Causal survival analysis

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

Charlotte Voinot[1] CMEI, Sanofi R&D
Premedical, INRIA
INSERM
Université de Montpellier
Clément Berenfeld[2] Universität Potsdam, Potsdam, Germany
Imke Mayer Charité – Universität Berlin, Berlin, Germany
Bernard Sebastien[3] CMEI, Sanofi R&D
Julie Josse[4] Premedical, INRIA
INSERM
Université de Montpellier

Last modified: 2024-11-23

Abstract

Estimating the Average Treatment Effect (ATE) is one of the fundamental measures in causal inference, aimed at assessing the causal impact of a treatment on an outcome variable. Causal survival analysis is at the heart of this approach, seeking to evaluate the effect of a treatment on patient survival over time. However, despite the abundance of literature on causal survival, the use of Cox methods remains predominant for assessing this effect. The main objective of this research is to estimate the causal effect of a treatment using survival data which not necessarily derived from randomized trials. Its main aim is to provide users with practical recommendations in the face of the multitude of information available, and to highlight the advantages and differences compared with the classic correlation approaches still widely used such as Hazard ratio to measure the impact of a treatment. To this end, we will begin by presenting the state of the art in causal survival methods, describing identifiable assumptions and the main estimators, including weighting, regression and triply/doubly robust approaches. Then, an extensive simulation study will then be carried out to compare the different estimators, their preferred regimes and illustrate their theoretical properties on finite sample sizes. Finally, we will examine how the addition of certain variables in the censoring, survival or treatment models can impact the variance of the estimators. Results will be discussed with a particular focus on the validity of the estimators and its robustness to misspecification on finite sample sizes simulated datasets for practical recommendations in non-randomized settings.

Keywords: Restricted Mean Survival Time, Randomized Control Trial, Observational Study, Censoring

[1]Corresponding author: charlotte.voinot@sanofi.com
[2]Corresponding author: clement.berenfeld@uni-potsdam.de
[3]Corresponding author: bernard.sebastien@sanofi.com
[4]Corresponding author: julie.josse@inria.fr

Contents

1 Introduction

1.1 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

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

ID	Covariates			Treatment	Censoring	Status	Potential outcome		True outcome	Observed outcome
	X_1	X_2	X_3	A	C	Δ	$T(0)$	$T(1)$	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

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).

1.2 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 $\Delta = \mathbb{I}\{T \leq C\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. Δ is 1 if the true outcome is observed, and 0 if it is censored.

We will assume observing a n-sample of variables $(X, A, \tilde{T}, \Delta)$ 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.

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 1.1. (Causal effect: Difference between Restricted Mean Survival Time)

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

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 $G(t) := \mathbb{P}(C \geq t)$ 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, θ_{RMST} can be interpreted as the mean difference between the survival function of treated and control until a fixed time horizon τ. A difference in RMST $\theta_{\text{RMST}} = 10$ days 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.

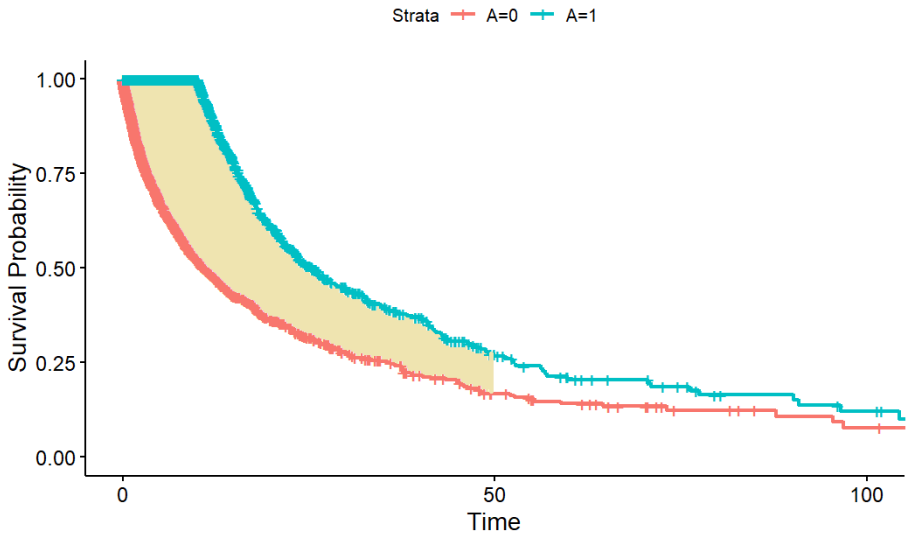


Figure 1: Plot of the estimated survival curves on synthetic toy-data. The θ_{RMST} at $\tau = 50$ corresponds to the yellow shaded area between the two survival curves. The curves have been estimated using Kaplan-Meier estimator, see Section 2.1.

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 τ, see for instance (Ozenne et al. 2020). As mentionned in Section 1.1, another widely used causal measure 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

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

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

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

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

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

where $\{t_1, \dots, t_K\}$ are the atoms of $T^{(a)}$. These hazard functions are classically used to model the survival times and the censoring times, see Section 2.2.1.

1.3 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 2) and observational data (Section 3). We give their statistical properties (consistency, asymptotic normality) along with proofs when possible. We then conduct in Section 5 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 6 with practical recommendations on estimator selection, based on criteria such as asymptotic behavior, computational complexity, and efficiency.

1.4 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 $\lambda^{(a)}(t) = \lim_{h \to 0^+} \mathbb{P}(T(a) \in [t, t+h) \vert T(a) \geq t)/h$ 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 $G(t) = \mathbb{P}(C \geq t)$ of the censoring time
$\tilde{T}$	Observed time ($T \wedge C$)
Δ	Censoring status $\mathbb{I}\{T \leq C\}$
Δ^τ	Censoring status of the restricted time $\Delta^\tau = \max\{\Delta, \mathbb{I}\{\tilde{T} \geq \tau\}\}$
$\{t_1, t_2, \dots, t_K\}$	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 $\mathbb{P}(C \geq t \vert X = x, A = a)$
$Q_S(t\vert x, a), a \in \{0, 1\}$	$\mathbb{E}[T \wedge \tau \mid X = x, A = a, T \wedge \tau > t]$

2 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 θ_{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 2.2), 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).

2.1 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.

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

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

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

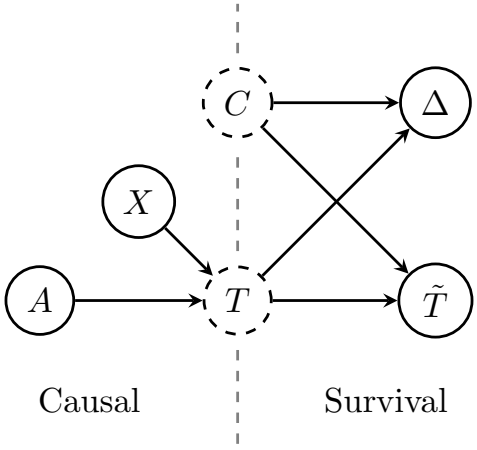


Figure 2: Causal graph in RCT survival data with independent censoring.

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 $\tilde{T}_i$ would be in $[0, t]$ almost surely. In practice, adjusting the threshold time τ 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, τ 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 $(\tilde{T}, \Delta, A = a)$, 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 $\{t_k\}_{k\geq 1}$ be a set of positive and increasing times and assume that $T \in \{t_k\}_{k\geq 1}$ almost surely. Then for any $t \in [0, \tau]$, it holds, letting $t_0 = 0$ by convention, thanks to Equation 4,

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

Using Assumptions 8 and 9, we find that

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

yielding the final identity

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

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

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

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

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

The assiociated RMST estimator is then simply defined as

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

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

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

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

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 7.1 a simple proof of our bound is our context.

Proposition 2.2. *Under Assumptions 1, 5, 6, 8 and 9, and for all $t \in [0, \tau]$, $\hat{S}_{\mathrm{KM}}(t|A=a)$ is asymptotically normal and $\sqrt{n}\left(\hat{S}_{\mathrm{KM}}(t|A=a) - S^{(a)}(t)\right)$ converges in distribution towards a centered Gaussian of variance*

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

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

The proof of Proposition 2.2 can be found in Section 7.1. 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):

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

We finally mention that the KM estimator as defined in Definition 2.1 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 ($\tilde{T}_i = t_k, \Delta_i = 1$). 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.

2.2 Conditionally independent censoring

An alternative hypothesis in survival analysis that relaxes the assumption of independent censoring is conditionally independent censoring, also refered sometimes as *non-informative censoring*. It allows to ;odel 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

$$G(t|A, X) \geq \epsilon. \tag{17}$$

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

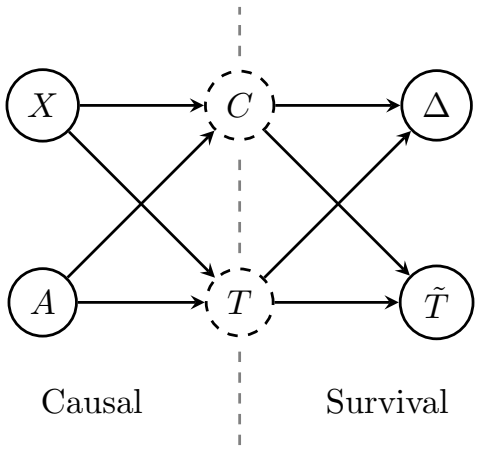


Figure 3: Causal graph in RCT survival data with dependent censoring.

Under dependent censoring, the Kaplan-Meier estimator as defined in Definition 2.1 can fail to estimate survival, in particular because Equation 10 does not hold anymore. Alternatives include plug-in G-formula estimators (Section 2.2.1) and unbiased transformations (Section 2.2.2).

2.2.1 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 μ:

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

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

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

We would like to stress that a G-formula approach works also in a observational context as the one introduced in Section 3.2. 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 μ, we decided to not delay the introduction of the G-formula any further, and we present below a few estimation methods for μ. These methods are sub-divised in two categories: *T-learners*, where $\mu(\cdot, 0)$ is estimated separately from $\mu(\cdot, 0)$, and _*S*-learners, where $\hat{\mu}$ is obtained by fitting a single model based on covariates (X, A).

Cox's Model. There are many ways to model μ 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):

$$\mathscr{L}(\beta) := \prod_{\Delta_i=1} \frac{\exp(X_i^\top \beta)}{\sum_{\tilde{T}_j \geq \tilde{T}_i} \exp(X_j^\top \beta)},$$

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

$$\widehat{\Lambda}_0^{(a)}(t) = \sum_{\Delta_i=1, \tilde{T}_i \leq t} \frac{1}{\sum_{\tilde{T}_j \geq \tilde{T}_i} \exp(X_j^\top \hat{\beta}^{(a)})}$$

where $\hat{\beta}^{(a)}$ is a partial likelihood maximizer. One can show that $(\hat{\beta}^{(a)}, \widehat{\Lambda}_0^{(a)})$ 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 $\hat{\beta}^{(a)}$ 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 $(\hat{\beta}^{(a)}, \widehat{\Lambda}_0^{(a)})$ 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 the treatment or 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 τ, 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.

