

# Multiply robust estimation for causal survival analysis with treatment noncompliance

Chao Cheng<sup>1</sup>, Yueqi Guo<sup>2</sup>, Bo Liu<sup>2</sup>, Lisa Wruck<sup>3,4</sup>, Fan Li<sup>2</sup> and Fan Li<sup>1\*</sup>

<sup>1</sup>Department of Biostatistics, Yale University

<sup>2</sup>Department of Statistical Science, Duke University

<sup>3</sup>Department of Biostatistics and Bioinformatics, Duke University

<sup>4</sup>Duke Clinical Research Institute

July 31, 2023

## Abstract

Comparative effectiveness research frequently addresses a time-to-event outcome and can require unique considerations in the presence of treatment noncompliance. Motivated by the challenges in addressing noncompliance in the ADAPTABLE pragmatic trial, we develop a multiply robust estimator to estimate the principal survival causal effects under the principal ignorability and monotonicity assumption. The multiply robust estimator involves several working models including that for the treatment assignment, the compliance strata, censoring, and time-to-event of interest. The proposed estimator is consistent even if one, and sometimes two, of the working models are misspecified. We apply the multiply robust method in the ADAPTABLE trial to evaluate the effect of low- versus high-dose aspirin assignment on patients' death and hospitalization from cardiovascular diseases. We find that, comparing to low-dose assignment, assignment to the high-dose leads to differential effects among always high-dose takers, compliers, and always low-dose takers. Such treatment effect heterogeneity contributes to the null intention-to-treatment effect, and suggests that policy makers should design personalized strategies based on potential compliance patterns to maximize treatment benefits to the entire study population. We further perform a formal sensitivity analysis for investigating the robustness of our causal conclusions under violation of two identification assumptions specific to noncompliance.

*Keywords:* Causal inference, pragmatic clinical trials, principal stratification, principal score, sensitivity analysis, time-to-event data.

---

\*Correspondence to Fan Li, Department of Biostatistics, Yale University (Email: fan.f.li@yale.edu).

# 1 Introduction

## 1.1 ADAPTABLE pragmatic trial and motivating question

ADAPTABLE (*Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-Term Effectiveness*) is an open-label, pragmatic, randomized trial to study the effectiveness of two strategies of aspirin dosing—325 mg versus 81 mg per day—for lowering risk on death and hospitalization among patients with existing cardiovascular diseases (Jones et al., 2021). A total of 15,076 participants were enrolled in study, where 7,540 were randomized to the 81-mg group (low dose) and 7,536 were randomized to the 325-mg group (high dose). The primary efficacy outcome is a composite of death from any cause and hospitalization for stroke or myocardial infarction, assessed by time to first event. Standard intention-to-treat (ITT) analysis suggested no statistically significant difference between the two dosing strategies in reducing patients’ risk on death and cardiovascular events (Jones et al., 2021).

As a pragmatic trial, the ADAPTABLE includes a substantial proportion of participants who did not adhere to their assigned aspirin dosage after randomization, which is likely to have been an important factor driving the null results in the ITT analysis. We are interested in evaluating the causal effects of aspirin dosing among subgroups of patients according to their treatment compliance status and separating the treatment efficacy from the direct effect of treatment assignment. We consider the principal stratification framework (Frangakis and Rubin, 2002) to subset patients into subgroups, including always high-dose takers who would stick to the 325-mg aspirin dosage regardless of their assignment, always low-dose takers who would stick to low-dose regardless of their assignment, and compliers who would comply with their assignment. The causal effect within each subgroup is referred to as the principal causal effect, which is a meaningful measure that can offer insights into whether one specific aspirin dosage benefits patients in each subgroup.

Several complications arise for estimating the principal causal effects in ADAPTABLE. First, addressing noncompliance requires identification assumptions. Since ADAPTABLE

is an open-label trial, these assumptions should ideally allow for possible direct effect from treatment assignment on the outcome among always high-dose and low-dose takers. Therefore, the *exclusion restriction* (ER), a conventional assumption in the existing noncompliance literature, may be questionable since it rules out any direct effect of assignment (Hirano et al., 2000). Second, identification assumptions are often unverifiable from the observed data, and sensitivity methods tailored for studying noncompliance are especially important to provide a context for interpreting the study results. Third, the primary outcome is time-to-event and requires unique handling as it is only partially observed due to right censoring from study dropout or end of follow-up. Despite randomization, patients’ censoring status may still depend on their baseline characteristics (Robins and Finkelstein, 2000), which must be accounted for. Last but not the least, the validity of existing methods for addressing noncompliance often heavily relies on the correct model specification. For example, to estimate the principal causal effects with a time-to-event outcome, one may need to specify multiple working models addressing the treatment assignment process, compliance status, censoring process, or the actual time-to-event outcome. Whereas the treatment assignment is known under randomization, all other models must be estimated from data and their potential misspecifications may lead to biased estimates (Jo and Stuart, 2009). Model-robust survival methods that can protect from working model misspecifications are desirable to address challenges in the ADAPTABLE study.

## 1.2 Related literature and our contributions

Several methods have been proposed to identify principal causal effects in the presence of noncompliance. Under *monotonicity* and the ER assumption, the instrumental variable (IV) approach was first used to identify the complier average causal effect (Imbens and Angrist, 1994; Baker and Lindeman, 1994; Angrist et al., 1996), which was later generalized to a principal stratification framework to deal with more general post-treatment confounding problems (Frangakis and Rubin, 2002). Despite the extensions of Angrist et al. (1996) and Frangakis and Rubin (2002)’s work in a variety of directions, there are only a few

methods that are geared toward studying noncompliance with a time-to-event outcome. Baker (1998) proposed a likelihood-based approach for the complier average causal effect on a hazard scale without covariates. Later, several authors considered semiparametric mixture models to estimate complier treatment effect defined on survival probabilities; for example, the proportional hazard model used in Loeys and Goetghebeur (2003) and Cuzick et al. (2007) and semiparametric transformation models used in Yu et al. (2015). A couple of nonparametric approaches also have been developed to address noncompliance with time-to-event outcomes; for example, the Kaplan–Meier estimator in Frangakis and Rubin (1999), the empirical likelihood approach in Nie et al. (2011), and the nonparametric estimator for complier quantile causal effect in Wei et al. (2021).

In the majority of these previous work, ER serves as a key assumption to identify the causal effect (see Web Table 1 in the Supplementary Material for a full survey of these existing developments), and assumes that the treatment assignment exerts no direct effect on patients who would receive the same treatment regardless of assignment. However, this assumption may not be easily justified, especially in open-label trials like ADAPTABLE. As an alternative to ER, we consider the principal ignorability assumption to identify the principal causal effect (Jo and Stuart, 2009; Ding and Lu, 2017). Principal ignorability assumes that the observed pre-treatment covariates are adequate to controlling for the confounding due to the patients’ compliance status without excluding the direct effect of assignment, and therefore allows us to estimate, in addition to the complier average causal effect, the causal effect of assignment among the always high-dose takers and always low-dose takers.

Despite advancement of causal inference methods under principal ignorability (Ding and Lu, 2017; Jiang et al., 2022), little development has been made to address survival outcomes, which require unique handling due to right-censoring. We therefore contribute a new multiply robust estimator for the principal survival causal effect, and implement this new estimator to estimate the principal causal estimands in ADAPTABLE. Although

this work is motivated by a randomized trial, the proposed estimator is also applicable to observational studies where the assignment process is ignorable. The core components of the multiply robust estimator include working models for the treatment propensity score, principal score, censoring process and survival outcome of interest. We show that it is multiply robust in the sense it is consistent to the principal causal effects even if one, and sometimes two, of the working models are misspecified, and thus our estimator provides stronger protection against model misspecification. Finally, because the validity of our estimator depends on the principal ignorability and monotonicity assumptions, which are empirically unverifiable, we develop a sensitivity analysis framework for their violations and operationalize this framework in ADAPTABLE.

## 2 Notation, estimands, and structural assumptions

Consider a comparative effectiveness study with  $n$  patients for comparing two treatments (high versus low aspirin dosage). For each patient  $i = 1, \dots, n$ , we record a vector of pre-treatment covariates  $\mathbf{X}_i$ . Let  $Z_i \in \{0, 1\}$  be the treatment assignment for patient  $i$ , with  $Z_i = 1$  if the patient is assigned to the high-dose group and  $Z_i = 0$  if assigned to the low-dose group. Let  $S_i \in \{0, 1\}$  be the actual treatment that patient  $i$  received, with  $S_i = 1$  if the patient received the high aspirin dose and  $S_i = 0$  if received the low aspirin dose. Each patient is associated with a failure time  $T_i$  that is incompletely observed due to right-censoring at time  $C_i$ . Therefore, we only have the observed failure time  $U_i = \min(T_i, C_i)$  and a censoring indicator  $\delta_i = \mathbb{I}(T_i \leq C_i)$ , where  $\mathbb{I}(\cdot)$  is the indicator function. The observed data consists of  $n$  independent and identically distributed copies of the quintuple  $\mathcal{O}_i = \{\mathbf{X}_i, Z_i, S_i, U_i, \delta_i\}$ .

We adopt the potential outcome framework to define the causal estimands (Rubin, 1974), and assume the Stable Unit Treatment Value Assumption (SUTVA) such that the assignment is defined unambiguously and there is no patient-level interference. For patient  $i$ , let  $S_i(z)$  be the potential treatment receipt if the patient is assigned to treatment  $z \in \{0, 1\}$ . Therefore,  $S_i(z) = 1$  indicates that the patient  $i$  would receive the high dose if

assigned to treatment  $z$  and  $S_i(z) = 0$  otherwise. Under SUTVA, we can connect the observed treatment receipt  $S_i$  and potential treatment receipt  $S_i(z)$  by recognizing  $S_i = Z_i S_i(1) + (1 - Z_i) S_i(0)$ . Similarly, we define the potential failure time and the potential censoring time for patient  $i$  under assignment  $z$  by  $T_i(z)$  and  $C_i(z)$ . Under SUTVA, we also have  $T_i = Z_i T_i(1) + (1 - Z_i) T_i(0)$  and  $C_i = Z_i C_i(1) + (1 - Z_i) C_i(0)$ .

Under the principal stratification framework (Frangakis and Rubin, 2002), we use the joint potential values of the treatment receipt,  $G_i = (S_i(1), S_i(0)) \in \{0, 1\}^{\otimes 2}$ , to define the principal strata. We refer to the four possible principal strata,  $G_i = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$ , as the *always high-dose takers* who take the 325-mg aspirin dosage regardless of randomized assignment, the *compliers* who comply with the randomized aspirin dosage, the *defiers* who take the the opposite aspirin dosage to their randomized assignment, and the *always low-dose takers* who take the 81-mg aspirin dosage regardless of randomized assignment. Hereafter, we abbreviate the four principal strata as  $G \in \{a, c, d, n\}$ . A central property of the principal strata is that it is unaffected by assignment and therefore can be considered as a pre-treatment covariate, conditional on which the subgroup causal effects are well-defined. Within each strata, our interest lies in the principal survival causal effect (PSCE):

$$\Delta_g(u) = \mathcal{S}_{1,g}(u) - \mathcal{S}_{0,g}(u) = \mathbb{P}(T(1) \geq u | G = g) - \mathbb{P}(T(0) \geq u | G = g), \quad (1)$$

for  $g \in \{n, c, d, a\}$ , where  $u$  is a pre-specified time point at which the counterfactual survival function is evaluated. In words, the PSCE estimand quantifies the survival benefit at time  $u$  when the subpopulation with the compliance pattern  $g$  is placed under the high-dose group versus when placed under the low-dose group. In particular,  $\Delta_c$  has been referred to as the complier average causal effect (CACE) in survival probability (Yu et al., 2015). In a similar fashion, if we consider the low-dose group as the “usual care” group, we can also refer to  $\Delta_a$  and  $\Delta_n$  as the always-taker average causal effect (AACE) and the never-taker average causal effect (NACE), which characterize the the direct effect of treatment assignment that could possibly be attributed to mechanisms that are unmeasured in the

study. With a survival outcome, identification of PSCE requires the following assumptions.

**Assumption 1** (*Unconfoundedness and overlap*)  $\{T(1), T(0), S(1), S(0)\} \perp Z | \mathbf{X}$ , where the symbol “ $\perp$ ” denotes independence. Furthermore, the propensity score  $\pi(\mathbf{X}) = \mathbb{P}(Z = 1 | \mathbf{X})$  is strictly bounded away from 0 and 1.

Assumption 1 is routinely invoked for observational comparisons of treatments and assumes away unmeasured baseline confounding. In a randomized trial such as ADAPTABLE, however, a stronger assumption,  $\{T(1), T(0), S(1), S(0), \mathbf{X}\} \perp Z$ , holds as a result of the study design. In what follows, we maintain this more general assumption 1 to unify developments for both randomized trials and observational studies.

**Assumption 2** (*Monotonicity*)  $S_i(1) \geq S_i(0)$  for all  $i$ .

The monotonicity assumption excludes defiers. The defiers in the ADAPTABLE trial, if exist, would consist of patients who take the 81-mg aspirin if assigned to the high-dose group, but take the 325-mg aspirin if assigned to the low-dose group. This type of patients is considered unlikely in the context of ADAPTABLE, though it might be applicable in other settings. However, monotonicity cannot be empirically verifiable based on the observed data alone, and we will consider sensitivity analysis strategies to assess assumed departure from this assumption in Section 4.2.

