# Causal survival analysis

## Estimation of the Average Treatment Effect (ATE): Practical Recommendations

Charlotte Voinot Clément Berenfeld Imke Mayer
Bernard Sebastien Julie Josse

2025-01-23

Causal survival analysis combines survival analysis and causal inference to evaluate the effect of a treatment or intervention on a time-to-event outcome, such as survival time. It offers an alternative to relying solely on Cox models for assessing these effects. In this paper, we present a comprehensive review of estimators for the average treatment effect measured with the restricted mean survival time, including regression-based methods, weighting approaches, and hybrid techniques. We investigate their theoretical properties and compare their performance through extensive numerical experiments. Our analysis focuses on the finite-sample behavior of these estimators, the influence of nuisance parameter selection, and their robustness and stability under model misspecification. By bridging theoretical insights with practical evaluation, we aim to equip practitioners with both state-of-the-art implementations of these methods and practical guidelines for selecting appropriate estimators for treatment effect estimation. Among the approaches considered, G-formula two-learners, AIPCW-AIPTW, Buckley-James estimators, and causal survival forests emerge as particularly promising.

## Introduction

### Context and motivations

Causal survival analysis is a growing field that integrates causal inference (D. B. Rubin 1974; Hernán and Robins 2010) with survival analysis (Kalbfleisch and Prentice 2002) to evaluate the impact of treatments on time-to-event outcomes, while accounting for censoring situations where only partial information about an event's occurrence is available.

Being a relatively new domain, the existing literature, though vast, remains fragmented. As a result, a clear understanding of the theoretical properties of various estimators is challenging to obtain. Moreover, the implementation of proposed methods is limited, leaving researchers confronted with a range of available estimators and the need to make numerous methodological decisions. There is a pressing need for a comprehensive survey that organizes the available methods, outlines the underlying assumptions, and provides an evaluation of estimator performance — particularly in finite sample settings. Such a survey also has the potential to help identify remaining methodological challenges that need to be addressed. This need becomes increasingly urgent as causal survival analysis gains traction in both theoretical and applied domains. For instance, its applications to external control arm analyses are particularly relevant in the context of single-arm clinical trials, where traditional comparator arms are unavailable. Regulatory guidelines have begun to acknowledge and support such semi-experimental approaches, reflecting the broader evolution of trial design and therapeutic innovation in precision medicine, see for instance (European Medecines Agency 2024).

By synthesizing the theoretical foundations, assumptions, and performance of various estimators, a survey on existing causal survival analysis methods would provide researchers and practitioners with the necessary tools to make informed methodological choices. This is crucial for fostering robust and reliable applications of causal survival analysis in both academic research and practical settings, where precise and valid results are paramount.

In this paper, we focus our attention to the estimation of the Restricted Mean Survival Time (RMST), a popular causal measure in survival analysis which offers an intuitive interpretation of the average survival time over a specified period. In particular, we decided to not cover the estimation of Hazard Ratio (HR), which has been prominently used but often questioned due to its potential non-causal nature (Martinussen, Vansteelandt, and Andersen 2020). Additionally, unlike the Hazard Ratio, the RMST has the desirable property of being a collapsible measure, meaning that the population effect can be expressed as a weighted average of subgroup effects, with positive weights that sum to one (Huitfeldt, Stensrud, and Suzuki 2019).

## Definition of the estimand: the RMST

We set the analysis in the potential outcome framework, where a patient, described by a vector of covariates $X \in \mathbb{R}^p$, either receives a treatment ($A = 1$) or is in the control group ($A = 0$). The patient will then survive up to a certain time $T(0) \in \mathbb{R}^+$ in the control group, or up to a time $T(1) \in \mathbb{R}^+$ in the treatment group. In practice, we cannot simultaneously have access to $T(0)$ and $T(1)$, as one patient is either treated or control, but only to $T$ defined as follows:

**Assumption.** (Stable Unit Treatment Value Assumption: SUTVA)

$$T = AT(1) + (1 - A)T(0). \tag{1}$$

Due to potential censoring, the outcome $T$ is not completely observed. The most common form of censoring is right-censoring (also known as type II censoring), which occurs when the event of interest has not taken place by the end of the observation period, indicating that it may have occurred later if the observation had continued (Turkson, Ayiah-Mensah, and Nimoh

2021). We focus in this study on this type of censoring only and we assume that we observe $\tilde{T} = T \wedge C = \min(T, C)$ for some censoring time $C \in \mathbb{R}^+$. When an observation is censored, the observed time is equal to the censoring time.

We also assume that we know whether an outcome is censored or not. In other words, we observe the censoring status variable $\Delta = \mathbb{I}\{T \leqslant C\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. $\Delta$ is 1 if the true outcome is observed, and 0 if it is censored.

We assume observing a $n$-sample of variables $(X, A, \widetilde{T}, \Delta)$ stemming from an $n$-sample of the partially unobservable variables $(X, A, T(0), T(1), C)$. A toy exemple of such data is given in Table 1.

Table 1: Example of a survival dataset. In practice, only $X, A, \widetilde{T}$ and $\Delta$ are observed.

| ID | Covariates | | | Treat-ment | Censo-ring | Status | Potential out-comes | | True out-come | Observed out-come |
|---|---|---|---|---|---|---|---|---|---|---|
| i | $X_1$ | $X_2$ | $X_3$ | $A$ | $C$ | $\Delta$ | $T(0)$ | $T(1)$ | $T$ | $\tilde{T}$ |
| 1 | 1 | 1.5 | 4 | 1 | ? | 1 | ? | 200 | 200 | 200 |
| 2 | 5 | 1 | 2 | 0 | ? | 1 | 100 | ? | 100 | 100 |
| 3 | 9 | 0.5 | 3 | 1 | 200 | 0 | ? | ? | ? | 200 |

Our aim is to estimate the Average Treatment Effect (ATE) defined as the difference between the Restricted Mean Survival Time of the treated and controls (Royston and Parmar 2013).

**Definition 0.1.** (Causal effect: Difference between Restricted Mean Survival Time)

$$\theta_{\mathrm{RMST}} = \mathbb{E}\left[T(1) \wedge \tau - T(0) \wedge \tau\right],$$

where $a \wedge b := \min(a, b)$ for $a, b \in \mathbb{R}$.

We define the survival functions $S^{(a)}(t) := \mathbb{P}(T(a) > t)$ for $a \in \{0, 1\}$, i.e., the probability that a treated or non-treated individual will survive beyond a given time $t$. Likewise, we let $S(t) := \mathbb{P}(T > t)$, and $S_C(t) := \mathbb{P}(C > t)$. We also let $G(t) := \mathbb{P}(C \geqslant t)$ be the left-limit of the survival function $S_C$. Because $T(a) \wedge \tau$ are non-negative random variables, one can easily express the restricted mean survival time using the survival functions:

$$\mathbb{E}(T(a) \wedge \tau) = \int_0^\infty \mathbb{P}(T(a) \wedge \tau > t)dt = \int_0^\tau S^{(a)}(t)dt. \tag{2}$$

Consequently, $\theta_{\mathrm{RMST}}$ can be interpreted as the mean difference between the survival function of treated and control until a fixed time horizon $\tau$. A difference in RMST $\theta_{\mathrm{RMST}} = 10$ days

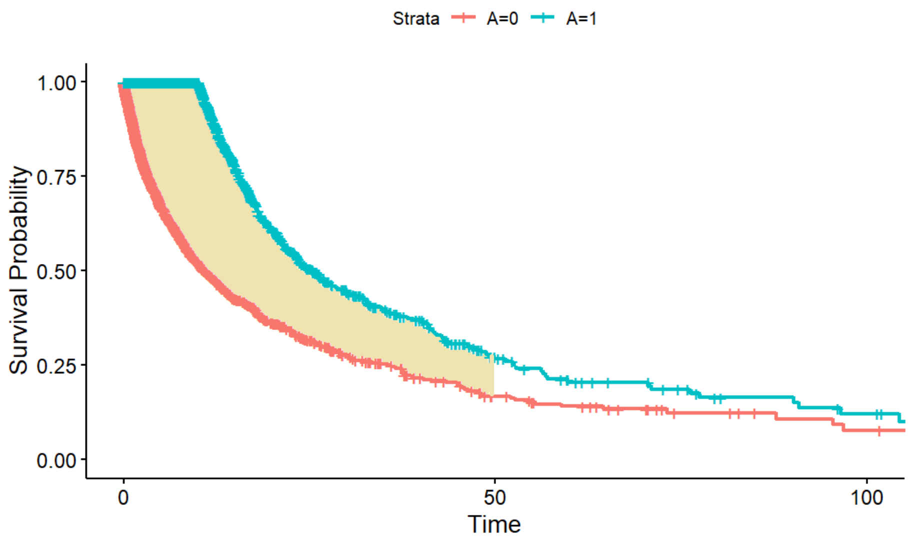



Figure 1: Plot of the estimated survival curves on synthetic toy-data. The $\theta_{\text{RMST}}$ at $\tau = 50$ corresponds to the yellow shaded area between the two survival curves. The curves have been estimated using Kaplan-Meier estimator, see Section .

with $\tau = 200$ means that, on average, the treatment increases the survival time by 10 days at 200 days. We give a visual interpretation of RMST in Figure 1.

Although the present work focuses on the estimation of the difference in RMST, we would like to stress that the causal effect can be assessed through other measures, such as for instance the difference of the survival functions

$$\theta_{\text{SP}} := S^{(1)}(\tau) - S^{(0)}(\tau)$$

for some time $\tau$, see for instance (Ozenne et al. 2020). As mentionned in Section , another widely used measure (though not necessarily causal) is the hazards ratio, defined as

$$\theta_{\text{HR}} := \frac{\lambda^{(1)}(\tau)}{\lambda^{(0)}(\tau)},$$

where the hazard function $\lambda^{(a)}$ is defined as

$$\lambda^{(a)}(t) := \lim_{h \to 0^+} \frac{\mathbb{P}(T(a) \in [t, t+j) | T(a) \geqslant t)}{h}.$$

in a continuous setting, or as $\lambda^{(a)}(t) := \mathbb{P}(T(a) = t | T(a) \geqslant t)$ when the survival times are discrete. The hazard functions and the survival functions are linked through the identities

$$S^{(a)}(t) = \exp\left(-\Lambda^{(a)}(t)\right) \quad \text{where} \quad \Lambda^{(a)}(t) := \int_0^t \lambda^{(a)}(s)\,\mathrm{d}s, \tag{3}$$

in the continuous case. The functions $\Lambda^{(a)}$ are call the cumulative hazard functions. In the discrete case, we have in turn

$$S^{(a)}(t) = \prod_{t_k \leqslant t} \left(1 - \lambda^{(a)}(t_k)\right), \tag{4}$$

where $\{t_1, \ldots, t_K\}$ are the atoms of $T^{(a)}$. These hazard functions are classically used to model the survival times and the censoring times, see Section .

## Organisation of the paper

In this paper, we detail the minimal theoretical framework that allows the analysis of established RMST estimators in the context of both Randomized Controlled Trials (Section ) and observational data (Section ). We give their statistical properties (consistency, asymptotic normality) along with proofs when possible. We then conduct in Section  a numerical study of these estimators through simulations under various experimental conditions, including independent and conditionally independent censoring and correct and incorrect model specifications. We conclude in Section  with practical recommendations on estimator selection, based on criteria such as asymptotic behavior, computational complexity, and efficiency.

## Notations

We provide in Table 2 a summary of the notation used throughout the paper.

Table 2: Summary of the notations.

| Symbol | Description |
|---|---|
| $X$ | Covariates |
| $A$ | Treatment indicator ($A = 1$ for treatment, $A = 0$ for control) |
| $T$ | Survival time |
| $T(a), a \in \{0, 1\}$ | Potential survival time respectively with and without treatment |
| $S^{(a)}, a \in \{0, 1\}$ | Survival function $S^{(a)}(t) = \mathbb{P}(T(a) > t)$ of the potential survival times |
| $\lambda^{(a)}, a \in \{0, 1\}$ | Hazard function $\lambda^{(a)}(t) = \lim_{h \to 0^+} \mathbb{P}(T(a) \in [t, t+h) \vert T(a) \geqslant t)/h$ of the potential survival times |
| $\Lambda^{(a)}, a \in \{0, 1\}$ | Cumulative hazard function of the potential survival times |
| $C$ | Censoring time |
| $S_C$ | Survival function $S_C(t) = \mathbb{P}(C > t)$ of the censoring time |
| $G$ | Left-limit of the survival function $G(t) = \mathbb{P}(C \geqslant t)$ of the censoring time |
| $\widetilde{T}$ | Observed time ($T \wedge C$) |
| $\Delta$ | Censoring status $\mathbb{I}\{T \leqslant C\}$ |
| $\Delta^\tau$ | Censoring status of the restricted time $\Delta^\tau = \max\{\Delta, \mathbb{I}\{\widetilde{T} \geqslant \tau\}\}$ |

| Symbol | Description |
|---|---|
| $\{t_1, t_2, \ldots, t_K\}$ | Discrete times |
| $e(x)$ | Propensity score $\mathbb{E}[A \vert X = x]$ |
| $\mu(x, a), a \in \{0, 1\}$ | $\mathbb{E}[T \wedge \tau \mid X = x, A = a]$ |
| $S(t \vert x, a), a \in \{0, 1\}$ | Conditional survival function, $\mathbb{P}(T > t \vert X = x, A = a)$. |
| $\lambda^{(a)}(t \vert x), a \in \{0, 1\}$ | Conditional hazard functions of the potential survival times |
| $G(t \vert x, a), a \in \{0, 1\}$ | left-limit of the conditional survival function of the censoring $\mathbb{P}(C \geqslant t \vert X = x, A = a)$ |
| $Q_S(t \vert x, a), a \in \{0, 1\}$ | $\mathbb{E}[T \wedge \tau \mid X = x, A = a, T \wedge \tau > t]$ |

## Causal survival analysis in Randomized Controlled Trials

Randomized Controlled Trials (RCTs) are the gold standard for establishing the effect of a treatment on an outcome, because treatment allocation is controlled through randomization, which ensures (asymptotically) the balance of covariates between treated and controls, and thus avoids problems of confounding between treatment groups. The core assumption in a classical RCT is the random assignment of the treatment (D. B. Rubin 1974).

**Assumption.** (Random treatment assignment) There holds:

$$A \perp\!\!\!\perp T(0), T(1), X \tag{5}$$

We also assume that there is a positive probability of receiving the treatment, which we rephrase under the following assumption.

**Assumption.** (Trial positivity)

$$0 < \mathbb{P}(A = 1) < 1 \tag{6}$$

Under Assumptions 5 and 6, classical causal identifiability equations can be written to express $\theta_{\text{RMST}}$ without potential outcomes.

$$\begin{aligned}
\theta_{\text{RMST}} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\
&= \mathbb{E}[T(1) \wedge \tau | A = 1] - \mathbb{E}[T(0) \wedge \tau | A = 0] \quad \text{(Random treatment assignment)} \\
&= \mathbb{E}[T \wedge \tau | A = 1] - \mathbb{E}[T \wedge \tau | A = 0]. \quad \text{(SUTVA)}
\end{aligned} \tag{7}$$

However, Equation 7 still depends on $T$, which remains only partially observed due to censoring. To ensure that censoring does not compromise the identifiability of treatment effects, we must impose certain assumptions on the censoring process, standards in survival analysis, namely, independent censoring and conditionally independent censoring. These assumptions lead to different estimation approaches. We focus on two strategies: those that aim to directly estimate $\mathbb{E}[T \wedge \tau | A = a]$ directly (e.g., through censoring-unbiased transformations, see Section ), and those that first estimate the survival curves to derive RMST via Equation 2 (such as the Kaplan-Meier estimator and all its variants, see the next Section).

### Independent censoring: the Kaplan-Meier estimator

In a first approach, one might assume that the censoring times are independent from the rest of the variables.

**Assumption.** (Independent censoring)

$$C \perp\!\!\!\perp T(0), T(1), X, A. \tag{8}$$

Under Equation 8, subjects censored at time $t$ are representative of all subjects who remain at risk at time $t$. Figure 2 represents the causal graph when the study is randomized and outcomes are observed under independent censoring.

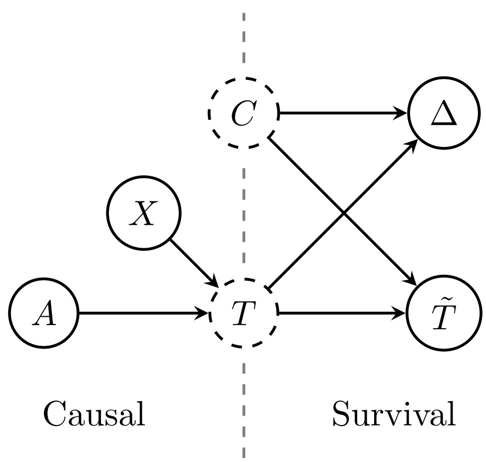



Figure 2: Causal graph in RCT survival data with independent censoring.

We also assume that there is no almost-sure upper bound on the censoring time before $\tau$, which we rephrase under the following assumption.

**Assumption.** (Positivity of the censoring process) There exists $\varepsilon > 0$ such that

$$G(t) \geqslant \varepsilon \quad \text{for all} \quad t \in [0, \tau). \tag{9}$$

If indeed it was the case that $\mathbb{P}(C < t) = 1$ for some $t < \tau$, then we would not be able to infer anything on the survival function on the interval $[t, \tau]$ as all observation times $\widetilde{T}_i$ would be in

$[0, t]$ almost surely. In practice, adjusting the threshold time $\tau$ can help satisfy the positivity assumption. For instance, in a clinical study, if a subgroup of patients has zero probability of remaining uncensored at a given time, $\tau$ can be modified to ensure that participants have a feasible chance of remaining uncensored up to the revised threshold.

The two Assumptions 8 and 9 together allow the distributions of $T(a)$ to be identifiable, in the sense that there exists an identity that expresses $S^{(a)}$ as a function of the joint distribution of $(\widetilde{T}, \Delta, A = a)$, see for instance Ebrahimi, Molefe, and Ying (2003) for such a formula in a non-causal framework. This enables several estimation strategies, the most well-known being the Kaplan-Meier product-limit estimator.

To motivate the definition of the latter and explicit the identifiability identity, we set the analysis in the discrete case. We let $\{t_k\}_{k\geqslant 1}$ be a set of positive and increasing times and assume that $T \in \{t_k\}_{k\geqslant 1}$ almost surely. Then for any $t \in [0, \tau]$, it holds, letting $t_0 = 0$ by convention, thanks to Equation 4,

$$\begin{aligned} S(t|A = a) = \mathbb{P}(T > t|A = a) &= \prod_{t_k \leqslant t} (1 - \mathbb{P}(T = t_k | T > t_{k-1}, A = a)) \\ &= \prod_{t_k \leqslant t} \left(1 - \frac{\mathbb{P}(T = t_k, A = a)}{\mathbb{P}(T \geqslant t_k, A = a)}\right). \end{aligned}$$

Using Assumptions 8 and 9, we find that

$$\frac{\mathbb{P}(T = t_k, A = a)}{\mathbb{P}(T \geqslant t_k, A = a)} = \frac{\mathbb{P}(T = t_k, C \geqslant t_k, A = a)}{\mathbb{P}(T \geqslant t_k, C \geqslant t_k, A = a)} = \frac{\mathbb{P}(\widetilde{T} = t_k, \Delta = 1, A = a)}{\mathbb{P}(\widetilde{T} \geqslant t_k, A = a)}, \tag{10}$$

yielding the final identity

$$S(t|A = a) = \prod_{t_k \leqslant t} \left(1 - \frac{\mathbb{P}(\widetilde{T} = t_k, \Delta = 1, A = a)}{\mathbb{P}(\widetilde{T} \geqslant t_k, A = a)}\right). \tag{11}$$

Notice that the right hand side only depends on the distribution of the observed tuple $(A, \widetilde{T}, \Delta)$. This last equation suggests in turn to introduce the quantities

$$D_k(a) := \sum_{i=1}^{n} \mathbb{I}(\widetilde{T}_i = t_k, \Delta_i = 1, A = a) \quad \text{and} \quad N_k(a) := \sum_{i=1}^{n} \mathbb{I}(\widetilde{T}_i \geqslant t_k, A = a), \tag{12}$$

which correspond respectively to the number of deaths $D_k(a)$ and of individuals at risk $N_k(a)$ at time $t_k$ in the treated group (a=1) or in the control group (a=0).

**Definition 0.1.** (Kaplan-Meier estimator, Kaplan and Meier (1958)) With $D_k(a)$ and $N_k(a)$ defined in Equation 12, we let

$$\widehat{S}_{\mathrm{KM}}(t|A = a) := \prod_{t_k \leqslant t} \left(1 - \frac{D_k(a)}{N_k(a)}\right). \tag{13}$$

The assiociated RMST estimator is then simply defined as

$$\widehat{\theta}_{\mathrm{KM}} = \int_0^\tau \widehat{S}_{KM}(t|A=1) - \widehat{S}_{KM}(t|A=0)dt. \tag{14}$$

The Kaplan-Meier estimator is the Maximum Likelihood Estimator (MLE) of the survival functions, see for instance Kaplan and Meier (1958). Furthermore, because $D_k(a)$ and $N_k(a)$ are sums of i.i.d. random variables, the Kaplan-Meier estimator inherits some convenient statistical properties.