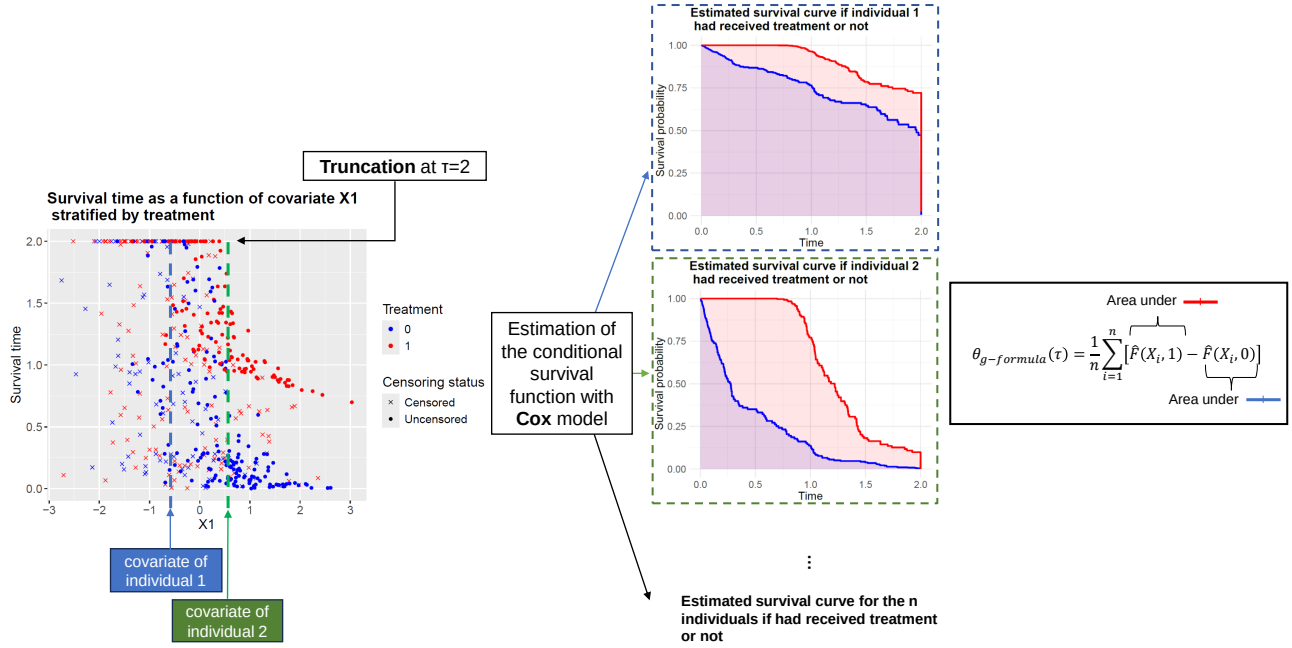


Figure 4: Illustration of the G-formula for estimating θ_{RMST} in an RCT when only one covariate X_1 influences the outcome.

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^γ 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).

2.2.2 Censoring unbiased transformations

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

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

for two wisely chosen transformations ϕ_0 and ϕ_1, and where

$$\Delta^\tau := \mathbb{I}\{T \wedge \tau \leq C\} = \Delta + (1-\Delta)\mathbb{I}(\tilde{T} \geq \tau) \tag{21}$$

is the indicator of the event where the individual is either uncensored or censored after time τ. The idea behind the introduction of Δ^τ is that because we are only interested in computed the expectation of the survival time thresholded by τ, any censored observation coming after time τ 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:

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

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

Proposition 2.3. *Under Assumptions 5 and 6, the estimator $\hat{\theta}$ 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 π is known. If not, it is usual to replace π by its empirical counterpart $\hat{\pi} = n_1/n$ where $n_a = \sum_i \mathbb{1}\{A = a\}$. The resulting estimator takes the form

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

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

Proposition 2.4. *Under Assumptions 5 and 6, the estimator $\hat{\theta}$ 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.

The Inverse-Probability-of-Censoring Weighted transformation

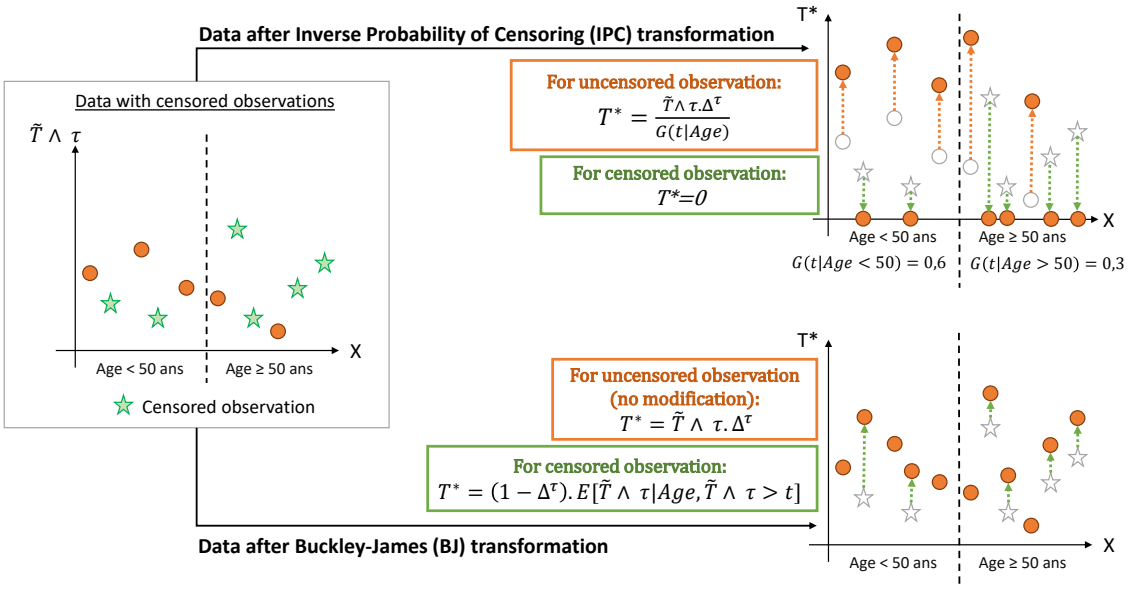


Figure 5: Illustration of Inverse-Probability-of-Censoring and Buckley-James transformations.

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

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

where we recall that $G(t|X, A) := \mathbb{P}(C \geq t|X, A)$ 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 2.5. *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 2.5 is in Section 7.2. 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 2.2.1 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

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

or

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

where we recall that $n_a := \#\{i \in [n] \mid A_i = a\}$. By Proposition 2.3, Proposition 2.4 and Proposition 2.5, we easily deduce that $\hat{\theta}_{\text{IPCW}}$ is asymptotically consistent as soon as $\widehat{G}$ is.

Corollary 2.1. *Under Assumptions1, 5, 6, 16 and 17, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\hat{\theta}_{\text{IPCW}}$, either as in defined in Equation 27 or in Equation 28.*

This result simply comes from the fact that $\hat{\theta}_{\text{IPCW}}$ 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 2.2. (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

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

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:

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

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

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

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

Proposition 2.6. *Under Assumptions 1, 5, 6, 16 and 17, and for all $t \in [0, \tau]$, the oracle estimator $S^*_{\text{IPCW}}(t|A = a)$ defined as in Definition 2.2 with $\widehat{G} = G$ is a stronlgy consistent and asymptotically normal estimator of $S^{(a)}(t)$.*

The proof of Proposition 2.6 can be found in Section 7.2. Because the evaluation of $N_k^{\text{IPCW}}(a)$ 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 2.2. *Under Assumptions 1, 5, 16 and 17, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\hat{S}_{\text{IPCW}}(t|A = a)$.*

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:

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

where, for $t \leq \tau$,

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

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 2.7. *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 2.7 can be found in Section 7.2. 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 2.2.1 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

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

or

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

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

Corollary 2.3. *Under Assumptions 1, 5, 16 and 17, if $\widehat{Q}_S$ is uniformly weakly (resp. strongly) consistent then so is $\hat{\theta}_{\mathrm{BJ}}$ defined as in Equation 31 or Equation 32.*

The proof is again a mere application of Propositions 2.3, 2.4 and 2.7, 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 2.1. *For any transformation T^* of the form 20, it holds*

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

This result is stated in Fan and Gijbels (1994) but without a proof. We detail it in Section 7.2 for completeness.

2.2.3 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 $\hat{S}$ (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),

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

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 2.8. *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.

3 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

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

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.

3.1 Independent censoring

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

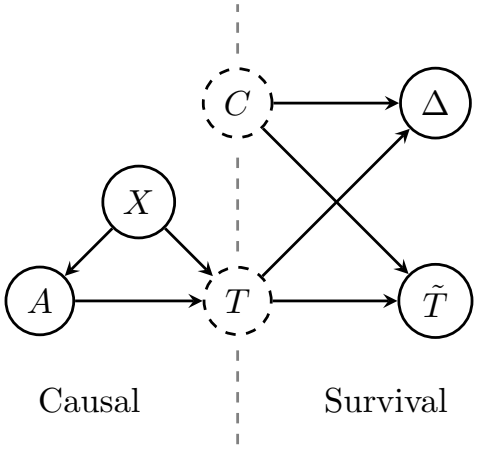


Figure 6: Causal graph in observational survival data with independent censoring.

Under Assumption 8, we saw in Section 2.1 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 3.1. (IPTW Kaplan-Meier estimator) We let $\hat{e}(\cdot)$ be an estimator of the propensity score $e(\cdot)$. We introduce

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

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:

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

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

Proposition 3.1. *Under Assumptions 1, 34, 35, 8 and 9 The oracle IPTW Kaplan-Meier estimator $S^*_{\text{IPTW}}(t|A=a)$ 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 δ-method and can be found in Section 7.3. Because now S^*_{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 3.1. *Under the same assumptions as Proposition 3.1, if $\hat{e}(\cdot)$ satisfies $\|\hat{e} - e\|_\infty \to 0$ almost surely (resp. in probability), then the IPTW Kaplan-Meier estimator $\hat{S}_{\text{IPTW}}(t|A=a)$ is a strongly (resp. weakly) consistent estimator of $S^{(a)}(t)$.*

The resulting RMST estimator simply takes the form:

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

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 $S^*_{\text{IPTW}}(t|A=a)$, and we refer to Xie and Liu (2005) for heuristics concerning these questions.

3.2 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 2.2.1.

A natural alternative approach is to weight the IPCW and BJ transformations from Section 2.2.2 by the propensity score to disentangle both confounding effects and censoring at the same time.

We mention that the G-formula approach developed in Section 2.2.1 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.

