Biometrics, 2024, 80(1), 1–14
https://doi.org/10.1093/biomtc/ujad016
Biometric Methodology

Principal stratification analysis of noncompliance with time-to-event outcomes

Bo Liu [1], Lisa Wruck [2,3], Fan Li [1,*]

[1]Department of Statistical Science, Duke University, Durham, NC 27708, United States, [2]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27708, United States, [3]Duke Clinical Research Institute, Durham, NC 27701, United States

[*]Corresponding author. Fan Li, Department of Statistical Science, Duke University, Durham, NC 27708, USA (fl35@duke.edu).

ABSTRACT

Post-randomization events, also known as intercurrent events, such as treatment noncompliance and censoring due to a terminal event, are common in clinical trials. Principal stratification is a framework for causal inference in the presence of intercurrent events. The existing literature on principal stratification lacks generally applicable and accessible methods for time-to-event outcomes. In this paper, we focus on the noncompliance setting. We specify 2 causal estimands for time-to-event outcomes in principal stratification and provide a nonparametric identification formula. For estimation, we adopt the latent mixture modeling approach and illustrate the general strategy with a mixture of Bayesian parametric Weibull-Cox proportional hazards model for the outcome. We utilize the Stan programming language to obtain automatic posterior sampling of the model parameters. We provide analytical forms of the causal estimands as functions of the model parameters and an alternative numerical method when analytical forms are not available. We apply the proposed method to the ADAPTABLE (**A**spirin **D**osing: **A** **P**atient-Centric **T**rial **A**ssessing **B**enefits and **L**ong-Term **E**ffectiveness) trial to evaluate the causal effect of taking 81 versus 325 mg aspirin on the risk of major adverse cardiovascular events. We develop the corresponding R package *PStrata*.

KEYWORDS: Bayesian; causal inference; mixture model; noncompliance; principal stratification; survival analysis; time-to-event outcomes.

1 INTRODUCTION

Randomized controlled trial is the gold standard in evaluating efficacy and safety of interventions in medicine. However, postrandomization events, also known as intercurrent events, such as treatment noncompliance, including discontinuation and switching, are prevalent in trials. Such intercurrent events are usually confounded and pose challenges to evaluating comparative effectiveness. Standard intention-to-treat (ITT) analysis ignores intercurrent events and compares outcomes between randomized arms. Intention-to-treat analysis preserves randomization, but it estimates effectiveness rather than efficacy of the intervention and fails to capture treatment effect heterogeneity. Naïve methods, such as comparing units per their actual treatment status or discarding subjects with intercurrent events, generally lead to biased causal estimates (Rosenbaum, 1984).

In a landmark paper, Angrist et al. (1996) proposed the instrumental variable (IV) approach to noncompliance, connecting the structural equation model based IV method to the potential outcome framework for causal inference. Here, randomization itself is an instrument, because it is unconfounded and usually has no direct effect on the outcome, but it affects the actual treatment received, which, in turn, affects the outcome. The key idea is to classify subjects into strata based on their joint potential compliance status under the treatment and control arms and then estimate causal effects specific to the subpopulation in each stratum. Under suitable assumptions, the effect of a stratum, namely, the compliers, is equivalent to the 2 stage least square IV estimand. Similar approach was also independently developed by Baker and Lindeman (1994). Frangakis and Rubin (2002) later extended the IV approach to noncompliance to the general framework of principal stratification to handle posttreatment confounding, which is applicable to a wide range of scenarios, including truncation by death (Zhang and Rubin, 2003), surrogate endpoints (Gilbert and Hudgens, 2008), and selection bias in cluster randomized trials (Li et al., 2022).

Though IV is well known in economics, statistics, and social sciences, principal stratification (PS), as a generalization of the IV approach, has been seldom applied in medicine. The ICH E9 guidelines for Statistical Principles for Clinical Trials [ICH E9 (R1), 2020] list PS as a statistically valid approach to analyze clinical trials with intercurrent events. This spurred an increasing interest in PS in regulatory agencies and industry (Lipkovich et al., 2022). However, there are several barriers to the wide adoption of PS in applications. First, survival and more broadly time-to-event outcomes are prevalent in medical research and require special handling of censoring, but PS with time-to-event outcomes has rarely been investigated. The few exceptions (Baker, 1998; Nie et al., 2011; Yu et al., 2015; Blanco et al., 2020; Wei et al., 2021) all rely on highly customized statistical models and methods such as empirical likelihood. Shepherd and co-authors have discussed time-to-event outcomes in the context of truncation

Received: January 18, 2023; Revised: October 2, 2023; Accepted: November 15, 2023

by death (Shepherd et al., 2007, 2011). Second, PS inherently involves latent mixture structures, and thus estimation is usually implemented via either the expectation-maximization algorithm or Bayesian method, both of which require substantial expertise in statistics and programming. Third, the lack of software package further limits the method's accessibility to applied researchers.

In this paper, we focus on the setting of noncompliance and provide a generally applicable method for time-to-event outcomes. Mathematically, the noncompliance setting is applicable to a variety of situations that differ in the substantive contexts, for example, randomized encouragement trials. The commonality in this setting is that the outcomes are potentially observable for all units. This is in contrast to the settings where outcomes of some units are not defined by design, such as truncation by death. For simplicity, we will generically refer to this setting as "noncompliance" and use the associated nomenclature throughout this paper. In Section 2, we specify 2 causal estimands for time-to-event outcomes and discuss nonparametric identification. In Section 3, we adopt the latent mixture modeling approach for estimation. For computational convenience, we illustrate the strategy with a mixture of Bayesian parametric Weibull-Cox proportional hazards model for the outcome (Section 3.2). We develop an R package **PStrata** (Liu and Li, 2023) utilizing the Stan programming language to obtain automatic posterior sampling of the model parameters. We provide the analytical forms of the causal estimands as functions of the model parameters and an alternative numerical method when analytical forms are not available (Section 3.3). We also extend the method to simultaneously accommodate noncompliance and discontinuation (Section 4). We conduct simulation studies to examine the performance of the proposed methods (Section 5). We apply the proposed method to the ADAPTABLE (**A**spirin **D**osing: **A** **P**atient-Centric **T**rial **A**ssessing **B**enefits and **L**ong-Term **E**ffectiveness) trial to evaluate the causal effect of taking 81 versus 325 mg aspirin on the risk of major adverse cardiovascular events (Section 6).

2 SETUP AND ESTIMANDS

To convey the main message, we set up PS with time-to-event outcomes in the classic setting of 2-arm randomized experiments with all-or-nothing (ie, binary) compliance. The methods discussed below can be readily adapted to more complex noncompliance settings, for example, multiple treatments and continuous compliance. Consider a sample of units $i\,(i = 1, \ldots, N)$ from a target population. For each unit i, we observe a set of p baseline covariates, X_i, and the randomized assigned treatment status, $Z_i(=z)$, with $z = 1$ indicating treatment and $z = 0$ control. Each unit has a failure time T_i, which is subject to right-censoring at time C_i. Therefore, we only observe the potentially censored time, $Y_i = \min(T_i, C_i)$, and the censoring indicator, $\delta_i = \mathbf{1}\{T_i \geq C_i\}$. To define causal effects, we adopt the potential outcomes framework. Assuming the stable unit treatment value assumption (SUTVA) (Rubin, 1980), each unit has 2 potential failure times $\{T_i(0), T_i(1)\}$ and 2 potential censoring times $\{C_i(0), C_i(1)\}$, one under each assignment status. Denote the potential survival distribution under assignment z as $S_z(t) = \Pr(T_i(z) > t)$ for $z = 1, 0$. The ITT survival probability causal effect (SPCE) is:

$$\tau^{SPCE}(t) = S_1(t) - S_0(t), \quad 0 \leq t \leq t_{\max}, \tag{1}$$

where $t_{\max}$ is the maximum follow-up time. This estimand represents the causal risk difference between the 2 arms. A second estimand is the ITT restricted average causal effect (RACE) in survival time,

$$\tau^{RACE}(t) = \mathbb{E}[T_i(1) \wedge t] - \mathbb{E}[T_i(0) \wedge t], \quad 0 \leq t \leq t_{\max}, \tag{2}$$

where $a \wedge b \equiv \min(a, b)$. The RACE compares the mean potential survival times restricted by a given time point of interest t.

A common estimand in survival analysis is hazard ratio, popularized by the Cox proportional hazards model (Cox, 1972). However, directly comparing the hazard between 2 treatment arms does not lead to a valid causal estimand (Hernán 2010) and thus defining causal contrasts of hazards is generally controversial. Therefore, in this paper, we focus on the SPCE and RACE estimands.

Noncompliance occurs when some units do not take the treatment they are assigned to. Let D_i denote the actual treatment received by unit i. Noncompliance implies $Z_i \neq D_i$ for some i. In clinical trials, this is often caused by patients switching treatments. For example, in our motivating application of the ADAPTABLE trial, 33.5% of the patients in the 325 mg arm switched to the 81 mg arm, while 6.1% of the patients in the 81 mg arm switched to 325 mg arm. Because the actual treatment D occurs after the randomization, it has 2 potential outcomes, $D_i(0)$ and $D_i(1)$, one of which is observed, denoted by $D_i = D_i(Z_i)$, and the other is missing. A principal stratification is the classification of units according to their joint potential treatment-receipt statuses, known as principal strata, $U_i = (D_i(0), D_i(1))$ (Angrist et al., 1996; Frangakis and Rubin, 2002). With binary assignment and treatment, there are 4 principal strata: (i) $U_i \equiv c = (0, 1)$, namely compliers, units that would take treatment if assigned to treatment and would take control if assigned to control; (ii) $U_i \equiv n = (0, 0)$, namely never-takers, units that would take control regardless of the initial assignment; (iii) $U_i \equiv a = (1, 1)$, namely always-takers, units that would take treatment regardless of the initial assignment; and (iv) $U_i \equiv d = (1, 0)$, namely defiers, units whose actual treatment status is the opposite to the assignment. Frangakis and Rubin (2002) generalized the noncompliance setting to any events occurring between treatment and outcome. We will use the above nomenclature of noncompliance throughout this paper but note that these strata are often termed differently according to the specific context. In the presence of noncompliance, strictly speaking, the potential failure time and censoring time should be denoted using double index as $T(z, D(z))$, $C(z, D(z))$, respectively. But because $D(z)$ is a function of the potential treatment value z, we can abbreviate the double-index notation to the single-index notation: $T(z) \equiv T(z, D(z))$, $C(z) \equiv C(z, D(z))$, which we will adopt hereafter.

