Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1–28

Assessing Causal Effects in the Presence of Treatment Switching Through Principal Stratification

Alessandra Mattei[*,†], Peng Ding[‡], Veronica Ballerini[*], and Fabrizia Mealli[*,†,§]

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 the synthetic data based on the Concorde study.

Keywords: Bayesian causal inference, ensoring, competing risks, noncompliance, potential outcomes, survival.

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 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

[*]Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy, alessandra.mattei@unifi.it; veronica.ballerini@unifi.it; fabrizia.mealli@unifi.it
[†]Florence Center for Data Science, Florence, Italy
[‡]Department of Statistics, University of California, Berkeley, CA, USA, pengdingpku@berkeley.edu
[§]Department of Economics, European University Institute, Florence, Italy, Fabrizia.Mealli@eui.eu

 https://doi.org/10.1214/24-BA1425

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 (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 AIDS.

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 we want to assess 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 immediate treatment with zidovudine has no or little effect on survival. Web Appendix A (Mattei et al., 2024) 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 treatment effects in the presence of intercurrent events, such as treatment switching (ICH, 2019). In Web Appendix B (Mattei et al., 2024), 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 through which the treatment affects the occurrence of the primary event, e.g., through other (secondary) events such as (in)tolerance to treatment, which may imply 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; Mattei et al., 2024). 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, the causal effects are well-defined for each principal stratum. See Web Appendix B (Mattei et al., 2024) 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 about the 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 versus deferred 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 AIDS 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^{obs} and S_i^{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 $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = \overline{\mathbb{S}}$, where the symbol "$\overline{\mathbb{S}}$" 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

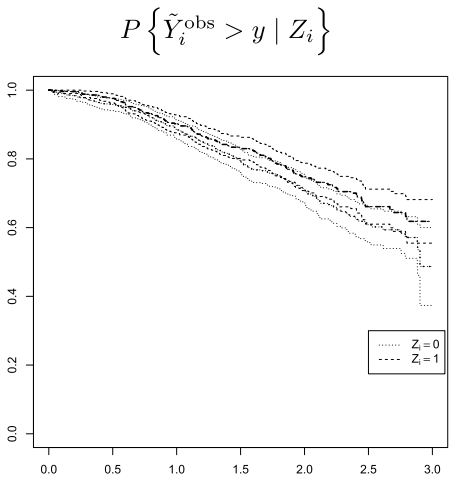


Figure 1: ITT analysis: Kaplan–Meier estimates of the survival functions with 95% confidence intervals. The solid line corresponds to the control, and the dashed line corresponds to the treatment.

censored switching time:

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

where we set $S_i^{\text{obs}} = \overline{\mathbb{S}}$ for control patients who progress the disease or 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.

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 ($\mathbb{I}\{\big(S_i^{\text{obs}} \in \mathbb{R}_+$ and $S_i^{\text{obs}} > C_i\big)$ or $S_i^{\text{obs}} = \overline{\mathbb{S}}\}$)	–	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	$Z_i = 0$		
	$Y_i^{\text{obs}} \leq C_i$ $\tilde{S}_i^{\text{obs}} = C_i$ ($n = 119$)	$S_i^{\text{obs}} \leq C_i$ ($n = 189$)	$\tilde{Y}^{\text{obs}} = C_i$ $\tilde{S}^{\text{obs}} = C_i$ ($n = 192$)
Indicator for the switching time taking on a non-real value ($\mathbb{I}\{S_i^{\text{obs}} = \overline{\mathbb{S}}\}$)	1	0	–
Indicator for the switching time being censored or taking on a non-real value ($\mathbb{I}\{\big(S_i^{\text{obs}} \in \mathbb{R}_+$ and $S_i^{\text{obs}} > C_i\big)$ or $S_i^{\text{obs}} = \overline{\mathbb{S}}\}$)	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

Table 1: Synthetic Concorde data: Descriptive statistics.