**Proposition 0.1.** *Under Assumptions 1, 5, 6, 8 and 9, and for all $t \in [0,\tau]$, the estimator $\widehat{S}_{\mathrm{KM}}(t|A=a)$ of $S^{(a)}(t)$ is strongly consistent and admits the following bounds for its bias:*

$$0 \leqslant S^{(a)}(t) - \mathbb{E}[\widehat{S}_{\mathrm{KM}}(t|A=a)] \leqslant O(\mathbb{P}(N_k(a)=0)),$$

*where $k$ is the greatest time $t_k$ such that $t \geqslant t_k$.*

Gill (1983) gives a more precise lower-bound on the bias in the case of continuous distributions, which was subsequently refined by Zhou (1988). The bound we give, although slightly looser, still exhibits the same asymptotic regime. In particular, as soon as $S^{(a)}(t) > 0$ (and Assumption 9 holds), then the bias decays exponentially fast towards 0. We give in Section  a simple proof of our bound is our context.

**Proposition 0.2.** *Under Assumptions 1, 5, 6, 8 and 9, and for all $t \in [0,\tau]$, $\widehat{S}_{\mathrm{KM}}(t|A=a)$ is asymptotically normal and $\sqrt{n}\left(\widehat{S}_{\mathrm{KM}}(t|A=a) - S^{(a)}(t)\right)$ converges in distribution towards a centered Gaussian of variance*

$$V_{\mathrm{KM}}(t|A=a) := S^{(a)}(t)^2 \sum_{t_k \leqslant t} \frac{1-s_k(a)}{s_k(a)r_k(a)},$$

*where $s_k(a) = S^{(a)}(t_k)/S^{(a)}(t_{k-1})$ and $r_k(a) = \mathbb{P}(\widetilde{T} \geqslant t_k, A=a)$.*

The proof of Proposition 0.2 can be found in Section . Because $D_k(a)/N_k(a)$ is a natural estimator of $1-s_k(a)$ and, $\frac{1}{n}N_k(a)$ a natural estimator for $r_k(a)$, the asymptotic variance of the Kaplan-Meier estimator can be estimated with the so-called Greenwood formula, as already derived heuristically in Kaplan and Meier (1958):

$$\widehat{\mathrm{Var}}\left(\widehat{S}_{\mathrm{KM}}(t|A=a)\right) := \widehat{S}_{\mathrm{KM}}(t|A=a)^2 \sum_{t_k \leqslant t} \frac{D_k(a)}{N_k(a)(N_k(a)-D_k(a))}. \tag{15}$$

We finally mention that the KM estimator as defined in Definition 0.1 still makes sense in a non-discrete setting, and one only needs to replace the fixed grid $\{t_k\}$ by the values at which we observed an event ($\widetilde{T}_i = t_k, \Delta_i = 1$). We refer to Breslow and Crowley (1974) for a study of this estimator in the continuous case and to Aalen, Borgan, and Gjessing (2008), Sec 3.2 for a general study of the KM estimator through the prism of point processes.

### Conditionally independent censoring

An alternative hypothesis in survival analysis that relaxes the assumption of independent censoring is conditionally independent censoring, also refered sometimes as *informative censoring.* It allows to model more realistic censoring processes, in particular in situations where there are reasons to believe that $C$ may be dependent from $A$ and $X$ (for instance, if patient is more like to leave the study when treated because of side-effects of the treatment).

**Assumption.** (Conditionally independent censoring)

$$C \perp\!\!\!\perp T(0), T(1) \mid X, A \tag{16}$$

Under Equation 16, subjects within a same stratum defined by $X = x$ and $A = a$ have equal probability of censoring at time $t$, for all $t$. In case of conditionally independent censoring, we also need to assume that all subjects have a positive probability to remain uncensored at their time-to-event.

**Assumption.** (Positivity / Overlap for censoring) There exists $\varepsilon > 0$ such that for all $t \in [0, \tau)$, it almost surely holds

$$G(t|A, X) \geqslant \varepsilon. \tag{17}$$

Figure 3 represents the causal graph when the study is randomized with conditionally independent censoring.

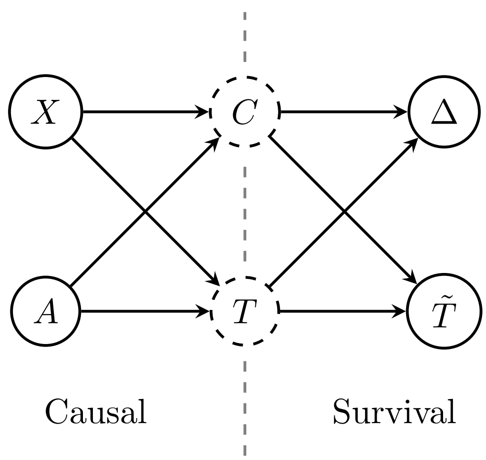



Figure 3: Causal graph in RCT survival data with dependent censoring.

Under dependent censoring, the Kaplan-Meier estimator as defined in Definition 0.1 can fail to estimate survival, in particular because Equation 10 does not hold anymore. Alternatives include plug-in G-formula estimators (Section ) and unbiased transformations (Section ).

### The G-formula and the Cox Model

Because the censoring is now independent from the potential outcome conditionally to the covariates, it would seem natural to model the response of the survival time conditionally to these covariates too:

$$\mu(x, a) := \mathbb{E}[T \wedge \tau | X = x, A = a].$$

Building on Equation 7, one can express the RMST as a function of $\mu$:

$$\theta_{\mathrm{RMST}} = \mathbb{E}\left[\mathbb{E}[T \wedge \tau | X, A = 1]\right] - \mathbb{E}\left[T \wedge \tau | X, A = 0\right]\right] = \mathbb{E}[\mu(X, 1) - \mu(X, 0)].$$

An estimator $\widehat{\mu}$ of $\mu$ would then straightforwardly yield an estimator of the difference in RMST through the so-called *G-formula* plug-in estimator:

$$\widehat{\theta}_{\mathrm{G-formula}} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}\left(X_i, 1\right) - \widehat{\mu}\left(X_i, 0\right). \tag{18}$$

We would like to stress that a G-formula approach works also in a observational context as the one introduced in Section . However, because the estimation strategies presented in the next sections relies on estimating nuisance parameters, and that this latter task is often done in the same way as we estimate the conditional response $\mu$, we decided to not delay the introduction of the G-formula any further, and we present below a few estimation methods for $\mu$. These methods are sub-divised in two categories: *T-learners*, where $\mu(\cdot, 1)$ is estimated separately from $\mu(\cdot, 0)$, and _$S$-learners, where $\widehat{\mu}$ is obtained by fitting a single model based on covariates $(X, A)$.

**Cox's Model**. There are many ways to model $\mu$ in a survival context, the most notorious of which being the Cox proportional hazards model (Cox 1972). It relies on a semi-parametric modelling the conditional hazard functions $\lambda^{(a)}(t|X)$ as

$$\lambda^{(a)}(t|X) = \lambda_0^{(a)}(t) \exp(X^\top \beta^{(a)}),$$

where $\lambda_0^{(a)}$ is a baseline hazard function and $\beta^{(a)}$ has the same dimension as the vector of covariate $X$. The conditional survival function then take the simple form (in the continuous case)

$$S^{(a)}(t|X) = S_0^{(a)}(t)^{\exp(X^\top \beta^{(a)})},$$

where $S_0^{(a)}(t)$ is the survival function associated with $\lambda_0^{(a)}$. The vector $\beta^{(a)}$ is classically estimated by maximizing the so-called *partial likelihood* function as introduced in the original paper of Cox (1972):

$$\mathcal{L}(\beta) := \prod_{\Delta_i = 1} \frac{\exp(X_i^\top \beta)}{\sum\limits_{\widetilde{T}_j \geqslant \widetilde{T}_i} \exp(X_j^\top \beta)},$$

while the cumulative baseline hazard function can be estimated through the Breslow's estimator (Breslow 1974):

$$\widehat{\Lambda}_0^{(a)}(t) = \sum_{\Delta_i=1, \widetilde{T}_i \leqslant t} \frac{1}{\sum_{\widetilde{T}_j \geqslant \widetilde{T}_i} \exp(X_j^\top \widehat{\beta}^{(a)})}$$

where $\widehat{\beta}^{(a)}$ is a partial likelihood maximizer. One can show that $(\widehat{\beta}^{(a)}, \widehat{\Lambda}_0^{(a)})$ is the MLE of the true likelihood, when $\widehat{\Lambda}_0^{(a)}$ is optimized over all step fonctions of the form

$$\Lambda_0(t) := \sum_{\Delta_i=1} h_i, \quad h_i \in \mathbb{R}^+.$$

This fact was already hinted in the original paper by Cox and made rigorous in many subsequent papers, see for instance Fan, Feng, and Wu (2010). Furthermore, if the true distribution follows a Cox model, then both $\widehat{\beta}^{(a)}$ and $\widehat{\Lambda}_0^{(a)}$ are strongly consistent and asymptotically normal estimator of the true parameters $\beta^{(a)}$ and $\Lambda^{(a)}$, see Kalbfleisch and Prentice (2002), Sec 5.7. When using a $T$-learner approach, one simply finds $(\widehat{\beta}^{(a)}, \widehat{\Lambda}_0^{(a)})$ for $a \in \{0, 1\}$ by considering the control group and the treated group separately. When using a $S$-learner approach, the treatment status $A$ becomes a covariate a the model becomes

$$\lambda(t|X, A) = \lambda_0(t) \exp(X^\top \beta + \alpha A). \tag{19}$$

for some $\alpha \in \mathbb{R}$. One main advantage of Cox's model is that it makes it very easy to interpret the effect of a covariate on the survival time. If indeed $\alpha > 0$, then the treatment has a negative effect of the survival times. Likewise, if $\beta_i > 0$, then the $i$-th coordinate of $X$ has a negative effect as well. We finally mention that the hazard ratio takes a particularly simple form under the later model since

$$\theta_{\mathrm{HR}} = e^\alpha.$$

In particular, it does not depends on the time horizon $\tau$, and is thus sometimes refered to as *proportional hazard*. Figure 4 illustrates the estimation of the difference in Restricted Mean Survival Time using G-formula with Cox models.

**Weibull Model**. A popular parametric model for survival is the Weibull Model, which amounts to assume that

$$\lambda^{(a)}(t|X) = \lambda_0^{(a)}(t) \exp(X^\top \beta)$$

where $\lambda_0^{(a)}(t)$ is the instant hazard function of a Weibull distribution, that is to say a function proportional to $t^\gamma$ for some shape parameter $\gamma > 0$. We refer to Zhang (2016) for a study of this model.

**Survival Forests**. On the non-parametric front, we mention the existence of survival forests (Ishwaran et al. 2008).

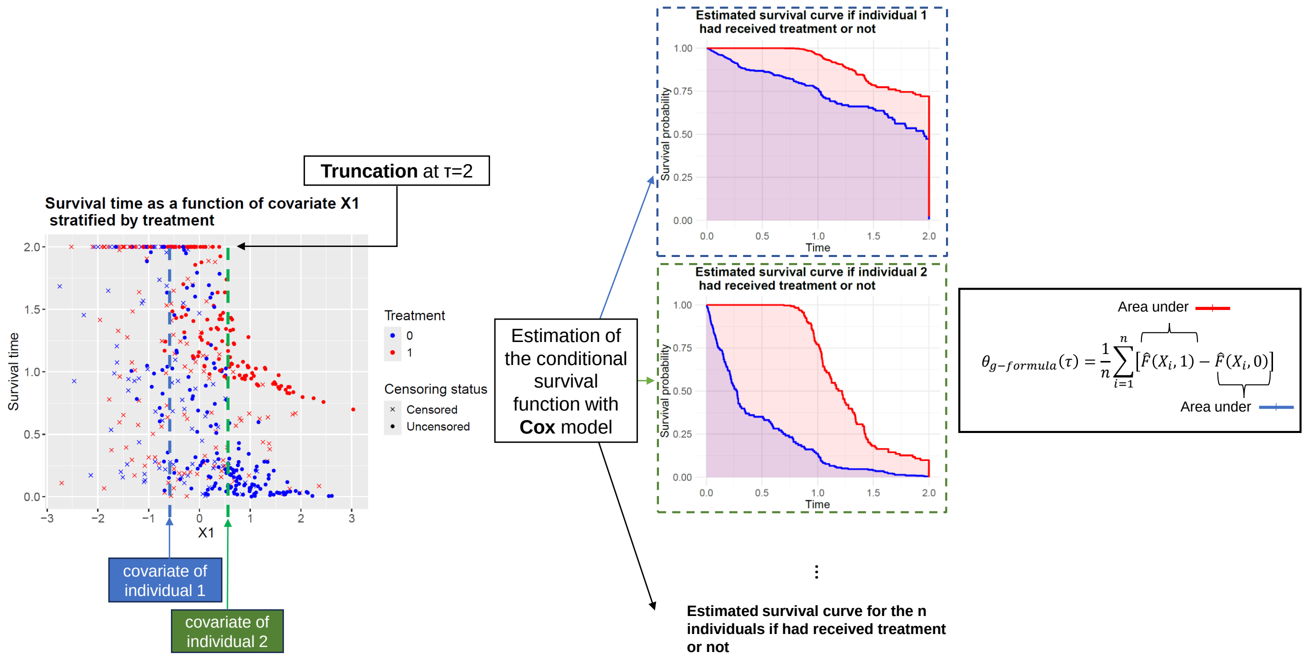



Figure 4: Illustration of the G-formula for estimating $\theta_{\mathrm{RMST}}$ in an RCT when only one covariate $X_1$ influences the outcome.

### Censoring unbiased transformations

Censoring unbiased transformations involve applying a transformation to $T$. Specifically, we compute a new time $T^*$ of the form

$$T^* := T^*(\widetilde{T}, X, A, \Delta) = \begin{cases} \phi_0(\widetilde{T} \wedge \tau, X, A) & \text{if} \quad \Delta^\tau = 0, \\ \phi_1(\widetilde{T} \wedge \tau, X, A) & \text{if} \quad \Delta^\tau = 1. \end{cases} \tag{20}$$

for two wisely chosen transformations $\phi_0$ and $\phi_1$, and where

$$\Delta^\tau := \mathbb{I}\{T \wedge \tau \leqslant C\} = \Delta + (1 - \Delta)\mathbb{I}(\widetilde{T} \geqslant \tau) \tag{21}$$

is the indicator of the event where the individual is either uncensored or censored after time $\tau$. The idea behind the introduction of $\Delta^\tau$ is that because we are only interested in computed the expectation of the survival time thresholded by $\tau$, any censored observation coming after time $\tau$ can in fact be considered as uncensored ($\Delta^\tau = 1$).

A censoring unbiased transformation $T^*$ shall satisfy: for $a \in \{0, 1\}$, it holds

$$\mathbb{E}[T^* | A = a, X] = \mathbb{E}[T(a) \wedge \tau | X] \quad \text{almost surely.} \tag{22}$$

A notable advantage of this approach is that it enables the use of the full transformed dataset $(X_i, A_i, T_i^*)$ as if no censoring occured. Because it holds

$$\mathbb{E}[\mathbb{E}[T^* | A = a, X]] = \mathbb{E}\left[\frac{\mathbb{I}\{A = a\}}{\mathbb{P}(A = a)} T^*\right], \tag{23}$$

there is a very natural way to derive an estimator of the difference in RMST from any censoring unbiased transformation $T^*$ as:

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{A_i}{\pi} - \frac{1 - A_i}{1 - \pi}\right) T_i^* \tag{24}$$

where $\pi = \mathbb{P}(A = 1) \in (0, 1)$ by Assumption 6 and where $T_i^* = T^*(\widetilde{T}_i, X_i, A_i, \Delta_i)$. We easily get the following result.

**Proposition 0.3.** *Under Assumptions 5 and 6, the estimator $\widehat{\theta}$ derived as in Equation 24 from a square integrable censoring unbiased transformations satisfying Equation 22 is an unbiased, strongly consistent, and asymptotically normal estimator of the difference in RMST.*

Square integrability will be ensured any time the transformnation is bounded, which will always be the case of the ones considered in this work. It is natural in a RCT setting to assume that probability of being treated $\pi$ is known. If not, it is usual to replace $\pi$ by its empirical counterpart $\widehat{\pi} = n_1/n$ where $n_a = \sum_i \nVdash\{A = a\}$. The resulting estimator takes the form

$$\widehat{\theta} = \frac{1}{n_1} \sum_{A_i = 1} T_i^* - \frac{1}{n_0} \sum_{A_i = 0} T_i^*. \tag{25}$$

Note however that this estimator is slighlty biased due to the estimation of $\pi$ (see for instance Colnet et al. (2022), Lemma 2), but it is still strongly consistent and asymptotically normal, and its biased is exponentially small in $n$.

**Proposition 0.4.** *Under Assumptions 5 and 6, the estimator $\widehat{\theta}$ derived as in Equation 25 from a square integrable censoring unbiased transformations satisfying Equation 22 is a strongly consistent, and asymptotically normal estimator of the difference in RMST.*

The two most popular transformations are Inverse-Probability-of-Censoring Weighting (Koul, Susarla, and Ryzin (1981)) and Buckley-James (Buckley and James (1979)), both illustrated in Figure 5 and detailed below. In the former, only non-censored observations are considered and they are weighted while in the latter, censored observations are imputed with an estimated survival time.

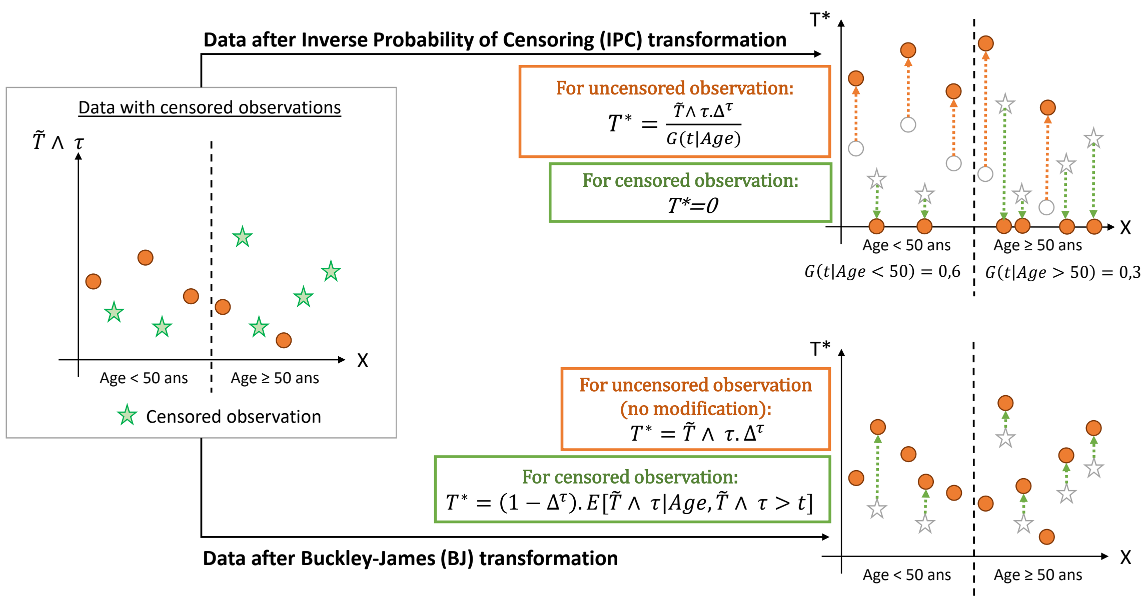



Figure 5: Illustration of Inverse-Probability-of-Censoring and Buckley-James transformations.

### The Inverse-Probability-of-Censoring Weighted transformation

The Inverse-Probability-of-Censoring Weighted (IPCW) transformation, introduced by (Koul, Susarla, and Ryzin (1981)) in the context of censored linear regression, involves discarding censored observations and applying weights to uncensored data. More precisely, we let

$$T^*_{\mathrm{IPCW}} = \frac{\Delta^\tau}{G(\widetilde{T} \wedge \tau | X, A)} \widetilde{T} \wedge \tau, \tag{26}$$

where we recall that $G(t|X,A) := \mathbb{P}(C \geqslant t|X,A)$ is the left limit of the conditional survival function of the censoring. This estimator assigns higher weights to uncensored subjects within a covariate group that is highly prone to censoring, thereby correcting for conditionally independent censoring and reducing selection bias (Howe et al. 2016).