Under Assumption 2, the principal strata membership for each individual only is partially identified. That is, the observed cells with  $(Z = 1, S = 0)$  and  $(Z = 0, S = 1)$  include only the always low-dose takers and the always high-dose takers, respectively. The observed cells with  $(Z = 0, S = 0)$  and  $(Z = 1, S = 1)$ , however, are a mixture of patients in two principal strata—the  $(Z = 0, S = 0)$  cell consists of the always low-dose takers and the compliers, while the  $(Z = 1, S = 1)$  cell consists of the always high-dose takers and the compliers. The following assumption thus is central to disentangle the latent membership within the  $(Z = 0, S = 0)$  and  $(Z = 1, S = 1)$  cells.

**Assumption 3** (*Principal ignorability*) For all  $u \geq 0$ , we have  $\mathbb{P}(T(1) \geq u|G = a, \mathbf{X}) = \mathbb{P}(T(1) \geq u|G = c, \mathbf{X})$  and  $\mathbb{P}(T(0) \geq u|G = c, \mathbf{X}) = \mathbb{P}(T(0) \geq u|G = n, \mathbf{X})$ .

Assumption 3 assumes that, conditional on pre-treatment covariates, the survival functions for  $T(1)$  remain the same between the always high-dose takers and the compliers conditional on covariates, and likewise, the survival functions for  $T(0)$  remain the same between the always low-dose takers and the compliers (Jo and Stuart, 2009; Ding and Lu, 2017). This implicitly requires a sufficient set of covariates,  $\mathbf{X}$ , have been collected to capture the confounding between noncompliance status and the survival outcome. As a pragmatic trial leveraging routinely-collected data, ADAPTABLE included patients' demographics, smoking status, medical history and aspirin use prior to randomization, which are critical in explaining the compliance patterns. However, since the principal ignorability is also not empirically verifiable based on the observed data itself, we will consider a sensitivity analysis strategy in Section 4.1 under assumed departure from this assumption.

Under Assumptions 1–2, Assumption 3 implies that

$$\mathbb{P}(T(1) \geq u|G = g, Z = 1, S = 1, \mathbf{X}) = \mathbb{P}(T \geq u|Z = 1, S = 1, \mathbf{X}), \text{ for } g \in \{a, c\}, \quad (2)$$

$$\mathbb{P}(T(0) \geq u|G = g, Z = 0, S = 0, \mathbf{X}) = \mathbb{P}(T \geq u|Z = 0, S = 0, \mathbf{X}), \text{ for } g \in \{n, c\}, \quad (3)$$

for all  $u \geq 0$ . These two equations allow us to identify a set of conditional survival functions for the potential survival times,  $T(1)$  and  $T(0)$ . For example, equation (2) suggests that, within the observed subpopulation who were assigned to the high-dose group and also received the high dose, the compliance strata variable does not further affect  $T(1)$  given pre-treatment covariates. Therefore the conditional survival function given the compliance status  $G$ ,  $\mathbb{P}(T(1) \geq u|G = g, Z = 1, S = 1, \mathbf{X})$ , simplifies to a function of the observed data only,  $\mathbb{P}(T \geq u|Z = 1, S = 1, \mathbf{X})$ .

Finally, noting that the survival outcome may be partially observable due to right censoring, we adopt the following conditionally independent censoring assumption:

**Assumption 4** (*Conditionally independent censoring*)  $T(z) \perp C(z)|\{Z = z, S, \mathbf{X}\}$ .



Assumption 4 stipulates that the potential censoring time does not provide any information about the potential failure time other than that the latter exceeds the censoring time conditional on pre-treatment covariates, and strata defined based on treatment assignment and the treatment receipt. Assumption 4 resembles the *independent censoring* or *coarsening at random* assumption in the standard survival analysis literature (Gill et al., 1997).

### 3 Estimating Principal Survival Causal Effects

#### 3.1 Working models for aspects of the data generating process

To analyze the PSCE estimands, there are four possible working models that can be estimated from the observed data, and each working model represents a distinct aspect of the data generating process. These models are:

- (a)  $\pi(\mathbf{X}) = \mathbb{P}(Z = 1|\mathbf{X})$ : the propensity score of assignment to the high-dose group. Also, we define  $\pi = \mathbb{E}[\pi(\mathbf{X})] = \mathbb{P}(Z = 1)$  as the marginal allocation proportion. In ADAPTABLE, we have  $\pi(\mathbf{X}) = \pi = 0.5$  by the study design, but more generally  $\pi(\mathbf{X}) \neq \pi$  in observational studies.
- (b)  $e_g(\mathbf{X}) = \mathbb{P}(G = g|\mathbf{X})$ , for  $g \in \{a, c, n\}$ : the principal score of being in the always high-dose, complier, and always low-dose strata, respectively. Similarly, we write  $e_g = \mathbb{E}[e_g(\mathbf{X})] = P(G = g)$  as the marginal strata proportion.
- (c)  $\mathcal{S}_{zs}^C(u|\mathbf{X}) = \mathbb{P}(C \geq u|Z = z, S = s, \mathbf{X})$ , for  $\{z, s\} \in \{0, 1\}^{\otimes 2}$ : the conditional survival function of the censoring time given the assignment, treatment receipt and covariates.
- (d)  $\mathcal{S}_{zs}(u|\mathbf{X}) = \mathbb{P}(T \geq u|Z = z, S = s, \mathbf{X})$ , for  $\{z, s\} \in \{0, 1\}^{\otimes 2}$ : the conditional survival function of the outcome given the assignment, treatment receipt and covariates.

A working logistic regression can be applied to model the propensity score such that  $\pi(\mathbf{X}; \boldsymbol{\alpha}) = 1/(1+e^{-\boldsymbol{\alpha}^T \mathbf{X}})$ , where  $\hat{\boldsymbol{\alpha}}$  can be obtained by solving the score equation  $\mathbb{P}_n[\mathbf{X}(Z - \pi(\mathbf{X}; \boldsymbol{\alpha}))] = \mathbf{0}$ . Here,  $\mathbb{P}_n[\cdot] = \frac{1}{n} \sum_{i=1}^n [\cdot]$  defines the empirical average. Although the true propensity score is known in randomized trials, one may preferably use the estimated propensity score to control for chance imbalance in covariates. As demonstrated by Zeng

et al. (2021), estimating the known propensity scores in a randomized trial can often increase the efficiency of a causal effect estimator as long as prognostic covariates are included in the working propensity score model.

To estimate the principal scores, we recognize the one-to-one relationship between  $\{e_a(\mathbf{X}), e_c(\mathbf{X}), e_n(\mathbf{X})\}$  and  $\{p_0(\mathbf{X}), p_1(\mathbf{X})\}$  that defined below (Ding and Lu, 2017):

$$e_a(\mathbf{X}) = p_0(\mathbf{X}), \quad e_n(\mathbf{X}) = 1 - p_1(\mathbf{X}), \quad e_c(\mathbf{X}) = p_1(\mathbf{X}) - p_0(\mathbf{X}). \quad (4)$$

Here,  $p_z(\mathbf{X}) = \mathbb{P}(S = 1|Z = z, \mathbf{X})$  is the observed probability of receiving the high dose conditional on assignment to group  $z \in \{0, 1\}$  and covariates. Therefore, one can specify working models  $p_z(\mathbf{X}; \gamma_z)$  for  $p_z(\mathbf{X})$  and then utilize the relationships in (4) to calculate  $e_g(\mathbf{X})$ . For the analysis of ADAPTABLE, we consider logistic regressions such that  $p_z(\mathbf{X}; \gamma_z) = 1/(1 + e^{-\gamma_z^T \mathbf{X}})$ , where  $\hat{\gamma}_z$  is obtained by maximum likelihood based on the subset with  $Z = z$ . Then,  $e_a(\mathbf{X})$ ,  $e_c(\mathbf{X})$ , and  $e_n(\mathbf{X})$  can be estimated by  $e_a(\mathbf{X}; \hat{\gamma}) = p_0(\mathbf{X}; \hat{\gamma}_0)$ ,  $e_c(\mathbf{X}; \hat{\gamma}) = p_1(\mathbf{X}; \hat{\gamma}_1) - p_0(\mathbf{X}; \hat{\gamma}_0)$ , and  $e_n(\mathbf{X}; \hat{\gamma}) = 1 - p_1(\mathbf{X}; \hat{\gamma}_1)$ , respectively.

The last two working models correspond to regression for the observed censoring time and survival time across each observed cell  $(Z, S) \in \{0, 1\}^{\otimes 2}$ . We consider the Cox proportional hazard model for the censoring process such that  $\mathcal{S}_{zs}^C(u|\mathbf{X}; \boldsymbol{\theta}_{zs}) = \exp\left(-\Lambda_{zs,0}^C(u)e^{-\tilde{\boldsymbol{\theta}}_{zs}^T \mathbf{X}}\right)$ , where  $\Lambda_{zs,0}^C(u)$  is the cumulative baseline hazard,  $\tilde{\boldsymbol{\theta}}_{zs}$  is the coefficient, and  $\boldsymbol{\theta}_{zs} = [\tilde{\boldsymbol{\theta}}_{zs}^T, \Lambda_{zs,0}^C]^T$ . The estimator  $\hat{\boldsymbol{\theta}}_{zs}$  can be obtained by maximum partial likelihood within the observed cell  $(Z = z, S = s)$  and  $\hat{\Lambda}_{zs,0}^C(u)$  can be obtained by the Breslow estimator (Breslow, 1972). We denote the estimated survival distribution of the censoring time is  $\mathcal{S}_{zs}^C(u|\mathbf{X}; \hat{\boldsymbol{\theta}}_{zs})$ . A similar working proportional hazards model for the survival outcome can be considered as  $\mathcal{S}_{zs}(u|\mathbf{X}; \boldsymbol{\beta}_{zs}) = \exp\left(-\Lambda_{zs,0}(u)e^{-\tilde{\boldsymbol{\beta}}_{zs}^T \mathbf{X}}\right)$  with unknown parameters  $\boldsymbol{\beta}_{zs} = [\tilde{\boldsymbol{\beta}}_{zs}^T, \Lambda_{zs,0}]^T$ ; the estimation of  $\boldsymbol{\beta}_{zs}$  is analogous to that of  $\boldsymbol{\theta}_{zs}$ .

Throughout, we use  $\mathcal{M}_\pi, \mathcal{M}_e, \mathcal{M}_C, \mathcal{M}_T$  to denote the working models  $\pi(\mathbf{X}; \boldsymbol{\alpha})$ ,  $e_g(\mathbf{X}; \boldsymbol{\gamma})$ ,  $\mathcal{S}_{zs}^C(u|\mathbf{X}; \boldsymbol{\theta}_{zs})$ , and  $\mathcal{S}_{zs}(u|\mathbf{X}; \boldsymbol{\beta}_{zs})$ , respectively. To facilitate presentation, we define  $p_z = \mathbb{E}[p_z(\mathbf{X})]$ , which is a key component for estimating the PSCE (1). One intuitive method to estimate  $p_z$  is by  $\mathbb{P}_n[p_z(\mathbf{X}; \hat{\gamma}_z)]$ , which is consistent only if  $\mathcal{M}_e$  is correctly specified. We

utilize a more robust estimator for  $p_z$ , suggested by Jiang et al. (2022):

$$\hat{p}_0 = \mathbb{P}_n \left[ \frac{(1-Z)(S - p_0(\mathbf{X}; \hat{\gamma}))}{1 - \pi(\mathbf{X}; \hat{\alpha})} + p_0(\mathbf{X}; \hat{\gamma}) \right], \quad \hat{p}_1 = \mathbb{P}_n \left[ \frac{Z(S - p_1(\mathbf{X}; \hat{\gamma}))}{\pi(\mathbf{X}; \hat{\alpha})} + p_1(\mathbf{X}; \hat{\gamma}) \right], \quad (5)$$

which leads to unbiased estimation under the union model  $\mathcal{M}_\pi \cup \mathcal{M}_e$ ; i.e., it is consistent if either  $\mathcal{M}_\pi$  or  $\mathcal{M}_e$  is correctly specified, but not necessary both. Throughout, we use the union notation, “ $\cup$ ”, to denote correct specification of at least one model. The estimators for proportion of each principal stratum are therefore  $\hat{e}_a = \hat{p}_0$ ,  $\hat{e}_n = 1 - \hat{p}_1$ , and  $\hat{e}_c = \hat{p}_1 - \hat{p}_0$ .

### 3.2 Identification formulas

Despite the fact that the data generating process includes multiple working models, identification of the PSCE estimands do not require the specification of all working models. We use the following result to summarize three different strategies to identify the target estimands within each principal strata.