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 the censoring time is not affected by the treatment; we denote it by $C_i(0) = C_i(1) = C_i$ 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 $S_i(1) = \overline{\mathbb{S}}$. Second, a patient i may not switch from the control to treatment no matter how long the follow-up is, implying $S_i(0) = \overline{\mathbb{S}}$. Third, a patient i can switch to the treatment arm only before her/his 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 patients with switching time, $S_i(0)$, censored by death, the switching time is not only not observed but also undefined; thus, $S_i(0) = \overline{\mathbb{S}}$. 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

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

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

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

Ju and Geng (2010) noted that $\text{ACE} = \int_0^{+\infty} \text{DCE}(y)\, \mathrm{d}y$. 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, $S_i(1) = \overline{\mathbb{S}}$ 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 $S_i(0)$ 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: $\{i : S_i(0) = s\}$, $s \in \{\overline{\mathbb{S}}\} \cup \mathbb{R}_+$. Throughout the paper, we refer to patients with $S_i(0) = \overline{\mathbb{S}}$ 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 $S_i(0) = \overline{\mathbb{S}}$, 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,

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

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

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

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

Because non-switchers would not switch to treatment if assigned to control, their treatment received coincides with their treatment assigned. Thus, the PCEs for non-switchers are attributable to treatment received, that is, $\text{ACE}(\overline{\mathbb{S}})$ and $\text{DCE}(y \mid \overline{\mathbb{S}})$ 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 the survival time 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

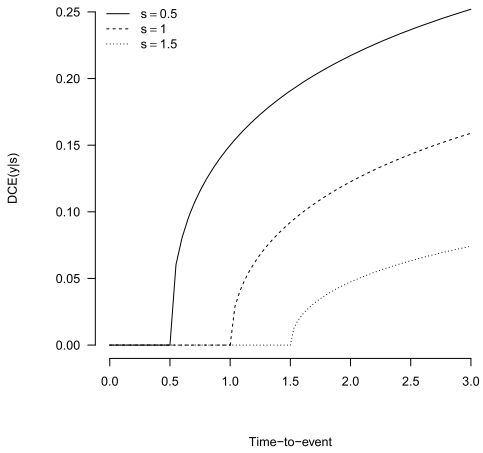


Figure 2: Examples of principal distributional causal effects under monotonicity ($c = 3$).

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 $Y_i(1) > S_i(0)$:

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

for $y, s \in \mathbb{R}_+$. For $y \leq s$, $\text{cDCE}(y \mid s) = 1 - 1 = 0$. The estimand $\text{cDCE}(y \mid s)$, $s \in \mathbb{R}_+$, 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

$$\begin{aligned} \text{ACE}(\mathcal{A}) &= \mathbb{E}\left\{Y_i(1) \mid S_i(0) \in \mathcal{A}\right\} - \mathbb{E}\left\{Y_i(0) \mid S_i(0) \in \mathcal{A}\right\}, &(6)\\ \text{DCE}(y \mid \mathcal{A}) &= P\left\{Y_i(1) > y \mid S_i(0) \in \mathcal{A}\right\} - P\left\{Y_i(0) > y \mid S_i(0) \in \mathcal{A}\right\}, &(7)\\ \text{cDCE}(y \mid \mathcal{A}) &= P\left\{Y_i(1) > y \mid Y_i(1) > S_i(0), S_i(0) \in \mathcal{A}\right\} &(8)\\ &\quad - P\left\{Y_i(0) > y \mid Y_i(1) > S_i(0), S_i(0) \in \mathcal{A}\right\}, \end{aligned}$$

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 (Mattei et al., 2024).

In general, we can discretize the switching time into several disjoint intervals $\mathbb{R}_+ = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_K$, and define $\text{ACE}(\mathcal{A}_k)$, $\text{DCE}(y \mid \mathcal{A}_k)$ and $\text{cDCE}(y \mid \mathcal{A}_k)$ for $k = 1, \ldots, K$. 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 studies involving survival analyses 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 (Mattei et al., 2024). 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. They 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 $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = S_i(1) = \overline{\mathbb{S}}$ under treatment and

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

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 $S_i^{\text{obs}} = S_i(0) = \overline{\mathbb{S}}$. 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 $\{i : S_i(0) = \overline{\mathbb{S}}\}$.

(b) The patient switches to the treatment arm, starting zidovudine before the onset of ARC or AIDS, at time S_i^{obs} and dies at time Y_i^{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 $S_i^{\text{obs}} \in \{\overline{\mathbb{S}}\} \cup (C_i, +\infty)$ 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 $S_i^{\text{obs}} = S_i(0) = \overline{\mathbb{S}}$ and $Y_i^{\text{obs}} = Y_i(0) > C_i$. This patient belongs to either the stratum $\{i : S_i(0) = \overline{\mathbb{S}}\}$ 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 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. $P\{Z_i \mid S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i\}$.