**Proposition 0.5.** *Under Assumptions 1, 5, 6, 16 and 17, the IPCW transform 26 is a censoring unbiased transformation in the sense of Equation 22.*

The proof of Proposition 0.5 is in Section . The IPCW depends on the unknown conditional survival function of the censoring $G(\cdot|X,A)$, which thus needs to be estimated. Estimating conditional censoring or the conditional survival function can be approached similarly, as both involve estimating a time—whether for survival or censoring. Consequently, we can use semi-parametric methods, such as the Cox model, or non-parametric approaches like survival forests, and we refer to Section  for a development on these approaches. Once an estimator $\widehat{G}(\cdot|A,X)$ of the later is provided, one can construct an estimator of the difference in RMST based on Equation 24 or Equation 25

$$\widehat{\theta}_{\text{IPCW}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{\pi} - \frac{1-A_i}{1-\pi}\right)T^*_{\text{IPCW},i}, \tag{27}$$

or

$$\widehat{\theta}_{\text{IPCW}} = \frac{1}{n_1}\sum_{A_i=1} T^*_{\text{IPCW},i} - \frac{1}{n_0}\sum_{A_i=0} T^*_{\text{IPCW},i}. \tag{28}$$

where we recall that $n_a := \#\{i \in [n] \mid A_i = a\}$. By Proposition 0.3, Proposition 0.4 and Proposition 0.5, we easily deduce that $\widehat{\theta}_{\text{IPCW}}$ is asymptotically consistent as soon as $\widehat{G}$ is.

**Corollary 0.1.** *Under Assumptions1, 5, 6, 16 and 17, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\widehat{\theta}_{\text{IPCW}}$, either as in defined in Equation 27 or in Equation 28.*

This result simply comes from the fact that $\widehat{\theta}_{\text{IPCW}}$ depends continuously on $\widehat{G}$ and that $G$ is lower-bounded (Assumption 17). Surprisingly, we found limited use of this estimator in the literature, beside its first introduction in Koul, Susarla, and Ryzin (1981). This could potentially be explained by the fact that, empirically, we observed that this estimator is highly variable. Consequently, we do not explore its properties further and will not include it in the numerical experiments. A related and more popular estimator is the IPCW-Kaplan-Meier, defined as follows.

**Definition 0.2.** (IPCW-Kaplan-Meier) We let again $\widehat{G}(\cdot|X,A)$ be an estimator of the (left limit of) the conditional censoring survival function and we introduce

$$D_k^{\mathrm{IPCW}}(a) := \sum_{i=1}^n \frac{\Delta_i^\tau}{\widehat{G}(\widetilde{T}_i \wedge \tau | X_i, A=a)} \mathbb{I}(\widetilde{T}_i = t_k, A_i = a)$$

$$\text{and} \quad N_k^{\mathrm{IPCW}}(a) := \sum_{i=1}^n \frac{\Delta_i^\tau}{\widehat{G}(\widetilde{T}_i \wedge \tau | X_i, A=a)} \mathbb{I}(\widetilde{T}_i \geqslant t_k, A_i = a),$$

be the weight-corrected numbers of deaths and of individual at risk at time $t_k$. The IPCW version of the KM estimator is then defined as:

$$\widehat{S}_{\mathrm{IPCW}}(t|A=a) = \prod_{t_k \leqslant t} \left(1 - \frac{D_k^{\mathrm{IPCW}}(a)}{N_k^{\mathrm{IPCW}}(a)}\right).$$

Note that the quantity $\pi$ is not present in the definition of $D_k^{\mathrm{IPCW}}(a)$ and $N_k^{\mathrm{IPCW}}(a)$ because it would simply disappear in the ratio $D_k^{\mathrm{IPCW}}(a)/N_k^{\mathrm{IPCW}}(a)$. The subsequent RMST estimator then take the form

$$\widehat{\theta}_{\mathrm{IPCW-KM}} = \int_0^\tau \widehat{S}_{\mathrm{IPCW}}(t|A=1) - \widehat{S}_{\mathrm{IPCW}}(t|A=0) dt. \tag{29}$$

Like before for the classical KM estimator, this new reweighted KM estimator enjoys good statistical properties.

**Proposition 0.6.** *Under Assumptions 1, 5, 6, 16 and 17, and for all $t \in [0, \tau]$, the oracle estimator $S^*_{\mathrm{IPCW}}(t|A=a)$ defined as in Definition 0.2 with $\widehat{G} = G$ is a stronlgy consistent and asymptotically normal estimator of $S^{(a)}(t)$ .*

The proof of Proposition 0.6 can be found in Section . Because the evaluation of $N_k^{\mathrm{IPCW}}(a)$ now depends on information gathered after time $t_k$ (through the computation of the weights), the previous proofs on the absence of bias and on the derivation of the asymptotic variance unfortunately do not carry over in this case. Whether its bias is exponentially small and whether the asymptotic variance can be derived in a closed form are questions left open. We are also not aware of any estimation schemes for the asymptotic variance in this context. In the case where we do not have access to the oracle survival function $G$, we can again still achieve consistency if the estimator $\widehat{G}(\cdot|A, X)$ that we provide is consistent.

**Corollary 0.2.** *Under Assumptions 1, 5, 16 and 17, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\widehat{S}_{\mathrm{IPCW}}(t|A=a)$.*

This corollary ensures that the corresponding RMST estimator defined in Equation 29 will be consistent as well.

**The Buckley-James transformation**

One weakness of the IPCW transformation is that it discards all censored data. The Buckley-James (BJ) transformation, introduced by (Buckley and James (1979)), takes a different path by leaving all uncensored values untouched, while replacing the censored ones by an extrapolated value. Formally, it is defined as follows:

$$T^*_{\mathrm{BJ}} = \Delta^\tau(\widetilde{T} \wedge \tau) + (1 - \Delta^\tau) Q_S(\widetilde{T} \wedge \tau | X, A), \tag{30}$$

where, for $t \leqslant \tau$,

$$Q_S(t|X, A) := \mathbb{E}[T \wedge \tau | X, A, T \wedge \tau > t] = \frac{1}{S(t|X, A)} \int_t^\tau S(u|X, A) \, \mathrm{d}u$$

where $S(t|X, A = a) := \mathbb{P}(T(a) > t|X)$ is the conditional survival function. We refer again to Figure 5 for a diagram of this transformation.

**Proposition 0.7.** *Under Assumptions 1, 5, 16 and 17, the BJ transform 30 is a censoring unbiased transformation in the sense of Equation 22.*

The proof of Proposition 0.7 can be found in Section . Again, the BJ transformation depends on a nuisance parameter (here $Q_S(\cdot|X, A)$) that needs to be estimated. We again refer to Section  for a brief overview of possible estimation strategies for $Q_S$. Once provided with an estimator $\widehat{Q}_S(\cdot|A, X)$, a very natural estimator of the RMST based on the BJ transformation and on Equation 24 or Equation 25 would be

$$\widehat{\theta}_{\mathrm{BJ}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{A_i}{\pi} - \frac{1 - A_i}{1 - \pi} \right) T^*_{\mathrm{BJ},i}, \tag{31}$$

or

$$\widehat{\theta}_{\mathrm{BJ}} = \frac{1}{n_1} \sum_{A_i=1} T^*_{\mathrm{BJ},i} - \frac{1}{n_0} \sum_{A_0=1} T^*_{\mathrm{BJ},i}. \tag{32}$$

Like for the IPCW transformation, the BJ transformation yields a consistent estimate of the RMST as soon as the model is well-specified.

**Corollary 0.3.** *Under Assumptions 1, 5, 16 and 17, if $\widehat{Q}_S$ is uniformly weakly (resp. strongly) consistent then so is $\widehat{\theta}_{\mathrm{BJ}}$ defined as in Equation 31 or Equation 32.*

The proof is again a mere application of Propositions 0.3, 0.4 and 0.7, and relies on the continuity of $S \mapsto Q_S$. The BJ transformation is considered as the best censoring transformation of the original response in the following sense.

**Theorem 0.1.** *For any transformation $T^*$ of the form 20, it holds*

$$\mathbb{E}[(T^*_{\mathrm{BJ}} - T \wedge \tau)^2] \leqslant \mathbb{E}[(T^* - T \wedge \tau)^2].$$

This result is stated in Fan and Gijbels (1994) but without a proof. We detail it in Section  for completeness.

### Augmented corrections

The main disadvantage of the two previous transformations is that they heavily rely on the specification of good estimator $\widehat{G}$ (for IPCW) or $\widehat{S}$ (for BJ). In order to circumvent this limitation, D. Rubin and Laan (2007) proposed the following transformations, whose expression is based on theory of semi-parametric estimation developed in Laan and Robins (2003),

$$T^*_{\mathrm{DR}} = \frac{\Delta^\tau \widetilde{T} \wedge \tau}{G(\widetilde{T} \wedge \tau | X, A)} + \frac{(1-\Delta^\tau) Q_S(\widetilde{T} \wedge \tau | X, A)}{G(\widetilde{T} \wedge \tau | X, A)} - \int_0^{\widetilde{T} \wedge \tau} \frac{Q_S(t|X,A)}{G(t|X,A)^2} \, \mathrm{d}\mathbb{P}_C(t|X,A), \tag{33}$$

where $\mathrm{d}\mathbb{P}_C(t|X,A)$ is the distribution of $C$ conditional on the covariates $X$ and $A$. We stress that this distribution is entirely determined by the $G(\cdot|X,A)$, so that this transformation only depends on the knowledge of both conditional survival functions $G$ and $S$, will be thus sometimes denoted $T^*_{\mathrm{DR}}(G,S)$ to underline this dependency. This transformation is not only a censoring unbiased transform in the sense of Equation 22, but is also doubly robust in the following sense.

**Proposition 0.8.** *We let $F, R$ be two conditional survival functions. Under Assumptions 1, 5, 6, 16 and 17, if $F$ also satisfies Assumption 17, and if $F(\cdot|X,A)$ is absolutely continuous wrt $G(\cdot|X,A)$, then the transformation $T^*_{\mathrm{DR}} = T^*_{\mathrm{DR}}(F,R)$ is a censoring unbiased transformation in the sense of Equation 22 whenever $F = G$ or $R = S$.*

The statement and proof of this results is found in D. Rubin and Laan (2007) in the case where $C$ and $T$ are continuous. A careful examination of the proofs show that the proof translates straight away to our discrete setting.

## Causal survival analysis in observational studies

Unlike RCT, observational data — such as from registries, electronic health records, or national healthcare systems — are collected without controlled randomized treatment allocation. In such cases, treated and control groups are likely unbalanced due to the non-randomized design, which results in the treatment effect being confounded by variables influencing both the survival outcome $T$ and the treatment allocation $A$. To enable identifiability of the causal effect, additional standard assumptions are needed.

**Assumption.** (Conditional exchangeability / Unconfoundedness) It holds

$$A \perp\!\!\!\perp T(0), T(1) | X \tag{34}$$

Under Equation 34, the treatment assignment is randomly assigned conditionally on the covariates $X$. This assumption ensures that there are no unmeasured confounder as the latter would make it impossible to distinguish correlation from causality.

**Assumption.** (Positivity / Overlap for treatment) Letting $e(X) := \mathbb{P}(A = 1|X)$ be the *propensity score*, there holds

$$0 < e(X) < 1 \quad \text{almost surely.} \tag{35}$$

The Equation 35 assumption requires adequate overlap in covariate distributions between treatment groups, meaning every observation must have a non-zero probability of being treated. Because Assumption 5 does not hold anymore, neither does the previous idenfiability Equation 7. In this new context, we can write

$$\begin{aligned}
\theta_{\mathrm{RMST}} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\
&= \mathbb{E}\left[\mathbb{E}[T(1) \wedge \tau|X] - \mathbb{E}[T(0) \wedge \tau|X]\right] \\
&= \mathbb{E}\left[\mathbb{E}[T(1) \wedge \tau|X, A = 1] - \mathbb{E}[T(0) \wedge \tau|X, A = 0]\right] \quad \text{(unconfoundness)} \\
&= \mathbb{E}\left[\mathbb{E}[T \wedge \tau|X, A = 1] - \mathbb{E}[T \wedge \tau|X, A = 0]\right]. \quad \text{(SUTVA)}
\end{aligned} \tag{36}$$

In another direction, one could wish to identify the treatment effect through the survival curve as in Equation 2:

$$S^{(a)}(t) = \mathbb{P}(T(a) > t) = \mathbb{E}\left[\mathbb{P}(T > t|X, A = a)\right]. \tag{37}$$

Again, both identities still depend on the unknown quantity $T$ and suggest two different estimation strategies. These strategies differ according to the censoring assumptions and are detailed below.

## Independent censoring

Figure 6 illustrates a causal graph in observational survival data with independent censoring (Assumption 8).

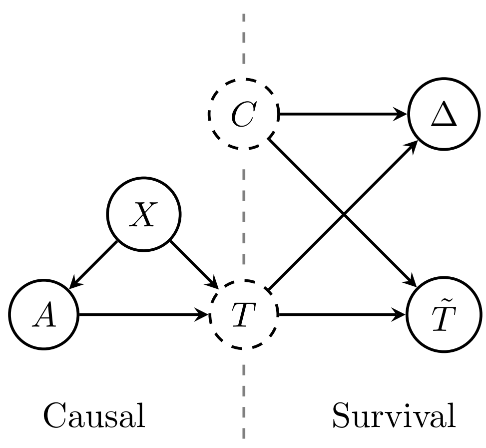



Figure 6: Causal graph in observational survival data with independent censoring.

Under Assumption 8, we saw in Section  that the Kaplan-Meier estimator could conveniently handle censoring. Building on Equation 37, we can write

$$S^{(1)}(t) = \mathbb{E}\left[\frac{\mathbb{E}[\mathbb{I}\{A=1, T>t\}|X]}{\mathbb{E}[\mathbb{I}\{A=1\}|X]}\right] = \mathbb{E}\left[\frac{A\mathbb{I}\{T>t\}}{e(X)}\right],$$

which suggests to adapt the classical KM estimator by reweighting it by the propensity score. The use of propensity score in causal inference has been initially introduced by Rosenbaum and Rubin (1983) and further developed in Hirano, Imbens, and Ridder (2003). It was extended to survival analysis by Xie and Liu (2005) through the adjusted Kaplan-Meier estimator as defined below.

**Definition 0.1.** (IPTW Kaplan-Meier estimator) We let $\widehat{e}(\cdot)$ be an estimator of the propensity score $e(\cdot)$. We introduce

$$\begin{aligned} D_k^{\mathrm{IPTW}}(a) &:= \sum_{i=1}^{n}\left(\frac{a}{\widehat{e}(X_i)} + \frac{1-a}{1-\widehat{e}(X_i)}\right)\mathbb{I}(\widetilde{T}_i = t_k, \Delta_i = 1, A_i = a) \\ \text{and} \quad N_k^{\mathrm{IPTW}}(a) &:= \sum_{i=1}^{n}\left(\frac{a}{\widehat{e}(X_i)} + \frac{1-a}{1-\widehat{e}(X_i)}\right)\mathbb{I}(\widetilde{T}_i \geqslant t_k, A_i = a), \end{aligned}$$

be the numbers of deaths and of individual at risk at time $t_k$, reweighted by the propensity score. The Inverse-Probability-of-Treatment Weighting (IPTW) version of the KM estimator is then defined as:

$$\widehat{S}_{\mathrm{IPTW}}(t|A=a) = \prod_{t_k \leqslant t}\left(1 - \frac{D_k^{\mathrm{IPTW}}(a)}{N_k^{\mathrm{IPTW}}(a)}\right). \tag{38}$$

We let $S^*_{\mathrm{IPTW}}(t|A=a)$ be the oracle KM-estimator defined as above with $\widehat{e}(\cdot) = e(\cdot)$. Although the reweighting slightly changes the analysis, the good properties of the classical KM carry on to this setting.

**Proposition 0.1.** *Under Assumptions 1, 34, 35, 8 and 9 The oracle IPTW Kaplan-Meier estimator $S^*_{\mathrm{IPTW}}(t|A=a)$ is a strongly consistent and asymptotically normal estimator of $S^{(a)}(t)$.*

The proof of this result simply relies again on the law of large number and the $\delta$-method and can be found in Section . Because now $S^*_{\mathrm{IPTW}}$ is a continuous function of $e(\cdot)$, and because $e$ and $1-e$ are lower-bounded as per Assumptions 35, we easily derive the following corollary.

**Corollary 0.1.** *Under the same assumptions as Proposition 0.1, if $\widehat{e}(\cdot)$ satisfies $\|\widehat{e} - e\|_\infty \to 0$ almost surely (resp. in probability), then the IPTW Kaplan-Meier estimator $\hat{S}_{\mathrm{IPTW}}(t|A=a)$ is a strongly (resp. weakly) consistent estimator of $S^{(a)}(t)$.*

The resulting RMST estimator simply takes the form:

$$\widehat{\theta}_{\mathrm{IPTW-KM}} = \int_0^\tau \widehat{S}_{\mathrm{IPTW}}(t|A=1) - \widehat{S}_{\mathrm{IPTW}}(t|A=0)dt. \tag{39}$$

which will be consistent under the same Assumptions as the previous Corollary. Note that, we are not aware of any formal results concerning the bias and the asymptotic variance of the oracle estimator $S^*_{\mathrm{IPTW}}(t|A=a)$, and we refer to Xie and Liu (2005) for heuristics concerning these questions.

## Conditional independent censoring

Under Assumptions 34 (uncounfoundedness) and 16 (conditional independent censoring), the causal effect is affected both by confounding variables and by conditional censoring. The associated causal graph is depicted in Figure 7. In this setting, one can still use the $G$-formula exactly as in Section .

A natural alternative approach is to weight the IPCW and BJ transformations from Section  by the propensity score to disentangle both confounding effects and censoring at the same time.

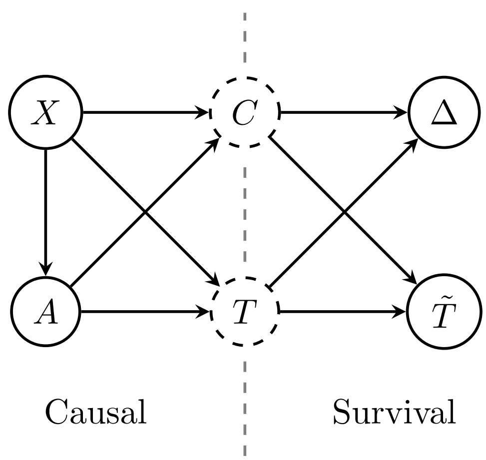



Figure 7: Causal graph in observational survival data with dependent censoring.

We mention that the G-formula approach developed in Section  does work in that context. In particular, Chen and Tsiatis (2001) prove consistency and asymptotic normality results for Cox estimators in a observational study, and they give an explicit formulation of the asymptotic variance as a function of the parameters of the Cox model. In the non-parametric world, Foster, Taylor, and Ruberg (2011) and Künzel et al. (2019) empirically study this estimator using survival forests, with the former employing it as a T-learner (referred to as *Virtual Twins*) and the latter as an S-learner.

### IPTW-IPCW transformations

One can check that the IPCW transformation as introduced in Equation 26 is also a censoring unbiased transformation in that context.

**Proposition 0.2.** *Under Assumptions 1, 34, 35, 16 and 17, the IPTW-IPCW transform 26 is a censoring unbiased transformation in the sense of Equation 22.*

The proof of Proposition 0.2 can be found in Section . Deriving an estimator of the difference in RMST is however slightly different in that context. In particular, Equation 23 rewrites

$$\mathbb{E}[\mathbb{E}[T^*|X, A=1]] = \mathbb{E}\left[\frac{A}{e(X)}T^*\right],$$

Which in turn suggests to define

$$\widehat{\theta}_{\mathrm{IPTW-IPCW}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A}{e(X)} - \frac{1-A}{1-e(X)}\right)T^*_{\mathrm{IPCW},i}. \tag{40}$$

This transformation now depends on two nuisance parameters, namely the conditional survival function of the censoring (through $T^*_{\mathrm{IPCW}}$) and the propensity score. Once estimators of these quantities are provided, one could look at the corresponding quantity computed with these estimators.

**Proposition 0.3.** *Under Assumptions 1, 34, 35, 16 and 17, and if $\widehat{G}(\cdot|X, A)$ and $\widehat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then estimator 40 defined with $\widehat{e}$ and $\widehat{G}$ is a weakly (resp. strongly) consistent estimator of the difference in RMST.*

The proof of Proposition 0.3 can be found in Section . We can also use the same strategy as for the IPCW transformation and incorporate the new weights into a Kaplan-Meier estimator.