**Result 1 (Identification)** *Suppose Assumptions 1–4 hold,  $e_g(\mathbf{X}) > 0$  for all  $g \in \{a, c, n\}$ , and  $S_{zs}^C(u|\mathbf{X}) > 0$  for all  $(z, s) \in \{0, 1\}^{\otimes 2}$ , the PSCEs are nonparametrically identified.*

(i) *Using propensity score, principal score, and censoring survival function, we have*

$$\begin{aligned} \Delta_c(u) &= \mathbb{E} \left[ \frac{e_c(\mathbf{X})}{p_1 - p_0} \frac{S}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X})} \right] - \mathbb{E} \left[ \frac{e_c(\mathbf{X})}{p_1 - p_0} \frac{(1-S)}{1 - p_0(\mathbf{X})} \frac{1-Z}{1 - \pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{00}^C(u|\mathbf{X})} \right], \\ \Delta_n(u) &= \mathbb{E} \left[ \frac{1-S}{1 - p_1} \frac{Z}{\pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{10}^C(u|\mathbf{X})} \right] - \mathbb{E} \left[ \frac{e_n(\mathbf{X})}{1 - p_1} \frac{1-S}{1 - p_0(\mathbf{X})} \frac{1-Z}{1 - \pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{00}^C(u|\mathbf{X})} \right], \\ \Delta_a(u) &= \mathbb{E} \left[ \frac{e_a(\mathbf{X})}{p_0} \frac{S}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X})} \right] - \mathbb{E} \left[ \frac{S}{p_0} \frac{1-Z}{1 - \pi(\mathbf{X})} \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{01}^C(u|\mathbf{X})} \right]. \end{aligned}$$

(ii) *Using principal score and outcome survival function, we have*

$$\begin{aligned} \Delta_c(u) &= \mathbb{E} \left[ \frac{e_c(\mathbf{X})}{p_1 - p_0} \{ \mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) \} \right], \\ \Delta_n(u) &= \mathbb{E} \left[ \frac{e_n(\mathbf{X})}{1 - p_1} \{ \mathcal{S}_{10}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) \} \right], \\ \Delta_a(u) &= \mathbb{E} \left[ \frac{e_a(\mathbf{X})}{p_0} \{ \mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{01}(u|\mathbf{X}) \} \right]. \end{aligned}$$

(iii) Using propensity score and outcome survival function, we have

$$\begin{aligned}\Delta_c(u) &= \frac{1}{p_1 - p_0} \mathbb{E} \left[ \left( \frac{SZ}{\pi(\mathbf{X})} - \frac{S(1-Z)}{1 - \pi(\mathbf{X})} \right) \{ \mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) \} \right], \\ \Delta_n(u) &= \frac{1}{1 - p_1} \mathbb{E} \left[ \left( 1 - \frac{SZ}{\pi(\mathbf{X})} \right) \{ \mathcal{S}_{10}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) \} \right], \\ \Delta_a(u) &= \frac{1}{p_0} \mathbb{E} \left[ \frac{S(1-Z)}{1 - \pi(\mathbf{X})} \{ \mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{01}(u|\mathbf{X}) \} \right].\end{aligned}$$

Result 1 extends the identification formulas in Ding and Lu (2017) and Jiang et al. (2022) to a right-censored time-to-event outcome. The proof is presented in Web Appendix A.1, where all appendices are shown in Supplementary Material. Using  $\Delta_c(u)$  as an example, we intuitively interpret the rationale underlying each identification formula below.

Result 1(i) suggests that  $\Delta_c(u)$  can be written as difference between two weighted averages of patients' observed survival status at  $u$  within each treatment group, which consist of 3 components. The first component,  $\frac{Z}{\pi(\mathbf{X})}$  for the high-dose group and  $\frac{1-Z}{1-\pi(\mathbf{X})}$  for the low-dose group, is the standard inverse probability weight to account for potential confounding due to treatment assignment. The second component,  $\frac{e_c(\mathbf{X})S}{(p_1-p_0)p_1(\mathbf{X})}$  for the high-dose group and  $\frac{e_c(\mathbf{X})(1-S)}{(p_1-p_0)(1-p_0(\mathbf{X}))}$  for the low-dose group, leverages the principal scores to account for the confounding due to noncompliance. The third component,  $\frac{I(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X})}$  for the high-dose group and  $\frac{I(U \geq u)}{\mathcal{S}_{00}^C(u|\mathbf{X})}$  for the low-dose group, addresses the selection bias associated with the right-censoring; this is also referred to as the inverse probability of censoring weight (Robins and Finkelstein, 2000).

Result 1(ii) suggests an alternative strategy to identify  $\Delta_c$  by using the principal score and the outcome survival function. The formulation is based on a combination of two components,  $\mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X})$  and  $\frac{e_c(\mathbf{X})}{p_1 - p_0}$ , where the second component equals to  $\frac{e_c(\mathbf{X})}{e_c}$  by noting  $p_1 - p_0 = e_c$  from (4). One can rewrite the first component to

$$\mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) = \mathbb{P}(T(1) \geq u | G = c, \mathbf{X}) - \mathbb{P}(T(0) \geq u | G = c, \mathbf{X}),$$

by applying principal ignorability, which is the PSCE for compliers conditional on covariates, referred to  $\Delta(u|G = c, \mathbf{X})$ . Let  $f(\mathbf{X})$  denote the density function of the covariates,

then taking expectation of the multiplication of the first and second component equals to the following integral over  $\mathbf{X}$ :

$$\mathbb{E} \left[ \frac{e_c(\mathbf{X})}{p_1 - p_0} \{ \mathcal{S}_{11}(u|\mathbf{X}) - \mathcal{S}_{00}(u|\mathbf{X}) \} \right] = \int \Delta(u|G=c, \mathbf{x}) \frac{\mathbb{P}(G=c|\mathbf{x})f(\mathbf{x})}{\mathbb{P}(G=c)} d\mathbf{x},$$

where by the Bayes' rule,  $\frac{\mathbb{P}(G=c|\mathbf{x})f(\mathbf{x})}{\mathbb{P}(G=c)}$  is the density of  $\mathbf{X}$  among compliers. It clearly suggests that the expression of  $\Delta_c(u)$  in Result 1(ii) integrates the conditional PSCE,  $\Delta(u|G=c, \mathbf{X})$ , over the distribution of  $\mathbf{X}$  defined over the compliers. Result 1(iii) bears resemblance to Result 1(ii), which replaces  $e_c(\mathbf{X})$  in the expression of  $\Delta_c(u)$  in Result 1(ii) by  $\frac{SZ}{\pi(\mathbf{X})} - \frac{S(1-Z)}{1-\pi(\mathbf{X})}$ . Under Assumptions 1-2, we can verify that  $\mathbb{E} \left[ \frac{SZ}{\pi(\mathbf{X})} - \frac{S(1-Z)}{1-\pi(\mathbf{X})} | \mathbf{X} \right] = e_c(\mathbf{X})$ . This reveals that the identification formula from Result 1(iii) only replaces  $e_c(\mathbf{X})$  in that from Result 1(ii) by its unconditional counterpart without compromising the estimation target.

Result 1 motivates three moment-type estimators of  $\Delta_g$ ,  $\widehat{\Delta}_g^1(u)$ ,  $\widehat{\Delta}_g^2(u)$ , and  $\widehat{\Delta}_g^3(u)$ , each corresponding to one identification strategy. We refer to these estimators as singly robust estimators as they are consistent to PSCEs when all associated working models are correctly specified and will be biased otherwise. To obtain these singly robust estimators, we can replace the true expectation “ $\mathbb{E}$ ” in each identification formula with the empirical average “ $\mathbb{P}_n$ ” operator after substituting the true probabilities,  $p_z$ ,  $p_z(\mathbf{X})$ ,  $\pi(\mathbf{X})$ ,  $e_g(\mathbf{X})$ ,  $\mathcal{S}_{zs}^C(u|\mathbf{X})$ , and  $\mathcal{S}_{zs}(u|\mathbf{X})$ , with their corresponding estimators introduced in Section 3.1,  $\widehat{p}_z$ ,  $p_z(\mathbf{X}; \widehat{\gamma})$ ,  $\pi(\mathbf{X}; \widehat{\alpha})$ ,  $e_g(\mathbf{X}; \widehat{\gamma})$ ,  $\mathcal{S}_{zs}^C(u|\mathbf{X}; \widehat{\theta}_{zs})$ , and  $\mathcal{S}_{zs}(u|\mathbf{X}; \widehat{\beta}_{zs})$ , respectively. Explicit expressions of these singly robust estimators are provided in Web Appendix A.2.

### 3.3 Multiply robust estimation of principal survival causal effects

The identification formulas in Result 1 motivate three singly robust estimators,  $\widehat{\Delta}_g^1(u)$ ,  $\widehat{\Delta}_g^2(u)$ , and  $\widehat{\Delta}_g^3(u)$ , which produce consistent estimation under  $\mathcal{M}_{\pi+e+C}$ ,  $\mathcal{M}_{e+T}$ , and  $\mathcal{M}_{\pi+T}$ , respectively. Here, we use the symbol “+” to denote the interaction of models such that  $\mathcal{M}_{\pi+e+C}$  denotes models with correctly specified  $\pi(\mathbf{X}; \alpha)$ ,  $e_g(\mathbf{X}; \gamma)$ , and  $\mathcal{S}_{zs}^C(t|\mathbf{X}; \theta_{zx})$ , and so on. Since the PSCE estimates may be different based on the use of different identification

results, we propose a multiply robust estimator  $\widehat{\Delta}_g^{mr}(u) = \widehat{\mathcal{S}}_{1,g}^{mr}(u) - \widehat{\mathcal{S}}_{0,g}^{mr}(u)$ , for  $g = a, c$  and  $n$ , that integrate these different identification results without the need to choose among the identification strategies. We use the superscript ‘ $mr$ ’ to highlight that this estimator is multiply robust in the sense that it does not require all of  $\mathcal{M}_\pi$ ,  $\mathcal{M}_e$ ,  $\mathcal{M}_C$ , and  $\mathcal{M}_T$  are correctly specified. To simplify the presentation, we only develop  $\widehat{\mathcal{S}}_{1,c}^{mr}(u)$  in this section, where explicit forms of  $\widehat{\mathcal{S}}_{0,c}^{mr}(u)$ ,  $\widehat{\mathcal{S}}_{z,n}^{mr}(u)$ , and  $\widehat{\mathcal{S}}_{z,a}^{mr}(u)$ , for  $z \in \{0, 1\}$ , can be similarly developed and are deferred to Web Appendix B.

We start by considering the complete data scenario without censoring such that  $C_i = \infty$  for all  $i$ . With complete data, we have  $\mathcal{A}_i = \{\mathbf{X}_i, Z_i, S_i, T_i\}$  for each patient, in contrast to the observed data  $\mathcal{O}_i = \{\mathbf{X}_i, Z_i, S_i, U_i, \delta_i\}$  in the presence of censoring. With complete data and under assumptions 1 to 3, Jiang et al. (2022) developed a multiply robust estimator of  $\mathcal{S}_{1,c}(u)$  based on solving the following unbiased estimating equation,

$$\mathbb{P}_n [\psi_{1,c}^{mr}(\mathcal{A})] = 0, \quad (6)$$

where

$$\begin{aligned} \psi_{1,c}^{mr}(\mathcal{A}) = & \frac{e_c(\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \frac{S}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} \mathbb{I}(T \geq u) - \mathcal{S}_{1,c}(u) + \frac{e_c(\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \left(1 - \frac{Z}{\pi(\mathbf{X})}\right) \mathcal{S}_{11}(u|\mathbf{X}) \\ & + \frac{\mathcal{S}_{11}(u|\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \left[ \frac{1 - Z}{1 - \pi(\mathbf{X})} (S - p_0(\mathbf{X})) + \frac{p_0(\mathbf{X})}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} (p_1(\mathbf{X}) - S) \right], \end{aligned}$$

is the estimating function,  $\tilde{p}_0 = \mathbb{P}_n \left[ \frac{(1-Z)(S-p_0(\mathbf{X}))}{1-\pi(\mathbf{X})} + p_0(\mathbf{X}) \right]$  and  $\tilde{p}_1 = \mathbb{P}_n \left[ \frac{Z(S-p_1(\mathbf{X}))}{\pi(\mathbf{X})} + p_1(\mathbf{X}) \right]$  are the doubly robust estimating function of  $p_0$  and  $p_1$  in equation (5). One can verify that the expectation of first two terms in  $\psi_{1,c}^{mr}(\mathcal{A})$  is 0 by using Result 1(i) without the censoring weighting. The third and fourth terms in  $\psi_{1,c}^{mr}(\mathcal{A})$  are two mean-zero augmentation terms to help achieve additional robustness if models in the first two terms, including  $\mathcal{M}_e$  and  $\mathcal{M}_\pi$ , are misspecified. Jiang et al. (2022) shows that the solution of (6) is a consistent and locally efficient estimator of  $\mathcal{S}_{1,c}(u)$  when  $\mathcal{M}_{\pi+e+T}$  is correctly specified and is also consistent under  $\mathcal{M}_{\pi+e} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$ ; that is, the corresponding estimator is also valid even if one of  $\mathcal{M}_\pi$ ,  $\mathcal{M}_e$ , and  $\mathcal{M}_T$  is incorrect.