3.2.1 IPTW-IPCW transformations

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

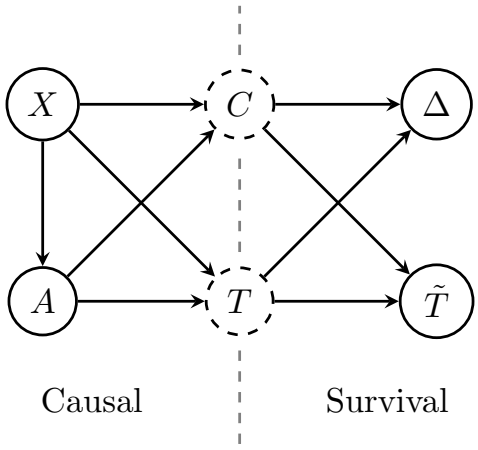


Figure 7: Causal graph in observational survival data with dependent censoring.

Proposition 3.2. *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 3.2 can be found in Section 7.4. 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

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

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

Proposition 3.3. *Under Assumptions 1, 34, 35, 16 and 17, and if $\widehat{G}(\cdot|X, A)$ and $\hat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then estimator 40 defined with $\hat{e}$ and $\widehat{G}$ is a weakly (resp. strongly) consistent estimator of the difference in RMST.*

The proof of Proposition 3.3 can be found in Section 7.4. We can also use the same strategy as for the IPCW transformation and incorporate the new weights into a Kaplan-Meier estimator.

Definition 3.2. (IPTW-IPCW-Kaplan-Meier) We let again $\widehat{G}(\cdot|X, A)$ and $\hat{e}(\cdot)$ be estimators of the conditional censoring survival function and of the propensity score. We introduce

$$D_k^{\text{IPTW–IPCW}}(a) := \sum_{i=1}^{n}\left(\frac{A_i}{\hat{e}(X_i)} + \frac{1-A_i}{1-\hat{e}(X_i)}\right)\frac{\Delta_i^\tau}{\widehat{G}(\tilde{T}_i \wedge \tau|X_i, A=a)}\mathbb{I}(\tilde{T}_i = t_k, A_i = a)$$

$$\text{and} \quad N_k^{\text{IPTW–IPCW}}(a) := \sum_{i=1}^{n}\left(\frac{A_i}{\hat{e}(X_i)} + \frac{1-A_i}{1-\hat{e}(X_i)}\right)\frac{\Delta_i^\tau}{\widehat{G}(\tilde{T}_i \wedge \tau|X_i, A=a)}\mathbb{I}(\tilde{T}_i \geq t_k, A_i = a),$$

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:

$$\hat{S}_{\text{IPTW–IPCW}}(t|A=a) = \prod_{t_k \leq t}\left(1 - \frac{D_k^{\text{IPTW–IPCW}}(a)}{N_k^{\text{IPTW–IPCW}}(a)}\right).$$

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

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

Proposition 3.4. *Under Assumptions 1, 34, 35, 16 and 17, and for all $t \in [0,\tau]$, if the oracle estimator $S^*_{\text{IPTW-IPCW}}(t|A=a)$ defined as in Definition 3.2 with $\widehat{G}(\cdot|A,X) = G(\cdot|A,X)$ and $\hat{e} = e$ is a strongly consistent and asymptotically normal estimator of $S^{(a)}(t)$.*

The proof of Proposition 3.4 can be found in Section 7.4. 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 $\hat{\theta}_{\text{IPTW-IPCW-KM}}$.

Corollary 3.2. *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 $\hat{e}$ are weakly (resp. strongly) uniformly consistent, then $\hat{S}_{\text{IPTW-IPCW}}(t|A=a)$ 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.

3.2.2 IPTW-BJ transformations

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

Proposition 3.5. *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 3.5 can be found in Section 7.4. The corresponding estimator of the difference in RMST is

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

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

Proposition 3.6. *Under Assumptions 1, 34, 35, 16 and 17, and if $\hat{S}(\cdot|X,A)$ and $\hat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then $\hat{\theta}_{\text{IPTW-BJ}}$ defined with $\hat{S}$ and $\hat{e}$ is a weakly (resp. strongly) consistent estimator of the RMST.*

The proof of Proposition 3.6 can be found in Section 7.4.

3.2.3 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 2.2.3 to obtain a *quadruply robuts* tranformation

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

where we recall that T^*_{DR} is defined in Section 2.2.3. This transformations depends on four nuisance parameters: G and S through T^*_{DR}, and now the propensity score e and the conditional response μ. This transformation doesn't really fall into the scope of censoring unbiased transform, but it is easy to show that Δ^*_{QR} is quadruply robust in the following sense.

Proposition 3.7. *Let F, R be two conditional survival function, p be a propensity score, and ν be a conditional response. Then, under the same assumption on F, R as in Proposition 2.8, and under Assumptions 1, 34, 35, 16 and 17, the transformations $\Delta^*_{\text{QR}} = \Delta^*_{\text{QR}}(F, R, p, \nu)$ satisfies, fo a $\in 0, 1$,*

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

This result is similar to Ozenne et al. (2020), Thm 1, and its proof can be found in Section 7.4. Based on estimators $(\widehat{G}, \hat{S}, \hat{\mu}, \hat{e})$ of (G, S, μ, e), one can then propose the following estimator of the RMST, coined the AIPTW-AIPCW estimator in Ozenne et al. (2020):

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

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

4 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 θ_{RMST}.

4.1 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.

4.2 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`).

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	Context	Outcome model	Censoring model	Treatment model	Robustness
Unadjusted KM	RCT + Indep. cens.				
IPCW-KM	RCT + Dep. cens.		G		No
BJ		S			No
IPTW-KM	Obs + Indep. cens.			e	No
IPCW-IPTW-KM	Obs + Dep. cens.		G	e	No
G-formula		μ			No
IPTW-BJ		S		e	No
AIPTW-AIPCW		S, μ	G	e	Yes (Prp 3.7)

- `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 previous `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.

```
source("utilitary.R")
# Kaplan-Meier estimator handmade implementation
```

```r
# The database 'data' must be in the same form as that shown in
# notation (Table 1) and with the same variable names (status, T_obs)
Kaplan_meier_handmade <- function(data,
                                  status = data$status,
                                  T_obs = data$T_obs) {
  # Sort unique observed times
  Y.grid <- sort(unique(T_obs))

  # Initialize vectors for number of events, number at risk, and survival
  # probability
  d <- rep(NA, length(Y.grid))  # Number of events at time Y.grid[i]
  n <- rep(NA, length(Y.grid))  # Number at risk just before time Y.grid[i]
  S <- rep(NA, length(Y.grid))  # Survival probability at time Y.grid[i]

  # Loop over each unique observed time
  for (i in 1:length(Y.grid)) {
    d[i] <- sum(T_obs == Y.grid[i] & status == 1, na.rm = TRUE)  # Count events
    n[i] <- sum(T_obs >= Y.grid[i])  # Count at risk

    # Calculate survival probability
    S[i] <- cumprod(1 - d / n)[i]
  }

  # Create a dataframe with the results
  df <- data.frame(d = d, n = n, S = S, T = Y.grid)

  return(df)
}


# Function to calculate RMST (Restricted Mean Survival Time):
# Two possibilities for computing RMST:
# - in using directly S_A1 and S_A0 (survival function of treated and control)
# - in using the dataframe and the function computes the survival functions
RMST_1 <- function(data = NULL, A1 = 1, A0 = 0, tau, S_A1 = NULL, S_A0 = NULL) {
  if (is.null(S_A1) & is.null(S_A0)) {
    # Subset data for treatment groups
    data1 <- data[data$A == A1,]
    data0 <- data[data$A == A0,]

    # Calculate Kaplan-Meier survival estimates
    S_A1 <- Kaplan_meier_handmade(data1, status = data1$status,
                                  T_obs = data1$T_obs)
    S_A0 <- Kaplan_meier_handmade(data0, status = data0$status,
                                  T_obs = data0$T_obs)

    # Restrict observations to those less than or equal to tau
    Y.grid1 <- data1$T_obs[data1$T_obs <= tau]
    Y.grid0 <- data0$T_obs[data0$T_obs <= tau]
  } else {
```

```
    # Restrict observations to those less than or equal to tau
    Y.grid1 <- S_A1$T[S_A1$T <= tau]
    Y.grid0 <- S_A0$T[S_A0$T <= tau]
  }

  # Filter survival estimates to restricted observations
  S_A1 <- S_A1 %>%
    dplyr::filter(T %in% Y.grid1)
  S_A0 <- S_A0 %>%
    dplyr::filter(T %in% Y.grid0)

  # Check if there is any event at tau for S_A1
  if (!any(S_A1$T == tau)) {
    new_row <- tibble(T = tau, S = S_A1$S[nrow(S_A1)])
    S_A1 <- dplyr::bind_rows(S_A1, new_row)
  }

  # Check if there is any event at tau for S_A0
  if (!any(S_A0$T == tau)) {
    new_row <- tibble(T = tau, S = S_A0$S[nrow(S_A0)])
    S_A0 <- dplyr::bind_rows(S_A0, new_row)
  }

  # Calculate integrals from 0 to tau of survival probabilities
  intA1 <- integral_rectangles(S_A1$T, S_A1$S)
  intA0 <- integral_rectangles(S_A0$T, S_A0$S)
  RMST1 <- intA1 - intA0

  return(list(RMST=RMST1, intA1=intA1,intA0=intA0))
}
```

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 τ in the corresponding summary function:

```
# Alternative code to estimate Kaplan-Meier estimator with survival package
# instead of handmade KM
RMST_alternative <- function(data, A1 = 1, A0 = 0, tau){
  # Estimate Kaplan-Meier estimator with survfit function on data subset
   # Groupe A = 0
  fit0 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A0,])
  # Groupe A = 1
  fit1 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A1,])

  # Estimate the RMST with rmean
  summary_fit0 <- summary(fit0, rmean = tau)  # RMST for A = 0
  summary_fit1 <- summary(fit1, rmean = tau)  # RMST for A = 1

  # Extract the RMST from the summary objects
  rmst0 <- summary_fit0$table["rmean"][[1]]
  rmst1 <- summary_fit1$table["rmean"][[1]]
```

```
  # Compute the difference of RMST between the two groups
  difference_rmst <- rmst1 - rmst0
return(difference_rmst)
}
```