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

Table 2: Observed data pattern and possible latent principal strata.

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

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

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 $P\{Z_i \mid S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i \mid X_i\}$, and ignorability of the censoring mechanism conditional on observed pretreatment variables X_i, $P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i\} = P\{C_i \mid X_i\}$. 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:

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

Under ignorability of the censoring mechanism, we can estimate the survival functions $P\left\{Y_i^{\text{obs}} > y \mid Z_i = z\right\}$ 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 $\text{ACE} = \int_0^c \text{DCE}(y)\text{d}y + \int_c^{+\infty} \text{DCE}(y)\text{d}y$ 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, $\text{DCE}(y \mid \overline{\mathbb{S}})$, is, in

general, different from the prima facie distributional effect,

$$\text{FDCE}(y \mid \overline{\mathbb{S}}) = P\left\{Y_i^{\text{obs}} > y \mid Z_i = 1\right\} - P\left\{Y_i^{\text{obs}} > y \mid Z_i = 0, S_i^{\text{obs}} = \overline{\mathbb{S}}\right\},$$

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

$$P\left\{Y_i(0) > y \mid S_i(0) = \overline{\mathbb{S}}\right\} = P\left\{Y_i^{\text{obs}} > y \mid Z_i = 0, S_i^{\text{obs}} = \overline{\mathbb{S}}\right\},$$

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

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

which implies that $\text{FDCE}(y \mid \overline{\mathbb{S}}) \leq \text{DCE}(y \mid \overline{\mathbb{S}})$.

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., 2019). 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 (Mattei et al., 2024) 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., 2019). The Weibull model has two positive parameters α and ξ. The parameter α 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 $\mathbb{I}\{S_i(0) = \overline{\mathbb{S}}\}$ 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 $S_i(0) = \overline{\mathbb{S}}$, we model $Y_i(0) \mid S_i(0) = \overline{\mathbb{S}}, X_i$ as a Weibull distribution with parameters $\left(\bar{\alpha}_Y, \log(\bar{\xi}_0) = \bar{\beta}_Y + X_i'\bar{\boldsymbol{\eta}}_Y\right)$, $\bar{\alpha}_Y > 0$, $\bar{\beta}_Y \in \mathbb{R}$ 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 $S_i(0) = \overline{\mathbb{S}}$, we model $Y_i(1)$ for non-switchers as a location shifted Weibull distribution:

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

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 (Mattei et al., 2024), we explicitly show the probability density functions, the survival functions, and the hazard functions corresponding to these model assumptions. The entire parameter vector is $\boldsymbol{\theta} = \big[(\eta_0, \boldsymbol{\eta}),\ (\alpha_S, \beta_S, \boldsymbol{\eta}_S),\ \left(\bar{\alpha}_Y, \bar{\beta}_Y, \bar{\boldsymbol{\eta}}_Y\right),$ $(\alpha_Y, \beta_Y, \lambda_0, \boldsymbol{\eta}_Y),\ \left(\bar{\nu}_Y, \bar{\gamma}_Y, \bar{\boldsymbol{\zeta}}_Y\right),\ (\nu_Y, \gamma_Y, \lambda_1, \boldsymbol{\zeta}_Y),\ \kappa\big]$.

4.2 Identification of some model parameters

The parameters λ_1 and κ deserve some discussion. The parameter κ characterizes the dependence between $Y_i(1)$ and $Y_i(0)$ given $\{S_i(0), X_i\}$. The observed data provide little information on κ because we can observe only one of the potential survival times for each patient. We can view κ 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 κ within the range $[0, 1]$.

The parameter λ_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 λ_1 to borrow information from other parameters and the modeling structure, leveraging information contained in the observed data. We assume equality of the association parameters λ_0 and λ_1: $\lambda \equiv \lambda_0 = \lambda_1$, so that a common parameter, λ, 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 λ. Conceptually, Bayesian principal stratification analysis does not require the assumption of equality of the association parameters λ_0 and λ_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 $\boldsymbol{\theta} = \big[(\eta_0, \boldsymbol{\eta}), (\alpha_S, \beta_S, \boldsymbol{\eta}_S), \big(\bar{\alpha}_Y, \bar{\beta}_Y, \bar{\boldsymbol{\eta}}_Y\big), (\alpha_Y, \beta_Y, \boldsymbol{\eta}_Y), \big(\bar{\nu}_Y, \bar{\gamma}_Y, \bar{\boldsymbol{\zeta}}_Y\big), (\nu_Y, \gamma_Y, \boldsymbol{\zeta}_Y), \lambda, \kappa\big]$.