A key property of the principal stratum S is that, by construction, it is not affected by the initial assignment, and thus can be viewed as a prerandomization variable, so we can define causal effects conditioning on U, that is, the principal causal effects. Therefore, for each of the ITT estimand in (1) and (2), we can

define its corresponding principal causal effect. Denote the potential survival distribution under assignment z for stratum u as $S_{z,u}(t) = \Pr(T_i(z) > t \mid U_i = u)$. Then the principal survival probability causal effect is

$$\tau_u^{SPCE}(t) = S_{1,u}(t) - S_{0,u}(t) \tag{3}$$

for $u \in \{c, n, a, d\}$. The effect for the compliers, $\tau_c^{SPCE}(t)$, is called the *complier average causal effect (CACE)*, which is an estimand of primary interest in the noncompliance setting. Similarly, we can define the principal RACE for $u \in \{c, n, a, d\}$:

$$\tau_u^{RACE}(t) = \mathbb{E}[T_i(1) \wedge t \mid U_i = u] - \mathbb{E}[T_i(0) \wedge t \mid U_i = u], \quad 0 \le t \le t_{\max}. \tag{4}$$

Because τ^{SPCE} and τ^{RACE} are additive estimands, it is straightforward to show that the ITT effect is a weighted average of the corresponding principal causal effects:

$$\tau^{SPCE}(t) = \sum_{u \in \{c,n,a,d\}} \tau_u^{SPCE}(t) \Pr(U = u),$$
$$\tau^{RACE}(t) = \sum_{u \in \{c,n,a,d\}} \tau_u^{RACE}(t) \Pr(U = u). \tag{5}$$

3 IDENTIFICATION AND ESTIMATION

3.1 Nonparametric identification

Due to the fundamental problem of causal inference, only 1 of the 2 potential treatment status, namely $D_i(Z_i)$, is observed for each unit, and thus the individual stratum membership U_i is not observed. In fact, without any assumptions, each observed cell of (Z, D) is composed of a mixture of 2 strata. For example, the untreated units in the control arm (ie, $Z_i = 0$, $D_i = 0$) can be either compliers or never-takers. Therefore, additional assumptions are required to identify the principal causal effects and the main task in estimation is to disentangle the latent mixtures from the observed data.

Below, we extend the framework of Angrist et al. (1996) to nonparametrically identify the causal estimands with censored outcomes. A similar strategy was proposed in Nie et al. (2011). We make the following assumptions.

Assumption 1 (*Unconfounded assignment*). $\{T_i(0), T_i(1), D_i(0), D_i(1)\} \perp\!\!\!\perp Z_i | X_i$.

Assumption 2 (*Monotonicity*): $D_i(1) \ge D_i(0)$.

Assumption 3 (*Exclusion restriction*): *for stratum u in which $D(0) = D(1)$, that is, always-takers and never-takers, $T_i(1)$ and $T_i(0)$ are identically distributed:* $\Pr(T_i(1) \mid U_i = u) = \Pr(T_i(0) \mid U_i = u)$.

Assumption 1 holds by design in randomized trials. Monotonicity (Assumption 2) rules out defiers. It is usually interpreted as subjects participating a trial in good faith, not intentionally defying the assignment, and is deemed reasonable in most clinical trials. Under monotonicity, the $\{Z = 0, D = 1\}$ cell consists of only always-takers and thus, the proportion of the always-takers is nonparametrically identified as $p_a = \Pr(D_i = 0 \mid Z_i = 1)$. Similarly, the $\{Z = 1, D = 0\}$ cell consists of only never-takers and thus, the proportion of the never-takers is nonparametrically identified as $p_n = \Pr(D_i = 1 \mid Z_i = 0)$. In contrast, the other 2 cells consist of mixture of 2 strata. Specifically, the cell $\{D = 0, Z = 0\}$ consists of never-takers and compliers, and the cell $\{D = 1, Z = 1\}$ consists of always-takers and compliers.

The exclusion restriction (ER) (Assumption 3) assumes away direct effects from the random assignment to the outcome for always-takers and never-takers. Assuming both ER and monotonicity, the only principal stratum with nonzero causal effect is the compliers, and the compliers average causal effects can be nonparametrically identified for both noncensored and survival outcome. Specifically, under ER (and monotonicity), we have

$$\mathbb{E}[T(0) \mid U = n] = \mathbb{E}[T(1) \mid U = n] = \mathbb{E}(T \mid D = 0, Z = 1), \tag{6}$$

$$\mathbb{E}[T(1) \mid U = a] = \mathbb{E}[T(0) \mid U = a] = \mathbb{E}(T \mid D = 1, Z = 0). \tag{7}$$

Therefore, the quantities $\mathbb{E}[T(z) \mid U = c]$ can be nonparametrically identified by

$$\mathbb{E}[T(0) \mid U = c] = \frac{1}{p_c}\left[(p_n + p_c)\,\mathbb{E}(T \mid D = 0, Z = 0) - p_n\,\mathbb{E}(T \mid D = 0, Z = 1)\right], \tag{8}$$

$$\mathbb{E}[T(1) \mid U = c] = \frac{1}{p_c}\left[(p_a + p_c)\,\mathbb{E}(T \mid D = 1, Z = 1) - p_a\,\mathbb{E}(T \mid D = 1, Z = 0)\right]. \tag{9}$$

When T is not censored, the expectation $\mathbb{E}(T \mid D = d, Z = z)$ can be estimated by the sample mean of the corresponding cell in the observed data. When T is censored, we need an additional assumption that the censoring process is independent of the potential survival outcomes given covariates and principal strata. This holds, for example, when the failure times are subject only to administrative right censoring.

Assumption 4 (*Conditional independent censoring*). $T_i(z) \perp\!\!\!\perp C_i(z) | \{Z_i, U_i, X_i\}$.

Noting that the survival probability $S_z(t) = \Pr[T(z) > t] = \mathbb{E}[1\{T(z) > t\}]$, it can be nonparametrically identified in a similar fashion under Assumptions 1-4 as

$$S_0(t \mid U = c) = \frac{1}{p_c}\left[(p_n + p_c)S(t \mid D = 0, Z = 0) - p_n S(t \mid D = 0, Z = 1)\right], \tag{10}$$

$$S_1(t \mid U = c) = \frac{1}{p_c}\left[(p_a + p_c)S(t \mid D = 1, Z = 1) - p_a S(t \mid D = 1, Z = 0)\right]. \tag{11}$$

Based on the above nonparametric identification formula, we can derive nonparametric estimators of the SPCE estimand for censored outcomes. Specifically, the stratum probabilities p_a and p_n can be estimated by the moments as $\widehat{p_a} = \sum_i 1\{Z_i = 1, D_i = 0\} / \sum_i 1\{Z_i = 1\}$ and

$\widehat{p}_n = \sum_i 1\{Z_i = 0, D_i = 1\}/\sum_i 1\{Z_i = 0\}$, respectively. It is immediate that $\widehat{p}_c = 1 - \widehat{p}_a - \widehat{p}_n$. The survival function $S(t|D = d, Z = z)$ can be nonparametrically estimated via the Kaplan-Meier estimator (Kaplan and Meier, 1958) from the observed data with $D_i = d, Z_i = z$. Plugging these estimates into (10) and (11), we obtain the nonparametric estimates

$$\widehat{S}_0(t \mid U = c) = \frac{1}{\widehat{p}_c}\left[(\widehat{p}_n + \widehat{p}_c)\widehat{S}(t \mid D = 0, Z = 0) - \widehat{p}_n\widehat{S}(t \mid D = 0, Z = 1)\right], \tag{12}$$

$$\widehat{S}_1(t \mid U = c) = \frac{1}{\widehat{p}_c}\left[(\widehat{p}_a + \widehat{p}_c)\widehat{S}(t \mid D = 1, Z = 1) - \widehat{p}_a\widehat{S}(t \mid D = 1, Z = 0)\right], \tag{13}$$

and thus the SPCE for the compliers is: $\widehat{\tau}_c^{SPCE}(t) = \widehat{S}_1(t \mid U = c) - \widehat{S}_0(t \mid U = c)$.

Similarly, the RACE can also be nonparametrically identified. Parallel to (8) and (9), the quantities $\mathbb{E}[T(z) \wedge t \mid U = c]$ can be nonparametrically identified by

$$\mathbb{E}[T(0) \wedge t \mid U = c] = [(p_n + p_c)\,\mathbb{E}(T \wedge t \mid D = 0, Z = 0) - p_n\,\mathbb{E}(T \wedge t \mid D = 0, Z = 1)]/p_c, \tag{14}$$

$$\mathbb{E}[T(1) \wedge t \mid U = c] = [(p_a + p_c)\,\mathbb{E}(T \wedge t \mid D = 1, Z = 1) - p_a\,\mathbb{E}(T \wedge t \mid D = 1, Z = 0)]/p_c, \tag{15}$$

where each of the expectation terms $\mathbb{E}(T \wedge t \mid D = d, Z = z) = \int (T \wedge t)\, \mathrm{d}S(T \mid D = d, Z = z)$ can be estimated via the Kaplan-Meier estimator $\widehat{S}$. Specifically, let $t_1, \ldots, t_n$ be the discontinuities of $\widehat{S}(T \mid D = d, Z = z)$, and let $p_i = \widehat{S}(t_i^- \mid D = d, Z = z) - \widehat{S}(t_i^+ \mid D = d, Z = z)$. Then, we obtain the nonparametric estimate of

$$\widehat{\mathbb{E}}(T \wedge t \mid D = d, Z = z) = \int (T \wedge t)\, \mathrm{d}\widehat{S}(T \mid D = d, Z = z) = \sum_{i=1}^{n}(t_i \wedge t)p_i. \tag{16}$$

By plugging (16) into (14) and (15), we obtain the nonparametric estimate of the RACE for the compliers as

$$\widehat{\tau}_c^{RACE}(t) = \widehat{\mathbb{E}}[T(1) \wedge t \mid U = c] - \widehat{\mathbb{E}}[T(0) \wedge t \mid U = c]. \tag{17}$$

The above nonparametric estimators are usually not efficient because they do not utilize the mixture structure of principal stratification or covariate information (Long and Hudgens, 2013; Mealli and Pacini, 2013). Alternatively, we can adopt a latent mixture model approach for estimation that flexibly incorporates covariates and accommodates different types of outcomes (Imbens and Rubin, 1997; Hirano et al., 2000; Mattei et al., 2013).