**Definition 0.2.** (IPTW-IPCW-Kaplan-Meier) We let again $\widehat{G}(\cdot|X, A)$ and $\widehat{e}(\cdot)$ be estimators of the conditional censoring survival function and of the propensity score. We introduce

$$\begin{aligned} D_k^{\mathrm{IPTW-IPCW}}(a) &:= \sum_{i=1}^{n}\left(\frac{A_i}{\widehat{e}(X_i)} + \frac{1-A_i}{1-\widehat{e}(X_i)}\right)\frac{\Delta_i^\tau}{\widehat{G}(\widetilde{T}_i \wedge \tau|X_i, A=a)}\mathbb{I}(\widetilde{T}_i = t_k, A_i = a) \\ \text{and} \quad N_k^{\mathrm{IPTW-IPCW}}(a) &:= \sum_{i=1}^{n}\left(\frac{A_i}{\widehat{e}(X_i)} + \frac{1-A_i}{1-\widehat{e}(X_i)}\right)\frac{\Delta_i^\tau}{\widehat{G}(\widetilde{T}_i \wedge \tau|X_i, A=a)}\mathbb{I}(\widetilde{T}_i \geqslant t_k, A_i = a), \end{aligned}$$

be the weight-corrected numbers of deaths and of individual at risk at time $t_k$. The IPTW-IPCW version of the KM estimator is then defined as:

$$\widehat{S}_{\mathrm{IPTW-IPCW}}(t|A=a) = \prod_{t_k \leqslant t}\left(1 - \frac{D_k^{\mathrm{IPTW-IPCW}}(a)}{N_k^{\mathrm{IPTW-IPCW}}(a)}\right).$$

The difference in RMST estimated with IPTW-IPCW-Kaplan-Meier survival curves is then simply as

$$\widehat{\theta}_{\mathrm{IPTW-IPCW-KM}} = \int_0^\tau \widehat{S}_{\mathrm{IPTW-IPCW}}(t|A=1) - \widehat{S}_{\mathrm{IPTW-IPCW}}(t|A=0)dt. \tag{41}$$

**Proposition 0.4.** *Under Assumptions 1, 34, 35, 16 and 17, and for all $t \in [0,\tau]$, if the oracle estimator $S^*_{\mathrm{IPTW-IPCW}}(t|A=a)$ defined as in Definition 0.2 with $\widehat{G}(\cdot|A,X) = G(\cdot|A,X)$ and $\widehat{e} = e$ is a strongly consistent and asymptotically normal estimator of $S^{(a)}(t)$ .*

The proof of Proposition 0.4 can be found in Section . Under consistency of the estimators of the nuisance parameters, the previous proposition implies that this reweighted Kaplan-Meier is a consistent estimator of the survival curve, which in turn implies consistency of $\widehat{\theta}_{\mathrm{IPTW-IPCW-KM}}$.

**Corollary 0.2.** *Under Assumptions 1, 34, 35, 16 and 17, and for all $t \in [0,\tau]$, if the nuisance estimators $\widehat{G}(\cdot|A,X)$ and $\widehat{e}$ are weakly (resp. strongly) uniformly consistent, then $\widehat{S}_{\mathrm{IPTW-IPCW}}(t|A=a)$ is a weakly (resp. strongly) consistent estimator of $S^{(a)}(t)$.*

We are not aware of general formula for the asymptotic variances in this context. We mention nonetheless that Schaubel and Wei (2011) have been able to derive asymptotic laws in this framework in the particular case of Cox-models.

### IPTW-BJ transformations

Like IPCW tranformation, BJ transformation is again a censoring unbiased transformation in an observational context.

**Proposition 0.5.** *Under Assumptions 1, 34, 35, 16 and 17, the IPTW-BJ transform 30 is a censoring unbiased transformation in the sense of Equation 22.*

The proof of Proposition 0.5 can be found in Section . The corresponding estimator of the difference in RMST is

$$\hat{\theta}_{\mathrm{IPTW-BJ}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A}{e(X)} - \frac{1-A}{1-e(X)}\right)T^*_{\mathrm{BJ},i}. \tag{42}$$

This transformation depends on the conditional survival function $S$ (through $T^*_{\mathrm{BJ}}$) and the propensity score. Consistency of the nuisance parameter estimators implies again consistency of the RMST estimator.

**Proposition 0.6.** *Under Assumptions 1, 34, 35, 16 and 17, and if $\widehat{S}(\cdot|X,A)$ and $\widehat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then $\widehat{\theta}_{\mathrm{IPTW-BJ}}$ defined with $\widehat{S}$ and $\widehat{e}$ is a weakly (resp. strongly) consistent estimator of the RMST.*

The proof of Proposition 0.6 can be found in Section .

### Double augmented corrections

Building on the classical doubly-robust AIPTW estimator from causal inference (Robins, Rotnitzky, and Zhao 1994), we could incorporate the doubly-robust transformations of Section  to obtain a *quadruply robust* tranformation

$$\Delta^*_{\mathrm{QR}} = \Delta^*_{\mathrm{QR}}(G,S,\mu,e) := \left(\frac{A}{e(X)} - \frac{1-A}{1-e(X)}\right)(T^*_{\mathrm{DR}}(G,S) - \mu(X,A)) + \mu(X,1) - \mu(X,0),$$

where we recall that $T^*_{\mathrm{DR}}$ is defined in Section . This transformations depends on four nuisance parameters: $G$ and $S$ through $T^*_{\mathrm{DR}}$, and now the propensity score $e$ and the conditional response $\mu$. This transformation doesn't really fall into the scope of censoring unbiased transform, but it is easy to show that $\Delta^*_{\mathrm{QR}}$ is quadruply robust in the following sense.

**Proposition 0.7.** *Let $F, R$ be two conditional survival function, $p$ be a propensity score, and $\nu$ be a conditional response. Then, under the same assumption on $F, R$ as in Proposition 0.8, and under Assumptions 1, 34, 35, 16 and 17, the transformations $\Delta^*_{\mathrm{QR}} = \Delta^*_{\mathrm{QR}}(F,R,p,\nu)$ satisfies, fo $a \in 0,1$,*

$$\mathbb{E}[\Delta^*_{\mathrm{QR}}|X] = \mathbb{E}[T(1)\wedge\tau - T(0)\wedge\tau|X] \quad \textit{if} \quad \begin{cases} F = G & \textit{or} \quad R = S \quad \textit{and} \\ p = e & \textit{or} \quad \nu = \mu. \end{cases}$$

This result is similar to Ozenne et al. (2020), Thm 1, and its proof can be found in Section . Based on estimators $(\widehat{G}, \widehat{S}, \widehat{\mu}, \widehat{e})$ of $(G,S,\mu,e)$, one can then propose the following estimator of the RMST, coined the AIPTW-AIPCW estimator in Ozenne et al. (2020):

$$\begin{aligned}\widehat{\theta}_{\mathrm{AIPTW-AIPCW}} &:= \frac{1}{n}\sum_{i=1}^n \Delta^*_{\mathrm{QR},i}(\widehat{G},\widehat{S},\widehat{\mu},\widehat{e}) \\ &= \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{\widehat{e}(X_i)} - \frac{1-A_i}{1-\widehat{e}(X_i)}\right)(T^*_{\mathrm{DR}}(\widehat{G},\widehat{S})_i - \widehat{\mu}(X_i,A_i)) + \widehat{\mu}(X_i,1) - \widehat{\mu}(X_i,0).\end{aligned} \tag{43}$$

This estimator enjoys good asymptotic properties under parametric models, as detailed in Ozenne et al. (2020).

## Implementation

In this section, we first summarize the various estimators and their properties. We then provide custom implementations for all estimators, even those already available in existing packages. These manual implementations serve two purposes: first, to make the methods accessible to the community when no existing implementation is available; and second, to facilitate a deeper understanding of the methods by detailing each step, even when a package solution exists. Finally, we present the packages available for directly computing $\theta_{\mathrm{RMST}}$.

### Summary of the estimators

Table 3 provides an overview of the estimators introduced in this paper, along with the corresponding nuisance parameters needed for their estimation and an overview of their statistical properties in particular regarding their sensitivity to misspecification of the nuisance parameters.

Table 3: Estimators of the difference in RMST and nuisance parameters needed to compute each estimator. Empty boxes indicate that the nuisance parameter is not needed in the estimator thus misspecification has no impact. Estimators in italic are those that are already implemented in available packages.

| Estimator | RCT | Obs | Ind Cens | Dep Cens | Outcome model | Censoring model | Treatment model | Robustness |
|---|---|---|---|---|---|---|---|---|
| *Unadjusted KM* | $X$ | | $X$ | | | | | |
| IPCW-KM | $X$ | | $X$ | $X$ | | $G$ | | |
| BJ | $X$ | | $X$ | $X$ | $S$ | | | |
| *IPTW-KM* | $X$ | $X$ | $X$ | | | | $e$ | |
| IPCW-IPTW-KM | $X$ | $X$ | $X$ | $X$ | | $G$ | $e$ | |
| IPTW-BJ | $X$ | $X$ | $X$ | $X$ | $S$ | | $e$ | |
| *G-formula* | $X$ | $X$ | $X$ | $X$ | $\mu$ | | | |
| AIPTW-AIPCW | $X$ | $X$ | $X$ | $X$ | $S, \mu$ | $G$ | $e$ | $X$ (Prp 0.7) |

### Implementation of the estimators

Across different implementations, we use the following shared functions provided in the `utilitary.R` file.

- `estimate_propensity_score`: function to estimate propensity scores $e(X)$ using either parametric (i.e. logistic regression with the argument `type_of_model = "glm"`) or non-parametric methods (i.e. probability forest with the argument `type_of_model ="probability forest"` based on the `probability_forest` function from the grf (Tibshirani et al. 2017) package). This latter can include cross-fitting (`n.folds 1`).

- `estimate_survival_function`: function to estimate the conditional survival model, which supports either Cox models (argument `type_of_model = "cox"`) or survival forests (argument `type_of_model = "survival forest"`) which uses the function `survival_forest` from the grf (Tibshirani et al. 2017) package. This latter can include cross-fitting (`n.folds 1`). The estimation can be done with a single learner (argument `learner = "S-learner"`) or two learners (argument `learner = "T-learner"`).
- `estimate_hazard_function`: function to estimate the instantaneous hazard function by deriving the cumulative hazard function at each time point. This cumulative hazard function is estimated from the negative logarithm of the survival function.
- `Q_t_hat`: function to estimate the remaining survival function at all time points and for all individuals which uses the former `estimate_survival_function`.
- `Q_Y`: function to select the value of the remaining survival function from `Q_t_hat` at the specific time-to-event.
- `integral_rectangles`: function to estimate the integral of a decreasing step function using the rectangle method.
- `expected_survival`: function to estimate the integral with $x, y$ coordinate (estimated survival function) using the trapezoidal method.
- `integrate`: function to estimate the integral at specific time points `Y.grid` of a given `integrand` function which takes initially its values on `times` using the trapezoidal method.

### Unadjusted Kaplan-Meier

Although Kaplan-Meier is implemented in the `survival` package (Therneau 2001), we provide a custom implementation, `Kaplan_meier_handmade`, for completeness. The difference in Restricted Mean Survival Time, estimated using Kaplan-Meier as in Equation 14 can then be calculated with the `RMST_1` function.

As an alternative, one can also use the `survfit` function in the survival package (Therneau 2001) for Kaplan-Meier and specify the `rmean` argument equal to $\tau$ in the corresponding summary function:

### IPCW Kaplan-Meier

We first provide an `adjusted.KM` function which is then used in the `IPCW_Kaplan_meier` function to estimate the difference in RMST $\hat{\theta}_{\text{IPCW}}$ as in Equation 29. The survival censoring function $G(t|X)$ is computed with the `estimate_survival_function` utility function from the `utilitary.R` file.

One could also use the `survfit` function in the survival package (Therneau 2001) in adding IPCW weights for treated and control group and specify the `rmean` argument equal to $\tau$ in the corresponding summary function:

This alternative approach for IPCW Kaplan-Meier would also be valid for IPTW and IPTW-IPCW Kaplan-Meier.

#### Buckley-James based estimator

The function `BJ` estimates $\theta_{\mathrm{RMST}}$ by implementing the Buckley-James estimator as in Equation 31. It uses two functions available in the `utilitary.R` file, namely `Q_t_hat` and `Q_Y`.

#### IPTW Kaplan-Meier

The function `IPTW_Kaplan_meier` implements the IPTW-KM estimator in Equation 39. It uses the `estimate_propensity_score` function from the `utilitary.R`.

#### G-formula

We implement two versions of the G-formula: `g_formula_T_learner` and `g_formula_S_learner`. In `g_formula_T_learner`, separate models estimate survival curves for treated and control groups, whereas `g_formula_S_learner` uses a single model incorporating both covariates and treatment status to estimate survival time. The latter approach is also available in the RISCA package but is limited to Cox models.

#### IPTW-IPCW Kaplan-Meier

The `IPTW_IPCW_Kaplan_meier` function implements the IPTW-IPCW Kaplan Meier estimator from Equation 41. It uses the utilitary functions from the `utilitary.R` file `estimate_propensity_score` and `estimate_survival_function` to estimate the nuisance parameters, and the function `adjusted.KM` which computes an adjusted Kaplan Meier estimator using the appropriate weight.

#### IPTW-BJ estimator

The `IPTW_BJ` implements the IPTW-BJ estimator in Equation 42. It uses the utilitary functions, from the `utilitary.R` file, `estimate_propensity_score`, `Q_t_hat` and `Q_Y` to estimate the nuisance parameters.

#### AIPTW-AIPCW

The `AIPTW_AIPCW` function implements the AIPTW_AIPCW estimator in Equation 43 using the utilitary function from the `utilitary.R` file `estimate_propensity_score`, `Q_t_hat`, `Q_Y`, and `estimate_survival_function` to estimate the nuisance parameters.

## Available packages

Currently, there are few sustained implementations available for estimating RMST in the presence of right censoring. Notable exceptions include the packages survRM2 (Hajime et al. 2015), grf (Tibshirani et al. 2017) and RISCA (Foucher, Le Borgne, and Chatton 2019). Those packages are implemented in the `utilitary.R` files.

### SurvRM2

The difference in RMST with Unadjusted Kaplan-Meier $\hat{\theta}_{KM}$ (Equation 14) can be obtained using the function `rmst2` which takes as arguments the observed time-to-event, the status, the arm which corresponds to the treatment and $\tau$.

### RISCA

The RISCA package provides several methods for estimating $\theta_{\mathrm{RMST}}$. The difference in RMST with Unadjusted Kaplan-Meier $\hat{\theta}_{KM}$ (Equation 14) can be derived using the `survfit` function from the the survival package (Therneau 2001) which estimates Kaplan-Meier survival curves for treated and control groups, and then the `rmst` function calculates the RMST by integrating these curves, applying the rectangle method (type="s"), which is well-suited for step functions.

The IPTW Kaplan-Meier (Equation 38) can be applied using the `ipw.survival` and `rmst` functions. The ipw.survival function requires user-specified weights (i.e. propensity scores). To streamline this process, we define the `RISCA_iptw` function, which combines these steps and utilizes the `estimate_propensity_score` from the `utilitary.R` file.

A single-learner version of the G-formula, as introduced in Section , can be implemented using the `gc.survival` function. This function requires as input the conditional survival function which should be estimated beforehand with a Cox model via the `coxph` function from the `survival` package (Therneau 2001). Specifically, the single-learner approach applies a single Cox model incorporating both covariates and treatment, rather than separate models for each treatment arm. We provide a function `RISCA_gf` that consolidates these steps.

### grf

The `grf` package (Tibshirani et al. 2017) enables estimation of the difference between RMST using the Causal Survival Forest approach (Cui et al. 2023), which extends the non-parametric causal forest framework to survival data. The RMST can be estimated with the `causal_survival_forest` function, requiring covariates $X$, observed event times, event status, treatment assignment, and the time horizon $\tau$ as inputs. The `average_treatment_effect` function then evaluates the treatment effect based on predictions from the fitted forest.

# Simulations

We compare the behaviors and performances of the estimators through simulations conducted across various experiemental contexts. These contexts include scenarios based on RCTs and observational data, with both independent and dependent censoring. We also use semi-parametric and non-parametric models to generate censoring and survival times, as well as logistic and nonlinear models to simulate treatment assignment.

## RCT

### Data Generating Process

We generate RCTs with independent censoring (Scenario 1) and conditionally independent censoring (Scenario 2). We sample $n$ iid datapoints $(X_i, A_i, C, T_i(0), T_i(1))_{i\in[n]}$ where $T_i(0), T_i(1)$ and $C$ follows Cox's models. More specifically, we set

- $X \sim \mathcal{N}(\mu, \Sigma)$ where $\mu = (1, 1, -1, 1)$ and $\Sigma = \mathrm{Id}_4$.
- The hazard function of $T(0)$ is

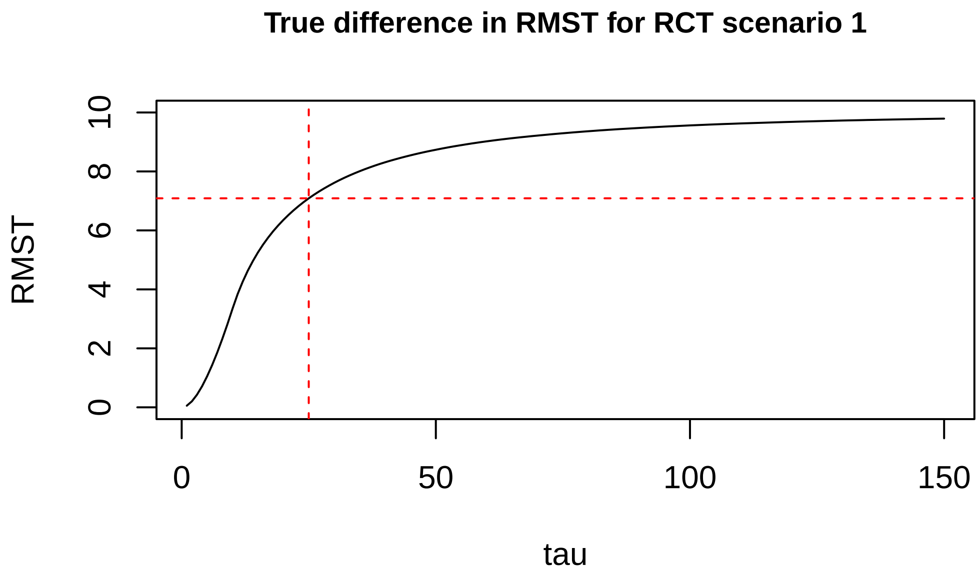
$$\lambda^{(0)}(t|X) = 0.01 \exp\left\{\beta_0^\top X\right\} \quad \text{where} \quad \beta_0 = (0.5, 0.5, -0.5, 0.5).$$

- The survival times in the treatment group are given by $T(1) = T(0) + 10$.
- The hazard function of the censoring time $C$ is simply taken as $\lambda_C(t|X) = 0.03$ in Scenario 1, and in Scenario 2 as

$$\lambda_C(t|X) = 0.03 \cdot \exp\left\{\beta_C^\top X\right\} \quad \text{where} \quad \beta_C = (0.7, 0.7, -0.25, -0.1).$$

- The treatment allocation is independent of $X$: $e(X) = 0.5$.
- The threshold time $\tau$ is set to 25.

The descriptive statistics of the two datasets are displayed in Annex (Section ). The graph of the difference in RMST as a function of $\tau$ for the two scenarii are displayed below; $\theta_{\mathrm{RMST}}$ is the same in both setting.




```
[1] "The ground truth for RCT scenario 1 and 2 at time 25 is 7.1"
```

**Estimation of the RMST**

For each Scenario, we estimate the difference in RMST using the methods summarized in Section . The methods used to estimate the nuisance components are indicated in brackets: either logistic regression or random forests for propensity scores and either cox models or survival random forests for survival and censoring models. A naive estimator where censored observations are simply removed and the survival time is averaged for treated and controls is also provided for a naive baseline.

Figure 8 shows the distribution of the difference in RMST for 100 simulations in Scenario 1 and different sample sizes: 500, 1000, 2000, 4000. The true value of $\theta_{\mathrm{RMST}}$ is indicated by red dotted line.

In this setting, and in accordance with the theory, the simplest estimator (unadjusted KM) performs just as well as the others, and presents an extremely small bias (as derived in Section ).

The naive estimator is biased, as expected, and the bias in both the G-formula (RISCA) and the manual G-formula S-learner arises because the treatment effect is additive $T(1) = T(0) + 10$ and violates the assumption that $T$ would follow a Cox model in the variables $(X, A)$. However, $T|A = a$ is a Cox-model for $a \in \{0, 1\}$, which explain the remarkable performance of G-formula (Cox/T-learners) and some of the other models based on a Cox estimation of $S$.

