameters, showing no evidence against convergence (see Table A.2). Figure A.1 shows the trace plots, which exhibit up-and-down variation with no long-term trends or drift, showing further evidence that convergence has been reached. Finally, Figure A.2 shows the posterior distributions of the model parameters, which are generally well-shaped.

### Distributional Causal Effects for Switchers

Figure A.3 shows the posterior median of the distributional causal effects for switchers. The distributional causal effects are almost always positive, with an increasing trend over time for switchers who would switch to zidovudine early after the assignment. For switchers who would switch to zidovudine between 0.25 and 1.25 years after the assignment, the distributional causal effects are negative for early durations greater than the switching time, and become positive for later durations. The later the switching time, the longer the durations until which the distributional causal effects are negative. For instance, the distributional causal effects for patients who would switch to zidovudine 0.25 years after the assignment are negative, ranging between  $-0.021$  and  $-0.009$ , for few durations longer than 0.25 years (between 0.25 and about 0.3 years). The dis-

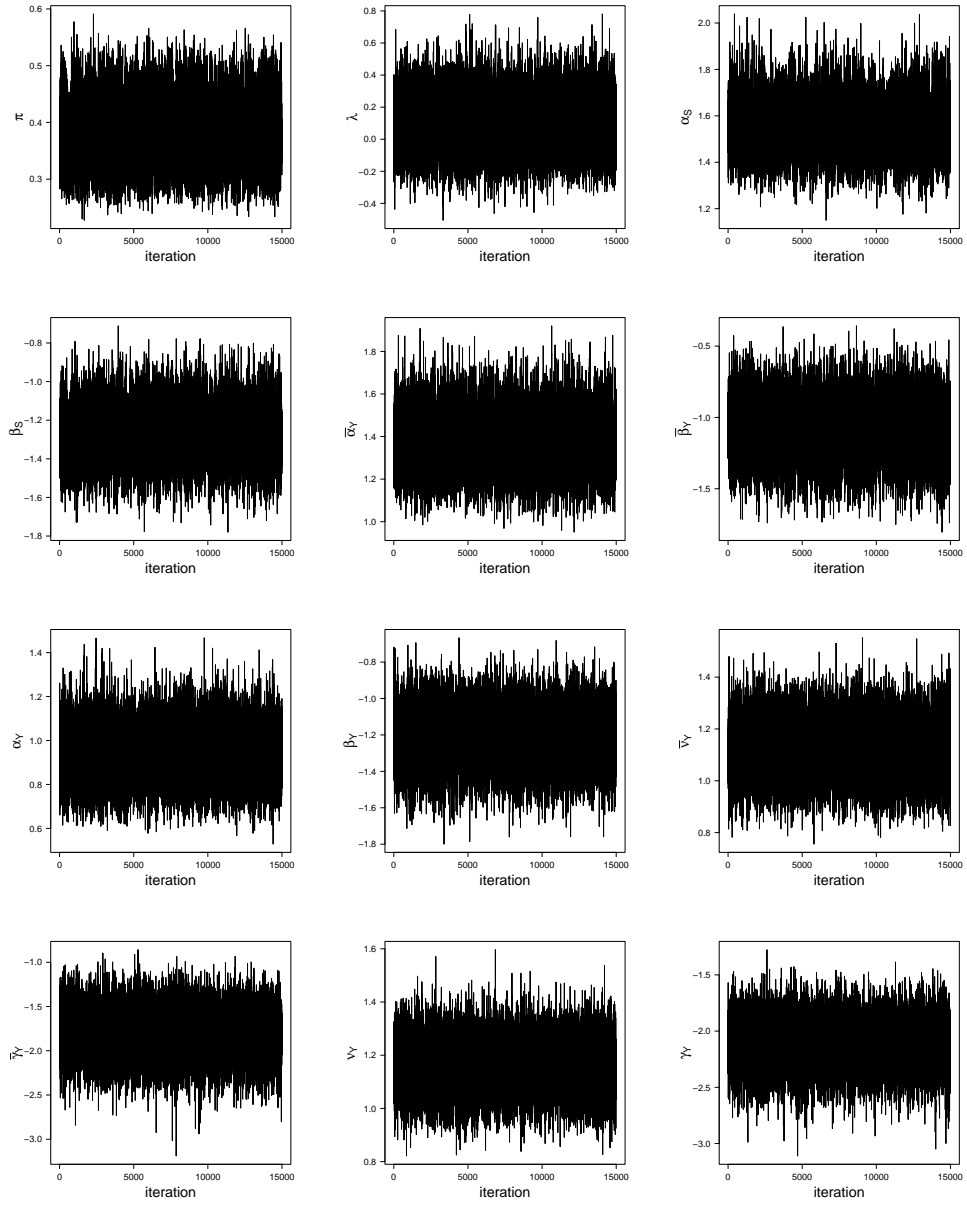


Figure A.1: Trace plots of the model parameters

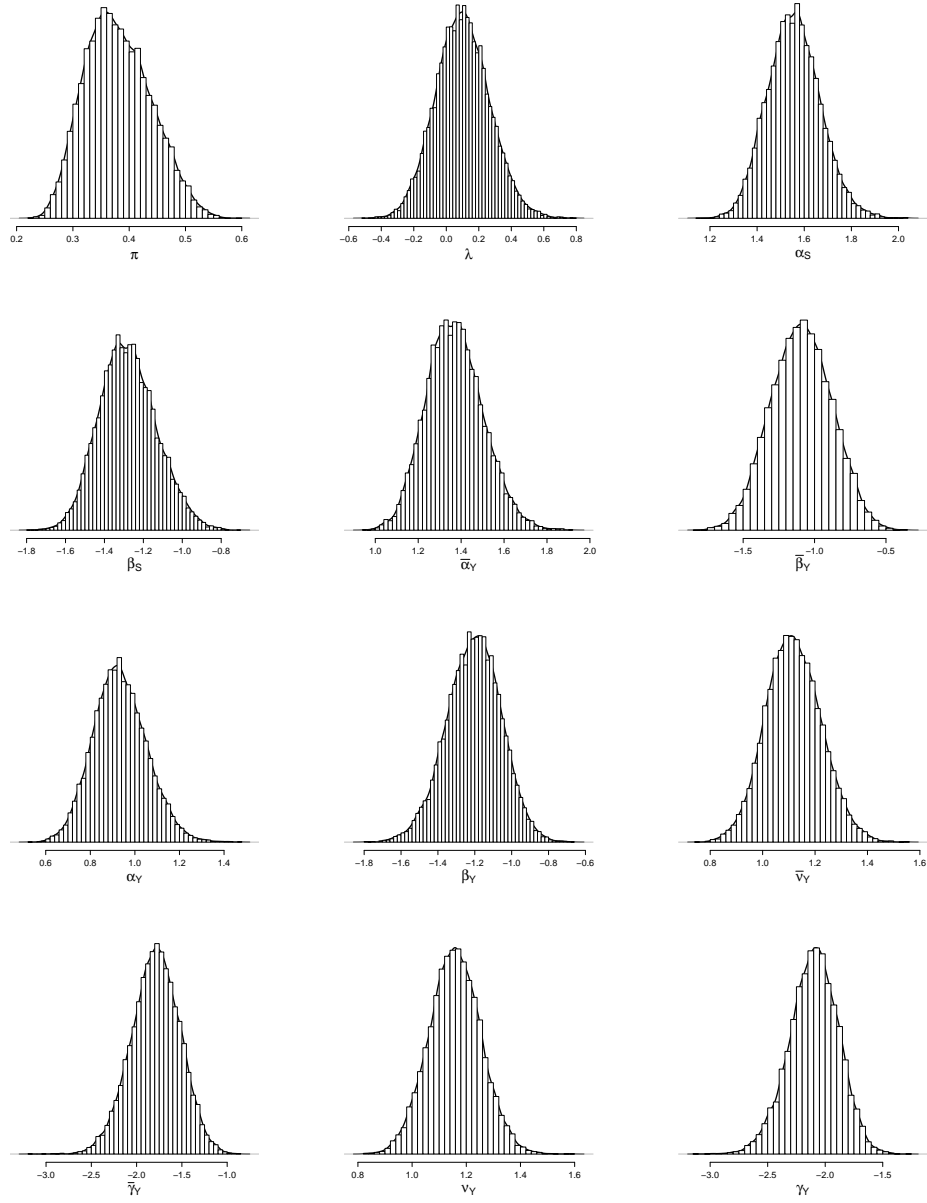


Figure A.2: Posterior density of the model parameters



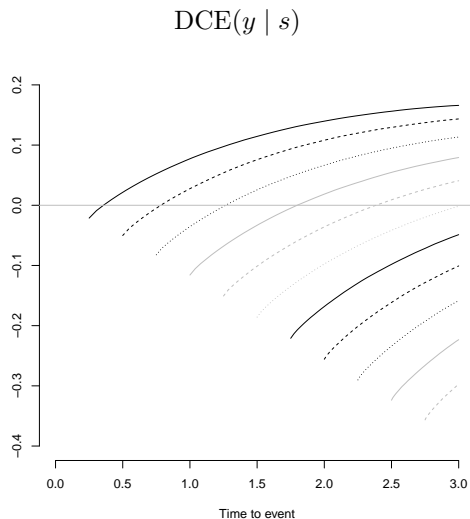


Figure A.3: Principal stratification analysis: Posterior median of the distributional causal effects for switchers,  $DCE(y | s)$ , for  $s = 0.25, 0.50, \dots, 2.50, 2.75$

tributional causal effects for patients who would switch to zidovudine 1.25 years after the assignment are negative, ranging between  $-0.151$  and  $-0.001$ , for durations between 1.25 and about 2.37 years. These results are, at least partially, driven by the natural constraint  $S_i(0) < Y_i(0)$ . Therefore, we need to interpret distributional causal effects for patients who would switch at a given time  $s$ , bearing in mind how such effects are defined. Indeed, a DCE is a comparison between the probability under assignment to immediate treatment with zidovudine that switchers at time  $s$  will survive beyond any specified time,  $y$ , and the probability under assignment to deferred treatment with zidovudine that those switchers will survive beyond  $y$ , given that they have survived beyond the time-to-switching,  $s$ . It is then sensible that some distributional causal effects are negative also for  $y \geq s$ , especially for long switching durations; in fact, immediate versus deferred treatment with zidovudine should have a very strong effect for making these distributional effects positive.