3.2 General structure of mixture-model-based estimation

With a slight abuse of notation, we generically denote a probability density or distribution function by $\Pr(\cdot \mid \cdot)$. For each unit i, we observe 5 random variables $\{X_i, Z_i, D_i, Y_i, \delta_i\}$. We assume the joint distribution of these variables of all units is governed by a generic parameter θ, conditional on which the random variables for each unit are *i.i.d.*. We denote the set of principal strata u that are compatible with each combination of $\{Z, D\}$ by $\mathcal{U}(Z, D)$. Then, we express the likelihood of the observed data, under Assumptions 1-4, as:

$$\begin{aligned}
&\prod_{i=1}^{N} \Pr(X_i, Z_i, D_i, Y_i, \delta_i) \\
&= \prod_{i=1}^{N} \sum_{u \in \mathcal{U}(Z_i, D_i)} \Pr(X_i)\Pr(Z_i \mid X_i)\Pr(U_i = u \mid Z_i, X_i)\Pr(D_i \mid U_i = u, Z_i, X_i)\Pr(Y_i, \delta_i \mid D_i, U_i = u, Z_i, X_i) \\
&\propto \prod_{i=1}^{N} \sum_{u \in \mathcal{U}(Z_i, D_i)} \Pr(U_i = u \mid Z_i, X_i)\Pr(Y_i, \delta_i \mid U_i = u, Z_i, X_i) \\
&\propto \prod_{i=1}^{N} \sum_{u \in \mathcal{U}(Z_i, D_i)} \Pr(U_i = u \mid Z_i, X_i)\Pr(T_i \geq Y_i \mid U_i = u, Z_i, X_i)^{\delta_i}\Pr(T_i = Y_i \mid U_i = u, Z_i, X_i)^{1-\delta_i}.
\end{aligned} \tag{18}$$

Three terms in (18) become constant with respect to the causal estimands and thus are absorbed by the proportional sign in the third line: (i) $\Pr(X)$, because we condition on the covariates X instead of specifying a distribution for X; (ii) $\Pr(Z_i \mid X_i)$, because of the randomization; and (iii) $\Pr(D_i \mid U_i = u, Z_i, X_i)$, because the summation is over the D_i values that are determined by the values of U_i and Z_i. The proportional sign in (18) holds because, under Assumption 4,

we have

$$\begin{aligned}
&\Pr(Y_i, \delta_i = 1 \mid U_i, Z_i, X_i) \\
&\quad = \Pr(T_i \geq Y_i, C_i = Y_i \mid U_i, Z_i, X_i) \\
&\quad \propto \Pr(T_i \geq Y_i \mid U_i, Z_i, X_i), \\
&\Pr(Y_i, \delta_i = 0 \mid U_i, Z_i, X_i) \\
&\quad = \Pr(T_i = Y_i, C_i > Y_i \mid U_i, Z_i, X_i) \\
&\quad \propto \Pr(T_i = Y_i \mid U_i, Z_i, X_i). \qquad (19)
\end{aligned}$$

Under unconfounded assignment, we have $\Pr(T_i(z) \mid U_i, X_i; \theta_T) = \Pr(T_i \mid U_i, Z_i, X_i; \theta_T)$, and thus the outcome model is equivalent to a model for the potential outcomes. Assumption 1 also implies that the principal strata are independent of the assignment given the covariates: $\{D_i(0), D_i(1)\} \perp\!\!\!\perp Z_i | X_i$, and thus $\Pr(U_i \mid Z_i, X_i) = \Pr(U_i \mid X_i)$. In summary, for model-based PS analysis under Assumption 1-4, we need to specify 2 models: (i) a principal strata model (*strata-model* hereafter): $\Pr(U_i \mid X_i; \theta_u)$, and (ii) an outcome model (*outcome-model* hereafter): $\Pr(T_i \mid U_i, Z_i, X_i; \theta_T)$. Monotonicity and ER are not required for the model-based inference of PS. But monotonicity reduces the number of strata and ER forces the causal effects to be 0 in always-takers and never-takers, and thus both reduce the components in the likelihood function (18) and help to reduce the variance of the causal estimates (Imbens and Rubin, 1997).

With the strata-model and outcome-model being specified, there are 2 common approaches to estimate the model parameters and consequently, the causal estimands. The first approach is to use the EM algorithm (Dempster et al. 1977) to integrate out the missing principal strata and then obtain the maximum-likelihood estimate (Zhang et al., 2009). The second approach is through the Bayesian paradigm, where we specify a prior distribution for all the parameters and obtain the posterior distribution of these parameters, given the data and the models. We choose the Bayesian approach for 2 reasons. First, for time-to-event outcomes, the likelihood function is complex due to the censoring, which renders the implementation of an EM algorithm challenging. Second, the Bayesian method enables straightforward uncertainty quantification of not only the model parameters but also their derived quantities, for example, causal estimands. Below, we describe the Bayesian models and posterior computation.

3.3 Model specification and posterior inference

Because the principal stratum U is a categorical variable, we specify a multinomial logistic regression strata-model with a reference stratum u_0, which we assume to be the same between the two arms:

$$\log\left\{\frac{\Pr(U = u \mid X)}{\Pr(U = u_0 \mid X)}\right\} = \eta_u + X'\xi_u. \qquad (20)$$

This model implies that the probability of each stratum is

$$\begin{aligned}
\Pr(U = u_0 \mid X) &= \frac{1}{1 + \sum_{l \neq u_0} \exp(\eta_l + X'\xi_l)}, \\
\Pr(U = u \mid X) &= \frac{\exp(\eta_u + X'\xi_u)}{1 + \sum_{l \neq u_0} \exp(\eta_l + X'\xi_l)}. \qquad (21)
\end{aligned}$$

For the time-to-event outcome, we impose the most popular Cox proportional hazard model (Cox, 1972); we note the following method can be readily modified to alternative outcome-models like the accelerated failure time (AFT) model (Wei, 1992). The Cox model assumes the hazard function for stratum u and assignment z to be

$$h(t; u, z) = h_0(t; u, z) \exp(X_i' \beta_{u,z}). \qquad (22)$$

where $h_0(t; u, z)$ is a baseline hazard function. In the classic Cox model, $h_0(t; u, z)$ is specified nonparametrically. However, in Bayesian inference, a parametric form of $h_0(t; u, z)$ is usually preferred for computational convenience. The most common parametric choice for $h_0(t; u, z)$ is the Weibull model (Abrams et al., 1996): $h_0(t; u, z) = \exp(\alpha_{u,z}) t^{\varphi_{u,z}-1}$ with $\varphi_{u,z} > 0$, based on which the Cox model becomes

$$h(t; u, z) = t^{\varphi_{u,z}-1} \exp(\alpha_{u,z} + X_i' \beta_{u,z}), \quad \varphi_{u,z} > 0. \qquad (23)$$

We allow the model parameters α, β, φ to differ between strata and arms to provide flexibility. The probability density function and the survival function are uniquely determined by the hazard function $h(t; u, z)$. Specifically, under the Weibull-Cox model (23), we can show that

$$\begin{aligned}
&\Pr(T_i \geq t \mid U_i = u, Z_i = z, X_i) = \exp\left(-\int_0^t h(w; u, z)\, \mathrm{d}w\right) \\
&= \exp\left\{-\frac{1}{\varphi_{u,z}} t^{\varphi_{u,z}} \exp(\alpha_{u,z} + X_i' \beta_{u,z})\right\}, \qquad (24) \\
&\Pr(T_i = t \mid U_i = u, Z_i = z, X_i) \\
&\quad = h(t; u, z) \exp\left(-\int_0^t h(w; u, z)\, \mathrm{d}w\right) \\
&\quad = t^{\varphi_{u,z}-1} \exp(\alpha_{u,z} + X_i' \beta_{u,z}) \\
&\quad \exp\left\{-\frac{1}{\varphi_{u,z}} t^{\varphi_{u,z}} \exp(\alpha_{u,z} + X_i' \beta_{u,z})\right\}. \qquad (25)
\end{aligned}$$

Then, we can plug Y_i in place of t into the above formulae to calculate the outcome-model component in the observed data likelihood (18).

For Bayesian inference, we need to specify prior distribution of the parameters in the strata-model and the outcome-model. We choose standard weakly informative priors: a flat prior for the intercepts, $p(\eta_u) \propto 1$ and $p(\alpha_{u,z}) \propto 1$; a Gaussian prior for the coefficients: $p(\xi_u) \sim \mathrm{N}(0, \sigma_\xi)$ and $p(\beta_{u,z}) \sim \mathrm{N}(0, \sigma_\beta)$, with large (eg, 100) prior variance σ_ξ and σ_β; a flat prior for the log shape parameter in the Weibull model $p(\log \varphi_{u,z}) \propto 1$.

Given the strata-model and outcome-model and the prior distributions of the model parameters θ, posterior sampling of θ is traditionally obtained via the Monte Carlo Markov chains (MCMC), for example, a Gibbs sampler (Geman and Geman, 1984) or a Metropolis-Hasting (Hastings, 1970) algorithm. In these methods, each unit's latent principal stratum U_i is sampled along with the model parameters. In this paper, we leverage the Stan programming language (Carpenter et al., 2017; Stan Development Team, 2022) to facilitate automatic sampling, bypassing the analytical derivation of MCMC. Posterior sampling in **Stan** is implemented via the Hamiltonian Monte Carlo (HMC) method (Neal, 2011), which uses the derivatives of the posterior

density function to sample efficiently across the parameter space. We provide the code of our simulations in the Online Supplementary Material.