In the presence of censoring, however, the estimating equation (6) is no longer directly applicable due to the partially observed outcome. To accomodate right-censored survival outcomes, we follow Chapter 9 of Tsiatis (2006) to formalize an augmented inverse probability of weighted complete-case (AIPWCC) estimator. Specifically, we can modify  $\psi_{1,c}^{mr}(\mathcal{A})$  to the following AIPWCC estimating score that depends only on observed data  $\mathcal{O}$ :

$$\psi_{1,c}(\mathcal{O}) = \frac{\delta\psi_{1,c}^{mr}(\mathcal{A})}{\mathcal{S}_{ZS}^C(U|\mathbf{X})} + \int h(r, Z, S, \mathbf{X}) dM_{ZS}^C(r|\mathbf{X}), \quad (7)$$

where  $h(t, Z, S, \mathbf{X})$  is an arbitrary function of  $(t, Z, S, \mathbf{X})$ ,  $dM_{zs}^C(t|\mathbf{X}) = dN^C(t) - \mathbb{I}(U \geq t)d\Lambda_{zs}^C(t|X)$  is a martingale for the censoring mechanism conditional on  $(Z = z, S = s, \mathbf{X})$ , and  $N^C(t) = I(U \leq t, \delta = 0)$  and  $\Lambda_{zs}^C(t|\mathbf{X}) = -\log \mathcal{S}_{zs}^C(t|\mathbf{X})$  are the counting process and cumulative hazard function relative to the censoring mechanism. The first term in (7) is a complete data estimating score weighted by the inverse probability of censoring. The second term in (7) is an augmentation term on the nuisance tangent space  $\mathcal{T}_C = \{\int h^*(r, Z, S, \mathbf{X})dM_{ZS}^C(r|\mathbf{X}) : \forall h^*\}$  of the censoring process under the conditional independent censoring assumption (Assumption 4). Thus, a class of estimating functions defined by  $\psi_{1,c}^{mr}(\mathcal{A})$  is index by each unique  $h(t, Z, S, \mathbf{X})$ , which is an element in the tangent space  $\mathcal{T}_C$ . In addition, the optimal choice of  $h(t, Z, S, \mathbf{X})$ , that is, the estimator with the smallest asymptotic variance, is given by  $h^{opt}(t, Z, S, \mathbf{X}) = \Pi \left\{ \frac{\delta\psi_{1,c}^{mr}(\mathcal{A})}{\mathcal{S}_{ZS}^C(U|\mathbf{X})} \middle| \mathcal{T}_C \right\}$ , where  $\Pi\{\cdot|\mathcal{T}_C\}$  is a projection operator onto the nuisance tangent space of censoring,  $\mathcal{T}_C$ . We show in Web Appendix B.1 that  $h^{opt}(t, Z, S, \mathbf{X})$  can be rewritten as  $h^{opt}(t, Z, S, \mathbf{X}) = \frac{\mathbb{E}[\psi_{1,c}^{mr}(\mathcal{A})|T \geq t, Z, S, \mathbf{X}]}{\mathcal{S}_{ZS}^C(t|\mathbf{X})}$ . Therefore, the optimal estimating function in the class of estimating scores that  $\psi_{1,c}^{mr}(\mathcal{A})$  belongs to is:

$$\psi_{1,c}^{opt}(\mathcal{O}) = \frac{\delta\psi_{1,c}^{mr}(\mathcal{A})}{\mathcal{S}_{ZS}^C(U|\mathbf{X})} + \int \frac{\mathbb{E}[\psi_{1,c}^{mr}(\mathcal{A})|T \geq r, Z, S, \mathbf{X}]}{\mathcal{S}_{ZS}^C(r|\mathbf{X})} dM_{ZS}^C(r|\mathbf{X}).$$

After some algebra (see Web Appendix B.1 for details), we obtain the following explicit expression of  $\psi_{1,c}^{opt}(\mathcal{O})$ :

$$\psi_{1,c}^{opt}(\mathcal{O}) = \frac{e_c(\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \frac{S}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} \left\{ \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X})} + \mathcal{S}_{11}(u|\mathbf{X}) \int_0^u \frac{dM_{11}^C(r|\mathbf{X})}{S_{11}(r|\mathbf{X})\mathcal{S}_{11}^C(r|\mathbf{X})} \right\} - \mathcal{S}_{1,c}(u)$$

$$\begin{aligned}
& + \frac{\mathcal{S}_{11}(u|\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \left[ \frac{1 - Z}{1 - \pi(\mathbf{X})} (S - p_0(\mathbf{X})) + \frac{p_0(\mathbf{X})}{p_1(\mathbf{X})} \frac{Z}{\pi(\mathbf{X})} (p_1(\mathbf{X}) - S) \right] \\
& + \frac{e_c(\mathbf{X})}{\tilde{p}_1 - \tilde{p}_0} \left( 1 - \frac{Z}{\pi(\mathbf{X})} \right) \mathcal{S}_{11}(u|\mathbf{X}).
\end{aligned} \tag{8}$$

Interestingly,  $\psi_{1,c}^{opt}(\mathcal{O})$  shares an similar form to its complete data counterpart,  $\psi_{1,c}^{mr}(\mathcal{A})$ , where their last three terms are identical but the first term in  $\psi_{1,c}^{opt}(\mathcal{O})$  has been specially to account for right censoring. Substituting the unknown probabilities in equation (8) by their estimators provided by the working models in Section 3.1, we arrive at the following optimal AIPWCC estimator,

$$\begin{aligned}
\widehat{\mathcal{S}}_{1,c}^{mr}(u) = & \mathbb{P}_n \left\{ \frac{e_c(\mathbf{X}; \hat{\gamma})}{\hat{p}_1 - \hat{p}_0} \frac{S}{p_1(\mathbf{X}; \hat{\gamma}_1)} \frac{Z}{\pi(\mathbf{X}; \hat{\alpha})} \left[ \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X}; \hat{\theta}_{11})} + \mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11}) \int_0^u \frac{d\widehat{M}_{11}^C(r|\mathbf{X})}{\mathcal{S}_{11}(r|\mathbf{X}; \hat{\beta}_{11}) \mathcal{S}_{11}^C(r|\mathbf{X}; \hat{\theta}_{11})} \right] \right. \\
& + \frac{\mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11})}{\hat{p}_1 - \hat{p}_0} \left[ \frac{1 - Z}{1 - \pi(\mathbf{X}; \hat{\alpha})} (S - p_0(\mathbf{X}; \hat{\gamma}_0)) + \frac{p_0(\mathbf{X}; \hat{\gamma})}{p_1(\mathbf{X}; \hat{\gamma}_1)} \frac{Z}{\pi(\mathbf{X}; \hat{\alpha})} (p_1(\mathbf{X}; \hat{\gamma}) - S) \right] \\
& \left. + \frac{e_c(\mathbf{X})}{\hat{p}_1 - \hat{p}_0} \left( 1 - \frac{Z}{\pi(\mathbf{X}; \hat{\gamma})} \right) \mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11}) \right\},
\end{aligned} \tag{9}$$

where  $d\widehat{M}_{zs}^C(t|\mathbf{X}) = dN^C(t) - \mathbb{I}(U \geq t) \frac{\widehat{\lambda}_{zs}^C(t|\mathbf{X})}{\mathcal{S}_{zs}^C(t|\mathbf{X}; \hat{\theta}_{zs})}$  and  $\widehat{\lambda}_{zs}^C(t|\mathbf{X}) = -\frac{d}{dt} \log \mathcal{S}_{zs}^C(t|\mathbf{X}; \hat{\theta}_{zs})$ . In Web Appendix B.3, we prove that expectation of the right hand side of (9) converges to  $\mathcal{S}_{1,c}(u)$  if either,  $\mathcal{M}_{\pi+e+C}$ ,  $\mathcal{M}_{\pi+T}$ , or  $\mathcal{M}_{e+T}$  is correctly specified, which is formalized in the following result.

**Result 2 (Multiple robustness)** *Suppose that assumptions 1–4 hold. For all  $z \in \{0, 1\}$  and  $g \in \{c, n, a\}$ ,  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  is a triply robust estimator for  $\mathcal{S}_{z,g}(u)$  in the sense that  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  is consistent to  $\mathcal{S}_{z,g}(u)$  under the union model  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$ , where  $\widehat{\mathcal{S}}_{1,c}^{mr}(u)$  is defined in (9) and other  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$ 's are defined in Web Appendix B.2. As a consequence,  $\widehat{\Delta}_g^{mr}(u)$  is also consistent to  $\Delta_g(u)$  under  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$  for all  $g \in \{c, n, a\}$ .*

Under randomized assignment as in ADAPTABLE, the true propensity score ( $\pi(\mathbf{X}) = \pi$ ) is known and estimating the known propensity score (through a working propensity score model) will at most improve the asymptotic efficiency of an estimator. Because the working propensity score is technically correctly specified,  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  and  $\widehat{\Delta}_g^{mr}(u)$  become doubly robust (rather than triply robust) estimators, in the sense that they are consistent when either



$\mathcal{M}_{e+C}$  (the weighting models) or  $\mathcal{M}_T$  (the outcome models) is correctly specified.

Standard error and confidence intervals for the multiply robust PSCE estimators can be constructed via nonparametric bootstrap. Specifically, we re-sample the entire dataset and re-estimate  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  for  $B$  iterations, where  $B$  is a large number (for example, 500 or more). Then, we re-estimate  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  within each bootstrap dataset. Finally, the empirical variance of  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  among the  $B$  bootstrap datasets is a valid estimation of  $\text{Var}(\hat{\mathcal{S}}_{z,g}^{mr}(u))$ . Also, the  $\frac{\alpha}{2} \times 100\%$  and  $(1 - \frac{\alpha}{2}) \times 100\%$  percentiles of empirical distribution of  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  among the  $B$  bootstrap datasets can be taken as its  $(1 - \alpha) \times 100\%$  confidence interval. Standard error and interval estimation of  $\hat{\Delta}_g^{mr}(u)$  can be similarly operationalized.

### 3.4 A simulation study

We conduct Monte Carlo simulations to evaluate the finite-sample performance of the singly robust estimators and the proposed triply robust estimator for estimating the PSCEs under ignorable treatment assignment. We consider two simulation settings where (i) the treatment assignment depends on baseline covariates to mimic an observational study and (ii) the treatment assignment process is randomized to mimic a randomized trial. We examine the bias, Monte Carlo standard error, and 95% bootstrap confidence interval converge rate under correct and incorrect specifications of the working models. The detail data generation process and the simulation results are provided in Web Appendix C. Overall, the simulation results are consistent with our theoretical prediction. In the observational study setting, the triply robust estimator exhibits minimal bias and close-to-nominal coverage when  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$  is correctly specified; the three singly robust estimators only present minimal bias when all of working models used for constructing the estimator are correctly specified. The pattern of simulation results under a randomized study setting is similar to that in the observational study setting, but now the multiply robust estimator is a doubly robust estimator and exhibits minimal bias with nominal coverage when the union model  $\mathcal{M}_{e+C} \cup \mathcal{M}_T$  is correctly specified.

## 4 Sensitivity analyses for identification assumptions

### 4.1 Violation of principal ignorability

We develop a sensitivity analysis strategy to evaluate robustness of the proposed estimators under departure from principal ignorability, which is a key assumption that enables the identification of PSCE estimands. We consider the following two sensitivity functions:

$$\varepsilon_1(t, \mathbf{X}) = \frac{\mathbb{P}(T(1) \geq t | G = c, \mathbf{X})}{\mathbb{P}(T(1) \geq t | G = a, \mathbf{X})}, \quad \varepsilon_0(t, \mathbf{X}) = \frac{\mathbb{P}(T(0) \geq t | G = c, \mathbf{X})}{\mathbb{P}(T(0) \geq t | G = n, \mathbf{X})},$$

which depend on the follow-up time  $t$  and covariates  $\mathbf{X}$ . Assumption 3 is then equivalent to  $\varepsilon_1(t, \mathbf{X}) = \varepsilon_0(t, \mathbf{X}) = 1$  for all  $t \geq 0$ ; and  $\varepsilon_1(t, \mathbf{X})$  or  $\varepsilon_0(t, \mathbf{X}) \neq 1$  implies that Assumption 3 does not hold. Therefore, the two sensitivity functions encode the magnitude and direction of departure from principal ignorability. For instance, in ADAPTABLE,  $\varepsilon_1(t, \mathbf{X})$  is the survival probability ratio between compliers and always high-dose takers if both groups of patients were assigned to high-dose group. If  $\varepsilon_1(t, \mathbf{X}) > 1$ , the compliers has a higher survival probability at a time  $t$  as compared to the always high-dose takers conditional on  $\mathbf{X}$ , which means that the compliers are still healthier than the always high-dose takers due to other unobserved factors. On the contrary, if  $\varepsilon_1(t, \mathbf{X}) < 1$ , then the always high-dose takers are healthier than the compliers due to unobserved factors. Similar interpretations extend to  $\varepsilon_0(t, \mathbf{X})$ .