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.

```
# Kaplan-Meier adjusted
# Times of event
# Failures:  1 if event, 0 if censored
# Variable:  1 if treated, 0 if control
# Weights:  Weight of the individual
adjusted.KM <- function(times, failures, variable, weights = NULL) {
  # Sanity checks
  if (sum(times < 0) > 0) {
    stop("Error: times must be positive")
  }
  if (!is.null(weights) && sum(weights < 0, na.rm = TRUE) > 0) {
    stop("Error: weights must be superior to 0")
  }
  if (sum(failures != 0 & failures != 1) > 0) {
    stop("Error: failures must be a vector of 0 or 1")
  }
  # If 'weights' is NULL, initialize 'w' with ones of the same length as 'times',
  # otherwise use 'weights'
  w <- if (is.null(weights)) rep(1, length(times)) else weights

  # Create a DataFrame 'data' with columns t (times), f (failures),
  # v (stratification variable: often treatment variable), and w (weights)
  data <- data.frame(t = times, f = failures, v = variable, w = w)

  # Remove rows from the DataFrame where the stratification variable is NA
  data <- data[!is.na(data$v),]

  # Initialize an empty DataFrame to store the Kaplan-Meier results
  table_KM <- data.frame(times = NULL, n.risk = NULL, n.event = NULL,
                         survival = NULL, variable = NULL)

  # Loop over each unique value of the stratification variable
  for (i in unique(variable)) {
    # Subset the data for the current stratification variable value
    d <- data[data$v == i,]

    # Create a sorted vector of unique event times, including time 0 and the
    # maximum time
```

```
    tj <- c(0, sort(unique(d$t[d$f == 1])), max(d$t))

    # Calculate the number of events at each time point
    dj <- sapply(tj, function(x) {
      sum(d$w[d$t == x & d$f == 1])
    })

    # Calculate the number of individuals at risk at each time point
    nj <- sapply(tj, function(x) {
      sum(d$w[d$t >= x])
    })

    # Compute the cumulative product for the survival probabilities
    st <- cumprod((nj - dj) / nj)

    # Append the results to the Kaplan-Meier table
    table_KM <- rbind(table_KM, data.frame(T = tj, n = nj, d = dj,
                                           S = st, variable = i))
  }
  return(table_KM)
}


# IPCW Kaplan-Meier estimator with restricted tau
IPCW_Kaplan_meier <- function(data, tau,
                              X.names.censoring,
                              nuisance_censoring = "cox",
                              n.folds = NULL) {

  # Compute of truncated T_obs, status and censored status
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)
  data$censor.status_tau <- 1 - as.numeric((data$T_obs >= tau) |
                                             (data$T_obs < tau & data$status == 1))
  data$status_tau <- as.numeric((data$T_obs >= tau) |
                                  (data$T_obs < tau & data$status == 1))
  Y.grid <- sort(unique(data$T_obs_tau))

  # Estimate probability of remaining uncensored based on nuisance model
  S_C_hat <- estimate_survival_function(data = data, X.names = X.names.censoring,
                                        Y.grid = Y.grid, T_obs = "T_obs_tau",
                                        status = "censor.status_tau",
                                        type_of_model = nuisance_censoring,
                                        n.folds = n.folds)

  # Select the probability of censoring for each observed T_obs_tau from the
  # curve
  data$S_C <- S_C_hat$S_hat[cbind(1:nrow(data), match(data$T_obs_tau, Y.grid))]

  # Compute IPC weights
  data$weights <- data$status_tau / data$S_C
```

```
  # Compute the adjusted IPCW Kaplan-Meier
  S <- adjusted.KM(times = data$T_obs, failures = data$status,
                   variable = data$A, weights = data$weights)

  # Compute differenceof RMST between the two groups
  RMST <- RMST_1(S_A1 = S[S$variable == 1,], S_A0 = S[S$variable == 0,], tau = tau)

  return(list(RMST = RMST$RMST,
              intA1 = RMST$intA1,
              intA0 = RMST$intA0,
              weights = data$weights))
}
```

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 τ in the corresponding summary function:

```
# Alternative code to estimate IPCW Kaplan-Meier, IPTW Kaplan-Meier or
# IPTW-IPCW Kaplan-Meier estimator with survival package instead of using
# handmade adjusted.KM function (the weights need to be calculated before).

# Weights0 corresponds to weights of the control and weights1 of treated
Adjusted_Kaplan_meier_alternative <- function(data, A1 = 1, A0 = 0, tau,
                                              weights0, weights1){
  # Estimate Kaplan-Meier estimator with survfit function on data subset
  fit0 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A0,], weights = weights0)  # Grou
  fit1 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A1,], weights = weights1)  # Grou

  # Estimate the RMST with rmean
  summary_fit0 <- summary(fit0, rmean = tau)  # RMST for A = 0
  summary_fit1 <- summary(fit1, rmean = tau)  # RMST for A = 1

  # Extract the RMST from the summary objects
  rmst0 <- summary_fit0$table["rmean"][[1]]
  rmst1 <- summary_fit1$table["rmean"][[1]]

  # Compute the difference in RMST between the two groups
  difference_rmst <- rmst1 - rmst0
return(difference_rmst)
}
```

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 θ_{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`.

```r
# Compute the Restricted Mean Survival Time (RMST) difference
BJ <- function(data, tau, X.names.outcome = c("X1", "X2", "X3", "X4"),
               nuisance = "cox", n.folds = NULL) {
  # Truncate observed times at tau
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)
  Y.grid <- sort(unique(data$T_obs_tau))

  # Censoring status at tau
  data$status_tau <- as.numeric((data$T_obs >= tau) |
                                (data$T_obs < tau & data$status == 1))

  # Compute Q_t for all time points
  Q_t <- Q_t_hat(data, tau, X.names.outcome, nuisance, n.folds)
  data$Q_y <- Q_Y(data, tau, Q_t)

  # Split data by treatment group
  data_treated <- data %>% dplyr::filter(A == 1)
  data_not_treated <- data %>% dplyr::filter(A == 0)

  # Calculate Restricted Survival Time (RST) for each group
  data_treated$RST <- data_treated$status_tau * data_treated$T_obs_tau +
                      (1 - data_treated$status_tau) * data_treated$Q_y

  data_not_treated$RST <- data_not_treated$status_tau * data_not_treated$T_obs_tau +
                          (1 - data_not_treated$status_tau) * data_not_treated$Q_y

  # Calculate RMST difference between treated and not treated
  RMST <- mean(data_treated$RST) - mean(data_not_treated$RST)

  # Return RMST and other relevant metrics
  return(list(
    RMST = RMST,
    ATE_treated = mean(data_treated$RST),
    ATE_not_treated = mean(data_not_treated$RST)
  ))
}
```

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`.

```r
# Function to calculate IPTW Kaplan-Meier
IPTW_Kaplan_meier <- function(data, tau, X.names.propensity,
                              nuisance_propensity = "glm", n.folds = NULL) {
  # Estimate propensity scores
  data$e_hat <- estimate_propensity_score(
    data,
    treatment_covariates = X.names.propensity,
    type_of_model = nuisance_propensity,
```

```
    n.folds = n.folds)

  # Truncate observed times at tau
  data$T_obs_tau <- pmin(data$T_obs, tau)

  # Define censoring status at tau
  data$status_tau <- as.numeric((data$T_obs >= tau) |
                                (data$T_obs < tau & data$status == 1))

  # Calculate weights
  data$weights <- (data$A) * (1 / data$e_hat) + (1 - data$A) / (1 - data$e_hat)

  # Adjusted Kaplan-Meier estimator
  S <- adjusted.KM(
    times = data$T_obs,
    failures = data$status,
    variable = data$A,
    weights = data$weights)

  # Calculate RMST from the adjusted survival curves
  RMST <- RMST_1(S_A1 = S[S$variable == 1,],
                 S_A0 = S[S$variable == 0,],
                 tau = tau)

  return(list("intA0" = RMST$intA0, "intA1" = RMST$intA1, "RMST" = RMST$RMST))
}
```

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.

```
# Function to estimate the g-formula Two-learner.
g_formula_T_learner <- function(data,
                                X.names.outcome,
                                tau,
                                nuisance_survival = "cox",
                                n.folds = NULL) {
  # Compute min(T_obs,tau)
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)

  # Y.grid is the grid of time points where we want to estimate the
  # survival function.
  Y.grid <- sort(unique(data$T_obs_tau))

  S_hat <- estimate_survival_function(data, X.names.outcome,
                                      Y.grid,
```

```
                                        type_of_model = nuisance_survival,
                                        T_obs = "T_obs",
                                        status = "status",
                                        n.folds = n.folds)

  # Compute the area under each survival curve up to max(Y.grid) = tau.
  E_hat1 <- expected_survival(S_hat$S_hat1, Y.grid)
  E_hat0 <- expected_survival(S_hat$S_hat0, Y.grid)

  # Calculate the mean difference.
  theta_g_formula <- mean(E_hat1 - E_hat0)

  return(theta_g_formula)
}

# Function to estimate the g-formula Single-learner.
g_formula_S_learner <- function(data,
                                X.names.outcome,
                                tau,
                                nuisance_survival = "cox",
                                n.folds = NULL) {
  # Compute min(T_obs,tau)
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)

  # Y.grid is the grid of time points where we want to estimate the
  # survival function.
  Y.grid <- sort(unique(data$T_obs_tau))

  S_hat <- estimate_survival_function(data, X.names.outcome,
                                      Y.grid,
                                      type_of_model = nuisance_survival,
                                      learner = "S-learner",
                                      T_obs = "T_obs",
                                      status = "status",
                                      n.folds = n.folds)

  # Compute the area under each survival curve until max(Y.grid) = tau.
  E_hat1 <- expected_survival(S_hat$S_hat1, Y.grid)
  E_hat0 <- expected_survival(S_hat$S_hat0, Y.grid)

  # Calculate the mean difference.
  theta_g_formula <- mean(E_hat1 - E_hat0)

  return(theta_g_formula)
}
```

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.

```r
IPTW_IPCW_Kaplan_meier <- function(data,
                                  X.names.propensity,
                                  X.names.censoring,
                                  tau,
                                  nuisance_propensity = "glm",
                                  nuisance_censoring = "cox",
                                  n.folds = NULL) {
  # Censoring time to tau if observed time exceeds tau
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)

  # Create censoring status for tau
  data$censor.status_tau <- 1 - as.numeric((data$T_obs >= tau) |
                                           (data$T_obs < tau & data$status == 1))

  # Create status at tau
  data$status_tau <- as.numeric((data$T_obs >= tau) |
                                (data$T_obs < tau & data$status == 1))

  # Grid of unique observed times truncated at tau
  Y.grid <- sort(unique(data$T_obs_tau))

  # Estimate propensity scores
  data$e_hat <- estimate_propensity_score(data,
                                          treatment_covariates = X.names.propensity,
                                          type_of_model = nuisance_propensity,
                                          n.folds = n.folds)

  # Estimate survival function for censoring
  S_C_hat <- estimate_survival_function(data, X.names = X.names.censoring,
                                        Y.grid = Y.grid, T_obs = "T_obs_tau",
                                        status = "censor.status_tau",
                                        type_of_model = nuisance_censoring,
                                        n.folds = n.folds)

  # Get estimated survival probabilities for censoring
  data$S_C <- S_C_hat$S_hat[cbind(1:nrow(data), match(data$T_obs_tau, Y.grid))]

  # Calculate weights
  data$weights <- data$status_tau / data$S_C *
                  (data$A * (1 / data$e_hat) +
                     (1 - data$A) * (1 / (1 - data$e_hat)))

  # Compute adjusted Kaplan-Meier estimator
  S <- adjusted.KM(times = data$T_obs,
                   failures = data$status,
                   variable = data$A,
                   weights = data$weights)
```

```
  # Compute Restricted Mean Survival Time (RMST)
  RMST <- RMST_1(S_A1 = S[S$variable == 1, ],
                 S_A0 = S[S$variable == 0, ],
                 tau = tau)

  # Return RMST and ATE for treated and not treated groups
  return(list(RMST = RMST$RMST, ATE_treated = RMST$intA1,
              ATE_not_treated = RMST$intA0))
}
```