Once the posterior samples of the parameters in models (20) and (23) are obtained, we can use them to calculate the posterior distribution of the principal causal estimands (3) and (4) based on their analytical relationship, which is derived below. Both estimands rely on the stratum-specific survival function, $S(t; u, z, \theta) = \Pr(T(z) > t \mid U = u, \theta) = \Pr(T > t \mid U = u, Z = z, \theta)$, which can be decomposed as follows

$$\begin{aligned}\Pr(T > t \mid U, Z, \theta) &= \frac{\int \Pr(T > t \mid U, Z, X = x, \theta)\Pr(U \mid X = x, \theta)\Pr(Z \mid X = x, \theta)\Pr(X = x)\,dx}{\int \Pr(U \mid X = x, \theta)\Pr(Z \mid X = x, \theta)\Pr(X = x)\,dx} \\ &= \frac{\int \Pr(T > t \mid U, Z, X = x, \theta)\Pr(U \mid X = x, \theta)\Pr(X = x)\,dx}{\int \Pr(U \mid X = x, \theta)\Pr(X = x)\,dx}. \end{aligned} \tag{26}$$

Given the randomization or the unconfoundedness assumption, the treatment assignment $\Pr(Z \mid X, \theta)$ is a constant with respect to the outcome and thus drops out from the expression as long as its parameters are *a priori* distinct and independent of the parameters in the strata-model and outcome-model. Denote $A_i(u, \theta) = \Pr(U = u \mid X_i, \theta)$, the analytical form of which under the strata-model (20) is given in (21); denote $B_i(t; u, z, \theta) = \Pr(T_i > t \mid U = u, Z = z, X_i, \theta)$, the analytical form of which under the outcome-model (23) is given in (24). Plugging the posterior samples of θ into the analytical form of $A(u, \theta)$ and $B(t; u, z, \theta)$ in Equation 26, we can obtain the posterior distribution of the survival function $S(t; u, z)$ for any stratum u and assignment z at any time t:

$$\widehat{S}(t; u, z, \theta) = \frac{\sum_{i=1}^{N} A_i(u, \theta)B_i(t; u, z, \theta)}{\sum_{i=1}^{N} A_i(u, \theta)}. \tag{27}$$

Then, we can obtain any summary statistic, for example, posterior mean or credible intervals, from the posterior distribution; for example, averaging over the posterior samples of θ provides the posterior mean of the survival function $\widehat{S}(t; u, z)$.

The posterior distribution of the principal causal effects on the survival probability and the restricted average causal effect are

$$\widehat{\tau}_u^{SPCE}(t) = \widehat{S}(t; u, 1) - \widehat{S}(t; u, 0), \tag{28}$$

$$\widehat{\tau}_u^{RACE}(t) = \int_0^t \widehat{S}(w; u, 1)\,dw - \int_0^t \widehat{S}(w; u, 0)\,dw, \tag{29}$$

respectively. Define $\tilde{B}_i(t; u, z, \theta) = \int_0^t B_i(w; u, z, \theta)\,dw$. Then, the integral in (29) can be expressed as

$$\int_0^t \widehat{S}(w; u, z)\,dw = \frac{\sum_{i=1}^{N} A_i(u, \theta)\tilde{B}_i(t; u, z, \theta)}{\sum_{i=1}^{N} A_i(u, \theta)}. \tag{30}$$

When the outcome-model is the Weibull-Cox model (23), $\tilde{B}_i$ can be derived analytically. Specifically, $\tilde{B}_i(t; u, z, \theta)$ is given by

$$\tilde{B}_i(t; u, z, \theta) = \left[\varphi_{u,z} c_{u,z}^{1/\varphi_{u,z}}\right]^{-1} \gamma\left(\frac{1}{\varphi_{u,z}}, c_{u,z} t^{\varphi}_{u,z}\right),$$

with $c_{u,z} = \frac{1}{\varphi_{u,z}} \exp(\alpha_{u,z} + X_i' \beta_{u,z})$ and $\gamma(\alpha, x) := \int_0^x w^{\alpha-1} \exp(-w)\,dw$ being the lower incomplete γ function, which can be readily calculated as a built-in function in most standard statistical software. Consequently, (30) can be easily evaluated; plugging it into (28) and (29) yields the posterior estimates of the two estimands.

When the outcome-model is not Weibull-Cox, the analytical forms of $\tilde{B}_i$ is usually intractable. Nonetheless, one can obtain the posterior estimates of the principal causal estimands numerically based on the posterior samples of the survival function $\widehat{S}(t; u, z, \theta)$ in (27). The basic routine for numerically computing the integral is as follows. Let $\{w_k = (k/K)t: k = 0, \ldots, K\}$ be $K + 1$ equally spaced points in $[0, t]$. With a sufficiently large K, the integral can be approximated by the Trapezoid rule as

$$\begin{aligned}&\int_0^t \widehat{S}(w; u, z)\,dw \\ &\quad\approx \frac{t}{K} \sum_{k=1}^{K} \left[\frac{1}{2}\widehat{S}(w_{k-1}; u, z) + \frac{1}{2}\widehat{S}(w_k; u, z)\right]. \end{aligned} \tag{31}$$

This can be readily modified to more efficient algorithms such as the Simpson's rule (Davis and Rabinowitz, 2007). Posterior inference of the causal estimands in (28) and (29) can be obtained by plugging in the numerical estimate (31) with the posterior samples of θ. We developed an R package **PStrata** (Liu and Li, 2023) to implement the estimation.

4 EXTENSION TO THE CASE WITH BOTH TREATMENT SWITCHING AND DISCONTINUATION

In real-world clinical trials, besides treatment switching, discontinuation is another common case of noncompliance. Discontinuation refers to the situation where the subjects stopped taking any intervention. We do not consider more complex situations such as some subjects stopped reporting their treatment status, that is, their intermediate variable D is also censored. In the ADAPTABLE trial, over 13% patients discontinued aspirin, with significantly higher occurrence in the 325 mg arm than the 81 mg arm. This indicates that discontinuation is likely confounded. Below, we extend principal stratification to incorporate both treatment switching and discontinuation, tailored to ADAPTABLE. Specifically, we treat discontinuation as an additional level of intermediate variable D, denoted by $*$. Now, D has 3 levels: $*$, 0, 1. Without restriction, this induces nine principal strata. In ADAPTABLE, discontinuation is likely caused by the side effects of aspirin and is more common in the 325 mg arm. Therefore, we make a set of new monotonicity assumptions.

TABLE 1 Composition of principal strata in observed cells of assigned and actual treatment (Z, D).

	$D = *$	$D = 0$	$D = 1$
$Z = 0$	$(*, *)$	never-takers, compliers	always-takers, defiers
$Z = 1$	$(*, *)$, $(0, *)$	never-takers, defiers	always-takers, compliers

TABLE 2 Parameters of the true outcome-model in simulations, where $\mu_{sz} = \alpha_{sz} + X'\beta_{sz}$.

	$\Pr(Y(0) \mid U = u)$		$\Pr(Y(1) \mid U = u)$		
u	φ_{s0}	μ_{s0}	φ_{s1}	μ_{s1}	Scenario
n	2.0	$-3 + 0.4X_1 + 0.1X_2$	2.0	$-3 + 0.4X_1 + 0.1X_2$	(i): ER
c	1.5	$-2.4 - 0.1X_1 + 0.3X_2$	1.5	$-1.8 + 0.2X_1 - 0.2X_2$	
a	1.0	$-1.2 + 0.1X_1$	1.0	$-1.2 + 0.1X_1$	
n	2.0	$-3 + 0.4X_1 + 0.1X_2$	2.0	$-2.4 + 0.2X_1 - 0.3X_2$	(ii): Non-ER
c	1.5	$-2.4 - 0.1X_1 + 0.3X_2$	1.5	$-1.8 + 0.2X_1 - 0.2X_2$	
a	1.0	$-1.2 + 0.1X_1$	1.0	$-0.6 + 0.2X_2$	

Assumption 5 *(i) If $D_i(0) = *$, then $D_i(1) = *$, namely, if a patient would discontinue under the 81 mg assignment, then he/she would also discontinue under the 325 mg assignment; (ii) if $D_i(0) = 1$, then $D_i(1) = 1$, namely, if a patient would take 325 mg under the 81 mg assignment, then he/she would also take 325 mg under the 325 mg assignment.*

We do not impose any restriction for $D_i(1)$ when $D_i(0) = 0$, namely, if a patient would take 81 mg under the 81 mg assignment, then he/she would either discontinue or take 81 mg or take 325 mg under the 325 mg assignment. Assumption 5(ii) implies the original monotonicity (Assumption 2) when there is no discontinuation. Under Assumption 5, there are 5 principal strata: $U \in \{(*, *), (0, *), (0, 0), (0, 1), (1, 1)\}$. The composition of principal strata in the observed cells of (Z, D) is listed in Table 1. We can also, but not necessarily, extend exclusion restriction accordingly. Regardless of whether we make these assumptions, we can readily adapt the mixture-model-based estimation strategies in Section 3 to estimate the principal causal effects, the technical details of which are thus omitted here.

5 SIMULATIONS

We conduct simulation studies to evaluate the performance of the proposed methods under a range of common settings. In particular, we examine (i) how exclusion restriction affects the inference, and (ii) how the methods perform under model misspecification. Throughout the simulations, we focus on the classic setting of binary noncompliance and maintain the simple monotonicity assumption (Assumption 2) to convey the main message.

We simulate a randomized experiment with $N = 2000$ units. The treatment assignment Z_i is independently drawn from a *Bernoulli*(0.5). Each unit i has a binary covariate $X_{1i} \sim$ *Bernoulli*(0.5) and a continuous covariate $X_{2i} \sim \mathcal{N}(0, 1)$. The principal stratum U_i is generated from the multinomial strata-model (20), with the never-takers as the reference stratum and

$$\log\left\{\frac{\Pr(U_i = c \mid X_i)}{\Pr(U_i = n \mid X_i)}\right\} = 1 + 0.1X_{1i} - 0.2X_{2i}, \quad (32)$$

$$\log\left\{\frac{\Pr(U_i = a \mid X_i)}{\Pr(U_i = n \mid X_i)}\right\} = -0.6X_{1i} - 0.8X_{2i}. \quad (33)$$

This model leads to a sample proportion of 0.2, 0.6, and 0.2 of never-takers, compliers, and always-takers, respectively. Each unit's actual treatment D_i is determined by Z_i and U_i. The true uncensored failure time is generated from the Weibull-Cox model (23), with separate parameters for each of the 6 combination of stratum and the treatment assignment (u, z) with $s \in \{n, c, a\}$ and $z \in \{0, 1\}$. We simulated 2 scenarios of the underlying truth: (i) ER holds; and (ii) ER does not hold. When ER holds, the outcome-model is the same between assignment z for always-takers and never-takers, and thus we need to specify 4 different outcome-models. When ER does not hold, we need to specify 6 outcome-models. Table 2 presents the outcome-model parameters for these 2 scenarios. Following the independent censoring assumption (Assumption 4), we draw the censoring time C_i independently from an exponential distribution with rate $\exp(-0.3 + 0.1X_{1i} - 0.2X_{2i})$, leading to a marginal event rate of approximately 15%.