### Sensitivity Analysis to $\kappa$

We conduct a sensitivity analysis to  $\kappa$ , the partial association between  $Y_i(1)$  and  $Y_i(0)$  given the switching status,  $S_i(0)$ . We derive the posterior distribution of the causal estimands for  $\kappa = 0, 0.25, 0.5, 0.75, 1$ , using the same priors for other parameters as in Section 5.2 (see Web-Appendix H for details). Table A.3 and Figures A.5-A.7 present the results. Results display some sensitivity to  $\kappa$ .

In Table A.3 and Figure A.4, the posterior distributions of the average causal

effects and the distributional causal effects for non-switchers suggest that the evidence in favor of beneficial effects of immediate versus deferred treatment with zidovudine on survival time for never-switchers weakens when the assumption of independence between the potential survival outcomes,  $Y_i(0)$  and  $Y_i(1)$  (i.e.,  $\kappa = 0$ ) is relaxed allowing for values of  $\kappa$  greater than zero. For  $\kappa = 1$  (which implies monotonicity, i.e.,  $Y_i(1) \geq Y_i(0)$ ), we still find evidence that immediate versus deferred treatment with zidovudine increases survival time for non-switchers, both on average and over time. However, the posterior distributions of the average causal effect and the distributional causal effects for non-switchers are centered on smaller values and have smaller posterior variances than those obtained for  $\kappa = 0$ , leading to tighter 95% posterior credible intervals. It is also worth noting that the distributional causal effects,  $DCE(y | \mathbb{S})$ , show a different time trend for  $\kappa = 1$  than for  $\kappa = 0$ : For  $\kappa = 1$  they increase over time from 0 to 0.109 up to  $y = 1.25$  years and then start to decrease, although they are always positive with rather tight 95% posterior credible intervals including only positive values. For  $\kappa = 0.25$ , the posterior distributions of the average causal effect and the distributional causal effects for non-switchers are still centered on positive values. Still, they have a rather large posterior variability, leading to 95% posterior credible intervals that cover zero except for the 95% posterior credible intervals for distributional causal effects,  $DCE(y | \mathbb{S})$  for  $y \leq 1.25$ . For  $\kappa = 0.5, 0.75$ , the posterior medians of the average causal effects for non-switchers are very close to zero, and the 95% posterior credible intervals cover zero. Therefore, there is no evidence that immediate versus deferred treatment with zidovudine increases survival time for non-switchers on average. For non-switchers, we find positive and statistically significant, even if small, distributional causal effects for times to event  $y \leq 0.95$  and  $y \leq 1.15$ , respectively, for  $\kappa = 0.5$  and  $\kappa = 0.75$ . Then, distributional causal effects start to decrease, also reaching negative values for  $y \geq 2.30$  ( $\kappa = 0.5$ ) and  $y \geq 2.40$  ( $\kappa = 0.75$ ); however, they are statistically negligible with 95% posterior credible intervals always covering zero.

Figure A.5 shows that the estimates of the average causal effects for switchers are statistically negligible as those we obtained for  $\kappa = 0$  for  $\kappa = 0.25, 0.5, 0.75$ . Instead, we find evidence that immediate versus deferred treatment with zidovudine increases the average survival time for switchers irrespective of the time to switching for  $\kappa = 1$ , under which monotonicity  $Y_i(1) \geq Y_i(0)$  holds.

Figure A.6 compares the posterior medians of  $cDCE(y | s)$  for  $s = 0.25, 0.50, \dots, 2.50, 2.75$ . From Figure A.6, the posterior medians of  $cDCE(y | s)$  show a trend increase throughout the years at  $\kappa = 0$ , but have an asymmetrical inverted U-shape skewed to the right at  $\kappa \in \{0.25, 0.5, 0.75, 1\}$ , at least for switchers who would switch relatively soon.

Figure A.7 compares the posterior medians of  $DCE(y | s)$  for  $s = 0.25, 0.50, \dots, 2.50, 2.75$ . Note that at  $\kappa = 1$ ,  $DCE(y | s) = cDCE(y | s)$ . Two major patterns appear in the posterior medians of  $DCE(y | s)$ . First, the posterior medians are negative for some durations greater than the switching time both at  $\kappa = 0$  and  $\kappa \in \{0.25, 0.5, 0.75\}$ , but the posterior medians of  $DCE(y | s)$  at  $\kappa \in \{0.25, 0.5, 0.75\}$  turn to be positive at earlier durations. Second, the pos-

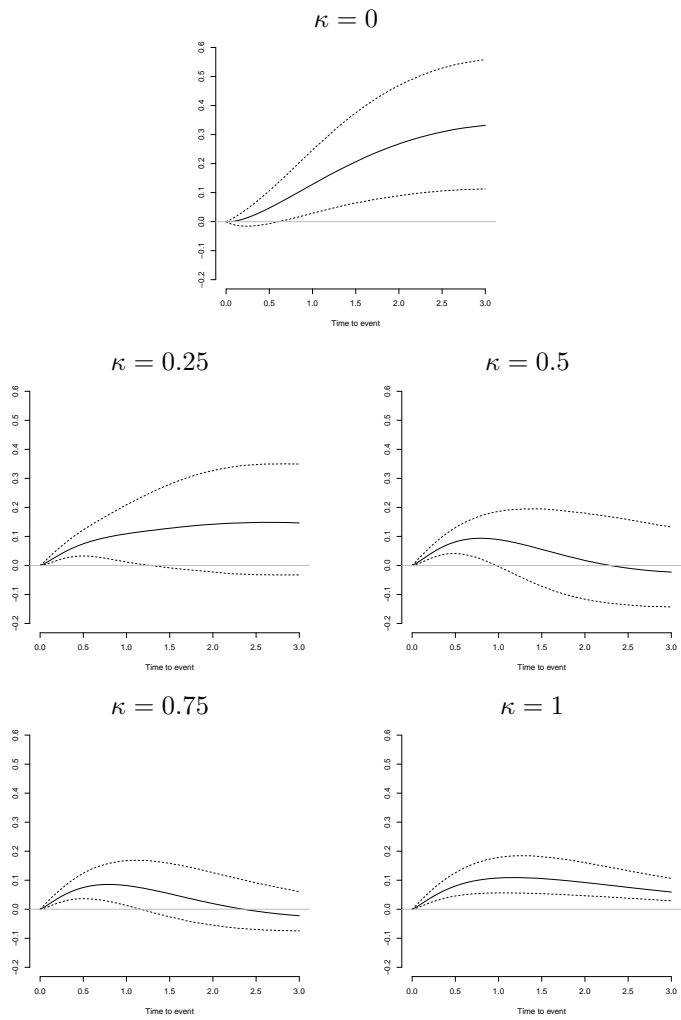


Figure A.4: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of distributional causal effects for never switchers,  $DCE(y | \bar{S})$ , for different values of  $\kappa$