To implement the sensitivity analysis method, we propose the following explicit forms of the two sensitivity functions:

$$\varepsilon_1(t, \mathbf{X}) = \exp \left\{ \xi_1(\mathbf{X}) \left( \frac{t}{t_{\max}} \right)^{\eta_1(\mathbf{X})} \right\}, \quad \varepsilon_0(t, \mathbf{X}) = \exp \left\{ \xi_0(\mathbf{X}) \left( \frac{t}{t_{\max}} \right)^{\eta_0(\mathbf{X})} \right\},$$

where  $t_{\max}$  is the maximum follow-up time or the maximum time of interest in defining the PSCE (1). Here,  $\{\xi_1(\mathbf{X}), \xi_0(\mathbf{X})\} \in \mathbb{R}^{\otimes 2}$  and  $\{\eta_1(\mathbf{X}), \eta_0(\mathbf{X})\} \in (0, \infty)^{\otimes 2}$  are user-specified parameters controlling the shape of the sensitivity functions, which, for simplicity, may be taken as constant values such as  $\{\xi_1, \xi_0\}$  and  $\{\eta_1, \eta_0\}$ , but can depend on covariates

more generally. The shape of  $\varepsilon_1(t, \mathbf{X})$  is visualized in Figure 1, where we recognize that  $\varepsilon_1(t, \mathbf{X})$  is a monotone function that increases (or decreases) from 1 at  $t = 0$  to  $\exp(\xi_1(\mathbf{X}))$  at  $t = t_{\max}$ , when  $\xi_1(\mathbf{X}) > 0$  (or  $< 0$ ). Therefore, the exponential transformation of  $\xi_1(\mathbf{X})$  defines the value of  $\varepsilon_1(t, \mathbf{X})$  at the maximum time of interest, and is referred to as the *extremum* parameter. The sensitivity parameter  $\eta_1(\mathbf{X})$  is referred to as the *curvature* parameter because it captures the rate of change in  $\varepsilon_1(t, \mathbf{X})$ . When  $\eta_1(\mathbf{X}) < 1$ ,  $\varepsilon_1(t, \mathbf{X})$  changes more slowly for small  $t$  but more dramatically when  $t$  close to  $t_{\max}$ ; in contrast, when  $\eta_1(\mathbf{X}) > 1$ ,  $\varepsilon_1(t, \mathbf{X})$  changes dramatically when  $t$  is close to 0 but more mildly when  $t$  gets bigger. When  $\eta_1(\mathbf{X}) = 1$ ,  $\varepsilon_1(t, \mathbf{X})$  changes as a linear function in the follow-up time. Interpretations of extreme parameter  $\xi_0(\mathbf{X})$  and curvature parameter  $\eta_0(\mathbf{X})$  on the sensitivity function  $\varepsilon_0(t, \mathbf{X})$  are similar to those for  $\varepsilon_1(t, \mathbf{X})$ .

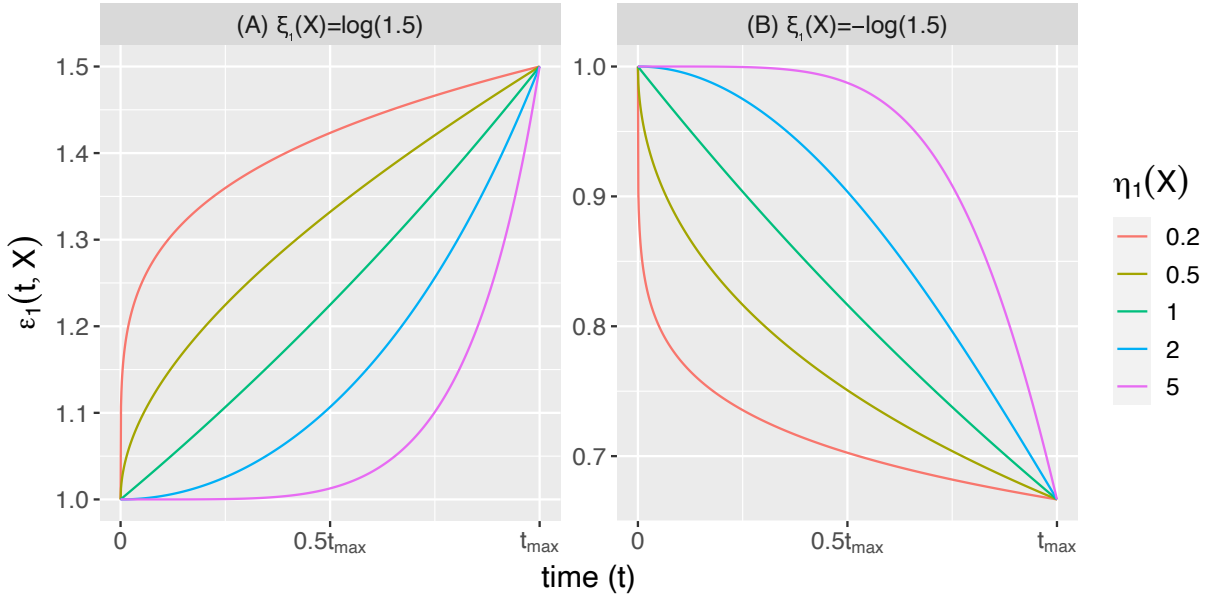


Figure 1: Graphical illustration of the sensitivity function  $\varepsilon_1(t, \mathbf{X})$ , when the  $\xi_1(\mathbf{X}) = \log(1.5)$  (Panel A) or  $\xi_1(\mathbf{X}) = -\log(1.5)$  (Panel B), with  $\eta_1(\mathbf{X}) \in \{0.2, 0.5, 1, 2, 5\}$ .

When  $\varepsilon_0(t, \mathbf{X}) \neq 1$  or  $\varepsilon_1(t, \mathbf{X}) \neq 1$ , the proposed single and multiply robust estimators in Section 3 will in theory be biased. However, we introduce the following weights based on the sensitivity functions to restore unbiased estimation of PSCE:

$$\hat{w}_{1,c}(t, \mathbf{X}) = \frac{\varepsilon_1(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + \varepsilon_1(t, \mathbf{X})e_a(\mathbf{X}; \hat{\gamma})}{\varepsilon_1(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + e_a(\mathbf{X}; \hat{\gamma})}, \quad \hat{w}_{1,a}(t, \mathbf{X}) = \frac{e_c(\mathbf{X}; \hat{\gamma}) + e_a(\mathbf{X}; \hat{\gamma})}{\varepsilon_1(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + e_a(\mathbf{X}; \hat{\gamma})},$$

$$\hat{w}_{0,n}(t, \mathbf{X}) = \frac{e_c(\mathbf{X}; \hat{\gamma}) + e_n(\mathbf{X}; \hat{\gamma})}{\varepsilon_0(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + e_n(\mathbf{X}; \hat{\gamma})}, \quad \hat{w}_{0,c}(t, \mathbf{X}) = \frac{\varepsilon_0(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + \varepsilon_0(t, \mathbf{X})e_n(\mathbf{X}; \hat{\gamma})}{\varepsilon_0(t, \mathbf{X})e_c(\mathbf{X}; \hat{\gamma}) + e_n(\mathbf{X}; \hat{\gamma})}.$$

In Web Appendix D.1, we generalize Result 1 to the scenario with  $\varepsilon_1(t, \mathbf{X})$  or  $\varepsilon_1(t, \mathbf{X}) \neq 1$  by including the additional weights  $\{\hat{w}_{1,c}(u, \mathbf{X}), \hat{w}_{0,c}(u, \mathbf{X})\}$  to remove the bias due to violation of Assumption 3. The explicit forms of the bias-corrected singly robust estimators are developed in Web Appendix D.1. In Web Appendix D.2, we generalize the multiply robust estimator,  $\hat{\mathcal{S}}_{z,g}^{mr}(\mathbf{X})$ , by reweighting the original estimator with the weights  $\hat{w}_{z,g}(t, \mathbf{X})$ ; for example, the new expression of  $\hat{\mathcal{S}}_{1,c}^{mr}(\mathbf{X})$ , i.e., the first component in  $\hat{\Delta}_g^{mr}(\mathbf{X})$ , is

$$\begin{aligned} & \hat{\mathcal{S}}_{1,c}^{mr}(u) \\ = & \mathbb{P}_n \left\{ \hat{w}_{1,c}(u, \mathbf{X}) \frac{e_c(\mathbf{X}; \hat{\gamma})}{\hat{p}_1 - \hat{p}_0} \frac{S}{p_1(\mathbf{X}; \hat{\gamma}_1)} \frac{Z}{\pi(\mathbf{X}; \hat{\alpha})} \left[ \frac{\mathbb{I}(U \geq u)}{\mathcal{S}_{11}^C(u|\mathbf{X}; \hat{\theta}_{11})} + \mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11}) \int_0^u \frac{d\widehat{M}_{11}^C(r|\mathbf{X})}{\mathcal{S}_{11}(r|\mathbf{X}; \hat{\beta}_{11})\mathcal{S}_{11}^C(r|\mathbf{X}; \hat{\theta}_{11})} \right] \right. \\ & + \frac{\hat{w}_{1,c}^2(u, \mathbf{X})}{\varepsilon_1(u, \mathbf{X})} \frac{\mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11})}{\hat{p}_1 - \hat{p}_0} \left[ \frac{1-Z}{1-\pi(\mathbf{X}; \hat{\alpha})} (S - p_0(\mathbf{X}; \hat{\gamma}_0)) + \frac{p_0(\mathbf{X}; \hat{\gamma})}{p_1(\mathbf{X}; \hat{\gamma}_1)} \frac{Z}{\pi(\mathbf{X}; \hat{\alpha})} (p_1(\mathbf{X}; \hat{\gamma}) - S) \right] \\ & \left. + \hat{w}_{1,c}(u, \mathbf{X}) \frac{e_c(\mathbf{X})}{\hat{p}_1 - \hat{p}_0} \left( 1 - \frac{Z}{\pi(\mathbf{X}; \hat{\gamma})} \right) \mathcal{S}_{11}(u|\mathbf{X}; \hat{\beta}_{11}) \right\}. \end{aligned} \quad (10)$$

The following result summarizes the properties of such bias-corrected multiply robust estimators, where the technical details are provided in Web Appendix D.3.

**Result 3 (Double robustness under violation of principal ignorability)** *Suppose that assumptions 1–2 and 4 hold and the sensitivity functions  $\{\varepsilon_1(t, \mathbf{X}), \varepsilon_0(t, \mathbf{X})\}$  are known. For all  $z \in \{0, 1\}$  and  $g \in \{c, n, a\}$ ,  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  developed in Web Appendix D.2 is a doubly robust estimator for  $\mathcal{S}_{z,g}(u)$  in the sense that  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  is consistent to  $\mathcal{S}_{z,g}(u)$  under the union model  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{e+T}$ . As a consequence,  $\hat{\Delta}_g^{mr}(u) = \hat{\mathcal{S}}_{1,g}^{mr}(u) - \hat{\mathcal{S}}_{0,g}^{mr}(u)$  is also doubly robust and consistent to  $\Delta_g(u)$  under  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{e+T}$  for all  $g \in \{c, n, a\}$ .*

Notice that  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  (and  $\hat{\Delta}_g^{mr}(u)$ ) is a doubly robust estimator in the sense that it can provide consistent estimation under either  $\mathcal{M}_{\pi+e+C}$  or  $\mathcal{M}_{e+T}$ . In a randomized trial such as ADAPTABLE, however, we can always consistently estimate  $\mathcal{M}_\pi$  by a working logistic regression  $\pi(\mathbf{X}; \hat{\alpha})$  due to randomization. In that case,  $\hat{\mathcal{S}}_{z,g}^{mr}(u)$  (and  $\hat{\Delta}_g^{mr}(u)$ ) remains doubly robust and is consistent under the union model  $\mathcal{M}_{e+C} \cup \mathcal{M}_{e+T}$ . To conduct the sensitivity

analysis, one can report  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  or  $\widehat{\Delta}_g^{mr}(u)$  over a range of values of  $\{\xi_1(\mathbf{X}), \xi_0(\mathbf{X})\}$ , which capture how the PSCE estimates are affected by deviations from Assumption 3.

## 4.2 Violation of monotonicity

If the monotonicity assumption (Assumption 2) does not hold, we cannot rule out defiers and we four principal strata  $G = \{c, n, a, d\}$  need to be considered for estimating PSCEs. We consider the sensitivity function proposed in Ding and Lu (2017) to capture the deviation from monotonicity:

$$\zeta(\mathbf{X}) = \frac{\mathbb{P}(G = d|\mathbf{X})}{\mathbb{P}(G = c|\mathbf{X})},$$

which takes values from 0 to  $\infty$ . The sensitivity function is  $\zeta(\mathbf{X})$  is the ratio between the probabilities of defiers and compliers conditional on  $\mathbf{X}$ . When  $\zeta(\mathbf{X}) = 0$ , then monotonicity holds, but  $\zeta(\mathbf{X}) > 0$  implies that the defier strata cannot be ignored. With a fixed value of  $\zeta(\mathbf{X})$ , we have then developed bias-corrected estimators for  $\mathcal{S}_{z,g}(u)$ ,  $z \in \{0, 1\}$  and  $g = \{c, n, a, d\}$  in Web Appendix E. Importantly, the multiply robust estimators,  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$ , given in Web Appendix E.2 remain triply robust under the union model  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$ , as summarized in the following result (proof provided in Web Appendix E.3).

**Result 4 (Multiple robustness in the presence of defiers)** *Suppose  $\zeta(\mathbf{X})$  is known and Assumptions 1, 4 and the generalized principal ignorability assumption (Assumption 5 in Web Appendix E) hold. For all  $z \in \{0, 1\}$  and  $g \in \{c, n, a, d\}$ ,  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  developed in Web Appendix E.2 is a triply robust estimator for  $\mathcal{S}_{z,g}(u)$  in the sense that  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  is consistent to  $\mathcal{S}_{z,g}(u)$  under the union model  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$ . As a consequence,  $\widehat{\Delta}_g^{mr}(u) = \widehat{\mathcal{S}}_{1,g}^{mr}(u) - \widehat{\mathcal{S}}_{0,g}^{mr}(u)$  is also consistent to  $\Delta_g(u)$  under  $\mathcal{M}_{\pi+e+C} \cup \mathcal{M}_{\pi+T} \cup \mathcal{M}_{e+T}$  for all  $g \in \{c, n, a\}$ .*

Similar to Section 4.1, one can report  $\widehat{\mathcal{S}}_{z,g}^{mr}(u)$  or  $\widehat{\Delta}_g^{mr}(u)$  over a range of values of  $\zeta(\mathbf{X})$ , which help assess how the PSCE estimates are affected by departures from Assumption 2.