Simulation 1: correct models. We first consider a setting where the strata- and outcome-models are corrected specified. For each simulation scenario, we conduct 2 analyses, with or without ER, so in total, we have 4 combinations of truth and model fit regarding ER. For each combination, we use our R package **PStrata** (Liu and Li, 2023) to run HMC with 6 chains, each chain with 1000 iterations including 500 run-in iterations, resulting in a total of 3000 nonrun-in iterations of parameters drawn from the posterior distribution. Mixing of the chains is deemed good under each scenario from the traceplots. The posterior mean and the 95% credible band of the survival probability curves under each combination are presented in Figure 1. We observed a few patterns from the simulations. First, in all scenarios, the distributions $\Pr(T(0) \mid U = a)$ and $\Pr(T(1) \mid U = n)$ are correctly estimated with low uncertainty. As discussed before, this is because both distributions are fully identified from the observed data. Second, as shown in the first 2 panels in Figure 1, when the underlying truth satisfies ER, our proposed method recovers the truth regardless of whether assuming ER, but not assuming ER leads to wider credible intervals, particularly for the

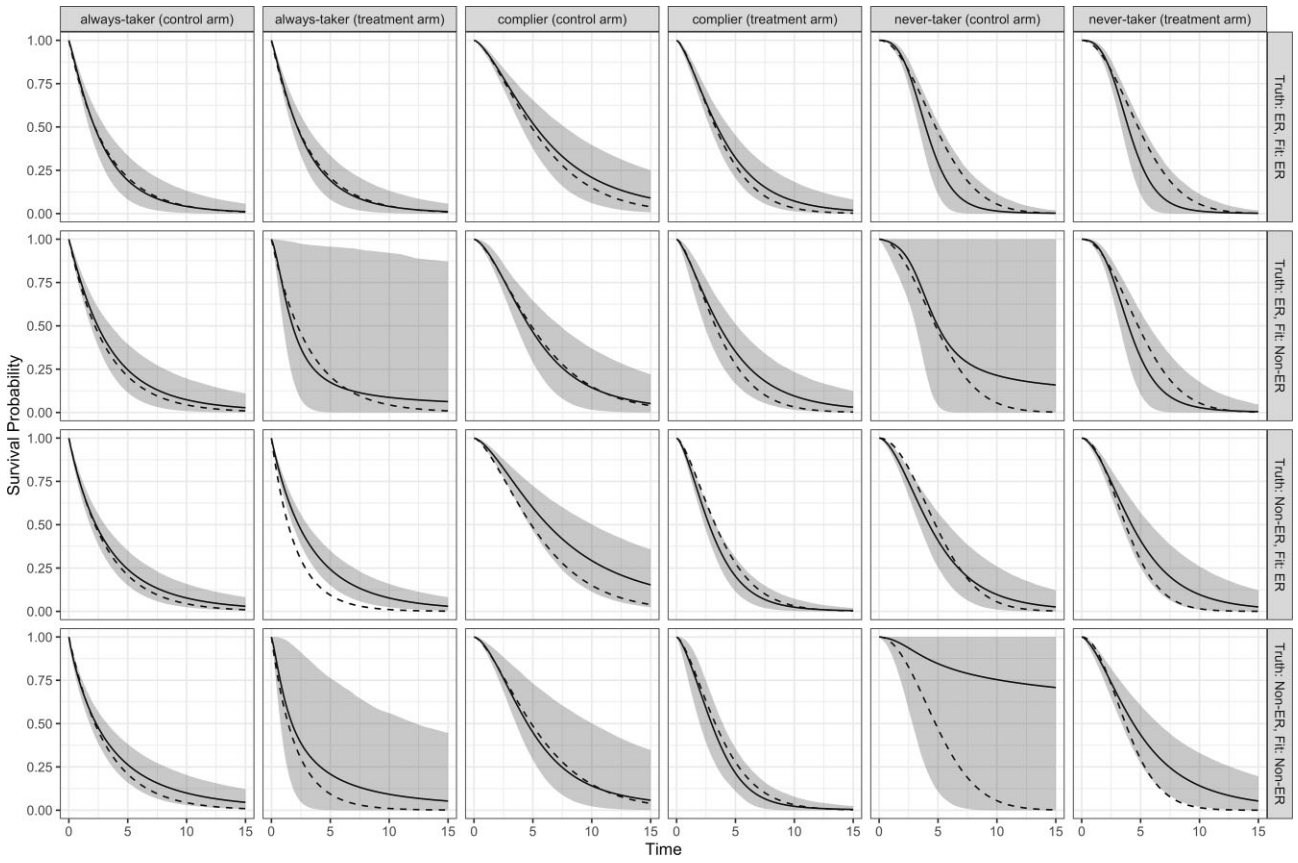

FIGURE 1 Posterior survival probability curves for each stratum-assignment combination with the true outcome being generated from a Weibull-Cox model with or without ER and fitted with a Weibull-Cox model with or without exclusion restriction (ER). The dashed line is the true survival curve. The solid line is the posterior mean survival curve with the associated 95% credible band being the shaded area.

distributions $\Pr(T(1) \mid U = a)$ and $\Pr(T(0) \mid U = n)$. This is partially due to the relative low proportion of always-takers and never-takers. Third, when the underlying truth does not satisfy ER but we fit the model with ER, all the 4 distributions that require disentangling latent mixtures (the middle 4 graphs in the third panel in Figure 1) are estimated with large bias and low uncertainty. In contrast, when the underlying truth does not satisfy ER and we fit the model without ER, the 4 distributions are covered by their corresponding credible intervals, which are wide but correctly reflect the inherent large uncertainties in disentangling mixtures. Similar patterns are observed in the model parameters and RACE, the results of which are thus delegated to the Supplementary Materials (Web Figure 1, Web Table 1, and Web Table 2).

Simulation 2: misspecified models. We now consider a more challenging case where the models are partially misspecified. We simulate from Scenario (i) in Simulation 1, which assumes a Weibull-Cox outcome model under ER, but fit the data with 2 types of misspecification: (i) the forms of the strata-model and the outcome-model are correctly specified, but both miss covariate X_2; and (ii) the strata-model is correctly specified, but the form of the outcome-model is misspecified as a log-linear AFT model with normally distributed error terms. The strata-model and the censoring model are simulated as before. Figure 2 presents the resulting estimated survival probability by stratum and arm. When ER is imposed, the estimated survival curves and the credible regions vary little across different model specification. This is not surprising because the survival probability can be nonparametrically identified in this case. Nevertheless, imposing an AFT model while the truth follows a Cox model leads to slightly larger bias and variance. When ER is not assumed, the fitted survival curves deviate from the truth and the credible regions become wide under misspecified models, especially for the 2 distributions $\Pr(T(1) \mid U = a)$ and $\Pr(T(0) \mid U = n)$ that require disentangling mixtures and with small sample size. We repeat the simulations under the truth that ER does not hold. Similar patterns are observed and thus the results are delegated to the Online Supplementary Material (Web Figure 2). It is worth noting that, as expected, in this case, the bias due to misspecification is generally larger than its counterpart when ER does hold.

Simulation 3: proportion of compliers. We also examine the effect of the proportion of compliers under correctly specified models. When the proportion of compliers is not too low (above 30%), the estimated survival curves and their confidence regions do not vary much with the proportion. As the proportion of compliers continues to decrease, the bias and the uncertainty of the complier survival curves increase dramatically. The detailed description and results are delegated to the Supplementary Materials (Web Figure 3).

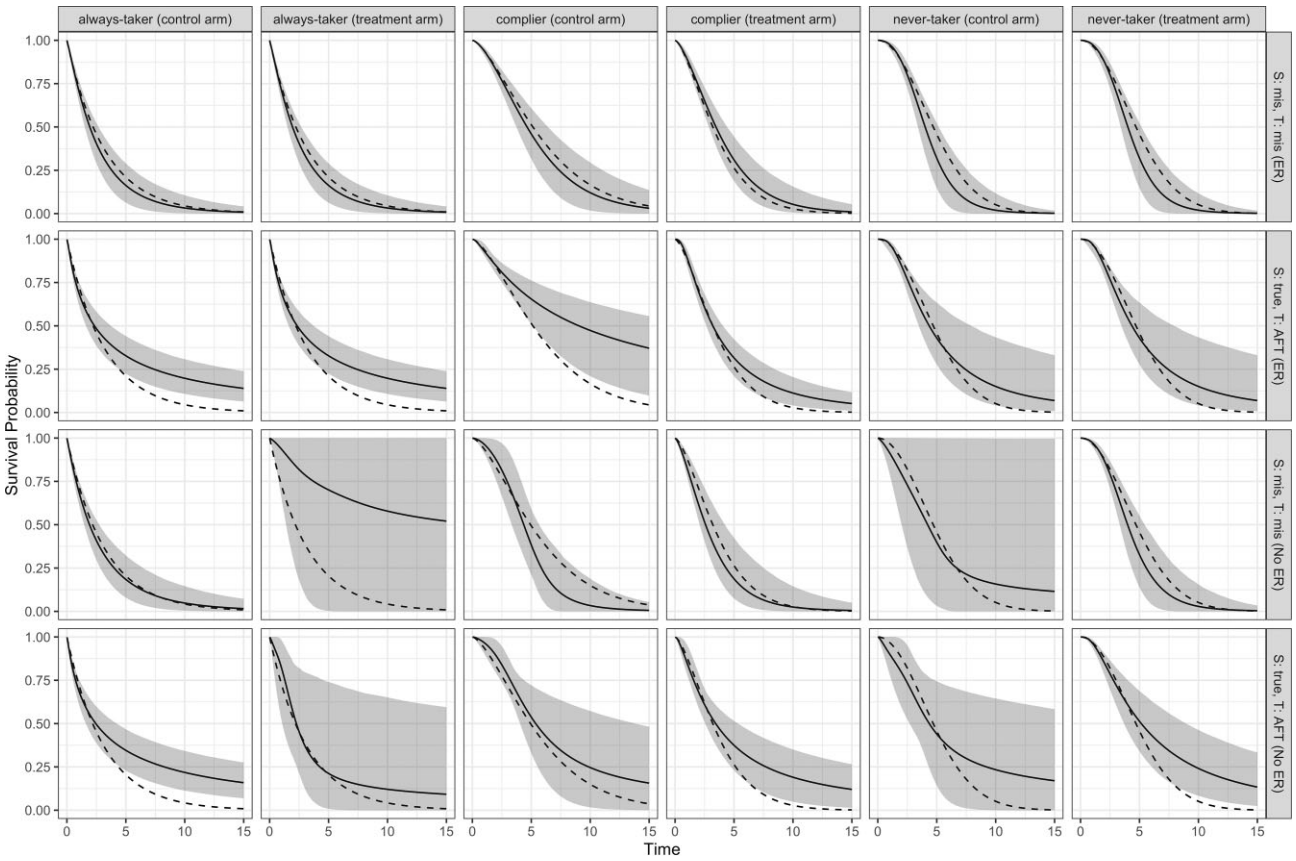

FIGURE 2 Posterior survival probability curves in Simulation 2 when exclusion restriction (ER) holds in truth. The top 2 panels show the results when the models are fit assuming ER, with misspecification type (i) (ie, both strata-model and outcome-model miss an important covariate) and (ii) [ie, the form of outcome-model is misspecified as a accelerated failure time (AFT) model], respectively. The bottom 2 panels show the corresponding results when the models are fit not assuming ER.