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.

```
IPTW_BJ <- function(data,
                    X.names.propensity,
                    X.names.outcome,
                    tau,
                    nuisance_propensity = "glm",
                    nuisance = "cox",
                    n.folds = NULL) {
  # Minimum of T_obs and tau
  data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)

  # Grid of unique observed times truncated at tau
  Y.grid <- sort(unique(data$T_obs_tau))

  # Indicator for min(T, tau) < C
  data$status_tau <- as.numeric((data$T_obs >= tau) |
                                (data$T_obs < tau & data$status == 1))


  # Estimate propensity scores
  data$e_hat <- estimate_propensity_score(data,
                                          treatment_covariates = X.names.propensity,
                                          type_of_model = nuisance_propensity,
                                          n.folds = n.folds)


  # Estimation of Q_s
  Q_t <- Q_t_hat(data, tau, X.names.outcome, nuisance, n.folds)
  data$Q_y <-  Q_Y(data,tau,Q_t)

  # BJ transformation
  data$Y <-  data$status_tau * data$T_obs_tau +
                            (1 - data$status_tau) * data$Q_y
```

```
  # IPTW on BJ transformation
  data$RST <- data$Y * (data$A/data$e_hat-(1-data$A)/(1-data$e_hat))

  RMST <- mean(data$RST)

  # Return RMST
  return(RMST)
}
```

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.

```
# DR censoring transformation
AIPCW <-function(data,
                 tau,
                 X.names.censoring = c("X1","X2","X3","X4"),
                 X.names.outcome = c("X1","X2","X3","X4"),
                 nuisance_Qt = "cox",
                 nuisance_censoring = "cox",
                 n.folds = NULL,
                 h_C_hat = NULL,
                 method_aipw = 1) {

  # Truncate observed times at tau
  data$T_obs_tau <- pmin(data$T_obs, tau)

  # Define status at tau
  data$status_tau <-  as.numeric((data$T_obs > tau) |
                                 (data$T_obs <= tau &  data$status == 1 ))

  data$censor.status_tau <- 1- as.numeric(
    (data$T_obs > tau) | (data$T_obs <= tau &  data$status == 1 ))

  Y.grid <- sort(unique(data$T_obs_tau))

  # Estimate survival function for censoring
  S_C_hat <- estimate_survival_function(data = data,X.names.censoring,
                                        type_of_model = nuisance_censoring,
                                        n.folds = n.folds,
                                        Y.grid = Y.grid,
                                        T_obs = "T_obs_tau",
                                        status = "censor.status_tau")

  Y.index <- findInterval(data$T_obs_tau, Y.grid)

  data$S_C_hat_T_obs_tau <- S_C_hat$S_hat[cbind(seq_along(Y.index), Y.index)]
```

```r
  if (is.null(h_C_hat)) {
      h_C_hat <- estimate_hazard_function(S_C_hat$S_hat,Y.grid)
  }

  # Compute Q.t.hat
  Q.t.hat <- Q_t_hat(data = data,
                     X.names = X.names.outcome,
                     tau = tau,
                     nuisance = nuisance_Qt,
                     n.folds = n.folds)

  # Compute Q.Y.hat
  data$Q.Y.hat <- Q_Y(data = data, tau, Q.t.hat)

  # Compute first term
  data$first_term <- (data$T_obs_tau * data$status_tau) /
    data$S_C_hat_T_obs_tau

  # Compute second term
  data$second_term <- (data$Q.Y.hat * (1 - data$status_tau)) /
    data$S_C_hat_T_obs_tau

  Y.diff <- diff(c(0, Y.grid))

  # Compute integrand for the third term
  integrand <- sweep( ( (h_C_hat) / S_C_hat$S_hat )* (Q.t.hat), 2, Y.diff, "*")

  # Compute third term
  data$third_term <- integrate(integrand, Y.grid, data$T_obs_tau)

  # Compute pseudo outcome
  pseudo_outcome <- data$first_term + data$second_term - data$third_term

  return(pseudo_outcome)
}


AIPTW_AIPCW <- function(data,
                        tau,
                        X.names.propensity = c("X1", "X2", "X3", "X4"),
                        X.names.censoring = c("X1", "X2", "X3", "X4"),
                        X.names.outcome = c("X1", "X2", "X3", "X4"),
                        nuisance_propensity = "glm",
                        nuisance_regression = "cox",
                        nuisance_censoring = "cox",
                        nuisance_Qt = "cox",
                        n.folds = NULL) {

  # Estimate propensity scores
```

```r
data$e_hat <- estimate_propensity_score(
  data = data,
  treatment_covariates = X.names.propensity,
  type_of_model = nuisance_propensity,
  n.folds = n.folds
)

# Prepare data for censoring model
data$T_obs_tau <- ifelse(data$T_obs >= tau, tau, data$T_obs)

data$censor.status_tau <- 1 - as.numeric((data$T_obs >= tau) |
                                          (data$T_obs < tau & data$status == 1))

data$status_tau <- as.numeric((data$T_obs >= tau) |
                                (data$T_obs < tau & data$status == 1))

# Create unique time grid
Y.grid <- sort(unique(data$T_obs_tau))

S_hat <- estimate_survival_function(data, X.names.outcome,
                                    type_of_model = nuisance_regression,
                                    Y.grid = Y.grid,
                                    T_obs= "T_obs",
                                    status = "status",
                                    n.folds = n.folds)

# Compute area under the survival curve up to tau
data$E_hat1 <- expected_survival(S_hat$S_hat1, Y.grid)
data$E_hat0 <- expected_survival(S_hat$S_hat0, Y.grid)

# Compute IPW-weighted residuals
data$IPW_res <- data$E_hat1 * (1 - data$A / data$e_hat) -
  data$E_hat0 * (1 - (1 - data$A) / (1 - data$e_hat))

# Compute AIPCW weights
TDR <- AIPCW(
  data = data,
  tau = tau,
  X.names.censoring = X.names.censoring,
  X.names.outcome = X.names.outcome,
  nuisance_Qt = nuisance_Qt,
  nuisance_censoring = nuisance_censoring,
  n.folds = n.folds
)

data$TDR <- TDR

# Compute AIPCW-weighted residuals
data$AIPCW_w <- data$TDR * (data$A / data$e_hat -
                              (1 - data$A) / (1 - data$e_hat))
```

```r
  # Compute regression residuals
  data$reg <- data$E_hat1 - data$E_hat0
  data$reg_res <- data$A / data$e_hat * (data$TDR - data$E_hat1) -
    (1 - data$A) / (1 - data$e_hat) * (data$TDR - data$E_hat0)

  # Compute estimators
  # na.rm = TRUE to remove NA for the mean calculation
  AIPTW_AIPCW_IPW_res <- mean(data$AIPCW_w + data$IPW_res, na.rm = TRUE)
  AIPTW_AIPCW_reg_res <- mean(data$reg + data$reg_res, na.rm = TRUE)

  return(list(AIPTW_AIPCW_reg_res = AIPTW_AIPCW_reg_res,
              AIPTW_AIPCW_IPW_res = AIPTW_AIPCW_IPW_res))
}
```

4.3 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 τ.

RISCA

The RISCA package provides several methods for estimating θ_{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 2.2.1, 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 τ as inputs. The `average_treatment_effect` function then evaluates the treatment effect based on predictions from the fitted forest.

5 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.

5.1 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 independant of X: $e(X) = 0.5$.
- The threshold time τ is set to 25.

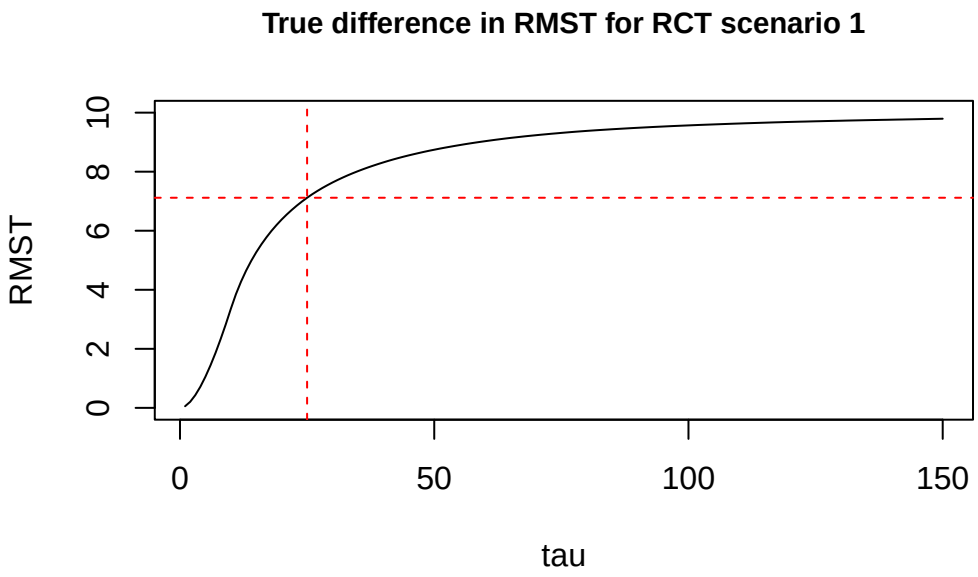
```
############ RCT
# RCT1:  Random treatment assignment + independent censoring
# RCT2:  Random treatment assignment + dependent censoring (conditional on X
# and A)
simulate_data_RCT <- function(n, mu = c(1, 1, -1, 1),
                              sigma = diag(4),
                              colnames_cov = c("X1", "X2", "X3", "X4"),
                              tau,
                              coefT0 = 0.01,
                              parsS = c(0.5, 0.5, -0.5, 0.5),
                              coefC = 0.03,
                              parsC = c(0.7, 0.3, -0.25, -0.1),
                              parsC_A = c(-0.2),
```

```
                              scenario = "RCT2",
                              mis_specification="none") {

# Generate X from a multivariate normal distribution
X <- MASS::mvrnorm(n, mu, sigma)
X <- as.data.frame(X)
colnames(X) <- colnames_cov

# Treatment variable selection: all X
X_treatment <- as.matrix(X)

# Propensity score: constant for random assignment
e <- rep(0.5, n)

# Random treatment assignment
A <- sapply(e, FUN = function(p) rbinom(1, 1, p))

# Outcome variable selection: all X
X_outcome <- as.matrix(X)

# Simulate the outcome using the cumulative hazard inversion method
epsilon <- runif(n, min = 1e-8, max = 1)
T0 <- -log(epsilon) / (coefT0 * exp(X_outcome %*% parsS))

if (scenario == "RCT1") {
  # Simulate independent censoring time
  epsilon <- runif(n, min = 1e-8, max = 1)
  C <- -log(epsilon) / coefC
}
else if (scenario == "RCT2") {
  # Simulate dependent censoring time
  X_censoring <- as.matrix(cbind(X,A))
  parsC <- c(parsC,parsC_A)

  epsilon <- runif(n, min = 1e-8, max = 1)
  C <- -log(epsilon) / (coefC * exp(rowSums(X_censoring %*% diag(parsC))))
}
# T(1) = T(0) + 10
T1 <- T0 + 10

# True survival time
T_true <- A * T1 + (1 - A) * T0

# Observed time
T_obs <- pmin(T_true, C)

# Status indicator
status <- as.numeric(T_true <= C)
censor.status <- as.numeric(T_true > C)
```

```r
  # Restricted survival time
  T_obs_tau <- pmin(T_obs, tau)
  status_tau <- as.numeric((T_obs > tau) | (T_obs <= tau & status == 1))

  # Combine all data into a single data frame
  data_target_population <- data.frame(X, tau, A, T0, T1, C, T_obs, T_obs_tau,
                                       status, censor.status, status_tau, e)

  return(data_target_population)
}
```