## 5 Applications to the ADAPTABLE pragmatic trial

### 5.1 Description of data and intention-to-treat analysis

We evaluate the comparative effectiveness of two aspirin dosing strategies, 325 mg versus 81 mg per day, in the ADAPTABLE trial. The ADAPTABLE trial began patient enrollment in 2016, and included 15,076 participants with established atherosclerotic cardiovascular disease that were randomly assigned in a 1:1 ratio to take the high dose and low dose. The participants were followed up until June 2020, with a median duration of follow-up time at 26 months and an interquartile range from 19 to 35 months. During follow up, 2,493 participants withdrew consent from study or discontinued aspirin intake, and were excluded from our analysis. Table 1 summarizes the baseline covariates among the  $n = 10,030$  complete-case participants that we used for our analysis to illustrate the main issue of noncompliance. Our data include all participants who neither have missing values in their baseline characteristics nor drop out from the study. In our dataset, all baseline characteristics are well balanced, ensuring to a certain extent that the exclusion of incomplete cases does not lead to systematic difference between the compliance groups.

Similar to the primary report of ADAPTABLE, we consider the primary outcome to be the time to first occurrence of death from any reason and hospitalization from stroke or myocardial infarction. We restrict the maximum follow-up time at  $t_{\max} = 38$  months after randomization, and censor any outcomes beyond this point. The outcome occurred in 358 participants in the low-dose arm (event rate 7.5%) and 386 participants in the high-dose arm (event rate 7.4%). We performed an ITT analysis to estimate the differences of counterfactual survival functions based on the doubly robust approach developed in Bai et al. (2013), where Cox proportional hazard regression was used to model the arm-specific time-to-censoring and time-to-event process conditional on all pre-randomization covariates in Table 1. The ITT results are provided in Web Figure 1 (all Web Figures are available in Supplementary Material), and suggest no differences between these two aspirin dosages on

Table 1: Mean and standard deviation of baseline characteristics stratified by the treatment assignment. ASD is the absolute standardized difference between patients assigned to the 325-mg group and those assigned to the 81-mg group. Given a specific baseline covariate, its ASD is calculated as  $|\bar{x}_1 - \bar{x}_0|/\sqrt{0.5(s_1^2 + s_0^2)}$ , where  $\bar{x}_z$  and  $s_z^2$  are the mean and variance of this covariate within the 325-mg group ( $z = 1$ ) or the 81-mg group ( $z = 0$ ).

Variable	81-mg Group (sample size 5239)	325-mg Group (sample size 4791)	ASD
Age (years)	67.00 (9.45)	66.77 (9.32)	0.02
Female sex	0.29 (0.46)	0.31 (0.46)	0.04
White race	0.84 (0.37)	0.85 (0.36)	0.02
Hispanic ethnicity	0.03 (0.18)	0.03 (0.17)	0.01
Non-internet users	0.15 (0.35)	0.14 (0.35)	0.02
Current smokers	0.09 (0.29)	0.10 (0.30)	0.01
P2Y12 inhibitor	0.23 (0.42)	0.23 (0.42)	0.00
<b>Medical History</b>			
Myocardial infarction	0.37 (0.48)	0.36 (0.48)	0.01
Atrial fibrillation	0.08 (0.27)	0.09 (0.28)	0.03
Percutaneous coronary intervention	0.42 (0.49)	0.42 (0.49)	0.01
Coronary artery disease	0.95 (0.21)	0.96 (0.20)	0.04
Coronary-artery bypass grafting	0.25 (0.44)	0.25 (0.43)	0.02
Hypertension	0.87 (0.34)	0.87 (0.33)	0.01
Hyperlipidemia	0.90 (0.30)	0.91 (0.29)	0.01
Peripheral artery disease	0.24 (0.43)	0.25 (0.44)	0.03
Congestive heart failure	0.23 (0.42)	0.25 (0.43)	0.04
History of bleeding	0.08 (0.28)	0.09 (0.29)	0.03
<b>Aspirin use prior to randomization</b>			
Prior dose: 81 mg	0.85 (0.36)	0.86 (0.35)	0.02
Prior dose: 325 mg	0.12 (0.33)	0.12 (0.33)	0.01

the time-to-event outcome (the two counterfactual survival functions almost overlapped). The ITT analysis, however, does not take treatment noncompliance into consideration, and offer no insights into potential treatment effect heterogeneity across subpopulations defined by compliance patterns. In ADAPTABLE, 38.0% of patients who were assigned to the high-dose arm took the low-dose and 6.5% of patients who were assigned to the low-dose arm took the high-dose, and we apply the proposed methods to investigate the PSCE estimands defined for compliers, always high-dose and always low-dose takers.

## 5.2 Model specification, balance check and strata characteristics

The four working models described in Section 3.1 are used to estimate the treatment assignment process, noncompliance status, the time to censoring, and the time to outcome in the ADAPTABLE trial, adjusting for all baseline characteristics in Table 1. In particular,

the working logistic propensity score model is fitted for adjusting any chance imbalance in baseline covariates, rather than addressing confounding due to assignment.

While the treatment is considered randomized, the noncompliance behavior is not randomized and the principal score models are fitted to address compliance-outcome confounding. To check adequacy of the principal score models, we empirically assess the balance for each baseline covariate across before and after principal score weighting; this step is analogous to balance check for propensity score weighted analysis of observational studies (Li et al., 2018). Specifically, we extend the strategy in Ding and Lu (2017) by defining the following three balancing metrics to quantify the weighted standardized mean differences (SMD) of a given covariate  $X$  across the four observed  $(Z, S)$  cell,  $(Z = 1, S = 1)$ ,  $(Z = 0, S = 1)$ ,  $(Z = 1, S = 0)$ , and  $(Z = 0, S = 0)$ :

$$\begin{aligned} \text{SMD}_c &= \frac{1}{s_c} \left| \mathbb{P}_n \left[ \frac{ZS\mathcal{W}_{1,c}(\mathbf{X})X}{\mathbb{P}_n[ZS]} - \frac{(1-Z)(1-S)\mathcal{W}_{0,c}(\mathbf{X})X}{\mathbb{P}_n[(1-Z)(1-S)]} \right] \right|, \\ \text{SMD}_n &= \frac{1}{s_n} \left| \mathbb{P}_n \left[ \frac{Z(1-S)\mathcal{W}_{1,n}(\mathbf{X})X}{\mathbb{P}_n[Z(1-S)]} - \frac{(1-Z)(1-S)\mathcal{W}_{0,n}(\mathbf{X})X}{\mathbb{P}_n[(1-Z)(1-S)]} \right] \right|, \\ \text{SMD}_a &= \frac{1}{s_a} \left| \mathbb{P}_n \left[ \frac{ZS\mathcal{W}_{1,a}(\mathbf{X})X}{\mathbb{P}_n[ZS]} - \frac{(1-Z)S\mathcal{W}_{0,a}(\mathbf{X})X}{\mathbb{P}_n[(1-Z)S]} \right] \right|, \end{aligned}$$

where  $\mathcal{W}_{z,g}(\mathbf{X})$  for  $z \in \{0, 1\}$  and  $g \in \{c, a, n\}$  are weights, the normalization factor is defined as  $s_c = \sqrt{\frac{s_{11}^2 + s_{00}^2}{2}}$ ,  $s_n = \sqrt{\frac{s_{10}^2 + s_{00}^2}{2}}$ ,  $s_a = \sqrt{\frac{s_{11}^2 + s_{01}^2}{2}}$ , and  $s_{zs}^2$  is the empirical variance of  $X$  within all subjects in the  $(Z = z, S = s)$  cell. By definition,  $\text{SMD}_c$  quantifies the weighted mean difference of  $X$  between the  $(Z = 1, S = 1)$  and  $(Z = 0, S = 0)$  cells,  $\text{SMD}_n$  quantifies the weighted mean difference of  $X$  between the  $(Z = 1, S = 0)$  and  $(Z = 0, S = 0)$  cells, and  $\text{SMD}_a$  quantifies the weighted mean difference of  $X$  between the  $(Z = 1, S = 1)$  and  $(Z = 0, S = 1)$  cells. When  $\mathcal{W}_{z,g}(\mathbf{X}) \equiv 1$ , the SMDs measure the systematic difference across different  $(Z, S)$ -strata, and therefore reflects the amount of confounding due to noncompliance. When  $\mathcal{W}_{z,g}(\mathbf{X})$  is specified as the corresponding principal score weights, that is,  $\mathcal{W}_{1,c}(\mathbf{X}) = \frac{e_c(\mathbf{X}; \hat{\gamma})}{p_1(\mathbf{X}; \hat{\gamma}_1)} \bigg/ \frac{\hat{e}_c}{\hat{p}_1}$ ,  $\mathcal{W}_{0,c}(\mathbf{X}) = \frac{e_c(\mathbf{X}; \hat{\gamma})}{1-p_0(\mathbf{X}; \hat{\gamma}_0)} \bigg/ \frac{\hat{e}_c}{1-\hat{p}_0}$ ,  $\mathcal{W}_{1,n}(\mathbf{X}) = 1$ ,  $\mathcal{W}_{0,n}(\mathbf{X}) = \frac{e_n(\mathbf{X}; \hat{\gamma})}{1-p_0(\mathbf{X}; \hat{\gamma}_0)} \bigg/ \frac{\hat{e}_n}{1-\hat{p}_0}$ , and  $\mathcal{W}_{1,a}(\mathbf{X}) = \frac{e_a(\mathbf{X}; \hat{\gamma})}{p_1(\mathbf{X}; \hat{\gamma}_0)} \bigg/ \frac{\hat{e}_a}{\hat{p}_1}$ ,  $\mathcal{W}_{0,a}(\mathbf{X}) = 1$ , the SMDs quantify the extent to which the confounding is addressed by weighting. In theory, if the



principal score models are correctly specified, the SMDs after the weighting converges to 0, and therefore can be used as a diagnostic check for  $e_g(\mathbf{X}; \gamma)$ .

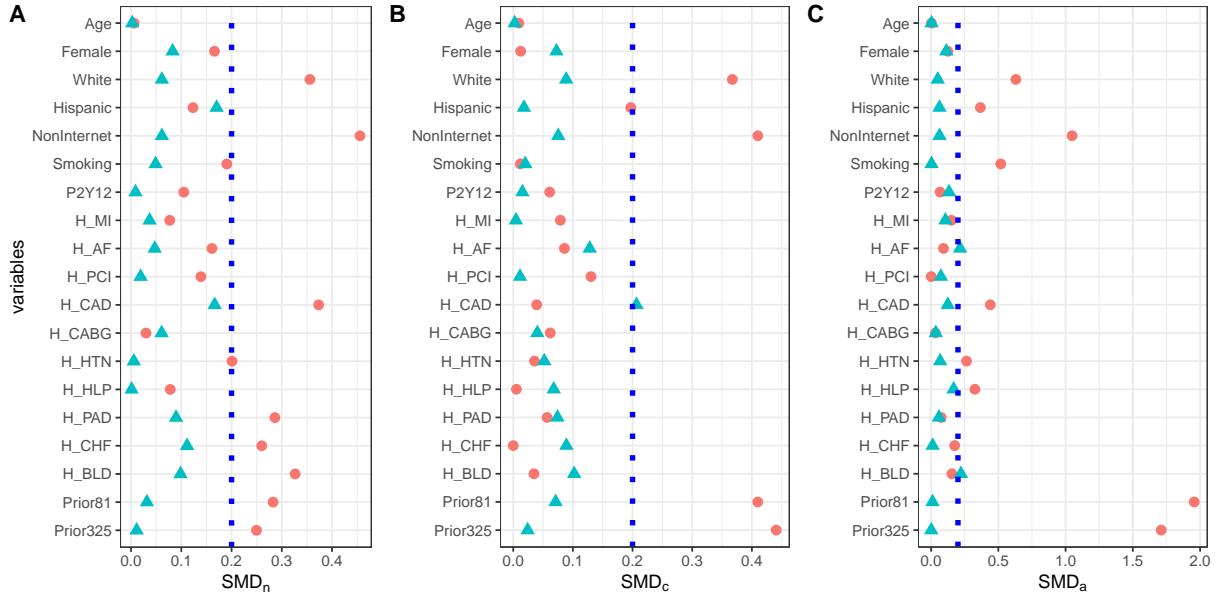


Figure 2: Balance check for baseline characteristics for  $SMD_n$  (panel A),  $SMD_c$  (panel B), and  $SMD_a$  (panel C). The red ‘●’ symbol indicates the unweighted SMDs and the blue ‘▲’ symbol indicates principal score weighted SMDs. The vertical axis shows the names of the variables. Abbreviations: NonInternet - non-internet users; Smoking: current smokers; P2Y12: P2Y12 inhibitor; H\_MI: Myocardial infarction; H\_AF: Atrial fibrillation; H\_PCI: Percutaneous coronary intervention; H\_CAD: Coronary artery disease; H\_CABG: coronary-artery bypass grafting; H\_HTN: Hypertension; H\_HLP: Hyperlipidemia; H\_PAD: peripheral artery disease; H\_CHF: congestive heart failure; H\_BLD: History of bleeding; Prior81: Prior aspirin dose - 81 mg; Prior325: Prior aspirin dose - 325 mg.