Other estimators (IPTW KM (Reg.Log), IPCW KM (Cox), IPTW-IPCW KM (Cox & Log.Reg), IPTW-BJ (Cox & Log.Reg), AIPTW-AIPCW (Cox & Cox & Log.Reg)) involve unnecessary nuisance parameter estimates, such as propensity scores or censoring models. Despite this,

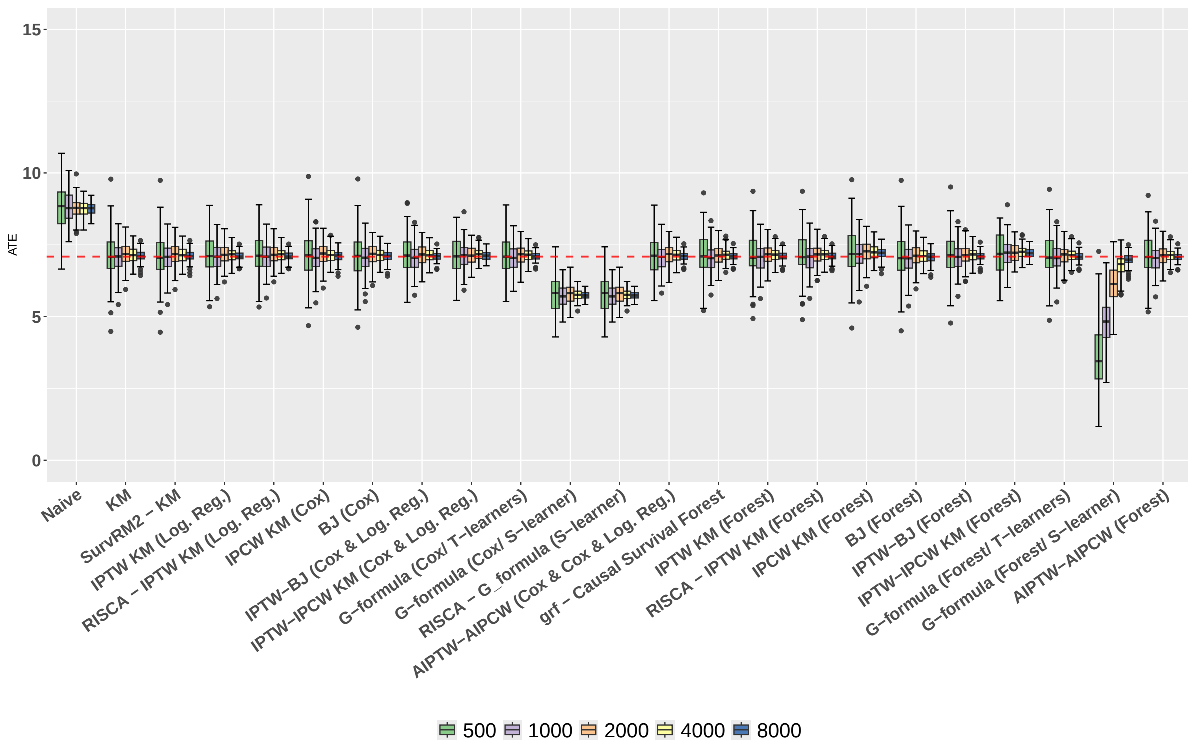



Figure 8: Results of the ATE for the simulation of a RCT with independent censoring.

their performance remains relatively stable in terms of variability, and there are roughly no differences between using (semi-)parametric or non-parametric estimation methods for nuisance parameters except for IPCW KM and IPTW-IPCW KM where there is a slight bias when using forest-based methods.

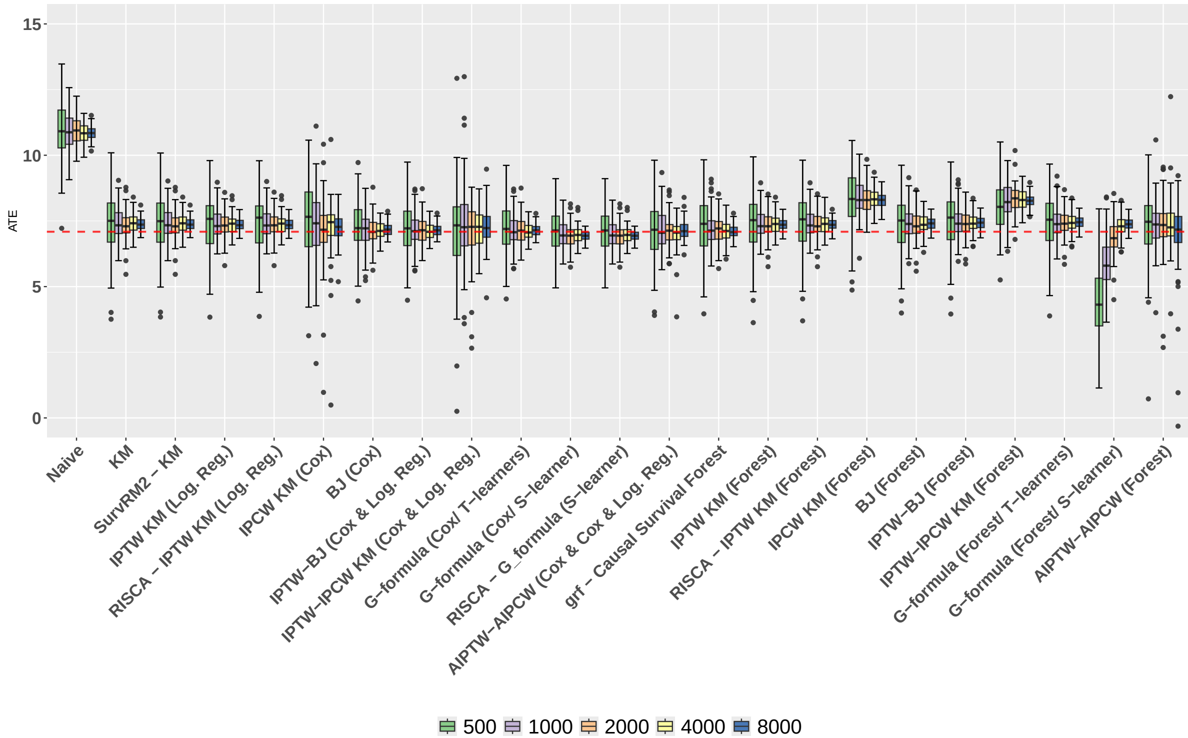



Figure 9: Estimation results of the ATE for the simulation of a RCT with dependent censoring.

Figure 9 shows the results for the RCT simulation with conditionally independent censoring (Scenario 2). In this setting, the Naive estimator remains biased. Similarly, both the unadjusted Kaplan-Meier (KM) and its SurvRM2 equivalent, as well as the treatment-adjusted IPTW KM and its RISCA equivalent, are biased due to their failure to account for dependent censoring. As in Scenario 1, G-formula (Cox/ S-learner) and its RISCA equivalent also remain biased. The IPCW KM (Cox) is slightly biased up to 4,000 observations and quite variable due to extreme censoring probabilities. IPTW-IPCW KM (Cox & Log.Reg.) is not biased but shows high variance. In contrast, the Buckley-James estimator BJ (Cox) is unbiased even with as few as 500 observations. The BJ estimator also demonstrates smaller variance than IPCW methods. G-formula (Cox/ T-learners) and AIPCW-AIPTW (Cox & Cox & Log.Reg.) estimators seem to perform well, even in small samples. The forest versions of these estimators seem more biased, except Causal Survival Forest and the AIPTW-AIPCW (Forest). Notably, all estimators exhibit higher variability compared to Scenario 1.

## Observational data

### Data Generating Process

As for Scenarii 1 and 2, we carry out two simulations of an observational study with both independent and conditional independent censoring. The only difference lies in the simulation of the propensity score, which is no longer constant. For the simulation, an iid $n$-sample $(X_i, A_i, C, T_i(0), T_i(1))_{i\in[n]}$ is generated as in Section , except for the treatment allocation process that is given by:

$$\text{logit}(e(X)) = \beta_A^\top X \quad \text{where} \quad \beta_A = (-1, -1, -2.5, -1),$$

where we recall that $\text{logit}(p) = \log(p/(1-p))$. The descriptive statistics for the two observational data with independent (`Obs1`) and conditionally independent censoring (`Obs2`) are displayed in Appendix (Section ). Note that we did not modify the survival distribution, the target difference in RMST is thus the same.

```
[1] "The ground truth for Obs scenario 1 at time 25 is 7.1"
```

```
[1] "The ground truth for Obs scenario 2 at #time 25 is 7.1"
```

### Estimation of the RMST

Figure 10 below shows the distribution of the estimators of $\theta_{\text{RMST}}$ for the observational study with independent censoring.

In the simulation of an observational study with independent censoring, confounding bias is introduced, setting it apart from RCT simulations. As expected, estimators that fail to adjust for this bias, such as unadjusted Kaplan-Meier (KM), IPCW KM (Cox), and their equivalents, are biased. However, estimators like IPTW KM (Log.Reg.), IPTW-IPCW KM (Cox & Log. Reg.) are unbiased, even if the latter estimate unnecessary nuisance components. Results with IPTW BJ (Cox & Log.Reg) are extremely variable.

The top-performing estimators in this scenario are G-formula (Cox/ T-learners) and AIPCW-AIPTW (Cox & Cox & Log.Reg.), which are unbiased even with 500 observations. The former has the lowest variance. All estimators that use forests to estimate nuisance parameters are biased across sample sizes from 500 to 8000. Although Causal Survival Forest and AIPW-AIPCW (Forest) are expected to eventually converge, they remain extremely demanding in terms of sample size. This setting thus highlights that one should either have an a priori knowledge on the specification of the models or large sample size.

Figure 11 below shows the distribution of the $\theta_{\text{RMST}}$ estimates for the observational study with conditionally independent censoring. The red dashed line represents the true $\theta_{\text{RMST}}$ for $\tau = 25$.

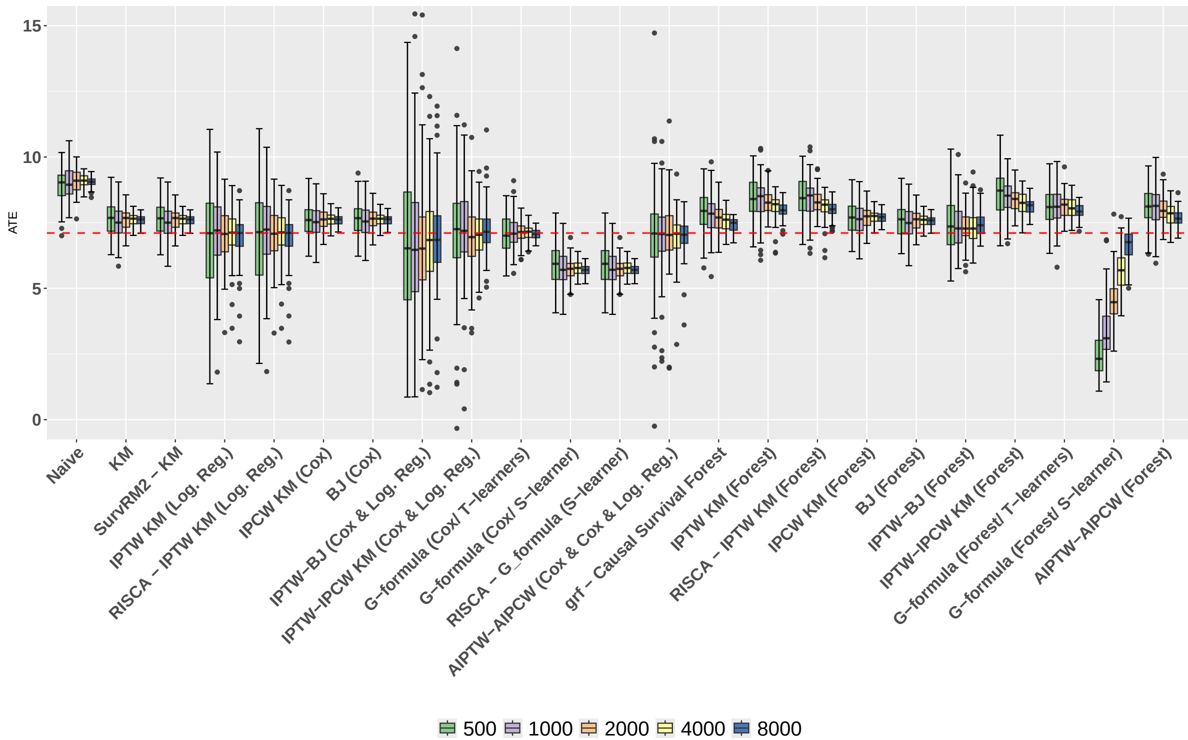



Figure 10: Estimation results of the ATE for the simulation of an observational study with independent censoring.

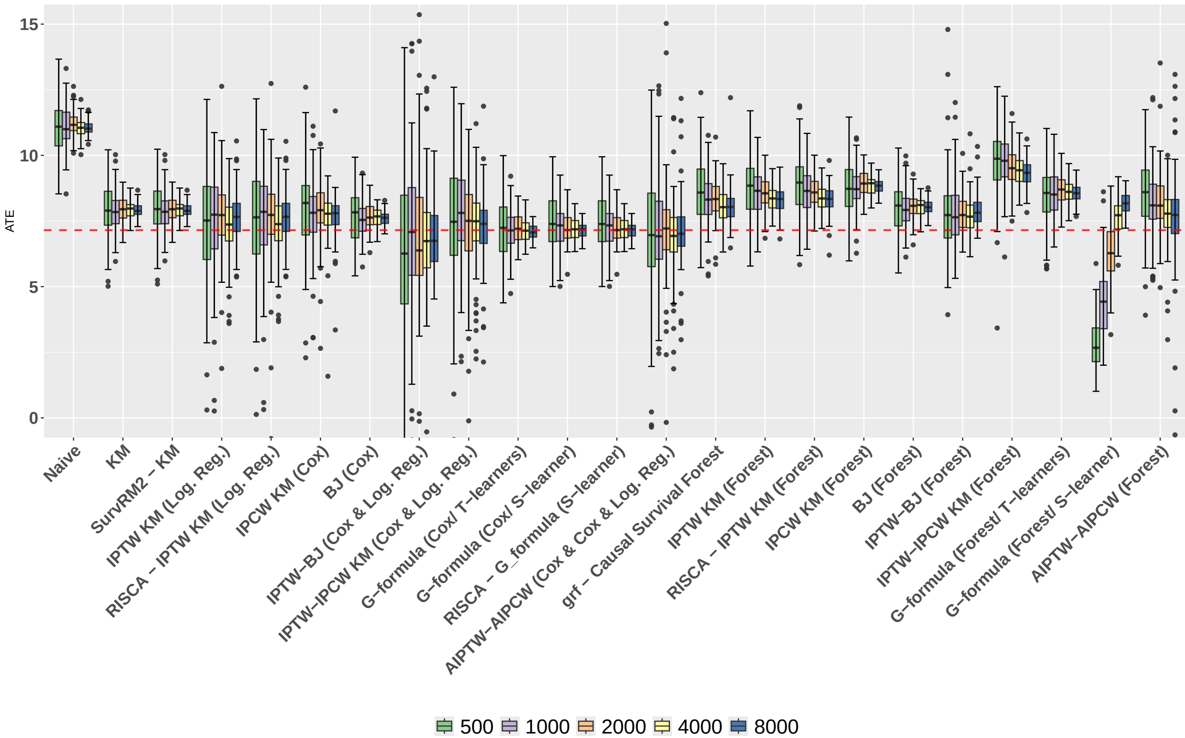



Figure 11: Estimation results of the ATE for the simulation of an observational study with dependent censoring.

In the simulation of an observational study with conditionally independent censoring, estimators that do not account for both censoring and confounding bias, such as KM, IPCW KM, IPTW KM, and their package equivalents, are biased. The top-performing estimators in this scenario are G-formula (Cox/ T-learners) and AIPCW-AIPTW (Cox & Cox & Log.Reg.), which are unbiased even with 500 observations. The former has the lowest variance as expected, see Section . Surprisingly, the G-formula (Cox/S-learner) and its equivalent from the RISCA package perform quite competitively, showing only a slight bias despite the violation of the proportional hazards assumption. All estimators that use forests to estimate nuisance parameters are biased across sample sizes from 500 to 8000. Although Causal Survival Forest and AIPTW-AIPCW (Forest) are expected to eventually converge, they remain extremely demanding in terms of sample size.

## Mispecification of nuisance components

### Data Generating Process

We generate an observational study with covariate interactions and conditionally independent censoring. The objective is to assess the impact of misspecifying nuisance components; specifically, we will use models that omit interactions to estimate these components. This approach enables us to evaluate the robustness properties of various estimators. In addition, in this setting forest based methods are expected to behave better.

We generate $n$ samples $(X_i, A_i, C, T_i(0), T_i(1))$ as follows:

- $X \sim \mathcal{N}(\mu, \Sigma)$ and $\mu = (0.5, 0.5, 0.7, 0.5)$, $\Sigma = \mathrm{Id}_4$.
- The hazard function of $T(0)$ is given by

$$\begin{aligned} \lambda^{(0)}(t|X) = \exp\{\beta_0^\top Y\} \quad \text{where} \quad & \beta_0 = (0.2, 0.3, 0.1, 0.1, 1, 0, 0, 0, 0, 1), \\ \text{and} \quad & Y = (X_1^2, X_2^2, X_3^2, X_4^2, X_1X_2, X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4). \end{aligned}$$

- The distribution of $T(1)$ is the one of $T(0)$ but shifted: $T(1) = T(0) + 1$.
- The hazard function of $C$ is given by

$$\lambda_C(t|X) = \exp\{\beta_C^\top Y\} \quad \text{where} \quad \beta_C = (0.05, 0.05, -0.1, 0.1, 0, 1, 0, -1, 0, 0).$$

- The propensity score is

$$\mathrm{logit}(e(x)) = \beta_A^\top Y \quad \text{where} \quad \beta_A = (0.05, -0.1, 0.5, -0.1, 1, 0, 1, 0, 0, 0).$$

  When the model is well-specified, the full vector $(X, Y)$ is given as an input of the nuisance parameter models. When it is not, only $X$ and the first half of $Y$ corresponding to $(X_1^2, X_2^2, X_3^2, X_4^2)$ is given as an input.

The descriptive statistics are given in Appendix (Section ).

```
[1] "The ground truth for mis scenario at time 0.45 is 0.26"
```

**Estimation of the RMST**

First, we estimate $\theta_{\text{RMST}}$ without any model misspecification to confirm the consistency of the estimators under correctly specified nuisance models. More specifically, it means that for parametric propensity score models, semi-parametric censoring and survival models, we use models including interactions and squared assuming knowledge on the data generating process.

Next, we introduce misspecification individually for the treatment model, censoring model, and outcome model (Figure 13), i.e., we use models without interaction to estimate parametric and semi-parametric nuisance components while the data are generated with interactions.
We further examine combined misspecifications for pairs of models: treatment and censoring, treatment and outcome, and outcome and censoring. Finally, we assess the impact of misspecifying all nuisance models simultaneously (Figure 14).

When there is no misspecification in Figure 12, as expected, IPTW-BJ (Cox & Log.Reg), G-formula (Cox/ T-learners) and AIPTW-AIPCW (Cox & Cox & Reg.Log) are unbiased. IPTW-IPCW KM (Cox & Log.Reg) exhibits a bias but seems to converge at larger sample size. Regarding forest-based methods, IPTW-BJ (Forest), AIPTW-AIPCW (Forest) and Causal Survival Forest estimate accurately the difference in RMST. Surprisingly, G-formula (Forest/ T-learners), G-formula (Forest/ S-learner) and IPTW-IPCW KM (Forest) exhibit small bias but are expected to eventually converge at large sample size.

Figure 13 shows that AIPTW-AIPCW (Cox & Cox & Reg.Log) is convergent when there is one nuisance misspecification. In contrary, the other estimators are biased when one of its nuisance parameter is misspecified.

Figure 14 shows that, as expected, when all nuisance models are misspecified, all estimators exhibit bias. AIPTW-AIPCW (Cox & Cox & Reg.Log) seems to converge in case where either the outcome and censoring models, or the treatment and censoring models are misspecificed which deviates from initial expectations. It was anticipated that AIPTW-AIPCW would converge solely when both the censoring and treatment models were misspecified.