The descriptive statistics of the two datasets are displayed in Annex (Section 8.1). The graph of the difference in RMST as a function of τ for the two scenarii are displayed below; θ_{RMST} is the same in both setting.

```r
# Function to calculate ground truth for RCT and Observational data
ground_truth <- function(tau,
                         data) {
  # Compute RMST with the true T1
  data$T1_tau <- ifelse(data$T1 >= tau, tau, data$T1)

  # Compute RMST with the true T0
  data$T0_tau <- ifelse(data$T0 >= tau, tau, data$T0)

  # Compute the difference in RMST if everyone had the treatment
  # and if everyone had the control
  truth <- mean(data$T1_tau) - mean(data$T0_tau)

  return(truth)
}
```




```
[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 4.1. 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 removing 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 θ_{RMST} is indicated by red dotted line.

```
# Update the theme to center the plot title
theme_update(plot.title = element_text(hjust = 0.5))

# Define the desired order of the estimators

desired_order <- c(
  "Naive",
  "KM",
  "SurvRM2 - KM",
  "IPTW KM (Log. Reg.)",
  "RISCA - IPTW KM (Log. Reg.)",
  "IPCW KM (Cox)",
  "BJ (Cox)",
  "IPTW-BJ (Cox & Log. Reg.)",
  "IPTW-IPCW KM (Cox & Log. Reg.)",
  "G-formula (Cox/ T-learners)",
  "G-formula (Cox/ S-learner)",
  "RISCA - G_formula (S-learner)",
  "AIPTW-AIPCW (Cox & Cox & Log. Reg.)",
  "grf - Causal Survival Forest",
  "IPTW KM (Forest)",
  "RISCA - IPTW KM (Forest)",
  "IPCW KM (Forest)",
  "BJ (Forest)",
  "IPTW-BJ (Forest)",
  "IPTW-IPCW KM (Forest)",
  "G-formula (Forest/ T-learners)",
  "G-formula (Forest/ S-learner)",
  "AIPTW-AIPCW (Forest)")

# Convert sample size to a factor with levels sorted in decreasing order
simulation_rct1$sample.size <- factor(
  simulation_rct1$sample.size,
  levels = sort(unique(simulation_rct1$sample.size), decreasing = TRUE)
)

# Convert estimator column to a factor with the specified order
simulation_rct1$estimator <- factor(simulation_rct1$estimator,
                                    levels = desired_order)
```

```r
# Create the plot for RCT + independent censoring
simulation_graph_rct1 <- simulation_rct1 %>%
  ggplot(aes(
    x = estimator, y = estimate,
    fill = factor(sample.size, levels = rev(levels(sample.size)))
  )) +
  scale_fill_brewer(palette = "Accent") +
  geom_boxplot(alpha = 0.9, show.legend = TRUE, position = "dodge") +
  xlab("") +  # Change x-axis label
  ylab("ATE") +  # Change y-axis label
  stat_boxplot(geom = "errorbar") +
  geom_hline(
    yintercept = truth_tau1, linetype = "dashed", color = "red",
    alpha = 0.8, size = 0.8
  ) +
theme(
    legend.title = element_blank(), legend.position = "bottom",
    legend.box = "vertical", legend.text = element_text(size = 18),
    axis.text.x = element_text(angle = 35, vjust = 1, hjust = 1),
    # Adjust text angle for better visibility
    axis.text = element_text(size = 15, face = "bold"),
    axis.title.x = element_text(size = 16, face = "bold"),
    plot.margin = margin(t = 10, r = 10, b = 50, l = 10)  # Add margin
  ) +
  coord_cartesian(ylim = c(0, 15))
```

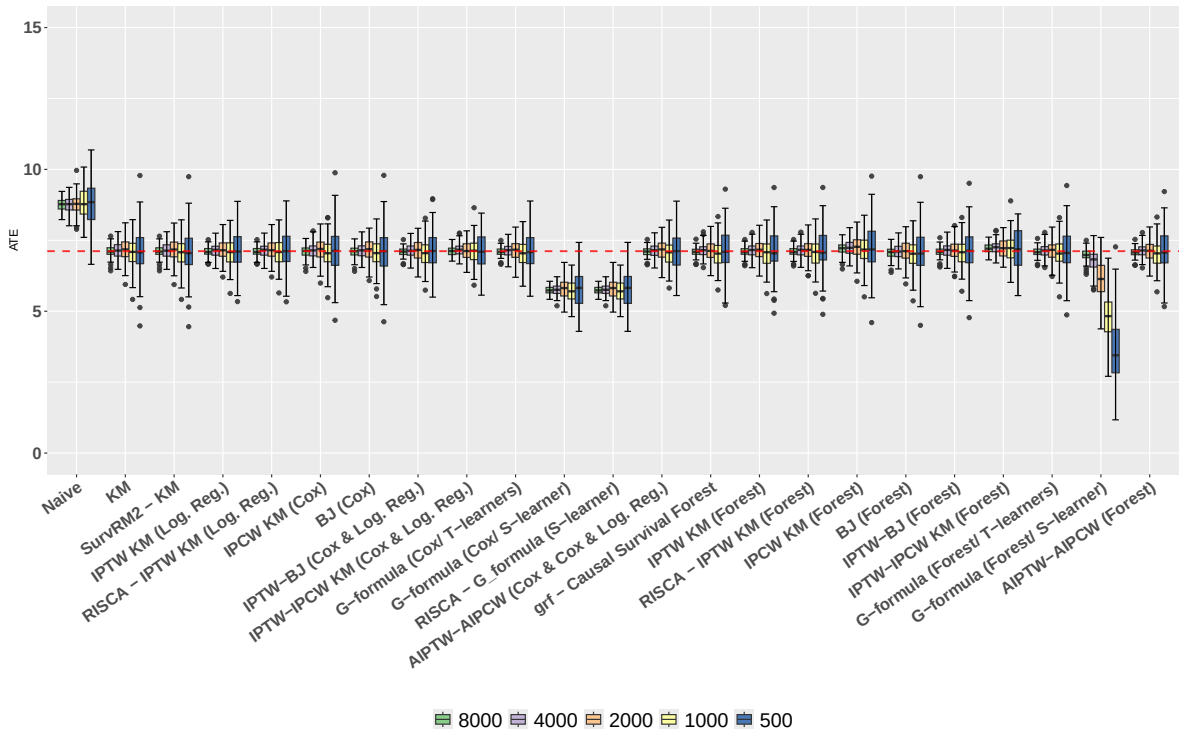


Figure 8: Results of the ATE for the simulation of a RCT with independent censoring.

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 2.1).

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, 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.

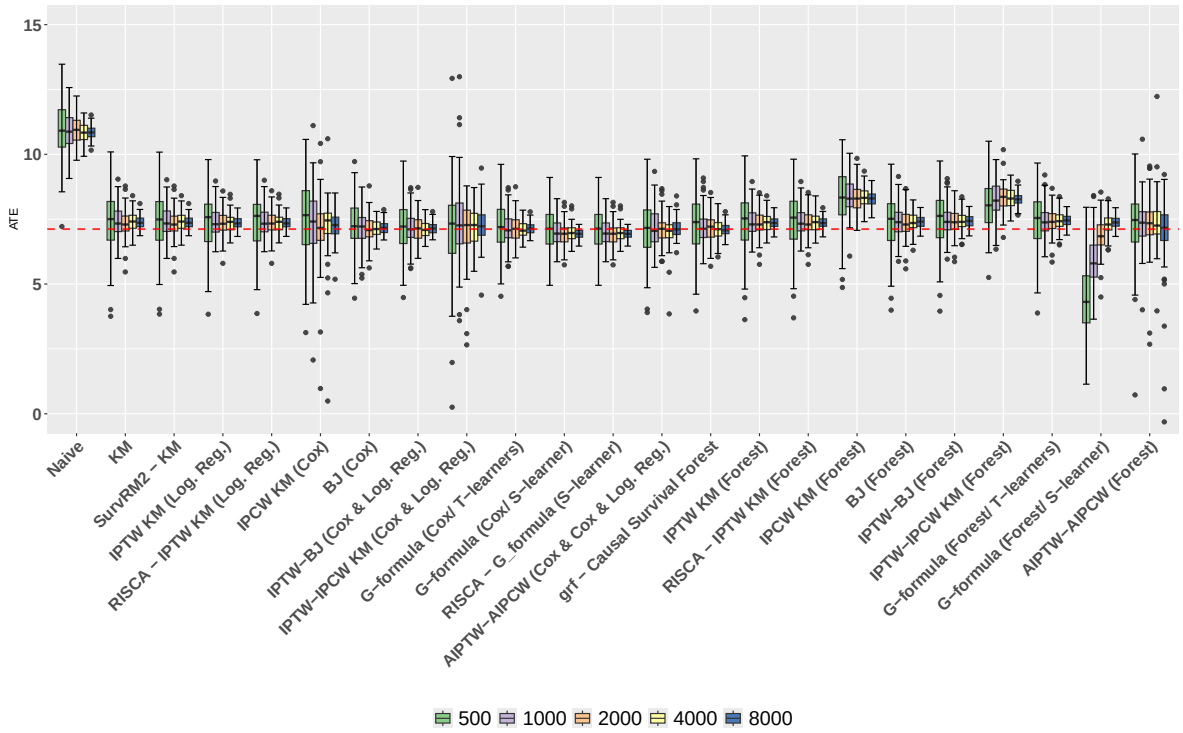


Figure 9: Estimation results of the ATE for the simulation of a RCT with dependent censoring.

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.

5.2 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 5.1, 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 8.2). Note that we did not modify the survival distribution, the target difference in RMST is thus the same.