It is worth noting that some values of the parameters (λ, κ) 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

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

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 (λ, κ). Under our parametric assumptions, (λ, κ) 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 $P\{Y_i(0), Y_i(1) \mid S_i(0), X_i\}$ into the product of $P\{Y_i(0) \mid S_i(0), X_i\}$ and $P\{Y_i(1) \mid Y_i(0), S_i(0), X_i\}$. The model for $P\{Y_i(0) \mid S_i(0), X_i\}$ characterizes the relationship between the survival outcome and the switching status under control. The model for $P\{Y_i(1) \mid Y_i(0), S_i(0), X_i\}$ 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 (Mattei et al., 2024) 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 (λ, κ). 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 (λ, κ), 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 κ concentrated at a pre-fixed value $\kappa_0 \in [0, 1]$, which is essentially the same as fixing κ at κ_0 a priori. See Web Appendix G (Mattei et al., 2024) 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 λ 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 λ_0 and λ_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

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

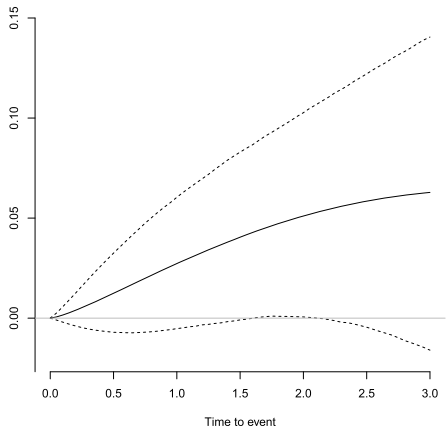


Figure 3: Bayesian ITT analysis using Weibull models. The solid line corresponds to the posterior median, and the dashed lines correspond to the 95% posterior credible interval.

to different assumptions on κ by examining how the posterior distributions of the causal estimands change with respect to different κ_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 (α_Y, β_Y), and (ν_Y, γ_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 α_Y and ν_Y, and Normal prior distributions with zero mean and variance 10 000 for β_Y and γ_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 me-

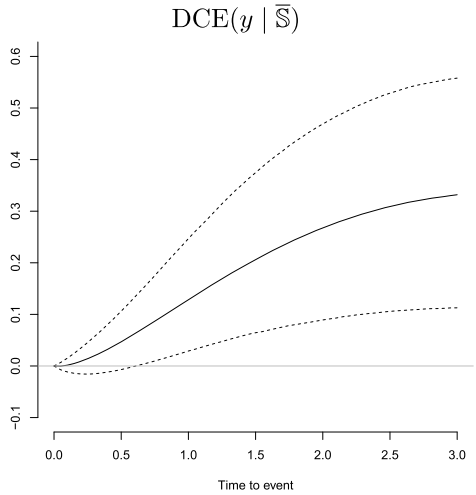


Figure 4: Principal stratification analysis: Posterior median (solid line) and 95% posterior credible interval (dashed lines) of the distributional causal effects for non-switchers.

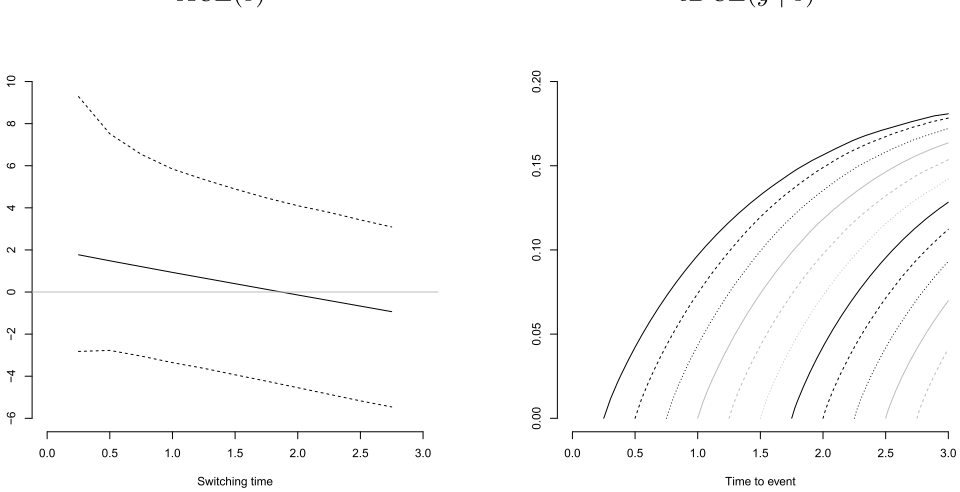
dians and 95% posterior credible intervals of the distributional causal effects. Thus, there is evidence that immediate treatment with zidovudine extends life in individuals infected with HIV, but the estimated effects are small and statistically negligible. Nevertheless, it is sensible to expect that 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 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 (Mattei et al., 2024) 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, DCE($y \mid \overline{\mathbb{S}}$),