## Conclusion

Based on the simulations and theoretical results, it might be advisable to stay away from the IPCW and IPTW-IPCW estimators, as they often exhibit excessive variability. Instead, we recommend implementing BJ which seems like a more stable transformation as IPCW, as well as Causal Survival Forest, G-formula (T-learners), IPTW-BJ, and AIPTW-AIPCW in both

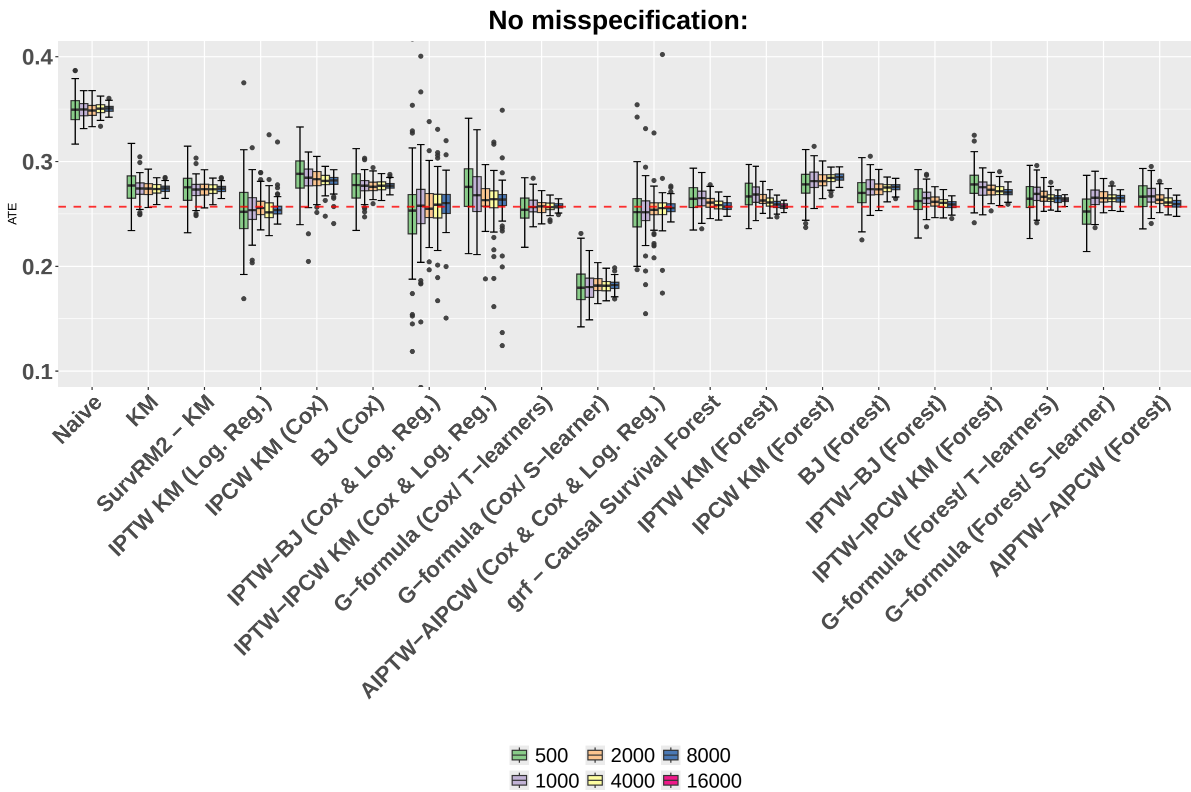



Figure 12: Estimation results of the ATE for the simulation of an observational study with dependent censoring and non linear relationships.

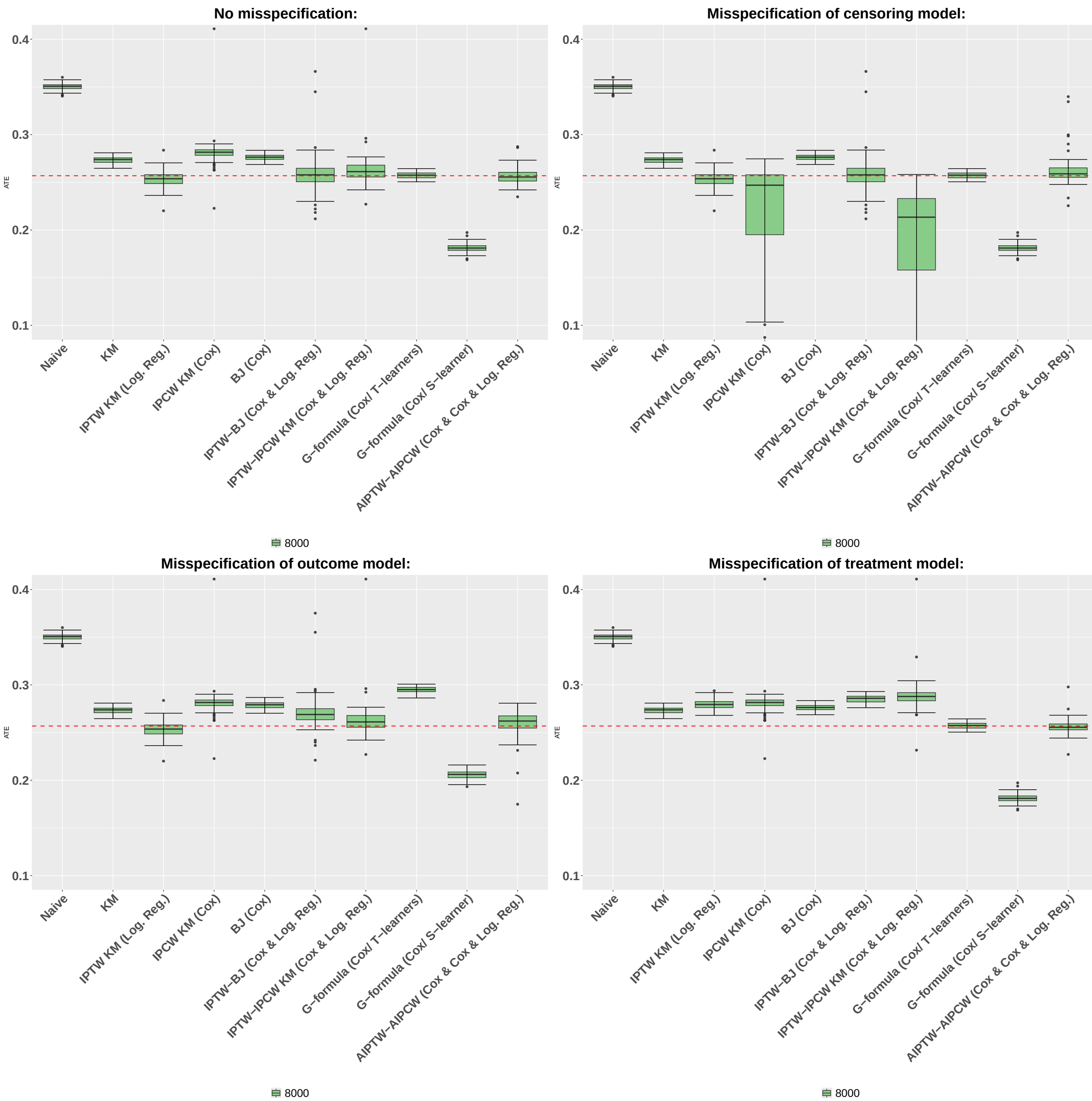



Figure 13: Estimation results of the ATE for an observational study with dependent censoring in case of a single misspecification.

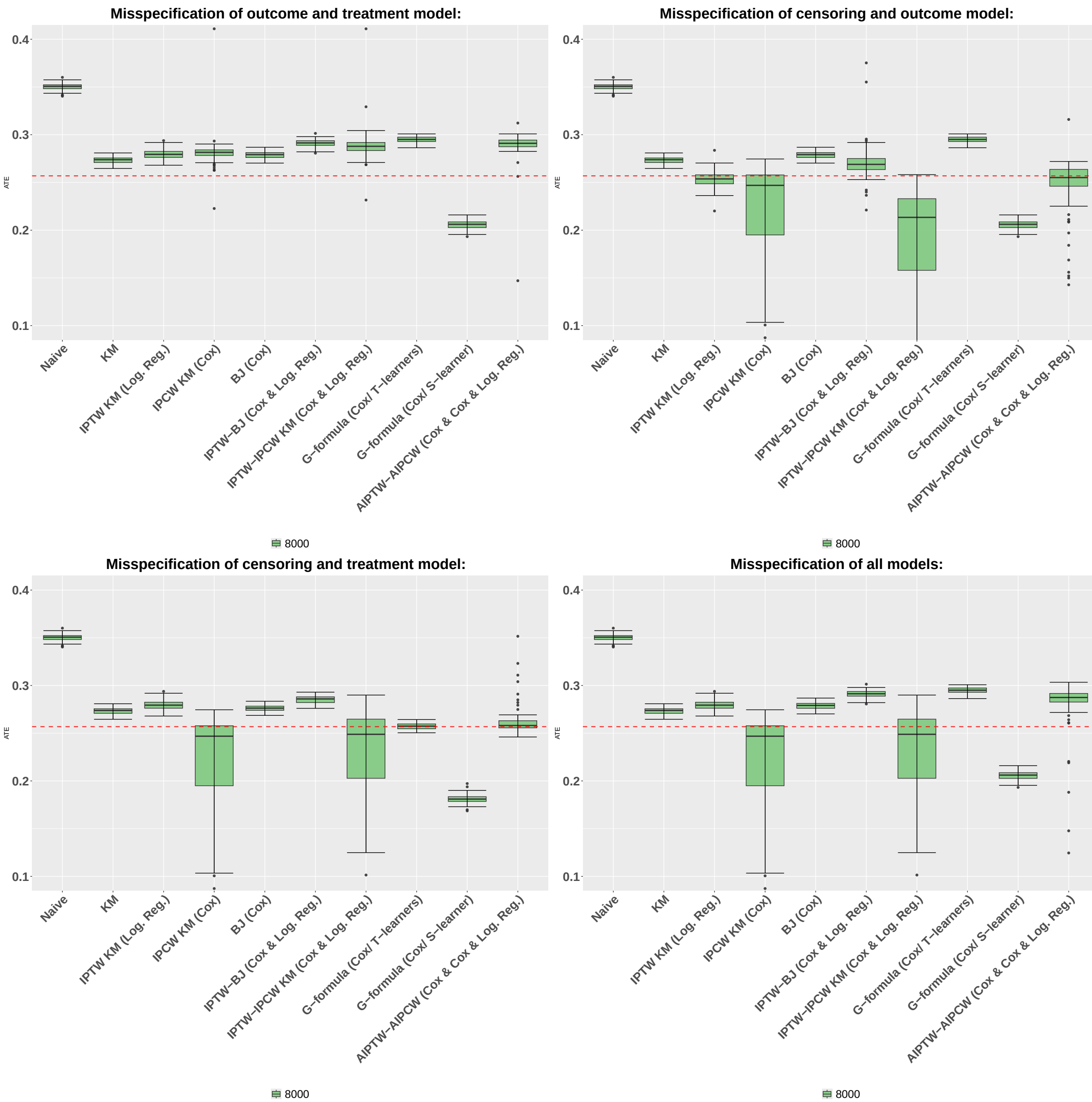



Figure 14: Estimation results of the ATE for an observational study with dependent censoring in case of a two or more misspecifications.

their Cox, Logistic Regression and forest versions. By qualitatively combining the results from these more robust estimators, we can expect to gain a fairly accurate understanding of the treatment effect.

It is important to note that our simulations utilize large sample sizes with relatively simple relationships, which may not fully capture the complexity of real-world scenarios. In practice, most survival analysis datasets tend to be smaller and more intricate, meaning the stability of certain estimators observed in our simulations may not generalize to real-world applications. Testing these methods on real-world datasets would provide a more comprehensive evaluation of their performance in practical settings.

An interesting direction for future work would be to focus on variable selection. Indeed, there is no reason to assume that the variables related to censoring should be the same as those linked to survival or treatment allocation. We could explore differentiating these sets of variables and study the impact on the estimators' variance. Similarly to causal inference settings without survival data, we might expect, for instance, that adding precision variables—-those solely related to the outcome—-could reduce the variance of the estimators.

Additionally, the estimators of the Restricted Mean Survival Time (RMST) provide a valuable alternative to the Hazard Ratio. The analysis and code provided with this article enables further exploration of the advantages of the estimators of RMST for causal analysis in survival studies. This could lead to a deeper understanding of how these estimators can offer more stable and interpretable estimates of treatment effects, particularly in complex real-world datasets.

## Disclosure

This study was funded by Sanofi. Charlotte Voinot and Bernard Sebastien are Sanofi employees and may hold shares and/or stock options in the company. Clément Berenfeld, Imke Mayer and Julie Josse have nothing to disclose.

## References

Aalen, O., O. Borgan, and H. Gjessing. 2008. *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health. Springer New York. https://books.google.de/books?id=wEi26X-VuCIC.

Breslow, N. 1974. "Covariance Analysis of Censored Survival Data." *Biometrics* 30 (1): 89–99. http://www.jstor.org/stable/2529620.

Breslow, N., and J. Crowley. 1974. "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship." *The Annals of Statistics* 2 (3): 437–53. https://doi.org/10.1214/aos/1176342705.

Buckley, and James. 1979. "Linear regression with censored data." *Biometrika* 66 (3): 429–36. https://doi.org/10.1093/biomet/66.3.429.

Chen, Pei-Yun, and Anastasios A. Tsiatis. 2001. "Causal Inference on the Difference of the Restricted Mean Lifetime Between Two Groups." *Biometrics* 57 (4): 1030–38. https://doi.org/10.1111/j.0006-341X.2001.01030.x.

Colnet, Bénédicte, Julie Josse, Gaël Varoquaux, and Erwan Scornet. 2022. "Reweighting the RCT for Generalization: Finite Sample Error and Variable Selection." *arXiv Preprint arXiv:2208.07614*.

Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society: Series B (Methodological)* 34 (2): 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

Cui, Yifan, Michael R Kosorok, Erik Sverdrup, Stefan Wager, and Ruoqing Zhu. 2023. "Estimating heterogeneous treatment effects with right-censored data via causal survival forests." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85 (2): 179–211. https://doi.org/10.1093/jrsssb/qkac001.

Ebrahimi, Nader, Daniel Molefe, and Zhiliang Ying. 2003. "Identifiability and Censored Data." *Biometrika* 90 (3): 724–27.

European Medecines Agency, EMA. 2024. "Reflection Paper on Use of Real-World Data in Non-Interventional Studies to Generate Real-World Evidence." https://www.ema.europa.eu/en/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence-scientific-guideline.

Fan, Jianqing, Yang Feng, and Yichao Wu. 2010. "High-Dimensional Variable Selection for Cox's Proportional Hazards Model." In *Borrowing Strength: Theory Powering Applications–a Festschrift for Lawrence d. Brown*, 6:70–87. Institute of Mathematical Statistics.

Fan, Jianqing, and Irène Gijbels. 1994. "Censored Regression: Local Linear Approximations and Their Applications." *Journal of the American Statistical Association* 89 (426): 560–70. https://doi.org/10.1080/01621459.1994.10476781.

Foster, Jared C., Jeremy M. G. Taylor, and Stephen J. Ruberg. 2011. "Subgroup Identification from Randomized Clinical Trial Data." *Statistics in Medicine* 30 (24): 2867–80. https://doi.org/10.1002/sim.4322.

Foucher, Yohann, Florent Le Borgne, and Arthur Chatton. 2019. "RISCA: Causal Inference and Prediction in Cohort-Based Analyses." *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.risca.

Gill, Richard. 1983. "Large Sample Behaviour of the Product-Limit Estimator on the Whole Line." *The Annals of Statistics*, 49–58.

Hajime, Uno, Tian Lu, Horiguchi Miki, Cronin Angel, Battioui Chakib, and Bell James. 2015. "survRM2: Comparing Restricted Mean Survival Time." *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.survrm2.

Hernán, Miguel A, and James M Robins. 2010. "Causal Inference." CRC Boca Raton, FL.

Hirano, Keisuke, Guido W Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Howe, Chanelle J, Stephen R. Cole, Bryan Lau, Sonia Napravnik, and Joseph J. Eron. 2016. "Selection Bias Due to Loss to Follow up in Cohort Studies." *Epidemiology* 27 1: 91–97.

https://api.semanticscholar.org/CorpusID:21993915.

Huitfeldt, A., M. Stensrud, and E. Suzuki. 2019. "On the Collapsibility of Measures of Effect in the Counterfactual Causal Framework." *Emerging Themes in Epidemiology* 16: 1–12.

Ishwaran, Hemant, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. 2008. "Random survival forests." *The Annals of Applied Statistics* 2 (3): 841–60. https://doi.org/10.1214/08-AOAS169.

Kalbfleisch, John D, and Ross L Prentice. 2002. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Kaplan, E. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53 (282): 457–81. https://doi.org/10.1080/01621459.1958.10501452.

Koul, H., V. Susarla, and J. Van Ryzin. 1981. "Regression Analysis with Randomly Right-Censored Data." *The Annals of Statistics* 9 (6): 1276–88. https://doi.org/10.1214/aos/1176345644.

Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. 2019. "Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning." *Proceedings of the National Academy of Sciences* 116 (10): 4156–65. https://doi.org/10.1073/pnas.1804597116.

Laan, Mark J. van der, and James M. Robins. 2003. "Introduction." In *Unified Methods for Censored Longitudinal Data and Causality*, 8–101. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-21700-0_1.

Martinussen, Torben, Stijn Vansteelandt, and Per Andersen. 2020. "Subtleties in the Interpretation of Hazard Contrasts." *Lifetime Data Analysis* 26 (October). https://doi.org/10.1007/s10985-020-09501-5.

Ozenne, Brice Maxime Hugues, Thomas Harder Scheike, Laila Stærk, and Thomas Alexander Gerds. 2020. "On the Estimation of Average Treatment Effects with Right-Censored Time to Event Outcome and Competing Risks." *Biometrical Journal* 62 (3): 751–63. https://doi.org/10.1002/bimj.201800298.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89 (427): 846–66.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1): 41–55. https://doi.org/10.1093/biomet/70.1.41.

Royston, Patrick, and Mahesh KB Parmar. 2013. "Restricted Mean Survival Time: An Alternative to the Hazard Ratio for the Design and Analysis of Randomized Trials with a Time-to-Event Outcome." *BMC Medical Research Methodology* 13: 1–15.

Rubin, Daniel, and Mark J. van der Laan. 2007. *The International Journal of Biostatistics* 3 (1). https://doi.org/10.2202/1557-4679.1052.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. https://doi.org/10.1037/h0037350.

Schaubel, Douglas E., and Guanghui Wei. 2011. "Double Inverse-Weighted Estimation of Cumulative Treatment Effects under Nonproportional Hazards and Dependent Censoring."

*Biometrics* 67 (1): 29–38. https://doi.org/10.1111/j.1541-0420.2010.01449.x.

Therneau, Terry M. 2001. "Survival: Survival Analysis." *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.survival.

Tibshirani, Julie, Susan Athey, Erik Sverdrup, and Stefan Wager. 2017. "Grf: Generalized Random Forests." *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.grf.

Turkson, Anthony, Francis Ayiah-Mensah, and Vivian Nimoh. 2021. "Handling Censoring and Censored Data in Survival Analysis: A Standalone Systematic Literature Review." *International Journal of Mathematics and Mathematical Sciences* 2021 (September): 1–16. https://doi.org/10.1155/2021/9307475.

Xie, Jun, and Chaofeng Liu. 2005. "Adjusted Kaplan–Meier Estimator and Log-Rank Test with Inverse Probability of Treatment Weighting for Survival Data." *Statistics in Medicine* 24 (20): 3089–3110. https://doi.org/10.1002/sim.2174.

Zhang, Zhongheng. 2016. "Parametric Regression Model for Survival Data: Weibull Regression Model as an Example." *Annals of Translational Medicine* 4 (24).

Zhou, M. 1988. "Two-Sided Bias Bound of the Kaplan-Meier Estimator." *Probability Theory and Related Fields* 79 (2): 165–73.