We draw a few conclusions from the simulations. First, for the distributions that need disentangling mixtures, the number of units in each stratum directly affects the estimating uncertainties. For example, in our simulations, there are fewest always-takers, followed by never-takers, and most compliers. The difference in the strata size is clearly reflected in the width of the corresponding credible intervals (Figure 1), with $\Pr(T(1) \mid U = a)$ consistently having the widest interval, followed by $\Pr(T(0) \mid U = n)$, whereas the distribution of the compliers having the tightest intervals. Second, ER plays an important role in the analysis. When correctly assumed, ER can protect estimation against model misspecification and significantly sharpen the inference; however, when incorrectly assumed, ER leads to large bias and sensitivity to model misspecification. In real applications, we suggest the analysts to first judge the plausibility of ER according to case-specific substantive knowledge and then perform the analysis with and without ER in parallel as a form of sensitivity analysis. Third, estimation of the nonparametric causal estimands (eg, survival probability) is largely robust to the specification of the strata and outcome model when ER is assumed. When ER is not assumed, the fully nonparametrically identified survival curves $\Pr(T(0) \mid U = a)$ and $\Pr(T(1) \mid U = 0)$ are still robust to the specification of the models, but the survival curves for other stratum-treatment combinations are prone to misspecification of models, particularly for strata with small sample sizes. Fourth, misspecification of the distributional form (eg, Weibull-Cox versus AFT) generally leads to larger bias than the misspecification of covariates (eg, whether including a covariate in the model).

6 APPLICATION TO THE ADAPTABLE TRIAL

The ADAPTABLE is a multicenter, open label, pragmatic randomized controlled trial designed to assess whether an aspirin intake of 325 mg per day, compared to that of 81 mg per day, would result in a lower risk of major adverse cardiovascular event (MACE), including death from any cause and hospitalization for myocardial infarction, or stroke, with patients with atherosclerotic cardiovascular disease (Jones et al. 2021). The trial recruited 15 076 participants, with prior aspirin use reported by 96% of them, among which the prior regular dose was 81 mg for 85.4% patients, 162 mg for 2.3% patients, and 325 mg for 12.2% patients. At the start of the trial, all patients were randomized to 81 mg ($N = 7540$) and 325 mg ($N = 7536$). The primary outcome of the trial is the time to the first occurrence of MACE. The overall rate of MACE during the study is 7.2%; the median time to event is 26.2 months, with the interquartile range being 19.0 to 34.9 months. During the follow up, 1977 (13.1%) patients discontinued aspirin, with 15.8% discontinued in the 325 mg arm and 10.4% in the 81 mg arm. Also, we restrict the maxi-

TABLE 3 Baseline characteristics by randomized arm and adherence status in the ADAPTABLE (**A**spirin **D**osing: **A** **P**atient-Centric **T**rial **A**ssessing **B**enefits and **L**ong-Term **E**ffectiveness) trial.

	Randomized dose: 81 mg			Randomized dose: 325 mg		
Characteristic	**Overall ($N = 6556$)**	**Adherent ($N = 6119$)**	**Nonadherent ($N = 437$)**	**Overall ($N = 6027$)**	**Adherent ($N = 3698$)**	**Nonadherent ($N = 2329$)**
Age (year)	68 (61-74)	68 (61-74)	66 (59-73)	67 (61-73)	67 (61-73)	68 (61-74)
Weight (kg)	91 (79-104)	91 (79-104)	92 (80-105)	91 (78-104)	92 (80-105)	89 (77-103)
Female	1942 (29.6)	1819 (29.7)	123 (28.1)	1856 (30.8)	1093 (29.6)	763 (32.8)
Race						
White	5375 (82.0)	5043 (82.4)	332 (76.0)	4979 (82.6)	3214 (86.9)	1765 (75.8)
Black or African American	588 (9.0)	537 (8.8)	51 (11.7)	533 (8.8)	238 (6.4)	295 (12.7)
Other	361 (5.5)	333 (5.4)	28 (6.4)	294 (4.9)	153 (4.1)	141 (6.1)
Hispanic ethnicity	211 (3.2)	194 (3.2)	17 (3.9)	176 (2.9)	96 (2.6)	80 (3.4)
Current smoker	627 (9.6)	561 (9.2)	66 (15.1)	588 (9.8)	334 (9.0)	254 (10.9)
Noninternet user	839 (12.8)	752 (12.3)	87 (19.9)	725 (12.0)	316 (8.5)	409 (17.6)
Medical history						
Coronary artery disease	5990 (91.4)	5590 (91.4)	400 (91.5)	5557 (92.2)	3395 (91.8)	2162 (92.8)
Myocardial infarction	2337 (35.6)	2165 (35.4)	172 (39.4)	2134 (35.4)	1266 (34.2)	868 (37.3)
Coronary-artery bypass grafting	1566 (23.9)	1464 (23.9)	102 (23.3)	1431 (23.7)	886 (24.0)	545 (23.4)
Percutaneous coronary intervention	2667 (40.7)	2494 (40.8)	173 (39.6)	2414 (40.1)	1399 (37.8)	1015 (43.6)
Cerebrovascular disease	1143 (17.4)	1046 (17.1)	97 (22.2)	1026 (17.0)	587 (15.9)	439 (18.8)
Hypertension	5458 (83.3)	5084 (83.1)	374 (85.6)	5041 (83.6)	3064 (82.9)	1977 (84.9)
Hyperlipidemia	5653 (86.2)	5284 (86.4)	369 (84.4)	5232 (86.8)	3224 (87.2)	2008 (86.2)
Atrial fibrillation	507 (7.7)	468 (7.6)	39 (8.9)	490 (8.1)	301 (8.1)	189 (8.1)
Congestive heart failure	1457 (22.2)	1345 (22.0)	112 (25.6)	1415 (23.5)	814 (22.0)	601 (25.8)
Peripheral artery disease	1487 (22.7)	1393 (22.6)	104 (23.8)	1432 (23.8)	803 (21.7)	629 (27.0)
Diabetes	2452 (37.4)	2285 (37.3)	167 (38.2)	2303 (38.2)	1402 (37.9)	901 (38.7)
Peptic ulcer disease	193 (2.9)	184 (3.0)	9 (2.1)	169 (2.8)	92 (2.5)	77 (3.3)
History of bleeding	518 (7.9)	479 (7.8)	39 (8.9)	530 (8.8)	289 (7.8)	241 (10.3)
Significant bleeding disorder	70 (1.1)	65 (1.1)	5 (1.1)	74 (1.2)	46 (1.2)	28 (1.2)
Significant GI bleed	391 (6.0)	369 (6.0)	22 (5.0)	389 (6.5)	210 (5.7)	179 (7.7)
Intracranial hemorrhage	84 (1.3)	68 (1.1)	16 (3.7)	95 (1.6)	48 (1.3)	47 (2.0)
Prior medications						
Prior aspirin use	6272 (95.7)	5866 (95.9)	406 (92.9)	5765 (95.7)	3541 (95.8)	2224 (95.5)
Prior dose: 81 mg	5320 (84.8)	5141 (87.6)	179 (44.1)	4906 (85.1)	2910 (82.2)	1996 (89.7)
Prior dose: 162 mg	156 (2.5)	123 (2.1)	33 (8.1)	130 (2.2)	94 (2.7)	36 (1.6)
Prior dose: 325 mg	783 (12.5)	594 (10.1)	189 (46.6)	721 (12.5)	536 (15.1)	185 (8.3)
P2Y12 inhibitor	1420 (21.7)	1325 (21.7)	95 (21.7)	1307 (21.7)	769 (20.8)	538 (23.1)

mum follow-up time 36 months because few events occurred after that.

Nonadherence to the randomized assignment is prevalent in ADAPTABLE: 33.5% of the patients in the 325 mg arm switched to low dose, and 6.1% of the patients in the 81 mg arm switched to high dose during the study. The high prevalence of nonadherence in the 325 mg arm is partially due to fact that the high dosage often causes gastrointestinal problems among patients. The timing of dose-switching varies across the patients: though a large percentage of switching occurred within the first 6 months after randomization, a non-negligible number of patients switched after that. To simplify the analysis, we dichotomize noncompliance using switching at 6 month as the cutoff. Table 3 presents the baseline characteristics by randomized arm and adherence behavior. Almost all covariates are well balanced between the arms, but marked difference in some covariates between adherent and nonadherent patients is observed. For example, there are higher proportion of African Americans, smokers, internet-users, previous gastrointestinal bleeding among the nonadherent patients in both arms, and generally nonadherent patients tend to switch to their previous aspirin dosage regardless of the randomized dosage.

The ITT Kaplan-Meier curves of the time to the first occurrence of death or hospitalization by the randomized arms are nearly identical throughout the study (Web Figure 4): the mean time is 26.3 months in each arm, with the standard deviation being 9.71 and 9.67 months for the 81 and 325 mg arm, respectively. The event rate is 7.3% in the 81 mg arm and 7.2% in the 325 mg arm. The ITT difference in survival probability between the 81 and 325 mg arms is 0.002 with 95% CI $(-0.001, 0.014)$ at 12 months, -0.001 with 95% CI $(-0.019, 0.017)$ at 24 months, and 0.001 with 95% CI $(-0.025, 0.027)$ at 36 months. Based on the ITT results, one would conclude that there is no clinical difference in the risk of major adverse cardiovascular event between taking 81 versus 325 mg aspirin among patients with atherosclerotic cardiovascular disease (Jones et al., 2021). However, the ITT analysis fails to accommodate the nuances of the large proportion of nonadherence and the potential treatment effect heterogeneity. Therefore, we applied the proposed PS method to re-analyze the data.

TABLE 4 Fitted coefficients $\beta_{u,z}$ of the Weibull-Cox outcome-model by stratum u and randomization arm z in the ADAPTABLE (**A**spirin **D**osing: **A** **P**atient-Centric **T**rial **A**ssessing **B**enefits and **L**ong-Term **E**ffectiveness) trial.