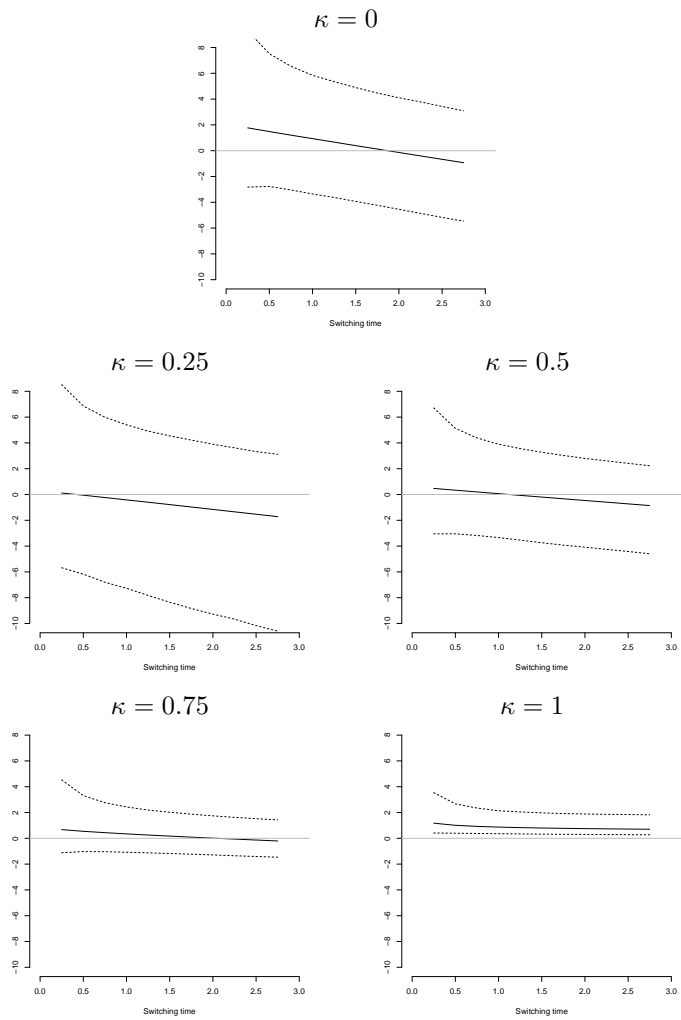


Figure A.5: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of average causal effects for switchers,  $ACE(s)$ ,  $s \in \mathbb{R}_+$  for different values of  $\kappa$

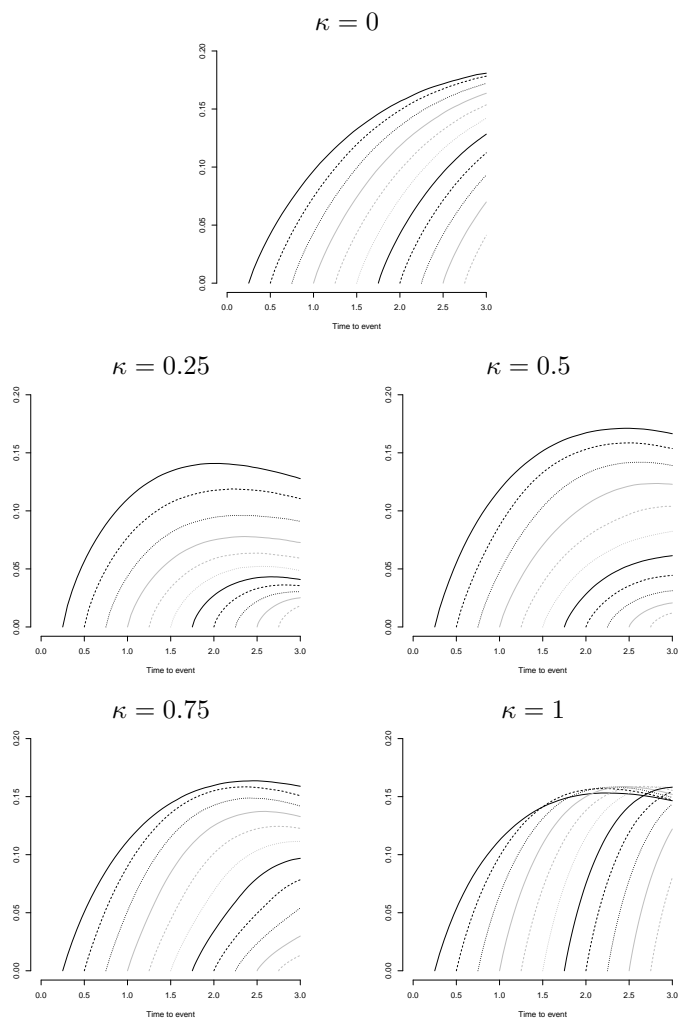


Figure A.6: Principal stratification analysis: Posterior median of conditional distributional causal effects for switchers,  $\text{cDCE}(y | s)$ , at time  $s = 0.25, 0.50, \dots, 2.50, 2.75$  for different values of  $\kappa$

Table A.3: Principal stratification analysis: Posterior median and 95% posterior credible interval for causal estimands for non-switchers for different values of  $\kappa$

$\kappa$	$\mathbb{E}[Y_i(0)   S_i(0) = \bar{S}]$	$\mathbb{E}[Y_i(1)   S_i(0) = \bar{S}]$	ACE( $\bar{S}$ )
$\kappa = 0$	2.05 (1.44; 2.99)	4.76 (2.80; 9.80)	2.66 (0.71; 7.73)
$\kappa = 0.25$	1.87 (1.38; 2.88)	2.78 (1.87; 4.58)	0.86 (-0.09; 2.49)
$\kappa = 0.50$	1.91 (1.37; 3.09)	1.98 (1.46; 2.78)	0.02 (-0.70; 0.72)
$\kappa = 0.75$	2.10 (1.48; 3.15)	2.11 (1.59; 2.94)	-0.01 (-0.36; 0.40)
$\kappa = 1$	2.06 (1.47; 2.96)	2.43 (1.81; 3.35)	0.36 (0.19; 0.59)

terior medians for switchers who would switch to zidovudine early after the assignment show an increasing trend over time at  $\kappa = 0$ , whereas those derived at  $\kappa \in \{0.25, 0.5, 0.75, 1\}$  follow an asymmetrical inverted U-shape skewed to the right.

### Sensitivity Analysis to the Prior Distribution for $\lambda$

Previous results are obtained using a weakly informative prior distribution for  $\lambda$ , namely,  $N(0, 10^4)$ . We assess the sensitivity of the results to the prior specification for  $\lambda$  by specifying three alternative priors. We consider two normal priors with smaller variances,  $N(0, 1)$  and  $N(0, 10)$ , and an improper prior uniformly over the whole real line. The hyperparameters of the prior distributions for the other model parameters are set to the same values as in Section H. We focus on the scenario with  $\kappa = 0$ . Table A.4 and Figures A.8-A.11 present the results, showing that inference is robust with respect to the prior specification for  $\lambda$ . We see that the posterior distribution of the causal estimands changes only slightly using different prior distributions for  $\lambda$ . Moreover, the posterior distribution of  $\lambda$  is robust to different prior specifications. The posterior mean of  $\lambda$  remains approximately 0.10, with a standard deviation of 0.17, irrespective of the prior specification. Although the 95% posterior credible intervals cover 0, the posterior probability that the parameter  $\lambda$  is positive ranges between 71.5% and 72.7% using different priors. Thus, there appears to be some evidence that the death hazard increases as the time of switching increases, suggesting that the residual lifetime after switching is shorter for patients who would switch later than for patients who would switch earlier.

### Sensitivity Analysis to the Parametric Assumption $\lambda_1 = \lambda_0$

We assess the sensitivity of the results to the parametric assumption  $\lambda_1 = \lambda_0$  by deriving the posterior distributions of the causal estimands of interest when we relax it.

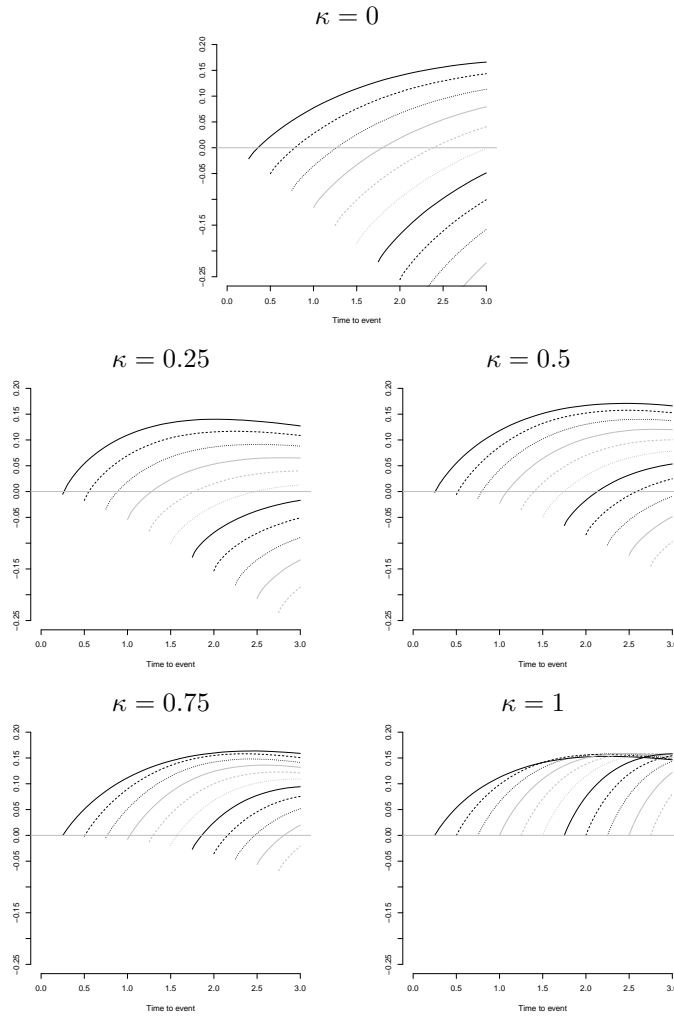


Figure A.7: Principal stratification analysis: Posterior median of distributional causal effects for switchers,  $DCE(y | s)$ , at time  $s = 0.25, 0.50, \dots, 2.50, 2.75$  for different values of  $\kappa$