# Appendix A: Proofs

## Proofs of Section

*Proof. (Proposition 0.1).* Consistency is a trivial consequence of the law of large number and the identity 11. To show that $\widehat{S}_{\mathrm{KM}}$ is unbiased, let us introduce $\mathcal{F}_k$ be the filtration generated by the set of variables

$$\{A_i, \mathbb{I}\{\widetilde{T}_i = t_j\}, \mathbb{I}\{\widetilde{T}_i = t_j, \Delta_i = 1\} \mid j \in [k], i \in [n]\}.$$

which corresponds to the known information up to time $t_k$, so that $D_k(a)$ is $\mathcal{F}_k$-measurable but $N_k(a)$ is $\mathcal{F}_{k-1}$-measurable. One can write that, for $k \geqslant 2$

$$\begin{aligned}
\mathbb{E}[\mathbb{I}\{\widetilde{T}_i = t_k, \Delta_i = 1, A_i = a\} \mid \mathcal{F}_{k-1}] &= \mathbb{E}[\mathbb{I}\{\widetilde{T}_i = t_k, \Delta_i = 1, A_i = a\} \mid \mathbb{I}\{\widetilde{T}_i \geqslant t_k\}, A_i] \\
&= \mathbb{I}\{A_i = a\}\mathbb{E}[\mathbb{I}\{T_i = t_k, C_i \geqslant t_k\} \mid \mathbb{I}\{T_i \geqslant t_k, C_i \geqslant t_k\}, A_i] \\
&= \mathbb{I}\{A_i = a\}\mathbb{I}\{C_i \geqslant t_k\}\mathbb{E}[\mathbb{I}\{T_i = t_k\} \mid \mathbb{I}\{T_i \geqslant t_k\}, A_i] \\
&= \mathbb{I}\{\widetilde{T}_i \geqslant t_k, A_i = a\}\left(1 - \frac{S^{(a)}(t_k)}{S^{(a)}(t_{k-1})}\right),
\end{aligned}$$

where we used that $T_i(a)$ is idependant from $A_i$ by Assumption 5. We then easily derive from this that

$$\mathbb{E}\left[\left(1 - \frac{D_k(a)}{N_k(a)}\right)\mathbb{I}\{N_k(a) > 0\}\middle|\mathcal{F}_{k-1}\right] = \frac{S^{(a)}(t_k)}{S^{(a)}(t_{k-1})}\mathbb{I}\{N_k(a) > 0\},$$

and then that

$$\mathbb{E}\left[\widehat{S}_{\mathrm{KM}}(t_k|A=a)\middle|\mathcal{F}_{k-1}\right] = \frac{S^{(a)}(t_k)}{S^{(a)}(t_{k-1})}\widehat{S}_{\mathrm{KM}}(t_{k-1}|A=a) + O(\mathbb{I}\{N_k(a)=0\}),$$

By induction, we easily find that

$$\mathbb{E}[\widehat{S}_{\mathrm{KM}}(t|A=a)] = \prod_{t_j \leqslant t} \frac{S^{(a)}(t_j)}{S^{(a)}(t_{j-1})} + O\left(\sum_{t_j \leqslant t} \mathbb{P}(N_j(a)=0)\right) = S^{(a)}(t) + O(\mathbb{P}(N_k(a)=0))$$

where $t_k$ is the greatest time such that $t_k \leqslant t$. □

*Proof. (Proposition 0.2).* The asymptotic normality is a mere consequence of the joint asymptotic normality of $(N_k(a), D_k(a))_{t_k \leqslant t}$ with an application of the $\delta$-method. To access the asymptotic variance, notice that, using a similar reasonning as in the previous proof:

$$\begin{aligned}\mathbb{E}[(1-D_k(a)/N_k(a))^2|\mathcal{F}_{k-1}] &= \mathbb{E}[1-D_k(a)/N_k(a)|\mathcal{F}_{k-1}(a)]^2 + \frac{1}{N_k(a)^2}\mathrm{Var}(D_k(a)|\mathcal{F}_{k-1}) \\ &= s_k^2(a) + \frac{s_k(a)(1-s_k(a))}{N_k(a)}\mathbb{I}\{N_k(a)>0\} + O(\mathbb{I}\{N_k(a)=0\}).\end{aligned}$$

Now we know that $N_k(a) = nr_k(a) + \sqrt{n}O_{\mathbb{P}}(1)$, with the $O_{\mathbb{P}}(1)$ having uniformly bounded moments. So that we deduce that

$$\mathbb{E}[(1-D_k(a)/N_k(a))^2|\mathcal{F}_{k-1}] = s_k^2(a) + \frac{s_k(a)(1-s_k(a))}{nr_k(a)} + \frac{1}{n^{3/2}}O_{\mathbb{P}}(1),$$

where $O_{\mathbb{P}}(1)$ has again bounded moments. Using this identity, we find that

$$\begin{aligned}n\mathrm{Var}\widehat{S}_{\mathrm{KM}}(t|A=a) &= n\left(\mathbb{E}S_{\mathrm{KM}}(t|A=a)^2 - S^{(a)}(t)^2\right) \\ &= nS^{(a)}(t)^2\left(\mathbb{E}\left[\prod_{t_k \leqslant t}\left(1 + \frac{1}{n}\frac{1-s_k(a)}{s_k(a)r_k(a)} + \frac{1}{n^{3/2}}O_{\mathbb{P}}(1)\right)\right] - 1\right).\end{aligned}$$

Expending the product and using that the $O_{\mathbb{P}}(1)$'s have bounded moments, we finally deduce that

$$\mathbb{E}\left[\prod_{t_k \leqslant t}\left(1 + \frac{1}{n}\frac{1-s_k(a)}{s_k(a)r_k(a)} + \frac{1}{n^{3/2}}O_{\mathbb{P}}(1)\right)\right] - 1 = \frac{1}{n}\sum_{t_k \leqslant t}\frac{1-s_k(a)}{s_k(a)r_k(a)} + \frac{1}{n^{3/2}}O(1),$$

$$n\mathrm{Var}\widehat{S}_{\mathrm{KM}}(t|A=a) = V_{\mathrm{KM}}(t|A=a) + O(n^{-1/2}),$$

which is what we wanted to show. □

## Proofs of Section

*Proof. (Proposition 0.5).* Assumption 17 allows the tranformation to be well-defined. Furthermore, it holds

$$\begin{aligned}
E[T^*_{\mathrm{IPCW}}|A=a,X] &= E\left[\frac{\Delta^\tau \times \widetilde{T}\wedge\tau}{G(\widetilde{T}\wedge\tau|A,X)}\middle|A=a,X\right]\\
&= E\left[\frac{\Delta^\tau \times T(a)\wedge\tau}{G(T(a)\wedge\tau|A,X)}\middle|A=a,X\right]\\
&= E\left[E\left[\frac{\mathbb{I}\{T(a)\wedge\tau\leqslant C\}\times T(a)\wedge\tau}{G(T(a)\wedge\tau|A,X)}\middle|A,X,T(1)\right]\middle|A=a,X\right]\\
&= E\left[T(a)\wedge\tau|A=a,X\right]\\
&= E\left[T(a)\wedge\tau|X\right].
\end{aligned}$$

We used in the second equality that on the event $\{\Delta^\tau=1, A=a\}$, it holds $\widetilde{T}\wedge\tau = T\wedge\tau = T(a)\wedge\tau$. We used in the fourth equality that $G(T(a)\wedge\tau|A,X) = E[\mathbb{I}\{T(a)\wedge\tau\leqslant C\}|X,T(a),A]$ thanks to Assumption 16, and in the last one that $A$ is idependent from $X$ and $T(a)$ thanks to Assumption 5. □

*Proof. (Proposition 0.6).* Similarly to the computations done in the proof of Proposition 0.5, it is easy to show that

$$\mathbb{E}\left[\frac{\Delta_i^\tau}{G(\widetilde{T}\wedge\tau|X,A)}\mathbb{I}(\widetilde{T}_i=t_k, A=a)\right] = \mathbb{P}(A=a)\mathbb{P}(T(a)=t_k),$$

and likewise that

$$\mathbb{E}\left[\frac{\Delta_i^\tau}{G(\widetilde{T}\wedge\tau|X,A)}\mathbb{I}(\widetilde{T}_i\geqslant t_k, A=a)\right] = \mathbb{P}(A=a)\mathbb{P}(T(a)\geqslant t_k),$$

so that $\widehat{S}_{\mathrm{IPCW}}(t)$ converges almost surely towards the product limit

$$\prod_{t_k\leqslant t}\left(1-\frac{\mathbb{P}(T(a)=t_k)}{\mathbb{P}(T(a)\geqslant t_k)}\right) = S^{(a)}(t),$$

yielding strong consistency. Asymptotic normality is straightforward. □

*Proof. (Proposition 0.7).* There holds

$$\begin{aligned}
\mathbb{E}[T^*_{\mathrm{BJ}}|X,A=a] &= \mathbb{E}\left[\Delta^\tau T(a)\wedge\tau + (1-\Delta^\tau)\frac{\mathbb{E}[T\wedge\tau\times\mathbb{I}\{T\wedge\tau>C\}|C,A,X]}{\mathbb{P}(T>C|C,A,X)}\middle|X,A=a\right]\\
&= \mathbb{E}[\Delta^\tau T(a)\wedge\tau|X] + \underbrace{\mathbb{E}\left[\mathbb{I}\{T\wedge\tau>C\}\frac{\mathbb{E}[T\wedge\tau\times\mathbb{I}\{T\wedge\tau>C\}|C,A,X]}{\mathbb{E}[\mathbb{I}\{T\wedge\tau\geqslant C\}|C,A,X]}\middle|X,A=a\right]}_{(\star)}.
\end{aligned}$$

Now we easily see that conditionning wrt $X$ in the second term yields

$$\begin{aligned}(\star) &= \mathbb{E}\left[\mathbb{E}[T\wedge\tau\times\mathbb{I}\{T\wedge\tau > C\}|C,A,X]|X,A=a\right]\\ &= \mathbb{E}[(1-\Delta^\tau)T\wedge\tau|X,A=a]\\ &= \mathbb{E}[(1-\Delta^\tau)T(a)\wedge\tau|X],\end{aligned}$$

ending the proof. □

*Proof. (Theorem 0.1).* We let $T^* = \Delta^\tau\phi_1 + (1-\Delta^\tau)\phi_0$ be a transformation of the form Equation 20. There holds

$$\mathbb{E}[(T^* - T\wedge\tau)^2] = \mathbb{E}[\Delta^\tau(\phi_1 - T\wedge\tau)^2] + \mathbb{E}[(1-\Delta^\tau)(\phi_0 - T\wedge\tau)^2].$$

The first term is non negative and is zero for the BJ transformation. Since $\phi_0$ is a function of $(\widetilde{T}, X, A)$ and that $\widetilde{T} = C$ on $\{\Delta^\tau = 0\}$, the second term can be rewritten in the following way. We let $R$ be a generic quantity that does not depend on $\phi_0$.

$$\begin{aligned}\mathbb{E}[(1-\Delta^\tau)(\phi_0 - T)^2] &= \mathbb{E}\left[\mathbb{I}\{T\wedge\tau > C\}\phi_0^2 - 2\mathbb{I}\{T\wedge\tau > C\}\phi_0 T\wedge\tau\right] + R\\ &= \mathbb{E}\left[\mathbb{P}(T\wedge\tau > C|C,A,X)\phi_0^2 - 2\mathbb{E}[T\wedge\tau\mathbb{I}\{T\wedge\tau > C\}|C,A,X]\phi_0\right] + R\\ &= \mathbb{E}\left[\mathbb{P}(T\wedge\tau > C|C,A,X)\left(\phi_0 - \frac{\mathbb{E}[T\wedge\tau\mathbb{I}\{T\wedge\tau > C\}|C,A,X]}{\mathbb{P}(T\wedge\tau > C|C,A,X)}\right)^2\right] + R.\end{aligned}$$

Now the first term in the right hand side is always non-negative, and is zero for the BJ tranformation. □

## Proofs of Section

*Proof. (Proposition 0.1).* The fact that it is strongly consistent and asymptotically normal is again a simple application of the law of large number and of the $\delta$-method. We indeed find that, for $t_k \leqslant \tau$

$$\begin{aligned}\mathbb{E}\left[\frac{1}{e(X_i)}\mathbb{1}\{\widetilde{T}_i = t_k, \Delta_i = 1, A_i = 1\}\right] &= \mathbb{E}\left[\frac{A_i}{e(X_i)}\mathbb{1}\{T_i = t_k, C_i \geqslant t_k\}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{A_i}{e(X_i)}\mathbb{1}\{T_i = t_k, C_i \geqslant t_k\}\middle|X_i\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{A_i}{e(X_i)}\middle|X_i\right]\mathbb{P}(T_i = t_k|X_i)\mathbb{P}(C_i \geqslant t_k)\right]\\ &= \mathbb{P}(T_i = t_k)\mathbb{P}(C_i \geqslant t_k),\end{aligned}$$

where we used that $A$ is independent from $T$ conditionnaly on $X$, and that $C$ is independent from everything. Likewise, one would get that

$$\mathbb{E}\left[\frac{1}{e(X_i)}\mathbb{1}\{\widetilde{T}_i \geqslant t_k, A_i = 1\}\right] = = \mathbb{P}(T_i \geqslant t_k)\mathbb{P}(C_i \geqslant t_k).$$

Similar computations hold for $A = 0$, ending the proof. □

## Proofs of Section

*Proof. (Proposition 0.2).* On the event $\{\Delta^\tau = 1, A = 1\}$, there holds $\widetilde{T} \wedge \tau = T \wedge \tau = T(1) \wedge \tau$, whence we find that,

$$\begin{aligned}
\mathbb{E}[T^*_{\text{IPCW}}|X, A=1] &= \mathbb{E}\left[\frac{A}{e(X)}\frac{\mathbb{I}\{T(1)\wedge\tau \leqslant C\}}{G(T(1)\wedge\tau|X,A)}T(1)\wedge\tau \middle| X\right] \\
&= \mathbb{E}\left[\frac{A}{e(X)}\mathbb{E}\left[\frac{\mathbb{I}\{T(1)\wedge\tau \leqslant C\}}{G(T(1)\wedge\tau|X,A)}\middle| X, A, T(1)\right]T(1)\wedge\tau \middle| X\right] \\
&= \mathbb{E}\left[\frac{A}{e(X)}T(1)\wedge\tau \middle| X\right] \\
&= \mathbb{E}\left[T(1)\wedge\tau|X\right],
\end{aligned}$$

and the same holds on the event $A = 0$. □

*Proof. (Proposition 0.3).* By consistency of $\widehat{G}(\cdot|X,A)$ and $\widehat{e}$ and by continuity, it suffices to look at the asymptotic behavior of the oracle estimator

$$\theta^*_{\text{IPTW-IPCW}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{e(X_i)} - \frac{1-A_i}{1-e(X_i)}\right)\frac{\Delta_i^\tau}{G(\widetilde{T}_i\wedge\tau|A_i,X_i)}\widetilde{T}_i\wedge\tau.$$

The later is converging almost towards its mean, which, following similar computations as in the previous proof, write

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{A}{e(X)} - \frac{1-A}{1-e(X)}\right)\frac{\Delta^\tau}{G(\widetilde{T}\wedge\tau|A,X)}\widetilde{T}\wedge\tau\right] &= \mathbb{E}\left[\left(\frac{A}{e(X)} - \frac{1-A}{1-e(X)}\right)T\wedge\tau\right] \\
&= \mathbb{E}\left[T(1)\wedge\tau\right] - \mathbb{E}\left[T(0)\wedge\tau\right].
\end{aligned}$$

□

*Proof. (Proposition 0.4).* Asymptotic normality comes from a mere application of the $\delta$-method, while strong consistency follows from the law of large number and the follozing computations. Like for the proof of Proposition 0.6, one find, by first conditionning wrt $X, A, T(a)$, that, for $t_k \leqslant \tau$,

$$\mathbb{E}\left[\left(\frac{A}{e(X)} + \frac{1-A}{1-e(X)}\right)\frac{\Delta^\tau}{G(\widetilde{T}\wedge\tau|A,X)}\mathbb{I}\{\widetilde{T} = t_k, A = a\}\right] = \mathbb{P}(T(a) = t_k)$$

and likewise that

$$\mathbb{E}\left[\left(\frac{A}{e(X)} + \frac{1-A}{1-e(X)}\right)\frac{\Delta^\tau}{G(\widetilde{T}\wedge\tau|A,X)}\mathbb{I}\{\widetilde{T} \geqslant t_k, A = a\}\right] = \mathbb{P}(T(a) \geqslant t_k)$$

so that indeed $S^*_{\text{IPTW-IPCW}}(t|A=a)$ converges almost surely towards $S^{(a)}(t)$. □

*Proof. (Proposition 0.5).* We write

$$\mathbb{E}[T^*_{\mathrm{IPTW-BJ}}|X, A=1] = \mathbb{E}\left[\frac{A}{e(X)}\Delta^\tau \times \widetilde{T}\wedge\tau \middle| X\right] + \mathbb{E}\left[\frac{A}{e(X)}(1-\Delta^\tau)Q_S(\widetilde{T}\wedge\tau|A,X)\middle| X\right].$$

On the event $\{\Delta^\tau = 1, A = 1\}$, there holds $\widetilde{T}\wedge\tau = T\wedge\tau = T(1)\wedge\tau$, whence we find that the first term on the the right hand side is equal to

$$\begin{aligned}\mathbb{E}\left[\frac{A}{e(X)}\Delta^\tau \times \widetilde{T}\wedge\tau \middle| X\right] &= \mathbb{E}\left[\frac{A}{e(X)}\Delta^\tau \times T(1)\wedge\tau \middle| X\right]\\ &= \mathbb{E}\left[\Delta^\tau \times T(1)\wedge\tau | X\right].\end{aligned}$$

For the second term in the right hand side, notice that on the event $\{\Delta^\tau = 0, A = 1\}$, there holds $\widetilde{T} = C < T(1)\wedge\tau$, so that

$$\begin{aligned}&\mathbb{E}\left[\frac{A}{e(X)}\mathbb{I}\{C < T(1)\wedge\tau\}\frac{\mathbb{E}[T(1)\wedge\tau \times \mathbb{I}\{C < T(1)\wedge\tau\}|X,A,C]}{\mathbb{P}(C < T(1)\wedge\tau|C,X,A)}\middle| X\right]\\ &= \mathbb{E}\left[\frac{A}{e(X)}\mathbb{E}[T(1)\wedge\tau \times \mathbb{I}\{C < T(1)\wedge\tau\}|X,A,C]\middle| X\right]\\ &= \mathbb{E}\left[T(1)\wedge\tau \times \mathbb{I}\{C < T(1)\wedge\tau\}|X\right]\\ &= \mathbb{E}\left[(1-\Delta^\tau)T(1)\wedge\tau|X\right],\end{aligned}$$

and the sane holds on the event $\{A = 0\}$, which ends the proof. □

*Proof. (Proposition 0.6).* By consistency of $\widehat{G}(\cdot|X, A)$ and $\widehat{e}$ and by continuity, it suffices to look at the asymptotic behavior of the oracle estimator

$$\theta^*_{\mathrm{IPTW-BJ}} = \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{e(X_i)} - \frac{1-A_i}{1-e(X_i)}\right)\left(\Delta_i^\tau \times \widetilde{T}_i\wedge\tau + (1-\Delta_i^\tau)Q_S(\widetilde{T}_i\wedge\tau|A_i,X_i)\right).$$

The later is converging almost towards its mean, which, following similar computations as in the previous proof, is simply equal to the difference in RMST. □

*Proof. (Proposition 0.7).* We can write that

$$\Delta^*_{\mathrm{QR}} = \underbrace{\frac{A}{p(X)}(T^*_{\mathrm{DR}}(F,R) - \nu(X,1)) + \nu(X,1)}_{\text{(A)}} - \left(\underbrace{\frac{1-A}{1-p(X)}(T^*_{\mathrm{DR}}(F,R) - \nu(X,0)) + \nu(X,0)}_{\text{(B)}}\right).$$

Focusing on term (A), we easily find that

$$\begin{aligned}\mathbb{E}[(\mathrm{A})|X] &= \mathbb{E}\left[\frac{A}{p(X)}(T^*_{\mathrm{DR}}(F,R) - \nu(X,1)) + \nu(X,1)\middle| X\right]\\ &= \frac{e(X)}{p(X)}(\mu(X,1) - \nu(X,1)) + \nu(X,1).\end{aligned}$$

Where we used that $T^*_{\mathrm{DR}}(F,R)$ is a censoring unbiased transform when $F = G$ or $R = S$. Now we see that if $p(X) = e(X)$, then

$$\mathbb{E}[(\mathrm{A})|X] = \mu(X,1) - \nu(X,1) + \nu(X,1) = \mu(X,1),$$

and if $\nu(X,1) = \mu(X,1)$, then

$$\mathbb{E}[(\mathrm{A})|X] = \frac{e(X)}{p(X)} \times 0 + \mu(X,1) = \mu(X,1).$$

Likewise, we would show that $\mathbb{E}[(\mathrm{B})|X] = \mu(X,0)$ under either alternative, ending the proof. □

## Appendix B: Descriptive statistics

### RCT

The summary by group of treatment of the generated (observed and unobserved) RCT with independent censoring is displayed below:

```
[1] "Descriptive statistics for group A=0:   994"

       X1                 X2                 X3                 X4         
 Min.   :-1.7289   Min.   :-1.6675   Min.   :-3.5536   Min.   :-2.5947  
 1st Qu.: 0.4049   1st Qu.: 0.3575   1st Qu.:-1.6691   1st Qu.: 0.2884  
 Median : 1.0551   Median : 1.0517   Median :-0.9817   Median : 0.9759  
 Mean   : 1.0749   Mean   : 1.0482   Mean   :-0.9637   Mean   : 0.9896  
 3rd Qu.: 1.7787   3rd Qu.: 1.7114   3rd Qu.:-0.2477   3rd Qu.: 1.6765  
 Max.   : 4.1047   Max.   : 4.2832   Max.   : 2.0039   Max.   : 4.6024  
       C                  T1                 T0                 status      
 Min.   :  0.0557   Min.   : 10.01   Min.   :  0.0055   Min.   :0.0000  
 1st Qu.: 10.0231   1st Qu.: 13.03   1st Qu.:  3.0341   1st Qu.:0.0000  
 Median : 24.0262   Median : 18.58   Median :  8.5822   Median :1.0000  
 Mean   : 33.9630   Mean   : 30.96   Mean   : 20.9574   Mean   :0.6781  
 3rd Qu.: 47.1684   3rd Qu.: 32.55   3rd Qu.: 22.5492   3rd Qu.:1.0000  
 Max.   :387.0902   Max.   :478.30   Max.   :468.2999   Max.   :1.0000  
    T_tild         
 Min.   :  0.00549  
 1st Qu.:  2.22200  
 Median :  5.86540  
 Mean   : 11.02970  
 3rd Qu.: 13.57754  
 Max.   :115.61110  
```

```
[1] "Descriptive statistics for group A=1:   1006"

      X1               X2               X3               X4        
 Min.   :-1.6761   Min.   :-2.4644   Min.   :-3.7535   Min.   :-2.1887  
 1st Qu.: 0.2868   1st Qu.: 0.2672   1st Qu.:-1.6744   1st Qu.: 0.3732  
 Median : 0.9781   Median : 0.9633   Median :-1.0249   Median : 1.0329  
 Mean   : 0.9494   Mean   : 0.9850   Mean   :-1.0098   Mean   : 1.0147  
 3rd Qu.: 1.5881   3rd Qu.: 1.6852   3rd Qu.:-0.3607   3rd Qu.: 1.7130  
 Max.   : 3.7951   Max.   : 3.7060   Max.   : 1.8770   Max.   : 3.9754  
      C                  T1               T0                status      
 Min.   :  0.05719   Min.   : 10.01   Min.   :  0.0145   Min.   :0.0000  
 1st Qu.:  9.86947   1st Qu.: 13.24   1st Qu.:  3.2382   1st Qu.:0.0000  
 Median : 24.82029   Median : 19.09   Median :  9.0866   Median :1.0000  
 Mean   : 36.18546   Mean   : 33.80   Mean   : 23.8003   Mean   :0.5159  
 3rd Qu.: 49.41240   3rd Qu.: 33.33   3rd Qu.: 23.3347   3rd Qu.:1.0000  
 Max.   :308.44118   Max.   :512.18   Max.   :502.1761   Max.   :1.0000  
    T_tild         
 Min.   :  0.05719  
 1st Qu.:  9.86947  
 Median : 13.26102  
 Mean   : 16.64539  
 3rd Qu.: 20.51981  
 Max.   :112.63104  
```

Covariates are balanced between groups, and censoring times are the same (independent censoring). However, there are more censored observations in the treated group ($A = 1$) than in the control group ($A = 0$). This is due to the higher instantaneous hazard of the event in the treated group (with $T_1 = T_0 + 10$) compared to the constant hazard of censoring.