	Always-low		Compliers		Always-high	
	81 mg	**325 mg**	**81 mg**	**325 mg**	**81 mg**	**325 mg**
Age (scaled)	0.09 (−0.03, 0.22)	0.08 (−0.04, 0.20)	0.17 (−0.67, 1.18)	0.28 (0.05, 0.53)∗	0.32 (0.05, 0.59)∗	0.42 (−0.54, 1.35)
Female	−0.04 (−0.28, 0.20)	0.02 (−0.25, 0.28)	−0.58 (−2.29, 0.93)	0.04 (−0.51, 0.42)	0.22 (−0.36, 0.78)	−0.38 (−2.19, 1.05)
White	−0.02 (−0.28, 0.25)	−0.08 (−0.37, 0.21)	−1.65 (−3.10, −0.26)∗	−0.39 (−0.83, 0.08)	−0.44 (−1.01, 0.16)	−0.67 (−2.29, 0.72)
Hispanic	0.14 (−0.40, 0.27)	0.20 (−0.39, 0.75)	1.94 (0.51, 3.43)∗	0.09 (−0.73, 0.84)	0.81 (−0.03, 1.58)	1.22 (−0.15, 2.80)
Myocardial infarction	0.31 (0.09, 0.53)∗	0.41 (0.13, 0.67)∗	−0.90 (−2.46, 0.60)	0.33 (−0.06, 0.69)	−0.07 (−0.60, 0.44)	−0.45 (−2.21, 0.97)
Atrial fibrillation	0.13 (−0.25, 0.49)	0.11 (−0.33, 0.49)	−0.29 (−2.04, 1.31)	0.08 (−0.63, 0.62)	0.34 (−0.43, 1.02)	0.05 (−1.84, 1.56)
P2Y12	0.44 (0.21, 0.68)∗	0.31 (0.04, 0.57)∗	−0.52 (−2.28, 1.08)	0.26 (−0.24, 0.65)	0.60 (0.04, 1.13)∗	−0.10 (−1.95, 1.41)

We first conduct a PS analysis excluding patients who discontinued aspirin during the follow up. In the PS analysis, we maintain the randomization, conditional independent censoring, and monotonicity assumptions, which are deemed plausible in ADAPTABLE. However, we do not impose ER here. As discussed earlier, though we dichotomized noncompliance in the analysis for simplicity, the timing of noncompliance varies among patients. In fact, most noncomplying patients complied to their assigned treatment at the beginning, but then switched to the other treatment after some time. Therefore, there could be a direct effect of the initial assignment on the outcome. Also, to reflect the specific context, we term the 3 strata as always-low $[U = (0, 0)]$, consisting of patients who would take 81 mg regardless of the initial randomization, compliers $[U = (0, 1)]$, consisting of patients who would take 81 mg if assigned to 81 mg and would take 325 mg if assigned to 325 mg, and always-high $[U = (1, 1)]$ patients, consisting of patients who would take 325 mg regardless of the initial randomization. These 3 strata correspond to never-taker, compliers, and always-taker, respectively, in the standard nomenclature.

We fit the multinomial strata model (20) and the Weibull-Cox outcome model (23) to the data. We include covariates that are clinically deemed predictive of the adherence behavior or the outcome. These include patient demographics (age, sex, race, ethnicity), whether the patient is a noninternet user, medical history (atrial fibrillation, percutaneous coronary intervention, bleeding, prior myocardial infarction), and prior medication use (baseline P2Y12, prior aspirin dose). We use **PStrata** (Liu and Li, 2023) to run 6 chains, each with 1000 iterations including 400 warm-up iterations, resulting in 3600 posterior draws in total. The estimated proportion of the always-high, compliers, and always-low stratum is 6.6% (95% CI: 6.1% to 7.2%), 54.2% (95% CI: 52.9% to 55.5%), and 39.2% (95% CI: 38.0% to 40.3%), respectively. The coefficients of the strata-model are delegated to the Supplementary Materials (Web Table 3). The coefficients of the outcome-model are displayed in Table 4, showing noticeable heterogeneity between the strata. We provide the stratum-specific summary of selected covariates as follows. In each HMC iteration, we calculate the probability that each patient belongs to each stratum $p_{i,s}$. Then, for a given stratum u, we calculate the weighted average of the covariates X_i over all patients in the population: $\bar{X}_s = \sum_i p_{i,s}X_i / \sum_i p_{i,s}$, as a summary of the covariates in stratum u in that specific HMC iteration. Repeating this for all iterations provides the posterior distribution of $\bar{X}_s$, from which we can calculate the posterior mean and 95% CIs, shown in Table 5. The strata appear different in many covariates. For example, the compliers consist of more White patients and fewer noninternet users. Moreover, patients with history of peripheral artery disease are more likely to take low dose regardless of the assignment. Consistent with the previous observation, the most notable difference lies in the prior aspirin dosage: patients tend to take the dose as they had previously taken.

Figure 3 presents the estimated curves of the potential survival probabilities under each treatment in each stratum. The most striking observation is the heterogeneous treatment effect between the strata. Specifically, for the compliers, being assigned to 81 mg increases their survival probability compared with being assigned to 325 mg throughout the follow-up period: the difference in survival probability between the 81 and 325 mg at 12, 24, and 36 months is −0.023 (95% CI: −0.030 to −0.014), −0.045 (95% CI: −0.057 to −0.030), and −0.067 (95% CI: −0.085 to −0.045), respectively. In contrast, for the patients in the always-low stratum, being assigned to 325 mg increases their survival probability comparing with being assigned to 81 mg throughout the follow-up period: the difference in survival probability between the 81 and 325 mg at 12, 24, and 36 months is 0.030 (95% CI: 0.017 to 0.042), 0.060 (95% CI: 0.040 to 0.078), and 0.087 (95% CI: 0.058 to 0.114), respectively. The effects are statistically significant as the credible intervals do not cover 0. In comparison, the effects of the assignment for the patients in the always-high stratum are estimated with much larger uncertainties: the difference in survival probability between being assigned to the 81 and 325 mg at 12, 24, and 36 months is 0.025 (95% CI: −0.040 to 0.067), −0.075 (95% CI: 0.042 to 0.113), and 0.056 (95% CI: −0.108 to 0.153). The wide intervals are likely due to the low proportion of the always-high stratum (6%)

TABLE 5 Stratum-specific summary of pretreatment variables.

	Always-low		Compliers		Always-high	
Variable	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Age (year)	67.48	(67.19, 67.76)	66.42	(66.20, 66.64)	65.64	(64.81, 66.46)
Female (%)	32.37	(31.03, 33.70)	28.86	(27.80, 29.97)	30.73	(26.82, 34.76)
White (%)	75.81	(74.63, 76.98)	87.74	(86.78, 88.73)	77.06	(73.59, 80.35)
Hispanic (%)	3.81	(3.29, 4.35)	2.59	(2.14, 3.02)	3.78	(2.31, 5.46)
Noninternet user (%)	18.98	(17.95, 19.99)	7.80	(6.96, 8.61)	20.24	(16.95, 23.62)
Myocardial infarction (%)	37.99	(37.45, 38.54)	34.92	(34.47, 35.36)	36.65	(35.10, 38.15)
Atrial fibrillation (%)	8.27	(7.45, 9.10)	7.86	(7.21, 8.51)	9.13	(6.91, 11.59)
Percutaneous coronary intervention (%)	44.80	(43.32, 46.24)	38.78	(37.60, 39.94)	39.74	(35.60, 43.80)
Bleeding (%)	10.19	(9.42, 10.97)	7.21	(6.59, 7.86)	9.75	(7.44, 12.37)
P2Y12 (%)	24.40	(23.13, 25.65)	21.04	(20.04, 22.02)	22.40	(18.99, 26.02)
Prior aspirin dose						
81 mg (%)	87.02	(86.02, 88.03)	83.86	(82.93, 84.79)	43.86	(39.66, 48.15)
162 mg (%)	1.79	(1.40, 2.21)	2.02	(1.61, 2.38)	7.32	(5.24, 9.53)
325 mg (%)	8.37	(7.53, 9.22)	10.62	(9.77, 11.43)	43.85	(39.75, 47.92)

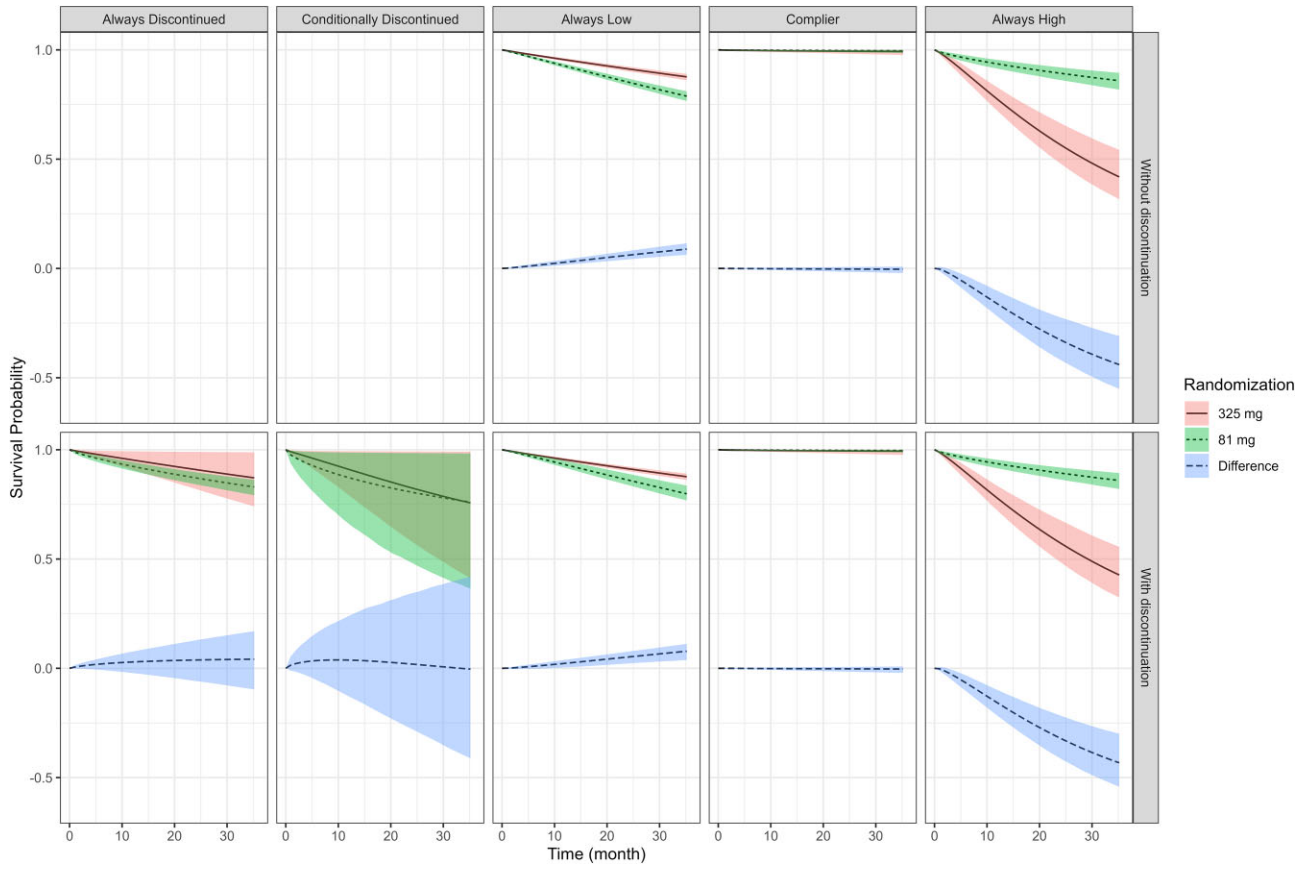

FIGURE 3 The posterior estimate of survival probability curve for the always-high, compliers and always-low strata under the randomized assignment to 81 and 325 mg aspirin with and without consideration of discontinuation.