Figure 2 provides the love plots of the  $SMD_c$ ,  $SMD_n$ , and  $SMD_a$  for all baseline characteristics before and after principal score weighting. A vertical line is superimposed on this figure to denote a standardized difference of 20%, which we chose as a threshold for adequate balance. It is evident that the principal score weighting improves the balance of baseline characteristics across the  $(Z, S)$ -strata. The SMDs after principal score weighting are generally controlled below 20%, where the only exceptions are the  $SMD_c$  of the coronary artery disease and the  $SMD_a$  of the atrial fibrillation and history of bleeding are slightly (at 21%, 22%, and 22%, respectively). Further inspections indicate that the prevalence of these 3 covariates are either close to 0 or 1, and additional attempts on including interac-

tions into the principal score models has limited improvements on the SMD. Therefore we consider the fitted principal score models adequate.

Table 2: Mean and standard deviation of baseline characteristics among the always low-dose takers, compliers, and always high-dose takers. Max ASD is the maximum pairwise absolute standardized difference across the three principal strata. Given a specific baseline covariate, its Max ASD is calculated as the maximum of  $|\bar{x}_g - \bar{x}_{g'}|/\sqrt{0.5(s_g^2 + s_{g'}^2)}$  for all  $g, g' \in \{c, n, a\}$ , where  $\bar{x}_g$  and  $s_g^2$  are the estimated mean and variance of this covariate in the stratum  $g$ .

Variable	Always low-dose takers	Compliers	Always high-dose takers	Max ASD
Expected sample size	3811	5577	642	
Age (years)	67.70 (9.34)	66.44 (9.36)	66.02 (9.56)	0.18
Female sex	0.32 (0.47)	0.29 (0.45)	0.28 (0.45)	0.09
White race	0.78 (0.41)	0.89 (0.31)	0.80 (0.40)	0.29
Hispanic ethnicity	0.04 (0.20)	0.03 (0.16)	0.04 (0.19)	0.08
Non-internet users	0.21 (0.41)	0.09 (0.28)	0.24 (0.42)	0.41
Current smokers	0.10 (0.31)	0.08 (0.28)	0.14 (0.35)	0.18
P2Y12 inhibitor	0.25 (0.43)	0.22 (0.41)	0.22 (0.42)	0.07
<b>Medical History</b>				
Myocardial infarction	0.39 (0.49)	0.35 (0.48)	0.38 (0.49)	0.09
Atrial fibrillation	0.09 (0.28)	0.08 (0.27)	0.10 (0.30)	0.09
Percutaneous coronary intervention	0.46 (0.50)	0.40 (0.49)	0.39 (0.49)	0.15
Coronary artery disease	0.96 (0.19)	0.95 (0.21)	0.94 (0.24)	0.12
Coronary-artery bypass grafting	0.26 (0.44)	0.25 (0.43)	0.25 (0.43)	0.02
Hypertension	0.89 (0.32)	0.85 (0.35)	0.90 (0.30)	0.13
Hyperlipidemia	0.91 (0.28)	0.90 (0.30)	0.88 (0.32)	0.10
Peripheral artery disease	0.29 (0.45)	0.22 (0.41)	0.25 (0.44)	0.16
Congestive heart failure	0.27 (0.44)	0.22 (0.41)	0.26 (0.44)	0.12
History of bleeding	0.11 (0.31)	0.07 (0.26)	0.10 (0.30)	0.12
<b>Aspirin use prior to randomization</b>				
Prior dose: 81 mg	0.90 (0.30)	0.87 (0.34)	0.45 (0.50)	1.11
Prior dose: 325 mg	0.08 (0.27)	0.11 (0.32)	0.47 (0.50)	0.97

Based on the principal score estimates, we calculate the proportion of the three principal strata. The always low-dose takers and compliers constitute the majority of study population (38.0% and 55.6%), and 6.4% of the study population are always high-dose takers. To provide further intuitions on each subpopulation, Table 2 summarizes the moments of baseline characteristic for each principal stratum. The estimated mean of a baseline covariate  $X \in \mathbf{X}$  in a specific stratum  $g \in \{a, c, n\}$  is calculated by  $\bar{x}_g = \mathbb{P}_n[\frac{e_g(\mathbf{X}; \hat{\gamma})}{\hat{e}_g} X]$ . Similarly, the estimated standard deviation of  $X$  in the stratum  $g$  is obtained by the square root of  $\mathbb{P}_n[\frac{e_g(\mathbf{X}; \hat{\gamma})}{\hat{e}_g} (X - \bar{x}_g)^2]$ . The baseline characteristics appear to be substantially different across

principal strata, as quantified by the maximum pairwise absolute standardized difference (Li and Li, 2019) in the last column of Table 2. For example, the always low-dose takers are older and include slightly more female patients than the compliers and the always high-dose takers. The compliers include fewer non-internet users and more white patients. The always high-dose takers are the youngest but more smokers. In addition, the always low-dose takers appear to be more likely to have cardiovascular disease histories, with higher prevalence of peripheral artery disease, percutaneous coronary intervention, and bleeding history (less healthier subpopulation). Moreover, aspirin use prior to randomization is strongly associated with the aspirin use during the trial and is a strong predictor for the principal strata membership.

### 5.3 Assessing principal survival causal effects

Figure 3 presents the PSCE estimates using the proposed multiply robust estimators, along with the counterfactual survival functions within each principal stratum. The point-wise 95% confidence bands are obtained by the nonparametric bootstrap with  $B = 1,000$  replicates. The results suggest the heterogeneity of causal effects across strata which contributes to the null ITT effect. In particular, the counterfactual survival functions are more separated (compared to ITT) once we stratify by noncompliance patterns. For the always low-dose takers, assignment to the high-dose group leads to a lower survival probability; the magnitude of PSCE peaks at around 33 months ( $\widehat{\Delta}_n^{mr}(33) = -0.025$  with 95% CI from -0.047 to -0.007). For the compliers, however, assignment to the high-dose group has a protective effect on survival during the entire follow-up time. The magnitude of PSCE increases over time but the 95% confidence band generally includes 0. For the always high-dose takers, assignment to the high-dose arm provides the strongest protective effect on patients' risk on death and hospitalization, and the magnitude of PSCE increases over time. For example, the PSCE among the always high-dose takers increases from 0.016 (95% CI: -0.011, 0.047) at 18 months follow-up to 0.059 (95% CI: 0.012, 0.145) at 36 months follow-up. However, the confidence interval for the always high-dose stratum is wider than

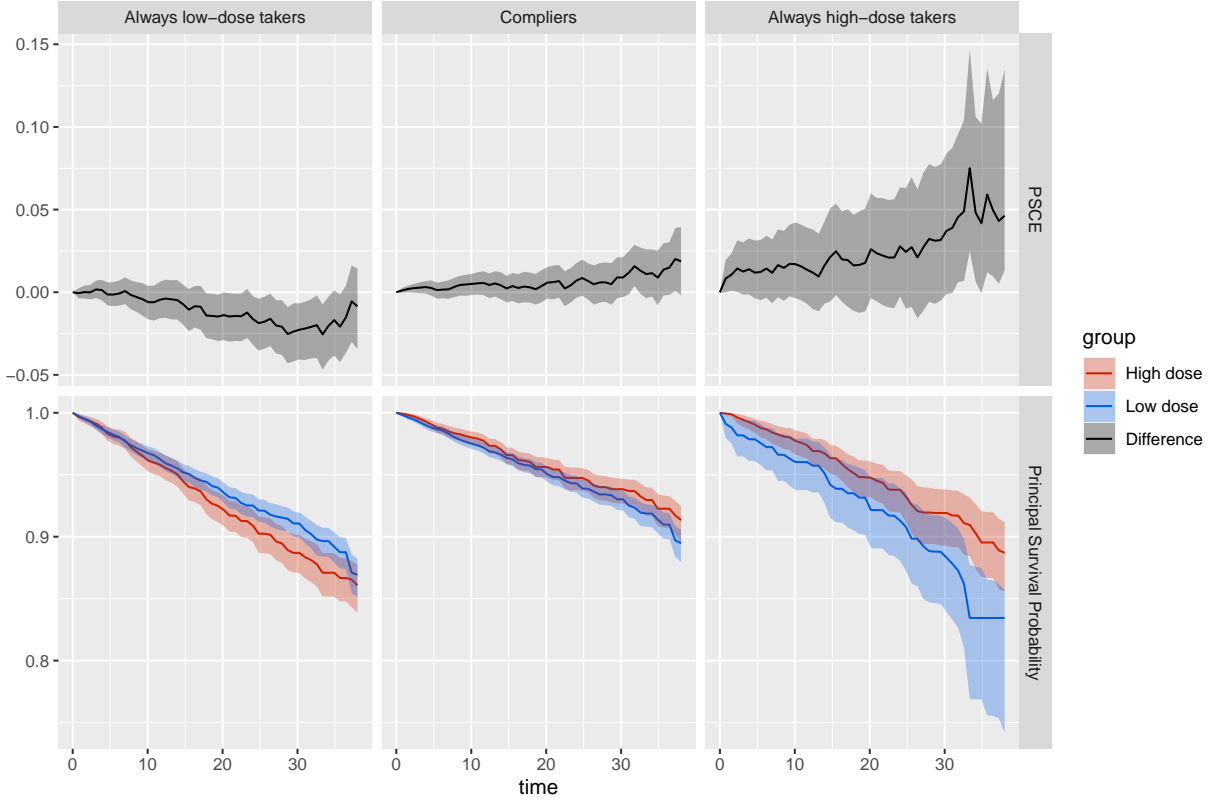


Figure 3: The principal survival causal effects (PSCEs) and the corresponding counterfactual survival functions among the always low-dose takers, compliers, and always high-dose takers, evaluated over 50 equally-spaced grid values. The 95% confidence bands are based on 1000 bootstrap replicates.

the other two strata, possibly due to its small sample size (6.4% of total population are always high-dose takers). In summary, we identify a protective effect due to the high dose strategy among the compliers, and provide preliminary evidence for efficacy. In addition, there appears to be an opposite effect due to assignment to the high-dose arm among the always low-dose takers and the always high-dose takers, possibly due to unmeasured factors and medications for patients in those strata. The nonzero direct effects among these two strata indicate that the commonly-use ER assumption may not be plausible in the ADAPTABLE trial.

## 5.4 Sensitivity analysis for principal ignorability

To better interpret the analysis in Section 5.3, we first examine the sensitivity of PSCE estimates under violations of the principal ignorability assumption. Resuming notation from Section 4.1, we consider constant values  $\{\xi_1, \xi_0\}$  for the extremum parameters and fix the two curvature parameters to 1 (i.e.,  $\eta_1(\mathbf{X}) = \eta_0(\mathbf{X}) = 1$ ); this assumes that the sensitivity functions  $\varepsilon_1(t, \mathbf{X})$  and  $\varepsilon_0(t, \mathbf{X})$  changes at a constant rate during the follow-up time. Due to low event rate (7.4% at  $t_{\max} = 38$  months), if the counterfactual survival function in one stratum is 10% larger or 10% lower than its counterpart in another stratum at the maximum of follow-up time, one of the counterfactual survival function may exceed 1 and hence becomes invalid. Therefore, we only allow  $(\xi_1, \xi_0)$  to vary within the region  $[\log(0.9), \log(1.1)] \times [\log(0.9), \log(1.1)]$  as an exploratory sensitivity analysis.

Table 2 suggests that the always low-dose takers are older and are generally the least healthiest, followed by the compliers and then the always high-dose takers. If this pattern persists and can be further attributed to unmeasured factors that affect survival, the extremum parameters should be restricted such that  $\xi_1 < 0$  and  $\xi_0 > 0$ . Intuitively, this would be the case if unhealthier patients at baseline concern about adverse effects that may occur due to taking the high dose, and therefore healthier patients are more willing to be adhere to the high-dose assignment. Under this *avoid-harm* scenario, Figure 4 presents the PSCE estimates at four values of  $(\xi_1, \xi_0)$  within the region  $[\log(0.9), 0) \times (0, \log(1.1)]$ . We observe that  $\widehat{\Delta}_a^{mr}(u)$  is more robust to such violation of principal ignorability, as  $\widehat{\Delta}_a^{mr}(u)$  is always positive (95% confidence bands exclude zero) for different values of  $(\xi_1, \xi_0)$ . In contrast,  $\widehat{\Delta}_c^{mr}(u)$  and  $\widehat{\Delta}_n^{mr}(u)$  are more sensitive to different values of  $(\xi_1, \xi_0)$  and there exists a tipping point. Specifically, the point estimates,  $\widehat{\Delta}_n^{mr}(u)$  and  $\widehat{\Delta}_c^{mr}(u)$  for  $u \leq t_{\max}$  generally attenuate toward the null when  $\xi_0 = \log(1.05)$  and can move to the opposite direction when  $\xi_0 = \log(1.1)$ . In other words, an unmeasured factor that contributes to 5% higher (on a relative scale) survival among compliers relative to the always low-dose takers (adjusting for all other observed covariates) can explain away the observed causal effects

among the compliers and the always low-dose takers strata.