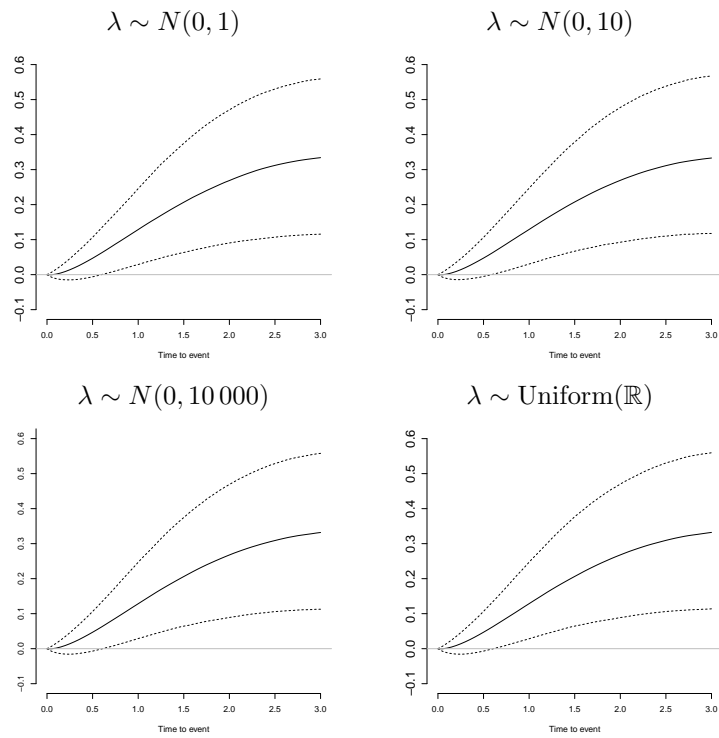


Figure A.8: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of distributional causal effects for never switchers,  $DCE(y | \bar{S})$ , for different prior distributions for  $\lambda$  with  $\kappa = 0$



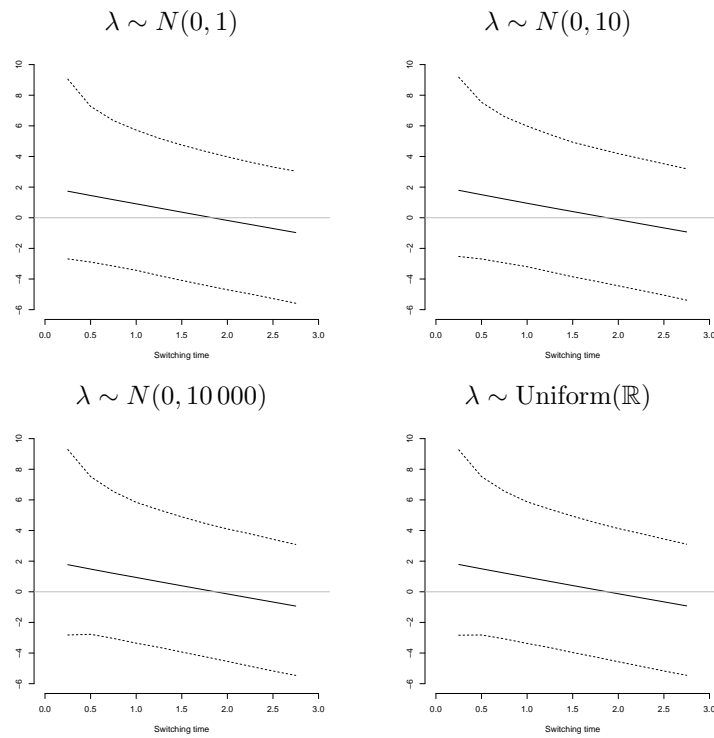


Figure A.9: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of average causal effects for switchers,  $ACE(s)$ ,  $s \in \mathbb{R}_+$ , for different prior distributions for  $\lambda$  with  $\kappa = 0$

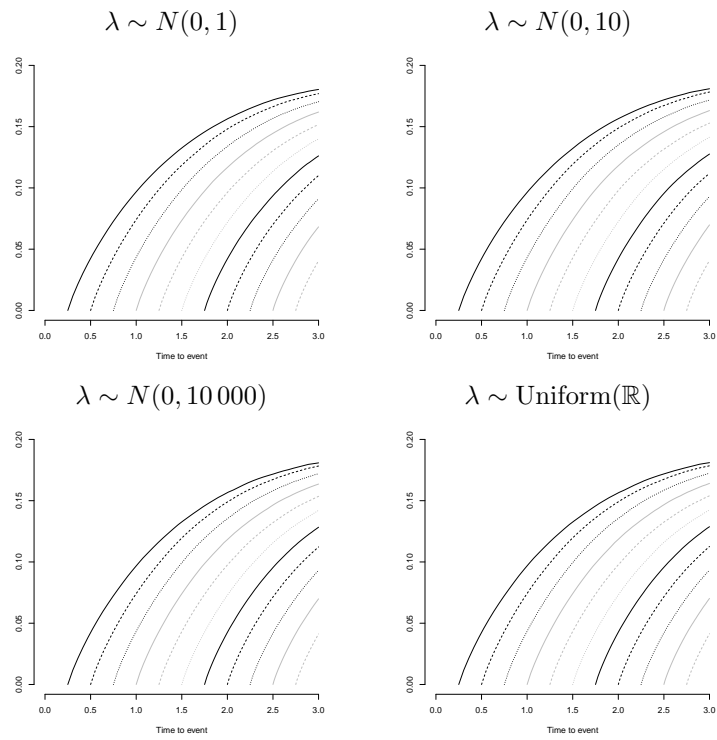


Figure A.10: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of conditional distributional causal effects for switchers at time  $s = 0.25, 0.50, \dots, 2.50, 2.75$ ,  $\text{cDCE}(y | s)$ , for different prior distributions for  $\lambda$  with  $\kappa = 0$

Table A.4: Principal stratification analysis: Summaries of posterior distributions of causal estimands for non-switchers for different prior distributions for  $\lambda$  ( $\kappa = 0$ )

Estimand	$\lambda \sim N(0, 1)$			$\lambda \sim N(0, 10)$		
	95% PCI			95% PCI		
	0.50	0.025	0.975	0.50	0.025	0.975
$\mathbb{E}[Y_i(0) \mid S_i(0) = \bar{S}]$	2.06	1.45	3.01	2.04	1.43	2.98
$\mathbb{E}[Y_i(1) \mid S_i(0) = \bar{S}]$	4.78	2.83	9.92	4.76	2.81	10.12
$\text{ACE}(\bar{S})$	2.68	0.72	7.79	2.66	0.72	8.02

Estimand	$\lambda \sim N(0, 10\,000)$			$\lambda \sim \text{Uniform}(\mathbb{R})$		
	95% PCI			95% PCI		
	0.50	0.025	0.975	0.50	0.025	0.975
$\mathbb{E}[Y_i(0) \mid S_i(0) = \bar{S}]$	2.05	1.44	2.99	2.04	1.44	3.00
$\mathbb{E}[Y_i(1) \mid S_i(0) = \bar{S}]$	4.76	2.80	9.80	4.75	2.81	9.80
$\text{ACE}(\bar{S})$	2.66	0.71	7.73	2.65	0.71	7.74

Table A.5 and Figure A.12 show the results for never-switchers and Figures A.13 and A.14 show the results for switchers.

Relaxing the parametric assumption  $\lambda_0 = \lambda_1$  does not affect the results for never-switchers (see Table A.5 and Figure A.12) and slightly changes the results for switchers, by leading to posterior distributions of the causal effects for switchers with a larger posterior variability (see Figures A.13 and A.14). The increased uncertainty in the causal estimands for switchers makes it more difficult to draw firm causal conclusions for them, especially for early switchers. For instance, for early switchers who would switch earlier than 1 year, we find positive and statistically significant distributional causal effects under the model with  $\lambda_0 = \lambda_1$  and statistically negligible distributional causal effects under the model with  $\lambda_0 \neq \lambda_1$  (see the graphs in the first row of Figure A.14).

## Posterior Predictive Checks

We evaluate the influence of the parametric assumptions using posterior predictive checks (e.g., Guttman, 1967; Rubin, 1984), by computing a Bayesian posterior predictive  $p$ -value ( $PPPV$ ) for various discrepancy measures (Meng, 1994; Gelman et al., 1996; Forastiere et al., 2018). A discrepancy measure is a known, real-valued function of the nuisance parameters, the imputed switching status, and the observed data. The corresponding Bayesian  $PPPV$  is defined as the integral average over the joint posterior distribution of the missing switching statuses and model parameters of the probability that the discrepancy measure calculated for replicated data is more extreme than the value for observed data.