```
# Obs1:  Treatment assignment dependent on X + independent censoring
# Obs2:  Treatment assignment dependent on X + dependent censoring (conditional
# on X)

# Function to simulate observational data for two scenarios: Obs1 and Obs2
simulate_data_obs <- function(n,
                              mu = c(1, 1, -1, 1),
                              sigma = diag(4),
                              colnames_cov = c("X1", "X2", "X3", "X4"),
                              tau,
                              coefT0 = 0.01,
                              parsS = c(0.5, 0.5, -0.5, 0.5),
                              parsA = c(-1, -1, -2.5, -1),
                              parsC_A = c(0),
                              coefC = 0.03,
                              parsC = c(0.7, 0.3, -0.25, -0.1),
                              scenario = "Obs2") {

  # Generate covariates X from a multivariate normal distribution
  X <- mvrnorm(n, mu, sigma)
  X <- as.data.frame(X)
  colnames(X) <- colnames_cov

  # Propensity score model based on X
  e <- rowSums(as.matrix(X) %*% diag(parsA))
  e <- plogis(e)  # Transform to probability scale

  # Treatment assignment based on the propensity score
  A <- sapply(e, FUN = function(p) rbinom(n = 1, size = 1, prob = p))

  # Outcome model based on X
  X_outcome <- as.matrix(X)
  epsilon <- runif(n, min = 0.00000001, max = 1)
```

```
  T0 <- -log(epsilon) / (coefT0 * exp(X_outcome %*% parsS))

  # Define treatment effect (shift in survival time due to treatment)
  T1 <- T0 + 10

  if (scenario == "Obs1") {
    # Scenario 1: Independent censoring
    C <- -log(runif(n, min = 0.00000001, max = 1)) / coefC

  } else if (scenario == "Obs2") {
    # Scenario 2: Dependent censoring based on X
    X_censoring <- as.matrix(cbind(X,A))
    parsC <- c(parsC,parsC_A)

    C <- -log(runif(n, min = 0.00000001, max = 1)) /
      (coefC * exp(rowSums(X_censoring %*% diag(parsC))))

  } else {
    stop("Invalid scenario. Choose 'Obs1' or 'Obs2'.")
  }

  # Determine the true survival time based on treatment
  T_true <- A * T1 + (1 - A) * T0

  # Observed time is the minimum of the true survival time and censoring time
  T_obs <- pmin(T_true, C)

  # Status indicator: 1 if the event (death) occurred, 0 if censored
  status <- as.numeric(T_true <= C)

  # Restricted survival time (censored at tau)
  T_obs_tau <- pmin(T_obs, tau)
  status_tau <- as.numeric((T_obs > tau) | (T_obs <= tau & status == 1))

  # Compile the simulated data into a data frame
  DATA_target_population <- data.frame(X, tau, A, T0, T1, C, T_obs, T_obs_tau,
                                       status, status_tau, e)

  return(DATA_target_population)
}
```

Estimation of the RMST

Figure 10 below shows the distribution of the estimators of θ_{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

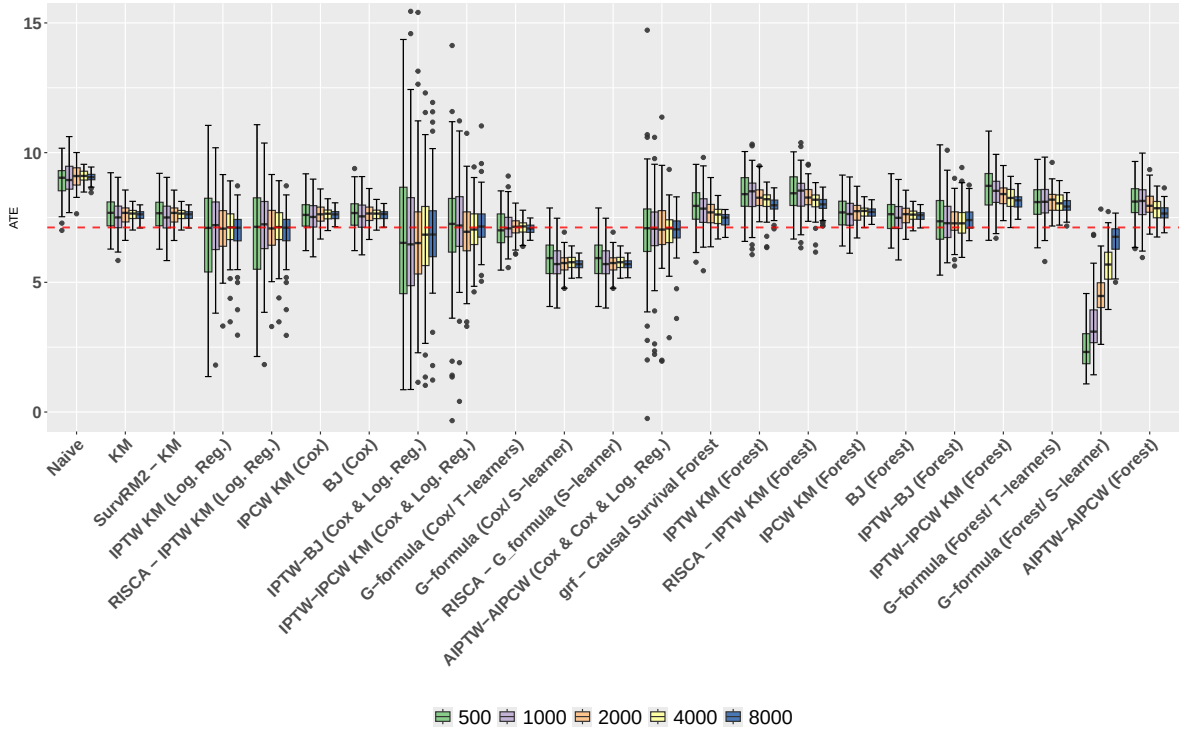


Figure 10: Estimation results of the ATE for the simulation of an observational study with independent censoring.

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 θ_{RMST} estimates for the observational study with conditionally independent censoring. The red dashed line represents the true θ_{RMST} for $\tau = 25$.

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 2.2.1. 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.

5.3 Mispecification of nuisance components

Data Generating Process

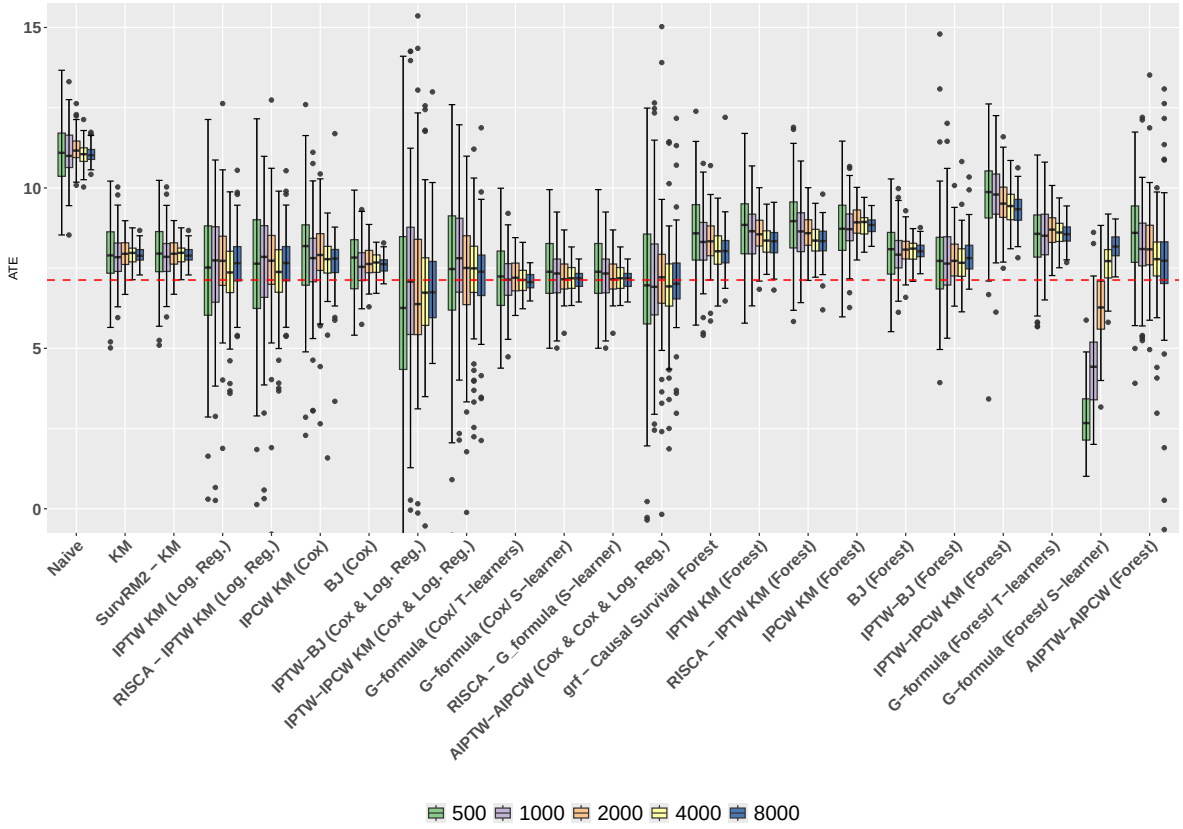


Figure 11: Estimation results of the ATE for the simulation of an observational study with dependent censoring.

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

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

- 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

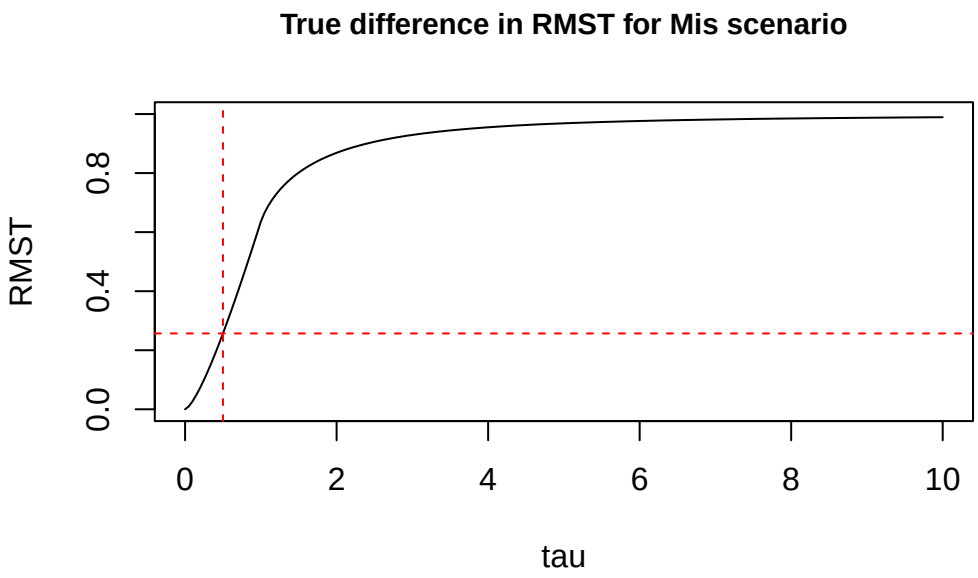
$$\mathrm{logit}(e(x)) = \beta_A^\top Y \quad \text{where} \quad \beta_A = (0.05, -0.1, 0.5, -0.1, 1, 0, 1, 0, 0, 0).$$

  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.