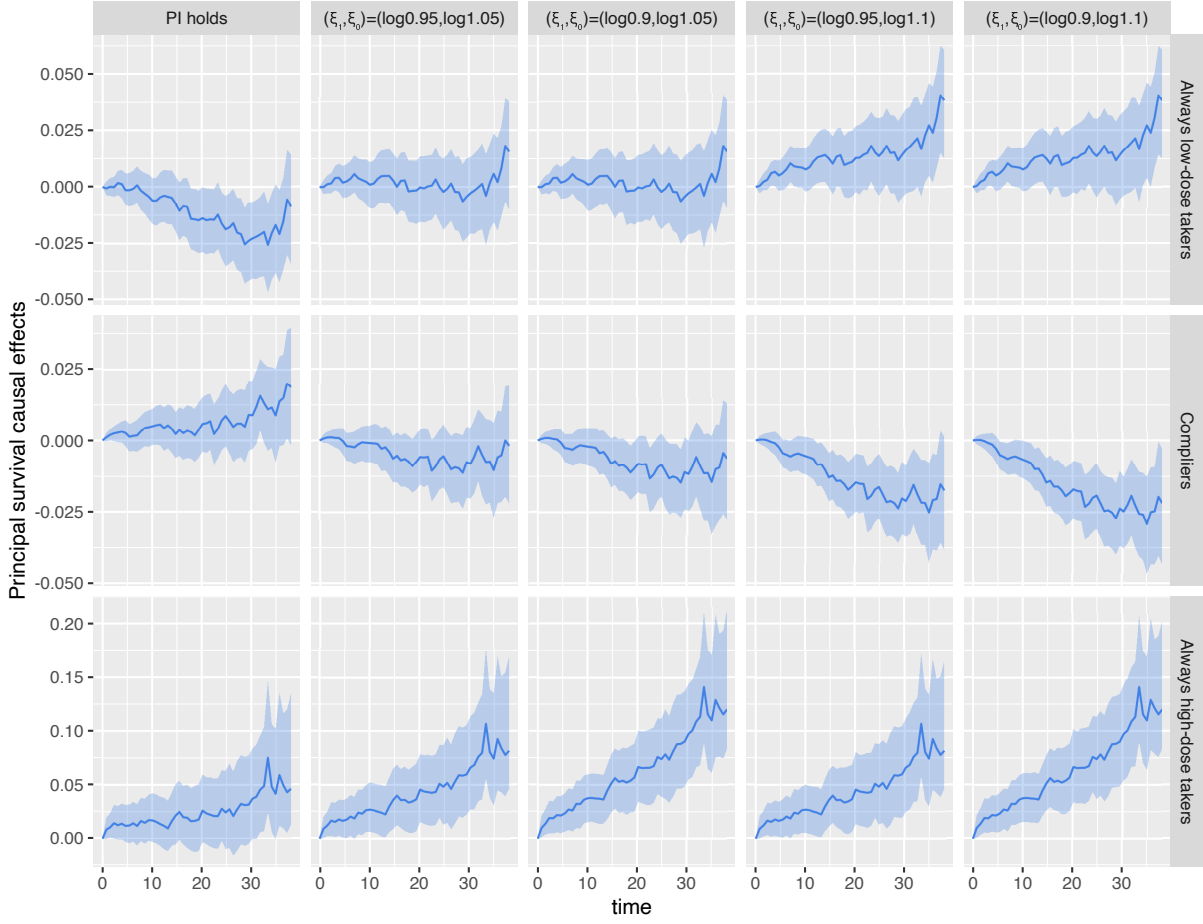


Figure 4: Sensitivity of the PSCEs to the violation of the principal ignorability assumption, under the hypothetical scenario that the always low-dose takers are the unhealthiest group and the always high-dose takers are the healthiest group. The PSCEs under principal ignorability assumption are included in the first column for ease of comparison.

We further consider an opposite scenario by restricting the extremum parameters within  $\xi_1 > 0$  and  $\xi_0 < 0$  such that the always low-dose takers are the healthiest and the always high-dose takers are the unhealthiest due to unmeasured factors. Intuitively, this scenario would occur if patients with more comorbidity consider the high-dose medication as the last resort and there are no better alternative medications. Under this *last-resort* scenario, the estimated PSCEs at four values of  $(\xi_1, \xi_0)$  within the region  $(0, \log(1.1)] \times [\log(0.9), 0)$  are summarized in Web Figure 2. We observe that  $\widehat{\Delta}_c^{mr}(u)$  and  $\widehat{\Delta}_n^{mr}(u)$  are generally robust to this specific type of violation of principal ignorability. The effect estimates  $\widehat{\Delta}_c^{mr}(u)$

and  $\widehat{\Delta}_n^{mr}(u)$  become even stronger as we vary the sensitivity parameters and their 95% confidence bands often exclude zero. However,  $\widehat{\Delta}_a^{mr}(u)$  attenuates toward the null when  $\xi_1$  and  $\xi_0$  deviates from 0, and the associated 95% confidence bands straddle the null.

## 5.5 Sensitivity analysis for monotonicity

While there is little empirical grounds to indicate the existence of defiers, the monotonicity assumption is unverifiable from the observed data alone. We therefore further consider a sensitivity function defined in Section 4.2. For simplicity, we assume  $\zeta(\mathbf{X}) = \zeta$  does not depend on baseline covariates, and choose values of  $\zeta$  under the range suggested in Proposition 2 of the Supplementary Material to ensure that the proportion of each principal stratum is not negative. Specifically, our previous analysis suggests that  $\widehat{p}_1 = 0.620$  and  $\widehat{p}_0 = 0.064$ , and therefore the range of  $\zeta$  are expected to lie between  $\left[0, 1 - \frac{\widehat{p}_1 - \widehat{p}_0}{\min(\widehat{p}_1, 1 - \widehat{p}_0)}\right] = [0, 0.103]$ . This range motivates us to explore  $\zeta = \{0, 0.05, 0.08, 0.09\}$ , and we present the results of  $\widehat{\Delta}_g^{mr}(u)$  under each  $\zeta$  in Web Figure 3. The results imply that the PSCE estimates among the always low-dose takers and compliers are robust even if there exist defiers. The PSCE among the always high-dose takers, however, are relatively sensitive of  $\zeta$ . As  $\zeta$  increases,  $\widehat{\Delta}_a^{mr}(u)$  is further attenuated toward the null and the 95% confidence bands also increase with a larger  $\zeta$ . This is likely because the existence of defiers reduce the sample size for the always high-dose takers, increasing the uncertainty for the PSCE among the always high-dose takers.

## 6 Discussion

We provide a general approach for identification of the principal causal effects in the presence of treatment noncompliance with a right-censored time-to-event outcome. Our approach is particularly useful when the exclusion restriction assumption is likely not plausible. Instead of using the exclusion restriction, we leverage the principal ignorability assumption and develop a multiply robust approach to identify the PSCE. The multiply robust estimator offers additional protection on model misspecification and represents a

principled approach to integrate different identification formulas for the PSCE. We implement the multiply robust estimator to the ADAPTABLE study, and our analysis of the ADAPTABLE trial reveals that the heterogeneity of the average causal effects across principal strata likely contributes to the null effect under the ITT analysis. In particular, we observe a small efficacy signal due to the high-dose aspirin medication among compliers who would always take the assigned treatment. However, the direct effect of assignment has an opposite direction among the always high-dose takers and always low-dose takers, possibly due to unmeasured factors. These results suggest the need to further investigate the mechanisms for such treatment effect heterogeneity across the study subpopulations.

We recognize that the analysis of ADAPTABLE trial relies on two key identification assumptions required for studying noncompliance. First, the principal ignorability assumption is only appropriate when sufficient baseline covariates have been collected to fully deconfound the relationship between noncompliance and the survival outcome of interest. Based on the assumed sensitivity parameters, we investigate to what extent the estimated PSCE may change under two types of violation of the principal ignorability assumption—the avoid-harm and the last-resort scenarios. We found that the causal effects in certain strata can be more sensitive to one type of violation than the other. For example, the avoid-harm scenario may arise when digestive disease is an unmeasured factor such that patients with digestive diseases may prefer not to take high aspirin dosage due to its risk on bleeding but having digestive disease can increase the risk of death. If the impact of such an unmeasured factor on survival leads to 5% more (on a relative scale) survival among the compliers relative to the always low-dose takers, then the observed PSCE within those two strata may be fully explained away. However, in the last-resort scenario, the PSCE within these two strata is robust to unmeasured factors and can even further deviate from the null. Second, we have also assessed the PSCE assuming the existence of defiers and the PSCE within the compliers and always low-dose takers strata are generally robust to violation of the monotonicity assumption, but the PSCE among the always high-dose tak-



ers are relatively sensitive to violation of the monotonicity, possibly due to its attenuated sample size in the presence of defiers.

By focusing on the noncompliance as a major factor that contributes to the null ITT effect, our analysis is currently restricted to a subset of the original participants in ADAPT-ABLE without missing baseline covariates or aspirin discontinuation. Addressing missing baseline covariates and treatment discontinuation necessitates additional structural assumptions and working models to point identify the PSCE, and it remains to be explored whether the multiply robustness property of the proposed estimator can hold in the presence of these additional complexities. While we recognize that randomization may not hold within the subset of participants without missing baseline covariates or aspirin discontinuation (say, for example, when treatment discontinuation is differential across arms), Table 1 suggests that the two arms are balanced in terms of all observed baseline covariates and does not provide clear evidence against randomization within this subset we analyzed. In addition, our development of the multiply robust estimator relies on the standard unconfoundedness assumption in observational studies, and by estimating a treatment propensity score given baseline covariates, our results should still be interpretable if conditional randomization (Assumption 1) holds.

Finally, the unbiased estimation of PSCE requires the correct specification of a certain subset of working models. We consider using finite-dimensional parametric models to estimate these nuisance models, as is commonly done in clinical research practice. When the vector of baseline covariates is high-dimensional, improved estimation of the PSCE may be possible with recent advances in the causal machine learning literature. For example, in the context of the non-censored outcomes, the multiply robust estimating function already provides a natural vehicle to integrate machine learning nuisance models, and admits an optimal estimator that achieve root- $n$  convergence rate under cross fitting (Chernozhukov et al., 2018). Given the recent advances on causal machine learning estimators for survival data in the absence of noncompliance (Westling et al., 2023; Vansteelandt et al., 2022),

expanding the current approach to integrate state-of-the-art machine learners represents a fruitful direction for future research.

## Acknowledgement

The authors gratefully acknowledge 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. We thank the clinical input and motivating questions from investigators of the ADAPTABLE study team, in particular, Sunil Kripalani, Schuyler Jones, Hillary Mulder.

## Supplementary Material

The supplementary material, including the web appendices, tables, and figures, is available at <https://sites.google.com/view/chaocheng>.

## References

- Angrist, J., Imbens, G., and Rubin, D. (1996), “Identification of causal effects using instrumental variables,” *Journal of the American Statistical Association*, 91, 444–455.
- Bai, X., Tsiatis, A. A., and O’Brien, S. M. (2013), “Doubly-Robust estimators of treatment-specific survival distributions in observational studies with stratified sampling,” *Biometrics*, 69, 830–839.
- 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.
- Breslow, N. E. (1972), “Discussion of Professor Cox’s paper,” *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 34, 216–217.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018), “Double/debiased machine learning for treatment and structural parameters,” *The Econometrics Journal*, 21.
- Cuzick, J., Sasieni, P., Myles, J., and Tyrer, J. (2007), “Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 565–588.

- Ding, P. and Lu, J. (2017), “Principal stratification analysis using principal scores,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 757–777.
- Frangakis, C. and Rubin, D. (2002), “Principal Stratification in Causal Inference,” *Biometrics*, 58, 21–29.
- Frangakis, C. E. and Rubin, D. B. (1999), “Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes,” *Biometrika*, 86, 365–379.
- Gill, R. D., Laan, M. J., and Robins, J. M. (1997), “Coarsening at random: Characterizations, conjectures, counter-examples,” in *Proceedings of the First Seattle Symposium in Biostatistics*, Springer, 255–294.
- 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.
- Imbens, G. and Angrist, J. (1994), “Identification and estimation of local average treatment effects,” *Econometrica*, 62, 467–475.
- 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. A. (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., Crenshaw, D. L., Effron, M. B., Re, R. N., Gupta, K., et al. (2021), “Comparative Effectiveness of Aspirin Dosing in Cardiovascular Disease,” *New England Journal of Medicine*, 384, 1981–1990.
- Li, F. and Li, F. (2019), “Propensity score weighting for causal inference with multiple treatments,” *The Annals of Applied Statistics*, 13, 2389–2415.
- Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018), “Balancing covariates via propensity score weighting,” *Journal of the American Statistical Association*, 113, 390–400.
- Loeys, T. and Goetghebeur, E. (2003), “A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance,” *Biometrics*, 59, 100–105.
- 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.
- Robins, J. M. and Finkelstein, D. M. (2000), “Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests,” *Biometrics*, 56, 779–788.

- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of Educational Psychology*, 66, 688.
- Tsiatis, A. A. (2006), *Semiparametric theory and missing data*, Springer.
- Vansteelandt, S., Dukes, O., Van Lancker, K., and Martinussen, T. (2022), “Assumption-lean Cox regression,” *Journal of the American Statistical Association*, 1–10.
- 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.
- Westling, T., Luedtke, A., Gilbert, P. B., and Carone, M. (2023), “Inference for treatment-specific survival curves using machine learning,” *Journal of the American Statistical Association*, 1–26.
- 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.
- Zeng, S., Li, F., Wang, R., and Li, F. (2021), “Propensity score weighting for covariate adjustment in randomized clinical trials,” *Statistics in Medicine*, 40, 842–858.