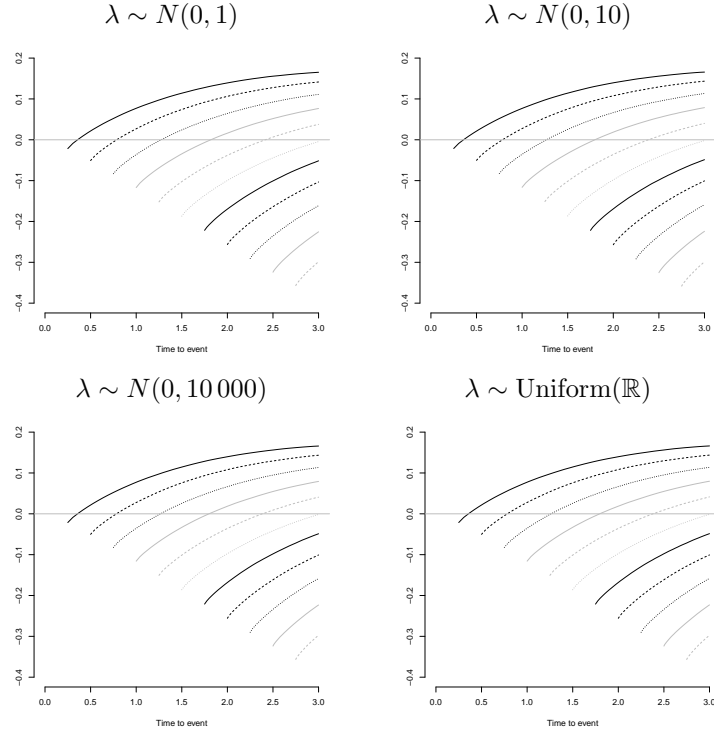


Figure A.11: Principal stratification analysis: Posterior median of the distributional causal effects for switchers, at time  $s = 0.25, 0.50, \dots, 2.50, 2.75$ ,  $DCE(y | s)$ , for different prior distributions for  $\lambda$  ( $\kappa = 0$ )

Replicated data are drawn from the posterior predictive distribution of the hypothesized model. A *PPPV* is a measure of model misfit, with the model including both the prior distribution and the likelihood. Extreme values (close to 0 or 1) of a *PPPV* would indicate that the model cannot adequately preserve features of the data reflected in the discrepancy measure.

We conduct model checking under conditionally independent potential survival outcomes with  $\kappa = 0$ . Let  $r$  be the study type indicator:  $r = \text{obs}$  for the observed study and  $r = \text{rep}$  for a replicated study. We generate the replicated data using the observed value of the assignment variable, entry, and censoring time, that is, we set  $Z_i^{\text{rep}} = Z_i^{\text{obs}} = Z_i$  and  $C_i^{\text{rep}} = C_i$  for all  $i = 1, \dots, n$ .

We measure the goodness-of-fit of the posited model using three types of posterior predictive discrepancy measures:

1. *BIC posterior predictive discrepancy measure.*

$$BIC^r = -2(\mathcal{L}\{\boldsymbol{\theta} | \kappa = 0, \mathbf{D}^r\} + \#\{\boldsymbol{\theta} \setminus \kappa\} \cdot \log(n)),$$

Table A.5: Principal stratification analysis: Summaries of posterior distributions of causal estimands for non-switchers under the model with  $\lambda_0 = \lambda_1$  and under the model with  $\lambda_0 \neq \lambda_1$

Estimand	Under the model with					
	$\lambda_0 = \lambda_1$			$\lambda_0 \neq \lambda_1$		
	95% PCI			95% PCI		
	0.50	0.025	0.975	0.50	0.025	0.975
$E[Y_i(0) \mid S_i(0) = \bar{S}]$	2.05	1.44	2.99	1.95	1.41	2.91
$E[Y_i(1) \mid S_i(0) = \bar{S}]$	4.76	2.80	9.80	4.39	2.63	9.18
$ACE(\bar{S})$	2.66	0.71	7.73	2.39	0.63	7.08

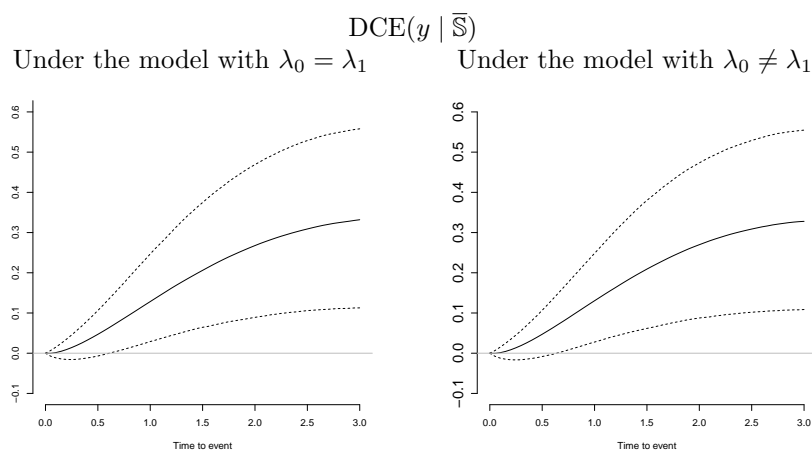


Figure A.12: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of the distributional causal effects for non-switchers under the model with  $\lambda_0 = \lambda_1$  (left) and under the model with  $\lambda_0 \neq \lambda_1$  (right)

ACE( $s$ ),  $s \in \mathbb{R}_+$

Under the model with  $\lambda_0 = \lambda_1$ 
Under the model with  $\lambda_0 \neq \lambda_1$

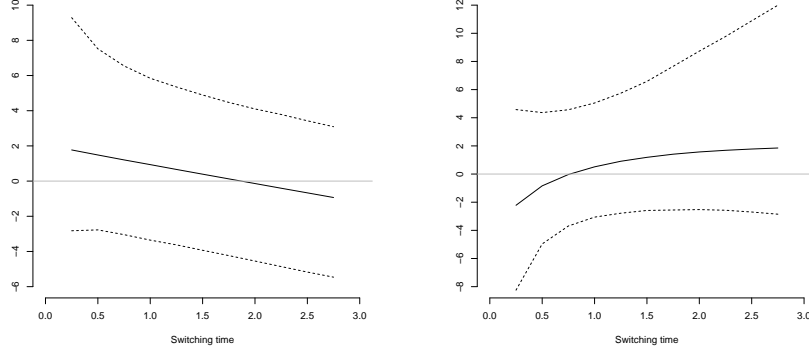


Figure A.13: Principal stratification analysis: Posterior medians (solid lines) and 95% posterior credible intervals (dashed lines) of average causal effects for switchers, under the model with  $\lambda_0 = \lambda_1$  (left) and under the model with  $\lambda_0 \neq \lambda_1$  (right)

where  $\mathbf{D}^r = [\mathbf{Z}, \mathbf{C}, \tilde{\mathbf{S}}^r, \mathbb{I}\{\mathbf{S}^r \leq \mathbf{C}\}, \tilde{\mathbf{Y}}^r, \mathbb{I}\{\mathbf{Y}^r \leq \mathbf{C}\}, \mathbf{S}^{*,r}(0)]$  is the  $n \times 7$  matrix of the complete switching status data,  $\mathcal{L}\{\boldsymbol{\theta} \mid \kappa = 0, \mathbf{D}^r\}$  is the complete switching status-data likelihood function for  $\kappa = 0$ , and  $\#\{\boldsymbol{\theta} \setminus \kappa\}$  is the number of parameters excluding  $\kappa$  ( $\#\{\boldsymbol{\theta} \setminus \kappa\} = 12$  in our study).

2. *Deviance posterior predictive discrepancy.* The deviance is defined as the sum of the deviance residuals for the Weibull model. We calculate the deviance posterior predictive discrepancy measure separately for the survival time and the switching time under control for switchers. For  $r = \text{obs, rep}$ ,

$$\begin{aligned} \text{Deviance}_{\tilde{Y}}^r(\mathbf{D}, \boldsymbol{\theta}) = & \\ -2 \sum_{i: Z_i=0,1} & \left\{ \sum_{i: Z_i=z, S_i^{*,r}(0)=\bar{S}} \left[ M_{\tilde{Y}(z)}^{\bar{S}}(\tilde{Y}_i^r) + \mathbb{I}\{Y_i^r \leq C_i\} \log \left( \mathbb{I}\{Y_i^r \leq C_i\} - M_{\tilde{Y}(z)}^{\bar{S}}(\tilde{Y}_i^r) \right) \right] + \right. \\ & \left. \sum_{i: Z_i=z, S_i^{*,r}(0) \in \mathbb{R}_+} \left[ M_{Y(z)}(\tilde{Y}_i^r \mid S_i^{*,r}(0)) + \mathbb{I}\{Y_i^r \leq C_i\} \log \left( \mathbb{I}\{Y_i^r \leq C_i\} - M_{Y(z)}(\tilde{Y}_i^r \mid S_i^{*,r}(0)) \right) \right] \right\}, \end{aligned}$$

where  $M_{\tilde{Y}(z)}^{\bar{S}}(\cdot)$  and  $M_{Y(z)}(\cdot \mid S_i(0))$  are the martingale residuals for non-switchers and switchers, respectively:  $M_{\tilde{Y}(z)}^{\bar{S}}(\tilde{Y}_i^r) = \mathbb{I}\{Y_i^r \leq C_i\} - \Lambda_{\tilde{Y}(z)}^{\bar{S}}(\tilde{Y}_i^r)$  if  $S_i(0) = \bar{S}$ , and  $M_{Y(z)}(\tilde{Y}_i^r \mid S_i(0)) = \mathbb{I}\{Y_i^r \leq C_i\} - \Lambda_{Y(z)}(\tilde{Y}_i^r \mid S_i(0))$  if  $S_i(0) \in \mathbb{R}_+$ , with  $\Lambda_{\tilde{Y}(z)}^{\bar{S}}(\cdot)$  and  $\Lambda_{Y(z)}(\cdot \mid S_i(0))$  denoting the cumulative hazards for the Weibull model (Therneau et al., 1990). Similarly,