(a) ACE(s), $s \in \mathbb{R}_+$ with 95% credible interval

(b) cDCE($y \mid s$) for $s = 0.25, 0.50, \ldots, 2.50, 2.75$

Figure 5: Posterior medians of the PCEs for switchers.

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 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 AIDS 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 AIDS 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 (Mattei et al., 2024). 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 sicklier patients, and spending even a short time under control may harm them. Under 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 λ, namely, $\mathrm{N}\,(0, 10^4)$. We assess the sensitivity of the results to κ, the partial association between $Y_i(1)$ and $Y_i(0)$ given the switching status, $S_i(0)$, and to the prior specification for λ, by investigating how the posterior distributions of causal estimands change under different κ values and different prior specifications for λ. Results appear robust to prior specifications for λ; different prior distributions for λ change the posterior distribution of the causal estimands only slightly. We find some sensitivity of inferences on the causal estimands to κ, especially for switchers.

We also investigate the robustness of the results with respect to the parametric assumption imposing prior equality of the association parameters λ_0 and λ_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 (Mattei et al., 2024) 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 combina-

tion 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 $\overline{\mathbb{S}}$, 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: $S_i(1) = \overline{\mathbb{S}}$; 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 $\overline{\mathbb{S}}$, so $S_i(0) = S_i(1) = \overline{\mathbb{S}}$ 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 $S_i(0) \in \mathbb{R}_+$ and $S_i(1) = \overline{\mathbb{S}}$; (*iii*) the principal strata of patients who would discontinue the active treatment but would not discontinue the placebo treatment. For this type of patients $S_i(0) = \overline{\mathbb{S}}$ and $S_i(1) \in \mathbb{R}_+$; (*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.

Funding

Alessandra Mattei and Fabrizia Mealli are supported by the Italian Ministry of University and Research (MUR), Department of Excellence project 2023-2027 ReDS "Rethinking Data Science" – Department of Statistics, Computer Science, Applications – University of Florence.

Peng Ding is supported by the U.S. National Science Foundation grant # 1945136.

Veronica Ballerini is supported by the European Union – Next GenerationEU, UNIFI Young Independent Researchers Call – BayesMeCOS Grant no. B008-P00634.

Supplementary Material

Supplementary Material for "Assessing Causal Effects in the Presence of Treatment Switching Through Principal Stratification" (DOI: 10.1214/24-BA1425SUPP; .pdf). Supplementary Material includes: a review of the existing literature about the treatment effect evaluation accounting for treatment switching (Web Appendix A); a discussion of a semi-competing risks approach to the analysis of randomized studies with survival outcomes in the presence of treatment switching (Web Appendix B); details about the connection to the noncompliance literature (Web Appendix C); the definition of coarsened principal causal effects introduced in Equations (6)-(8) (Web Appendix D); the detailed description of the Bayesian principal stratification approach introduced in Section 4 (Web Appendix E); the parametric assumptions, namely the probability density functions, the survival functions, and the hazard functions corresponding to the model assumptions (Web Appendix F); the prior specifications (Web Appendix G); the details of the data augmentation algorithm (Web Appendix H); additional results (Web Appendix I).