coupled with the low event rate (around 7%) in the trial. Similar patterns are observed in the RACE, the details of which are delegated to the Supplementary Materials (Web Figure 5). We conducted a few sensitivity analyses, for example, fitting an accelerated failure time model, or varying the cutoff time of switching in defining the actual treatment. The results are similar and thus are omitted.

We also apply the method proposed in Section 4 to handle discontinuation. Besides the 3 strata defined for patients who would continue the treatment regardless of assignment, there are 2 additional strata for discontinued patients: always-discontinued $[U = (*, *)]$ and conditionally discontinued $[U = (0, *)]$. We do not impose ER for the always-discontinued stratum, because most patients took aspirin as assigned prior to discontinu-

ation, and thus the assigned treatment may have a direct effect on the outcome. We fit the multinomial strata model (with 5 categories) and Weibull-Cox model with the same covariates as in the previous analysis. The estimated proportion of the always-discontinued and conditionally discontinued strata is 6.6% (95% CI: 6.0% to 7.3%) and 2.6% (95% CI: 1.5% to 3.7%), respectively, both of which are small. The estimated proportion for the always-low, compliers, and always-high strata is 35.2% (95% CI: 34.0% to 36.3%), 49.4% (95% CI: 48.1% to 50.7%), and 6.2% (95% CI: 5.7% to 6.8%), respectively. The relative composition of these 3 strata is nearly identical to that in the case when discontinuation was ignored. As shown in Figure 3, the estimated survival curves associated credible regions are also similar in the three strata not concerning discontinuation. Therefore, discontinued aspirin intake in patients in ADAPTABLE does not appear to affect the conclusion.

Comparing with the ITT analysis, the PS analysis of the ADAPTABLE trial provides a more refined picture of heterogeneous treatment effects between subpopulations. However, the PS analysis requires strong and often untestable assumptions like exclusion restriction. Therefore, we shall interpret the clinical meaning of these results with caution. As discussed earlier, there may be a direct effect from the initial assigned treatment on the outcome due to the different timing of treatment switching. Therefore, for the patients in the always-high and always-low strata, the estimated effects for these 2 strata may be either attributed to the direct effect or a placebo effect of the assignment, which we cannot differentiate without further assumptions. This could also attenuate the effects of the treatment for the compliers.

7 DISCUSSION

We systematically investigated the PS analysis of noncompliance with time-to-event outcomes. We provided a generally applicable method based on the latent mixture models to obtain valid inference of several common causal estimands in the scale of survival probabilities and restricted mean survival time. For inference, we adopted the Bayesian approach, leveraging the automatic posterior sampling of the **Stan** computing platform. Though we focused on the parametric Weibull-Cox model for the outcome, the same inferential framework is applicable to any parametric outcome model specification, such as the AFT model. The Bayesian approach provides a unified framework for inferring all model parameters and derived parameters with automatic uncertainty quantification. It is also straightforward to extend the Bayesian approach to complex cases, such as multiple intermediate variables (Mealli et al., 2004), cluster treatments (Frangakis et al., 2002), and quantile effects (Wei et al., 2021).

Besides the mixture modeling approach, another identification and estimation strategy of PS is based on a weighting method under the assumption of *principal ignorability* (Jo and Stuart, 2009), which assumes the distributions of the outcomes between strata are the same conditional on the covariates. Jiang et al. (2022) proposed a semiparametrically efficient multiply robust weighting estimator under principal ignorability for principal causal effects with noncensored data; Cheng et al. (2023) extended that method to time-to-event outcomes. The mixture modeling and the weighting approaches rely on different set of identification assumptions and thus can lead to much different results when applied to the same data. Choosing between them critically depends on which set of corresponding identification assumptions is deemed more plausible in a specific application based on substantive knowledge.

ACKNOWLEDGMENTS

We thank the clinical input and motivating questions from investigators of ADAPTABLE, in particular, Sunil Kripalani, Schuyler Jones, and Hillary Mulder. We also thank the associate editor and two reviewers for their constructive comments.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Table 1-3 and Web Figures 1-5 referenced in Sections 5 and 6 are available with this paper at the Biometrics website on Oxford Academic.

The reproducible **R** code for simulations in this paper is provided at the Biometrics website on Oxford Academic and the GitHub repository: https://github.com/LauBok/PS_survival.

FUNDING

This research is supported by the Patient-Centered Outcomes Research Institute (PCORI) contract ME-2019C1-16146. The contents of this article are solely the responsibility of the authors and do not necessarily represent the view of PCORI.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

Research data are not shared due to confidentiality.

REFERENCES

Abrams, K., Ashby, D. and Errington, D. (1996). A Bayesian approach to Weibull survival models—application to a cancer clinical trial. *Lifetime Data Analysis*, 2, 159–174.

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

Baker, S. G. (1998). Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association*, 93, 929–934.

Baker, S. G. and Lindeman, K. S. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*, 13, 2269–2278.

Blanco, G., Chen, X., Flores, C. A. and Flores-Lagunes, A. (2020). Bounds on average and quantile treatment effects on duration outcomes under censoring, selection, and noncompliance. *Journal of Business & Economic Statistics*, 38, 901–920.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M. et al. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software*, 76, 1–32.

Cheng, C., Guo, Y., Liu, B., Wruck, L., Li, F. and Li, F. (2023). Multiply robust estimation for causal survival analysis with treatment noncompliance. arXiv:2305.13443.
Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34, 187–202.
Davis, P. J. and Rabinowitz, P. (2007). *Methods of Numerical Integration*, 2nd Edition. New York: Dover Publishers.
Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
Frangakis, C. and Rubin, D. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21–29.
Frangakis, C. E., Rubin, D. B. and Zhou, X.-H. (2002). Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms. *Biostatistics*, 3, 147–164.
Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721–741.
Gilbert, P. B. and Hudgens, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics*, 64, 1146–1154.
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.
Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*, 21, 13.
Hirano, K., Imbens, G., Rubin, D. and Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1, 69–88.
ICH E9 (R1) (2020). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Technical Report EMA/CHMP/ICH/436221/2017, Step 5 (17 February 2020).
Imbens, G. and Rubin, D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25, 305–327.
Jiang, Z., Yang, S. and Ding, P. (2022). Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84, 1423–1445.
Jo, B. and Stuart, E. (2009). On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine*, 28, 2857–2875.
Jones, W. S., Mulder, H., Wruck, L. M., Pencina, M. J., Kripalani, S., Muñoz, D., et al. (2021). Comparative effectiveness of aspirin dosing in cardiovascular disease. *New England Journal of Medicine*, 384, 1981–1990.
Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457–481.
Li, F., Tian, Z., Bobb, J., Papadogeorgou, G. and Li, F. (2022). Clarifying selection bias in cluster randomized trials. *Clinical Trials*, 19, 33–41.
Lipkovich, I., Ratitch, B., Qu, Y., Zhang, X., Shan, M. and Mallinckrodt, C. (2022). Using principal stratification in analysis of clinical trials. *Statistics in Medicine*, 41, 3837–3877.
Liu, B. and Li, F. (2023). *PStrata: Principal Stratification Analysis in R, R Package Version 3.5.0*. https://cran.r-project.org/web/packages/PStrata/index.html. [Accessed 17 October 2023].
Long, D. M. and Hudgens, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics*, 69, 812–819.
Mattei, A., Li, F. and Mealli, F. (2013). Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program. *The Annals of Applied Statistics*, 7, 2360–2013.
Mealli, F., Imbens, G. W., Ferro, S. and Biggeri, A. (2004). Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, 5, 207–222.
Mealli, F. and Pacini, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *Journal of the American Statistical Association*, 108, 1120–1131.
Neal, R. M. (2011). *Handbook of Markov Chain Monte Carlo, Chapter 5, MCMC Using Hamiltonian Dynamics*. New York: Chapman and Hall/CRC.
Nie, H., Cheng, J. and Small, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics*, 67, 1397–1405.
Rosenbaum, P. (1984). The consquences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series B*, 147, 656–666.
Rubin, D. (1980). Comment on 'Randomization analysis of experimental data: the Fisher randomization test' by D. Basu. *Journal of the American Statistical Association*, 75, 591–593.
Shepherd, B. E., Gilbert, P. B. and Dupont, C. T. (2011). Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity. *Biometrics*, 67, 1100–1110.
Shepherd, B. E., Gilbert, P. B. and Lumley, T. (2007). Sensitivity analyses comparing time-to-event outcomes existing only in a subset selected postrandomization. *Journal of the American Statistical Association*, 102, 573–582.
Stan Development Team (2022). *RStan: The R Interface to Stan. R Package Version 2.32.3*. https://cran.r-project.org/web/packages/rstan/index.html. [Accessed 2 December 2023].
Wei, B., Peng, L., Zhang, M.-J. and Fine, J. P. (2021). Estimation of causal quantile effects with a binary instrumental variable and censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83, 559–578.
Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871–1879.
Yu, W., Chen, K., Sobel, M. E. and Ying, Z. (2015). Semiparametric transformation models for causal inference in time to event studies with all-or-nothing compliance. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 77, 397.
Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28, 353–368.
Zhang, J. L., Rubin, D. B. and Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104, 166–176.

Received: January 18, 2023; Revised: October 2, 2023; Accepted: November 15, 2023