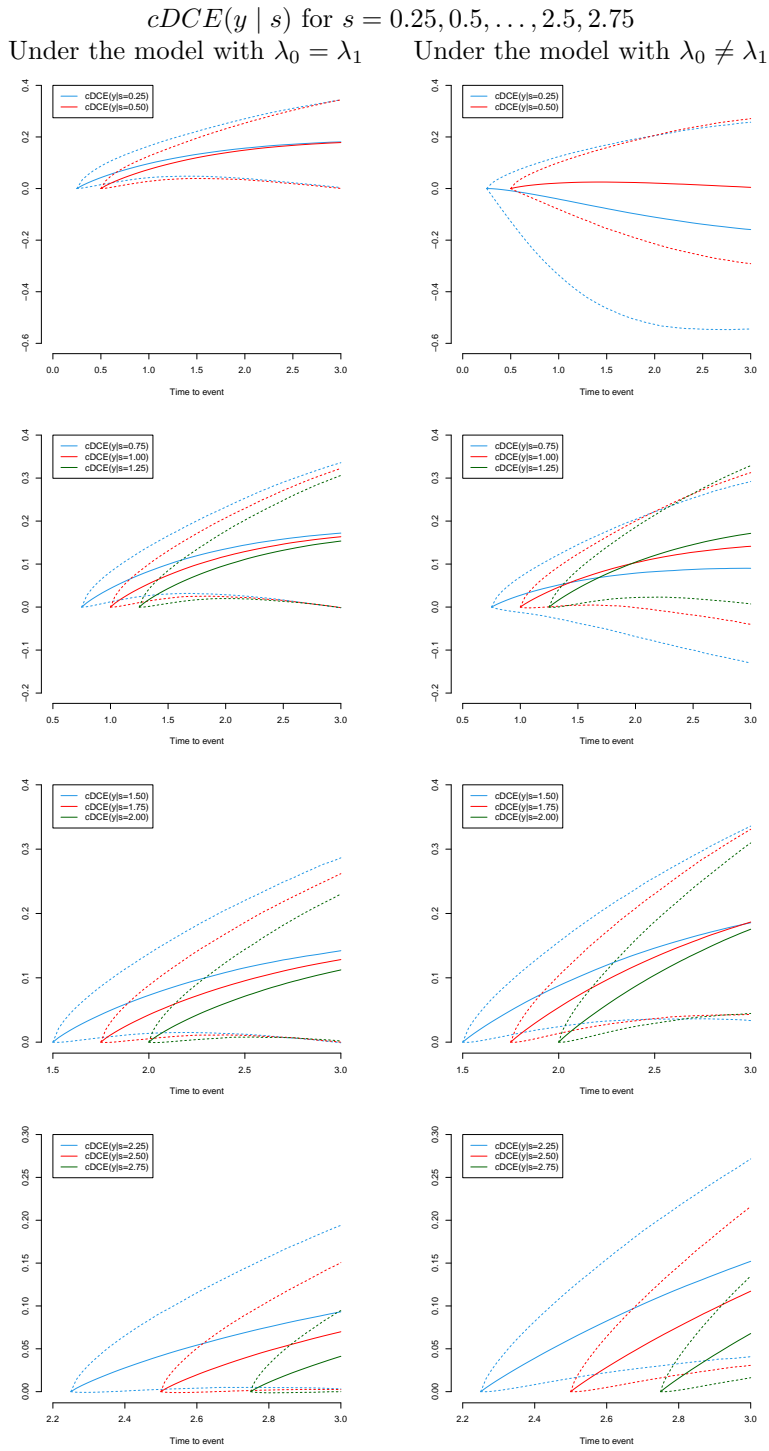


Figure A.14: Principal stratification analysis: Posterior medians (solid lines) and 95% posterior credible intervals (dashed lines) of the conditional distributional causal effects for switchers under the model with  $\lambda_0 = \lambda_1$  (first column) and under the model with  $\lambda_0 \neq \lambda_1$  (second column).

we define

$$\begin{aligned} \text{Deviance}_S^r(\mathbf{D}, \boldsymbol{\theta}) = & \\ -2 \sum_{i: Z_i=0, S_i^{*,r}(0) \in \mathbb{R}_+} & M_{S(0)}(\tilde{S}_i^r) + \mathbb{I}\{S_i^{*,r}(0) \leq C_i\} \log \left( \mathbb{I}\{S_i^{*,r}(0) \leq C_i\} - M_{S(0)}(\tilde{S}_i^r) \right). \end{aligned}$$

3. *Kaplan–Meier posterior predictive discrepancy.* We calculate Kaplan–Meier estimates of the survival curves for the time-to-death/disease progression, separately for non-switchers and switchers, and for the time-to-switching under control for switchers.

For data from study type  $r$ ,  $r = \text{obs}, \text{rep}$ , for each time-point  $t$ , let  $d_Y^r(t \mid S_i^{*,r}(0) \in \mathcal{A})$  be the number of events (deaths or disease progressions) at time  $t$  among patients with  $S_i^{*,r}(0) \in \mathcal{A}$ ,  $\mathcal{A} = \{\bar{\mathbb{S}}\}$  (non-switchers) and  $\mathcal{A} = \mathbb{R}_+$  (switchers), and let  $d_S^r(t \mid Z_i = 0, S_i^{*,r}(0) \in \mathbb{R}_+)$  be the number of switchers assigned to the control treatment who switch at time  $t$ . Let  $R_Y^r(t \mid S_i^{*,r}(0) \in \mathcal{A})$  denote the number of subjects with  $S_i^{*,r}(0) \in \mathcal{A}$ ,  $\mathcal{A} = \{\bar{\mathbb{S}}\}, \mathbb{R}_+$ , at risk of death or disease progression at time  $t$ , and let  $R_S^r(t \mid Z_i = 0, S_i^{*,r}(0) \in \mathbb{R}_+)$  denote the number of switchers assigned to the control treatment at risk of switching at time  $t$ . We define

$$KM_{\mathcal{A}}^r(t; \mathbf{D}, \boldsymbol{\theta}) = \prod_{i: S_i^{*,r}(0) \in \mathcal{A}} \left[ 1 - \frac{d_Y^r(t \mid S_i^{*,r}(0) \in \mathcal{A})}{R_Y^r(t \mid S_i^{*,r}(0) \in \mathcal{A})} \right] \quad \mathcal{A} = \{\bar{\mathbb{S}}\}, \mathbb{R}_+;$$

and

$$KM^r(t; \mathbf{D}, \boldsymbol{\theta}) = \prod_{i: Z_i=0, S_i^{*,r}(0) \in \mathbb{R}_+} \left[ 1 - \frac{d_S^r(t \mid Z_i = 0, S_i^{*,r}(0) \in \mathbb{R}_+)}{R_S^r(t \mid Z_i = 0, S_i^{*,r}(0) \in \mathbb{R}_+)} \right].$$

Following [Barnard et al. \(2003\)](#), we then consider posterior predictive discrepancy measures aimed to assess the ability of the model to preserve features in the outcome distributions of non-switchers and switchers that we think can be very influential in estimating the average and distributional causal effects.