The summary of the generated (observed and unobserved) RCT with conditionally independent censoring stratified by treatment is displayed below.

```
[1] "Descriptive statistics for group A=0:   997"

      X1               X2               X3               X4        
 Min.   :-2.3950   Min.   :-2.2070   Min.   :-4.1034   Min.   :-1.7474  
 1st Qu.: 0.3124   1st Qu.: 0.3943   1st Qu.:-1.6598   1st Qu.: 0.3885  
 Median : 0.9498   Median : 1.0679   Median :-0.9881   Median : 0.9871  
 Mean   : 0.9782   Mean   : 1.0600   Mean   :-0.9936   Mean   : 1.0248  
 3rd Qu.: 1.6811   3rd Qu.: 1.6960   3rd Qu.:-0.3340   3rd Qu.: 1.6931  
 Max.   : 4.6384   Max.   : 4.6064   Max.   : 2.3274   Max.   : 4.2426  
      C                 T1               T0                status      
```

```
 Min.   :  0.0091   Min.   : 10.00   Min.   :  0.0018   Min.   :0.0000
 1st Qu.:  2.5676   1st Qu.: 13.12   1st Qu.:  3.1208   1st Qu.:0.0000
 Median :  7.5821   Median : 18.26   Median :  8.2641   Median :0.0000
 Mean   : 15.9557   Mean   : 30.60   Mean   : 20.5991   Mean   :0.4524
 3rd Qu.: 16.9992   3rd Qu.: 32.36   3rd Qu.: 22.3636   3rd Qu.:1.0000
 Max.   :627.9902   Max.   :469.92   Max.   :459.9153   Max.   :1.0000
   status_tau         T_tild
 Min.   :0.0000   Min.   :  0.00178
 1st Qu.:0.0000   1st Qu.:  1.29789
 Median :0.0000   Median :  3.90189
 Mean   :0.4855   Mean   :  8.13142
 3rd Qu.:1.0000   3rd Qu.:  9.11235
 Max.   :1.0000   Max.   :221.32293

[1] "Descriptive statistics for group A=1:   1003"

       X1                 X2                 X3                 X4
 Min.   :-2.3018   Min.   :-2.0328   Min.   :-3.8875   Min.   :-2.5845
 1st Qu.: 0.3753   1st Qu.: 0.2979   1st Qu.:-1.6300   1st Qu.: 0.2945
 Median : 1.0592   Median : 0.9514   Median :-0.9889   Median : 0.9468
 Mean   : 1.0415   Mean   : 0.9750   Mean   :-0.9670   Mean   : 0.9724
 3rd Qu.: 1.7389   3rd Qu.: 1.6713   3rd Qu.:-0.2790   3rd Qu.: 1.6869
 Max.   : 4.4888   Max.   : 4.1004   Max.   : 2.1800   Max.   : 4.1333
       C                  T1                 T0               status
 Min.   :  0.0045   Min.   : 10.00   Min.   :  0.0006   Min.   :0.0000
 1st Qu.:  3.1334   1st Qu.: 12.99   1st Qu.:  2.9893   1st Qu.:0.0000
 Median :  8.7082   Median : 19.34   Median :  9.3374   Median :0.0000
 Mean   : 18.8248   Mean   : 33.19   Mean   : 23.1888   Mean   :0.2114
 3rd Qu.: 21.1503   3rd Qu.: 35.69   3rd Qu.: 25.6902   3rd Qu.:0.0000
 Max.   :517.0102   Max.   :416.46   Max.   :406.4628   Max.   :1.0000
   status_tau         T_tild
 Min.   :0.0000   Min.   :  0.0045
 1st Qu.:0.0000   1st Qu.:  3.1334
 Median :0.0000   Median :  8.7082
 Mean   :0.2772   Mean   : 13.0888
 3rd Qu.:1.0000   3rd Qu.: 15.2133
 Max.   :1.0000   Max.   :396.5864
```

Covariates are balanced between the two groups. However, censoring times differ between groups due to conditionally independent censoring based on covariates and treatment group. Indeed, the distribution of $C$ is different between the treatment group.

### Observational study with linear relationship

The summary of the generated (observed and unobserved) data set observational study with independent censoring stratified by treatment is displayed below to enhance the difference with the other scenario.

```
[1] "Descriptive statistics for group A=0:   1107"

       X1                 X2                 X3                 X4
 Min.   :-1.5415   Min.   :-1.5836   Min.   :-3.5725   Min.   :-2.382
 1st Qu.: 0.5226   1st Qu.: 0.5977   1st Qu.:-1.0392   1st Qu.: 0.515
 Median : 1.1336   Median : 1.2308   Median :-0.5184   Median : 1.215
 Mean   : 1.1680   Mean   : 1.2294   Mean   :-0.4639   Mean   : 1.190
 3rd Qu.: 1.8406   3rd Qu.: 1.8727   3rd Qu.: 0.1137   3rd Qu.: 1.855
 Max.   : 4.1065   Max.   : 4.2164   Max.   : 2.6845   Max.   : 4.295
       C                 T1                T0                status
 Min.   :  0.1119   Min.   : 10.00   Min.   :  0.0006   Min.   :0.0000
 1st Qu.:  9.2470   1st Qu.: 12.56   1st Qu.:  2.5637   1st Qu.:0.0000
 Median : 23.5543   Median : 17.75   Median :  7.7471   Median :1.0000
 Mean   : 33.4391   Mean   : 31.25   Mean   : 21.2524   Mean   :0.6811
 3rd Qu.: 45.5396   3rd Qu.: 31.11   3rd Qu.: 21.1062   3rd Qu.:1.0000
 Max.   :250.8151   Max.   :588.38   Max.   :578.3805   Max.   :1.0000
     T_tild
 Min.   :  0.00057
 1st Qu.:  1.90447
 Median :  5.26114
 Mean   : 10.21345
 3rd Qu.: 13.60035
 Max.   :139.57810

[1] "Descriptive statistics for group A=1:   893"

       X1                 X2                  X3                 X4
 Min.   :-3.0669   Min.   :-2.32992   Min.   :-4.6819   Min.   :-2.4816
 1st Qu.: 0.1596   1st Qu.: 0.07743   1st Qu.:-2.2080   1st Qu.: 0.1109
 Median : 0.7912   Median : 0.71186   Median :-1.5932   Median : 0.7625
 Mean   : 0.7804   Mean   : 0.72313   Mean   :-1.6210   Mean   : 0.7762
 3rd Qu.: 1.3751   3rd Qu.: 1.39288   3rd Qu.:-1.0415   3rd Qu.: 1.4506
 Max.   : 3.7677   Max.   : 3.92047   Max.   : 0.5198   Max.   : 3.6989
       C                  T1                T0                status
 Min.   :  0.07883   Min.   : 10.01   Min.   :  0.0107   Min.   :0.0000
```

```
 1st Qu.:  8.76083   1st Qu.: 12.99   1st Qu.:  2.9876   1st Qu.:0.0000
 Median : 24.34523   Median : 18.76   Median :  8.7567   Median :0.0000
 Mean   : 34.47734   Mean   : 32.00   Mean   : 21.9972   Mean   :0.4994
 3rd Qu.: 46.52534   3rd Qu.: 34.31   3rd Qu.: 24.3055   3rd Qu.:1.0000
 Max.   :261.56279   Max.   :748.90   Max.   :738.9036   Max.   :1.0000
     T_tild
 Min.   :  0.07883
 1st Qu.:  8.76083
 Median : 12.69039
 Mean   : 16.12371
 3rd Qu.: 20.31389
 Max.   :104.90774
```

The covariates between the two groups of treatment are unbalanced because of dependent treatment assignation. The mean of $X1$, $X2$, $X3$ and $X4$ is bigger in the control group than in the treated group. The censoring times have the same distribution (independent censoring). There are more censored observation in the treated group (A=1) than in the control group (A=0) for the same reason than in the RCT scenario.

The summary of the generated (observed and unobserved) data set observational study with conditionally independent censoring stratified by treatment is displayed below.

```
[1] "Descriptive statistics for group A=0:   1162"

       X1                  X2                  X3                   X4
 Min.   :-2.4545   Min.   :-1.5684   Min.   :-2.895571   Min.   :-1.570
 1st Qu.: 0.4962   1st Qu.: 0.5413   1st Qu.:-1.016918   1st Qu.: 0.562
 Median : 1.1778   Median : 1.1685   Median :-0.552030   Median : 1.269
 Mean   : 1.1934   Mean   : 1.1853   Mean   :-0.493956   Mean   : 1.231
 3rd Qu.: 1.8634   3rd Qu.: 1.8418   3rd Qu.: 0.002819   3rd Qu.: 1.857
 Max.   : 4.1974   Max.   : 4.4669   Max.   : 2.323784   Max.   : 4.647
       C                   T1                T0                status
 Min.   :  0.00523   Min.   : 10.01   Min.   :  0.0128   Min.   :0.0000
 1st Qu.:  2.61917   1st Qu.: 12.64   1st Qu.:  2.6420   1st Qu.:0.0000
 Median :  6.64848   Median : 18.03   Median :  8.0294   Median :0.0000
 Mean   : 13.84299   Mean   : 30.72   Mean   : 20.7181   Mean   :0.4639
 3rd Qu.: 16.33470   3rd Qu.: 32.10   3rd Qu.: 22.0968   3rd Qu.:1.0000
 Max.   :222.59564   Max.   :564.55   Max.   :554.5536   Max.   :1.0000
   status_tau          T_obs                 e
 Min.   :0.0000   Min.   :  0.00523   Min.   :0.0000263
 1st Qu.:0.0000   1st Qu.:  1.33612   1st Qu.:0.0222259
 Median :1.0000   Median :  3.60392   Median :0.1107057
```

```
 Mean   :0.5052   Mean   :  7.70079   Mean   :0.2020889  
 3rd Qu.:1.0000   3rd Qu.:  8.78692   3rd Qu.:0.3214706  
 Max.   :1.0000   Max.   :163.07354   Max.   :0.9550110  
```

```
[1] "Descriptive statistics for group A=1:   838"
```

```
       X1                  X2                 X3               X4          
 Min.   :-1.822557   Min.   :-2.53859   Min.   :-4.344   Min.   :-2.92866  
 1st Qu.: 0.007232   1st Qu.: 0.07453   1st Qu.:-2.061   1st Qu.: 0.07862  
 Median : 0.672269   Median : 0.74975   Median :-1.531   Median : 0.69405  
 Mean   : 0.655351   Mean   : 0.72827   Mean   :-1.551   Mean   : 0.68255  
 3rd Qu.: 1.303977   3rd Qu.: 1.40323   3rd Qu.:-1.004   3rd Qu.: 1.29402  
 Max.   : 3.881681   Max.   : 3.74087   Max.   : 1.232   Max.   : 3.80721  
       C                   T1                T0               status      
 Min.   :  0.0083   Min.   : 10.01   Min.   :  0.0132   Min.   :0.0000  
 1st Qu.:  2.9285   1st Qu.: 13.56   1st Qu.:  3.5645   1st Qu.:0.0000  
 Median :  8.5018   Median : 20.24   Median : 10.2359   Median :0.0000  
 Mean   : 18.8262   Mean   : 39.05   Mean   : 29.0495   Mean   :0.2088  
 3rd Qu.: 19.7484   3rd Qu.: 36.73   3rd Qu.: 26.7296   3rd Qu.:0.0000  
 Max.   :491.3237   Max.   :800.81   Max.   :790.8084   Max.   :1.0000  
   status_tau         T_obs                e          
 Min.   :0.0000   Min.   :  0.00829   Min.   :0.01437  
 1st Qu.:0.0000   1st Qu.:  2.92849   1st Qu.:0.60360  
 Median :0.0000   Median :  8.50175   Median :0.84475  
 Mean   :0.2745   Mean   : 12.64204   Mean   :0.74858  
 3rd Qu.:1.0000   3rd Qu.: 15.13575   3rd Qu.:0.95586  
 Max.   :1.0000   Max.   :148.36151   Max.   :0.99995  
```

The covariates between the two groups are unbalanced. The censoring time is dependent on the covariates also, as the covariates are unbalanced between the two groups, the censoring time is also unbalanced. In particular, the mean of $X1$, $X2$, $X3$ and $X4$ is bigger in the control group than in the treated group. Also, the number of events is bigger in the control than treated group.

### Observational study with interaction

```
       X1                X2                 X3                  X4         
 Min.   :-2.6282   Min.   :-3.3541   Min.   :-3.07680   Min.   :-3.5034  
 1st Qu.:-0.2492   1st Qu.:-0.1818   1st Qu.: 0.07369   1st Qu.:-0.1005  
 Median : 0.4252   Median : 0.4893   Median : 0.72716   Median : 0.5533  
 Mean   : 0.4530   Mean   : 0.4790   Mean   : 0.74757   Mean   : 0.5328  
```

```
 3rd Qu.: 1.1444   3rd Qu.: 1.1479   3rd Qu.: 1.43490   3rd Qu.: 1.1951
 Max.   : 3.8543   Max.   : 4.2205   Max.   : 3.79897   Max.   : 4.7476
      tau              A                 T0                    T1
 Min.   :0.5   Min.   :0.0000   Min.   :  0.00000   Min.   :  1.000
 1st Qu.:0.5   1st Qu.:0.0000   1st Qu.:  0.03138   1st Qu.:  1.031
 Median :0.5   Median :1.0000   Median :  0.17577   Median :  1.176
 Mean   :0.5   Mean   :0.5995   Mean   :  0.89130   Mean   :  1.891
 3rd Qu.:0.5   3rd Qu.:1.0000   3rd Qu.:  0.72433   3rd Qu.:  1.724
 Max.   :0.5   Max.   :1.0000   Max.   :108.83207   Max.   :109.832
       C                 T_obs            T_obs_tau            status
 Min.   :    0.000   Min.   :0.000001   Min.   :0.0000007   Min.   :0.000
 1st Qu.:    0.131   1st Qu.:0.045934   1st Qu.:0.0459341   1st Qu.:0.000
 Median :    0.548   Median :0.230326   Median :0.2303264   Median :0.000
 Mean   :   12.783   Mean   :0.520115   Mean   :0.2624220   Mean   :0.448
 3rd Qu.:    1.863   3rd Qu.:0.941402   3rd Qu.:0.5000000   3rd Qu.:1.000
 Max.   :14881.665   Max.   :8.590410   Max.   :0.5000000   Max.   :1.000
   status_tau     censor.status         e
 Min.   :0.0000   Min.   :0.000   Min.   :0.0000854
 1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.4149439
 Median :1.0000   Median :1.000   Median :0.6012078
 Mean   :0.6005   Mean   :0.552   Mean   :0.5941768
 3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.8179769
 Max.   :1.0000   Max.   :1.000   Max.   :0.9999800
```

The summary of the generated (observed and unobserved) data set complex observational study (conditionally independent censoring) stratified by treatment is displayed below.

```
[1] "Descriptive statistics for group A=0:   801"

       X1                 X2                  X3                 C
 Min.   :-2.3890   Min.   :-3.09873   Min.   :-1.8748   Min.   :    0.000
 1st Qu.:-0.2655   1st Qu.:-0.01631   1st Qu.:-0.1215   1st Qu.:    0.173
 Median : 0.3891   Median : 0.67151   Median : 0.4489   Median :    0.626
 Mean   : 0.4261   Mean   : 0.63974   Mean   : 0.4494   Mean   :   22.594
 3rd Qu.: 1.1143   3rd Qu.: 1.34162   3rd Qu.: 1.0486   3rd Qu.:    1.880
 Max.   : 3.8543   Max.   : 4.22046   Max.   : 3.2792   Max.   :14881.665
       T1                 T0                  status            T_obs
 Min.   :  1.000   Min.   :  0.00000   Min.   :0.0000   Min.   :0.000002
 1st Qu.:  1.029   1st Qu.:  0.02863   1st Qu.:0.0000   1st Qu.:0.019174
 Median :  1.146   Median :  0.14637   Median :1.0000   Median :0.087344
 Mean   :  1.754   Mean   :  0.75433   Mean   :0.6916   Mean   :0.258288
 3rd Qu.:  1.617   3rd Qu.:  0.61651   3rd Qu.:1.0000   3rd Qu.:0.263492
```

```
 Max.   :109.832   Max.   :108.83207   Max.   :1.0000   Max.   :8.590410  
   status_tau          e            
 Min.   :0.000   Min.   :0.0000854  
 1st Qu.:1.000   1st Qu.:0.2383314  
 Median :1.000   Median :0.4388064  
 Mean   :0.764   Mean   :0.4239425  
 3rd Qu.:1.000   3rd Qu.:0.5929849  
 Max.   :1.000   Max.   :0.9991406  

[1] "Descriptive statistics for group A=1:   1199"

       X1                 X2                 X3                 C            
 Min.   :-2.6282   Min.   :-3.3541   Min.   :-3.0768   Min.   :   0.0000  
 1st Qu.:-0.2323   1st Qu.:-0.2765   1st Qu.: 0.2594   1st Qu.:   0.1143  
 Median : 0.4507   Median : 0.3942   Median : 0.9545   Median :   0.4651  
 Mean   : 0.4709   Mean   : 0.3716   Mean   : 0.9468   Mean   :   6.2293  
 3rd Qu.: 1.1546   3rd Qu.: 0.9919   3rd Qu.: 1.6995   3rd Qu.:   1.8400  
 Max.   : 3.6645   Max.   : 3.7623   Max.   : 3.7990   Max.   :1649.8305  
       T1               T0              status          T_obs         
 Min.   : 1.000   Min.   : 0.00000   Min.   :0.0000   Min.   :0.000001  
 1st Qu.: 1.033   1st Qu.: 0.03324   1st Qu.:0.0000   1st Qu.:0.114330  
 Median : 1.201   Median : 0.20110   Median :0.0000   Median :0.465132  
 Mean   : 1.983   Mean   : 0.98280   Mean   :0.2852   Mean   :0.695030  
 3rd Qu.: 1.835   3rd Qu.: 0.83494   3rd Qu.:1.0000   3rd Qu.:1.058750  
 Max.   :54.512   Max.   :53.51198   Max.   :1.0000   Max.   :8.467359  
   status_tau           e          
 Min.   :0.0000   Min.   :0.07066  
 1st Qu.:0.0000   1st Qu.:0.54606  
 Median :0.0000   Median :0.74257  
 Mean   :0.4912   Mean   :0.70790  
 3rd Qu.:1.0000   3rd Qu.:0.89906  
 Max.   :1.0000   Max.   :0.99998  
```

The observations are the same than the previous scenario: The covariates and the censoring time between the two groups are unbalanced. To be able to evaluate the estimators, we need to know the true $\theta_{\mathrm{RMST}}$ at time $\tau$.