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

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. MR3068438. doi: https://doi.org/10.1093/biomet/ass091. 2

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*. 3

Concorde Coordinating Committee (1994). "Concorde: MRC / ANRS randomised dou-

ble blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection." *The Lancet*, 343: 871–881. 2, 4

Ding, P. and Li, F. (2018). "Causal inference: A missing data perspective." *Statistical Science*, 33(2): 214–237. MR3797711. doi: https://doi.org/10.1214/18-STS645. 13

Frangakis, C. E. and Rubin, D. B. (2002). "Principal stratification in causal inference." *Biometrics*, 58: 191–199. MR1891039. doi: https://doi.org/10.1111/j.0006-341X.2002.00021.x. 3, 7, 8, 9

Gustafson, P. (2010). "Bayesian inference for partially identified models." *International Journal of Biostatistics*, 2(2): Article 17. MR2602560. doi: https://doi.org/10.2202/1557-4679.1206. 13, 16

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. doi: https://doi.org/10.1016/S0140-6736(12)60868-X. 2, 21

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer. MR1876598. doi: https://doi.org/10.1007/978-1-4757-3447-8. 14

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 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. MR1429927. doi: https://doi.org/10.1214/aos/1034276631. 13

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. MR2463484. doi: https://doi.org/10.1198/016214507000000347. 13

Jin, H. and Rubin, D. B. (2009). "Public schools versus private schools: Causal inference with partial compliance." *Journal of Educational and Behavioral Statistics*, 34(1): 24–45. doi: https://doi.org/10.3102/1076998607307475. 13

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. MR2751247. doi: https://doi.org/10.1111/j.1467-9868.2009.00729.x. 7

Kim, C., Daniels, M. J., Hogan, J. W., Choirat, C., and Zigler, C. M. (2019). "Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls." *Annals of Applied Statistics*, 13(3): 1927–1956. MR4019162. doi: https://doi.org/10.1214/19-AOAS1260. 13, 14

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. doi: https://doi.org/10.1056/NEJMoa1504030. 2, 21

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. doi: https://doi.org/10.1634/theoncologist.2014-0429. 21

Lindley, D. V. (1972). *Bayesian Statistics: A review*. SIAM. MR0329081. 13

Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020). "Causal inference and estimands in clinical trials." *Statistics in Biopharmaceutical Research*, 12(1): 54–67. doi: https://doi.org/10.1080/19466315.2019.1697739. 23

Mattei, A., Ding, P., Ballerini, V., and Mealli, F. (2024). "Supplementary Material for "Assessing Causal Effects in the Presence of Treatment Switching Through Principal Stratification"." *Bayesian Analysis*. doi: https://doi.org/10.1214/24-BA1425SUPP. 2, 3, 10, 13, 14, 16, 18, 20

Mattei, A. and Mealli, F. (2007). "Application of the prinicipal stratification approach to the Faenza randomized experiment on breast self-examination." *Biometrics*, 63(2): 437–446. MR2370802. doi: https://doi.org/10.1111/j.1541-0420.2006.00684.x. 13

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*. MR4187187. doi: https://doi.org/10.1080/07350015.2019.1647843. 3

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. MR1978855. doi: https://doi.org/10.1002/0470867205.ch16. 14

Robins, J. M. (1994). "Correcting for non-compliance in randomized trials using structural nested mean models." *Communications in Statistics-Theory and methods*, 23(8): 2379–2412. MR1293185. doi: https://doi.org/10.1080/03610929408831393. 2

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. MR1144866. doi: https://doi.org/10.1080/03610929108830654. 2

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. doi: https://doi.org/10.1200/JCO.2017.73.2289. 22

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. MR2896839. doi: https://doi.org/10.1198/jasa.2011.ap10425. 13, 14

Stensrud, M. J. and Dukes, O. (2022). "Translating questions to estimands in randomized clinical trials with intercurrent events." *Statistics in Medicine*, 41(16): 3211–3228. MR4444895. doi: https://doi.org/10.1002/sim.9398. 10

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. doi: https://doi.org/10.1002/(SICI)1097-0258(19991015)18:19%3C2617::AID-SIM187%3E3.0.CO;2-E. 2

White, I. R., Walker, S., and Babiker, A. (2002). "strbee: Randomization-based efficacy estimator." *The Stata Journal*, 2(2): 140–150. doi: https://doi.org/10.1177/1536867X0200200203. 4

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 nonparametric approach for evaluating the causal effect of treatment in randomized trials with semicompeting risks." *Biostatistics*, 23(1): 34–49. MR4366034. doi: https://doi.org/10.1093/biostatistics/kxaa008. 3

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. MR4075855. doi: https://doi.org/10.1002/sim.8471. 10

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. MR2899671. doi: https://doi.org/10.1093/biomet/asr062. 2

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. MR3055197. doi: https://doi.org/10.1111/j.1541-0420.2011.01736.x. 13