Define the following subsets of units in the study of type  $r$  ( $r = \text{obs}, \text{rep}$ ):

$$\mathcal{I}_{\mathcal{A},z}^r = \{i : \mathbb{I}\{Y_i^r \leq C_i\} \mathbb{I}\{S_i^{*,r}(0) \in \mathcal{A}\} \mathbb{I}\{Z_i = z\} = 1\}$$

for  $\mathcal{A} = \{\bar{\mathbb{S}}\}$  and  $\mathcal{A} = \mathbb{R}_+$ , and  $z = 0, 1$ ; and

$$\mathcal{I}^r = \{i : \mathbb{I}\{S_i^{*,r}(0) \leq C_i\} \mathbb{I}\{S_i^{*,r}(0) \in \mathbb{R}_+\} \mathbb{I}\{Z_i = 0\} = 1\}$$

Let  $\bar{Y}_{\mathcal{A},z}^r$  and  $s_{Y,\mathcal{A},z}^{2,r}$  be the mean and the variance of the survival outcome,  $Y_i$ , for units belonging to  $\mathcal{I}_{\mathcal{A},z}^r$ , for which we observe  $Y_i(z)$  in study  $r$ . Similarly, let  $\bar{S}^r$  and  $s_S^{2,r}$  the mean and the variance of the switching time,  $S_i^*(0)$ , for units



Table A.6: Bayesian Posterior Predictive  $p$ -values

Variable	Deviance	Signal	Noise	Signal to noise
Survival time	0.917			
<i>Non-Switchers</i>		0.258	0.595	0.251
<i>Switchers</i>		0.621	0.773	0.542
Switching time	0.485	0.378	0.343	0.549

*PPP*V for BIC : 0.497

belonging to  $\mathcal{I}^r$ , for which  $S_i^*(0)$  is observed in study  $r$ . Then,

$$\begin{aligned}
 \text{Signal}_{\mathcal{A}}^r(\mathbf{D}, \boldsymbol{\theta}) &= \left| \bar{Y}_{\mathcal{A},1}^r - \bar{Y}_{\mathcal{A},0}^r \right| & \text{Signal}^r(\mathbf{D}, \boldsymbol{\theta}) &= \bar{S}^r \\
 \text{Noise}_{\mathcal{A}}^r(\mathbf{D}, \boldsymbol{\theta}) &= \sqrt{\frac{s_{Y,\mathcal{A},0}^{2,r}}{\#\mathcal{I}_{\mathcal{A},0}^r} + \frac{s_{Y,\mathcal{A},1}^{2,r}}{\#\mathcal{I}_{\mathcal{A},1}^r}} & \text{Noise}^r(\mathbf{D}, \boldsymbol{\theta}) &= \sqrt{\frac{s_S^{2,r}}{\#\mathcal{I}^r}} \\
 \text{Ratio}_{\mathcal{A}}^r(\mathbf{D}, \boldsymbol{\theta}) &= \frac{\text{Signal}_{\mathcal{A}}^r(\mathbf{D}, \boldsymbol{\theta})}{\text{Noise}_{\mathcal{A}}^r(\mathbf{D}, \boldsymbol{\theta})} & \text{Ratio}^r(\mathbf{D}, \boldsymbol{\theta}) &= \frac{\text{Signal}^r(\mathbf{D}, \boldsymbol{\theta})}{\text{Noise}^r(\mathbf{D}, \boldsymbol{\theta})}
 \end{aligned}$$

where  $\#\mathcal{I}_{\mathcal{A},z}^r = \sum_{i=1}^n \mathbb{I}\{i \in \mathcal{I}_{\mathcal{A},z}^r\}$  and  $\#\mathcal{I}^r = \sum_{i=1}^n \mathbb{I}\{i \in \mathcal{I}^r\}$  are the number of units in the  $r$  data belonging to the  $\mathcal{I}_{\mathcal{A},z}^r$  and  $\mathcal{I}^r$  group, respectively.

It is worth noting that these measures are not treatment effects, but they provide information on whether the model can preserve broad features of signal, noise, and signal-to-noise ratio in the survival time distributions for non-switchers and switchers and in the switching time distribution for switchers assigned to the control arm.

Table A.6 shows the Bayesian *PPP*Vs. The *PPP*V for the BIC is 0.497, and the *PPP*Vs for the deviance posterior predictive discrepancy measures are 0.485 for the switching time and 0.917 for the survival time, suggesting that our model fits the data pretty well. The *PPP*Vs for the Kaplan-Meier posterior predictive discrepancy measures are also sufficiently far away from 0 and 1 for all time points  $t$ ; the only exceptions are the Kaplan-Meier posterior predictive discrepancy for the time-to-switching under control for switchers for times shorter than 0.06 (approximately 22 days) and times between 0.17 and 0.54 (approximately between 2 and 6.5 months) for the time-to-death/disease progression for switchers.

It is worth noting that, in the observed data, no patient assigned to the control treatment is observed to switch to the active treatment within 22 days, and only 28 patients are observed to either die or experience a progression of the disease between 2 and 6.5 months. Results provide no special evidence for specific influences of the model too. The estimated Bayesian *PPP*Vs for the

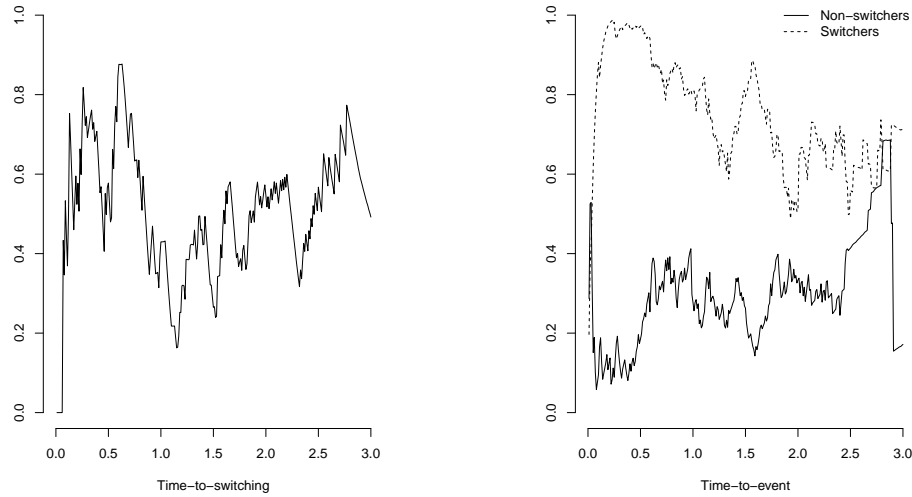


Figure A.15: Bayesian posterior predictive  $p$ -values for Kaplan-Meier posterior predictive discrepancy measures ( $t = 0.01, 0.02, \dots, 2.00, 3.00$ )

signal, noise, and signal-to-noise ratio posterior predictive discrepancy measures range between 0.251 and 0.773, suggesting that our model successfully replicates the corresponding measure of location, dispersion, and their relative magnitude.

## References

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). “Identification of causal effects using instrumental variables.” *Journal of the American Statistical Association*, 91: 444–455. [3](#), [4](#), [38](#)
- Barnard, J., Frangakis, C. E., Hill, J. L., and Rubin, D. B. (2003). “Principal Stratification Approach to Broken Randomized Experiments.” *Journal of the American Statistical Association*, 98: 299–323. [72](#)
- Bartolucci, F. and Grilli, L. (2011). “Modeling Partial Compliance Through Copulas in a Principal Stratification Framework.” *Journal of the American Statistical Association*, 106(494): 469–479. [38](#), [39](#)
- Chen, Q., Zeng, D., Ibrahim, J. G., Akacha, M., and Schmidli, H. (2013). “Estimating time-varying effects for overdispersed recurrent events data with treatment switching.” *Biometrika*, 100(2): 339–354. [2](#), [34](#)

- Comment, L., Mealli, F., Haneuse, S., and Zigler, C. (2019). “Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks.” *arXiv preprint arXiv:1902.09304*. 4, 37
- Concorde Coordinating Committee (1994). “Concorde: MRC / ANRS randomised double blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection.” *Lancet*, 343: 871–881. 2, 4
- De Finetti, B. (1937). “La prévision: ses lois logiques, ses sources subjectives.” *Annales de l’institut Henri Poincaré*, 7: 1–68. 40
- Ding, P. and Li, F. (2018). “Causal inference: A missing data perspective.” *Statistical Science*, 33(2): 214–237. 14
- Ding, P. and Lu, J. (2017). “Principal stratification analysis using principal scores.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3): 757–777. 37
- Feller, A., Grindal, T., Miratrix, L., and Page, L. C. (2016). “Compared to what? Variation in the impacts of early childhood education by alternative care type.” *The Annals of Applied Statistics*, 10(3): 1245–1285. 37
- Fisher, L. D. and Lin, D. Y. (1999). “Time-dependent covariates in the Cox proportional-hazards regression model.” *Annual Review of Public Health*, 20: 145–157. 33
- Forastiere, L., Mealli, F., and Miratrix, L. (2018). “Posterior Predictive  $p$  - Values with Fisher Randomization Tests in Noncompliance Settings: Test Statistics vs Discrepancy Measures.” *Bayesian Analysis*. 67
- Frangakis, C. E. and Rubin, D. B. (1999). “Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes.” *Biometrika*, 86(2): 365–379. 38
- (2002). “Principal stratification in causal inference.” *Biometrics*, 58: 191–199. 3, 8, 10, 32
- Gelman, A. E., Meng, X.-L., and Stern, H. S. (1996). “Posterior predictive assessment of model fitness via realized discrepancies (with discussion).” *Statistica Sinica*, 6: 733–807. 67
- Gelman, A. E. and Rubin, D. R. (1992). “Inference from Iterative Simulation Using Multiple Sequences (with discussion).” *Statistica Science*, 7: 457–472. 54
- Geskus, R. B. (2016). *Data analysis with competing risks and intermediate states*. CRC Press Boca Raton. 35
- Gilbert, P. B. and Hudgens, M. G. (2008). “Evaluating candidate principal surrogate endpoints.” *Biometrics*, 64: 1146–1154. 39

- Gustafson, P. (2010). “Bayesian inference for partially identified models.” *International Journal of Biostatistics*, 2. [14](#), [17](#)
- Guttman, I. (1967). “The use of the concept of a future observation in goodness-of-fit problems.” *Journal of the Royal Statistical Society B*, 29(1): 83–100. [67](#)
- Hauschild, A., Grob, J.-J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., Rutkowski, P., Blank, C. U., Miller Jr, W. H., Kaempgen, E., et al. (2012). “Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial.” *The Lancet*, 380(9839): 358–365. [2](#), [22](#)
- Hernán, M. and Robins, J. (2006). “Instruments for causal inference: An epidemiologist’s dream?” 17: 360–372. [32](#)
- Hernan, M. A., Brumback, B., and Robins, J. M. (2000). “Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.” *Epidemiology*, 1: 561–570. [33](#)
- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer. [15](#)
- ICH (2019). “Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, E9(R1).”  
URL [https://database.ich.org/sites/default/files/E9-R1\\_Step4\\_Guideline\\_2019\\_1203.pdf](https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf) [3](#)
- Imbens, G. W. and Rubin, D. B. (1997). “Bayesian inference for causal effects in randomized experiments with noncompliance.” *The Annals of Statistics*, 25: 305–327. [14](#)
- Jin, H. and Rubin, D. B. (2008). “Principal stratification for causal inference with extended partial compliance.” *Journal of the American Statistical Association*, 103(481): 101–111. [14](#), [38](#), [39](#)
- (2009). “Public schools versus private schools: Causal inference with partial compliance.” *Journal of Educational and Behavioral Statistics*, 34(1): 24–45. [14](#), [39](#)
- Jo, B. and Stuart, E. A. (2009). “On the use of propensity scores in principal causal effect estimation.” *Statistics in Medicine*, 28: 2857–2875. [37](#)
- Ju, C. and Geng, Z. (2010). “Criteria for surrogate end points based on causal distributions.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72: 129–142. [8](#)
- Kim, C., Daniels, M. J., Hogan, J. W., Choirat, C., and Zigler, C. M. (2017). “Bayesian methods for multiple mediators: Principal stratification and causal mediation analysis of power plant emission controls.” *Working Paper*. [14](#), [15](#), [38](#), [39](#)

- Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., et al. (2015). “Combined nivolumab and ipilimumab or monotherapy in untreated melanoma.” *New England journal of medicine*, 373(1): 23–34. 2, 22
- Latimer, N. R., Abrams, K. R., Amonkar, M. M., Stapelkamp, C., and Swann, R. S. (2015). “Adjusting for the confounding effects of treatment switching—the BREAK-3 trial: dabrafenib versus dacarbazine.” *The Oncologist*, 20(7): 798–805. 22
- Lindley, D. V. (1972). *Bayesian Statistics: A review*. SIAM. 14
- Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020). “Causal inference and estimands in clinical trials.” *Statistics in Biopharmaceutical Research*, 12(1): 54–67. 25
- Ma, Y., Roy, J., and Marcus, B. (2011). “Causal models for randomized trials with two active treatments and continuous compliance.” *Statistics in Medicine*, 30(19): 2349–2362. 38, 39
- Mattei, A., Forastiere, L., and Mealli, F. (2023). “Assessing Principal Causal Effects Using Principal Score Methods.” In Zubizarreta, J., Stuart, E. A., Small, D. S., and R., R. P. (eds.), *Handbook of Matching and Weighting Adjustments for Causal Inference*, chapter 17, 313–348. Chapman and Hall/CRC. 37
- Mattei, A. and Mealli, F. (2007). “Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination.” *Biometrics*, 63(2): 437–446. 14
- Mattei, A., Mealli, F., and Ding, P. (2020). “Assessing causal effects in the presence of treatment switching through principal stratification.” *arXiv preprint arXiv:2002.11989v1*. 4
- Mealli, F. and Pudney, S. (2003). *Applying heterogeneous transition models in labour economics: the role of youth training in labour market transitions*, Chapter 16. Wiley. 15
- Meng, X. L. (1994). “Posterior predictive p-values.” *Annals of Statistics*, 22: 1142–1160. 67
- Morden, J. P., Lambert, N., Paul C. and Latimer, Abrams, K. R., and Wailoo, A. J. (2011). “Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study.” *Medical Research Methodology*, 1(4). 32
- Nevo, D. and Gorfine, M. (2022). “Causal inference for semi-competing risks data.” *Biostatistics*, 23(4): 1115–1132. 37, 38
- Pearl, J. (2001). “Direct and indirect effects.” In Breese, J. S. and Koller, D. (eds.), *17th Conference on Uncertainty in Artificial Intelligence*, 411–420. Morgan Kaufmann. 35

- Robins, J. M. (1989). *The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies*, 113–159. Wiley. [32](#)
- (1994). “Correcting for non-compliance in randomized trials using structural nested mean models.” *Communications in Statistics-Theory and methods*, 23(8): 2379–2412. [2](#)
- Robins, J. M. and Finkelstein, D. M. (2000). “Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests.” *Biometrics*, 56(3): 779–788. [33](#)
- Robins, J. M. and Greenland, S. (1992). “Identifiability and exchangeability for direct and indirect effects.” *Epidemiology*, 143–155. [35](#)
- Robins, J. M. and Tsiatis, A. A. (1991). “Correcting for non-compliance in randomized trials using rank preserving structural failure time models.” *Communications in Statistics-Theory and Methods*, 20(8): 2609–2631. [2](#), [33](#)
- Rubin, D. B. (1984). “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.” *The Annals of Statistics*, 12(4): 1151–1172. [67](#)
- Schadendorf, D., Wolchok, J. D., Hodi, F. S., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.-J., Cowey, C. L., Lao, C. D., Chesney, J., et al. (2017). “Efficacy and safety outcomes in patients with advanced melanoma who discontinued treatment with nivolumab and ipilimumab because of adverse events: a pooled analysis of randomized phase II and III trials.” *Journal of Clinical Oncology*, 35(34): 3807. [23](#)
- Schwartz, S., Li, F., and Mealli, F. (2011). “A Bayesian semiparametric approach to intermediate variables in causal inference.” *Journal of the American Statistical Association*, 31(10): 949–962. [14](#), [15](#), [38](#), [39](#)
- Shao, J., Chang, M., and Chow, S.-C. (2005). “Statistical inference for cancer trials with treatment switching.” *Statistics in Medicine*, 24(12): 1783–1790. [34](#)
- Stensrud, M. J. and Dukes, O. (2022). “Translating questions to estimands in randomized clinical trials with intercurrent events.” *Statistics in Medicine*. [11](#), [36](#)
- Therneau, T., Grambsch, P., and Fleming, T. (1990). “Martingale based residuals for survival models.” *Biometrika*, 77: 147–160. [70](#)
- Varadhan, R., Xue, Q.-L., and Bandeen-Roche, K. (2014). “Semicompeting risks in aging research: methods, issues and needs.” *Lifetime data analysis*, 20(4): 538–562. [36](#)

- Walker, A. S., White, I. R., and Babiker, A. G. (2004). “Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment.” *Statistics in Medicine*, 23(4): 571–590. [34](#)
- White, I. R. (2006). “Estimating treatment effects in randomized trials with treatment switching.” *Statistics in Medicine*, 25(9): 1619–1622. [34](#)
- White, I. R., Babiker, A. G., Walker, S., and Darbyshire, J. H. (1999). “Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial.” *Statistics in Medicine*, 18(19): 2617–2634. [2](#), [33](#)
- White, I. R., Walker, S., and Babiker, A. (2002). “strbee: Randomization-based efficacy estimator.” *Stata Journal*, 2(2): 140–150. [5](#)
- White, I. R., Walker, S., Babiker, A. G., and Darbyshire, J. H. (1997). “Impact of treatment changes on the interpretation of the Concorde trial.” *AIDS*, 11(8): 999–1006. [2](#)
- Xu, Y., Scharfstein, D., Müller, P., and Daniels, M. (2022). “A Bayesian non-parametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks.” *Biostatistics*, 23(1): 34–49. [4](#), [37](#)
- Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., and Hernán, M. A. (2020). “A causal framework for classical statistical estimands in failure-time settings with competing events.” *Statistics in Medicine*, 39(8): 1199–1236. [11](#), [35](#)
- Zeng, D., Chen, Q., Chen, M.-H., Ibrahim, J. G., and Groups, A. R. (2011). “Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study.” *Biometrika*, 99(1): 167–184. [2](#), [34](#)
- Zhang, J. and Chen, C. (2016). “Correcting treatment effect for treatment switching in randomized oncology trials with a modified iterative parametric estimation method.” *Statistics in Medicine*, 35(21): 3690–3703. [34](#)
- Zigler, C. M. and Belin, T. R. (2012). “A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint.” *Biometrics*, 68(3): 922–932. [14](#), [38](#), [39](#)