```r
# DGP for misspecification
simulate_data_mis <- function(n,
                              mu = c(0.5, 0.5, 0.7, 0.5),
                              sigma =  matrix(c(1, 0, 0, 0,
                                                0, 1, 0, 0,
                                                0, 0, 1, 0,
                                                0, 0, 0, 1),
                                              nrow = 4, byrow = TRUE),
                              colnames_cov = c("X1", "X2", "X3", "X4"),
                              parsA =  c(0.05, -0.1, 0.5, -0.1),
                              tau){

  # Generate X from a multivariate normal distribution
  X <- MASS::mvrnorm(n, mu, sigma)
  X <- as.data.frame(X)
  colnames(X) <- colnames_cov

  # Treatment variable selection: all X
  X_treatment <- as.matrix(X)

  # Propensity score model based on X
  e <- parsA[1]*X_treatment[, "X1"]^2 + parsA[2]*X_treatment[, "X2"]^2 +
```

```
    parsA[3]*X_treatment[, "X3"]^2 + parsA[4]*X_treatment[, "X4"]^2-
    X_treatment[, "X1"]*X_treatment[, "X2"] +
    X_treatment[, "X1"]*X_treatment[, "X4"]

  # Logistic regression
  e <- plogis(e)

  # Treatment assignment based on the propensity score
  A <- sapply(e, FUN = function(p) rbinom(n = 1, size = 1, prob = p))

  # Outcome variable selection: all X
  X_outcome <- as.matrix(X)

  lambda <- exp(0.2*X[,1]^2 + 0.3*X[,2]^2 + 0.1*X[,3]^2 + 0.1*X[,4]^2 +
    X[,1] * X[,2] + X[,3] * X[,4])
  # Simulate the outcome using the cumulative hazard inversion method
  epsilon <- runif(n, min = 1e-8, max = 1)
  T0 <- -log(epsilon) / lambda

  # Simulate independent censoring time
  censoring_lambda <- exp(0.05*X[,1]^2 + 0.05*X[,2]^2-0.1*X[,3]^2 + 0.1*X[,4]^2 +
    X[,3] * X[,1] - X[,2]*X[,4])
  epsilon <- runif(n, min = 1e-8, max = 1)
  C <- -log(epsilon) / censoring_lambda


  # T(1) = T(0) + 1
  T1 <- T0 + 1

  # True survival time
  T_true <- A * T1 + (1 - A) * T0

  # Observed time
  T_obs <- pmin(T_true, C)

  # Status indicator
  status <- as.numeric(T_true <= C)
  censor.status <- as.numeric(T_true > C)

  # Restricted survival time
  T_obs_tau <- pmin(T_obs, tau)
  status_tau <- as.numeric((T_obs > tau) | (T_obs <= tau & status == 1))
  # Compile the simulated data into a data frame
  DATA_target_population <- data.frame(X, tau, A, T0, T1, C, T_obs, T_obs_tau,
                                       status, status_tau, censor.status, e)

  return(DATA_target_population)
}
```

The descriptive statistics are given in Appendix (Section 8.3).




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

Estimation of the RMST

First, we estimate θ_{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.

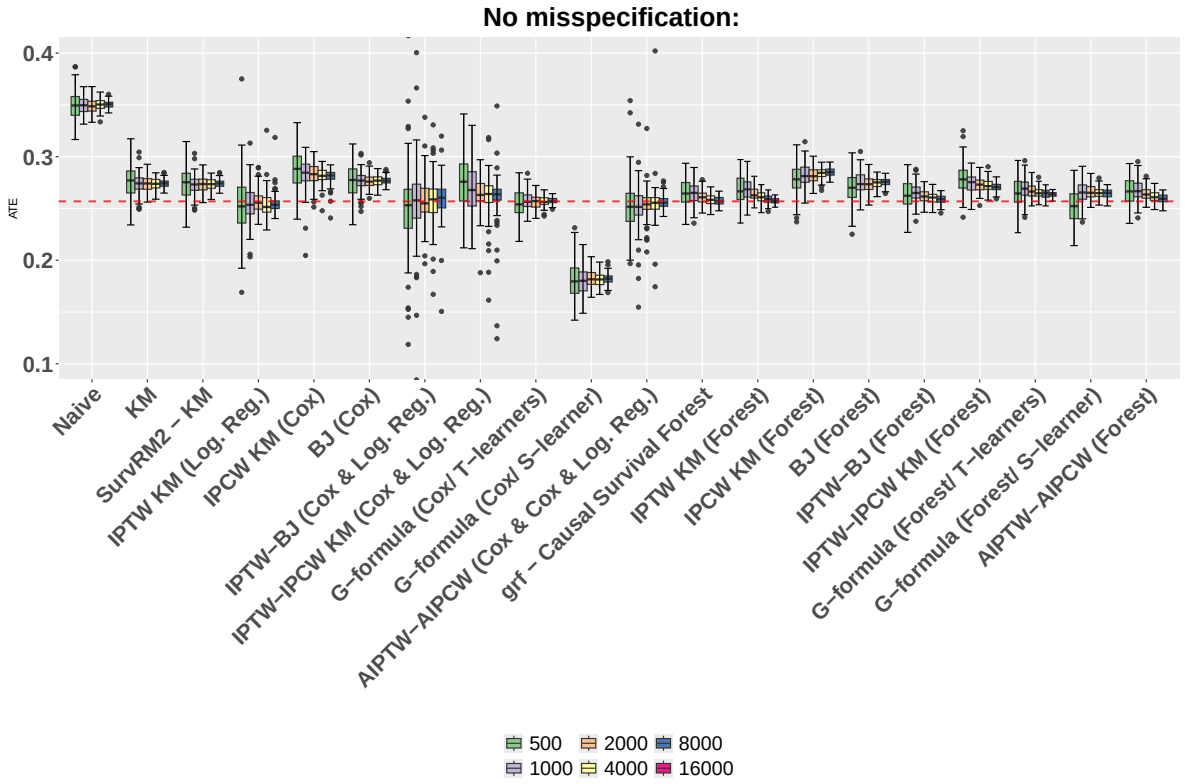


Figure 12: Estimation results of the ATE for the simulation of an observational study with dependent censoring and non linear relationships.

6 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 their Cox 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

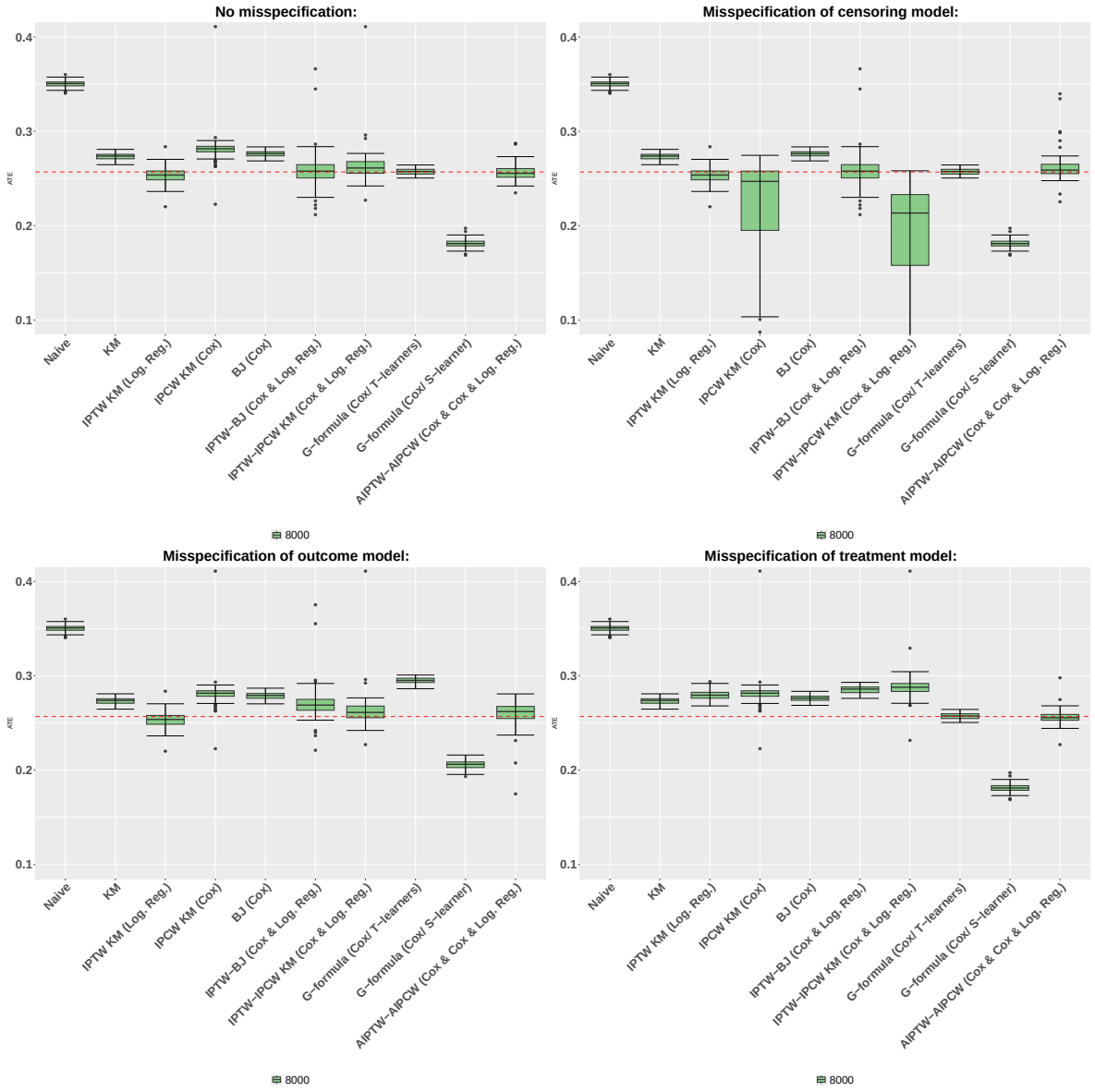


Figure 13: Estimation results of the ATE for an observational study with dependent censoring in case of a single misspecification.

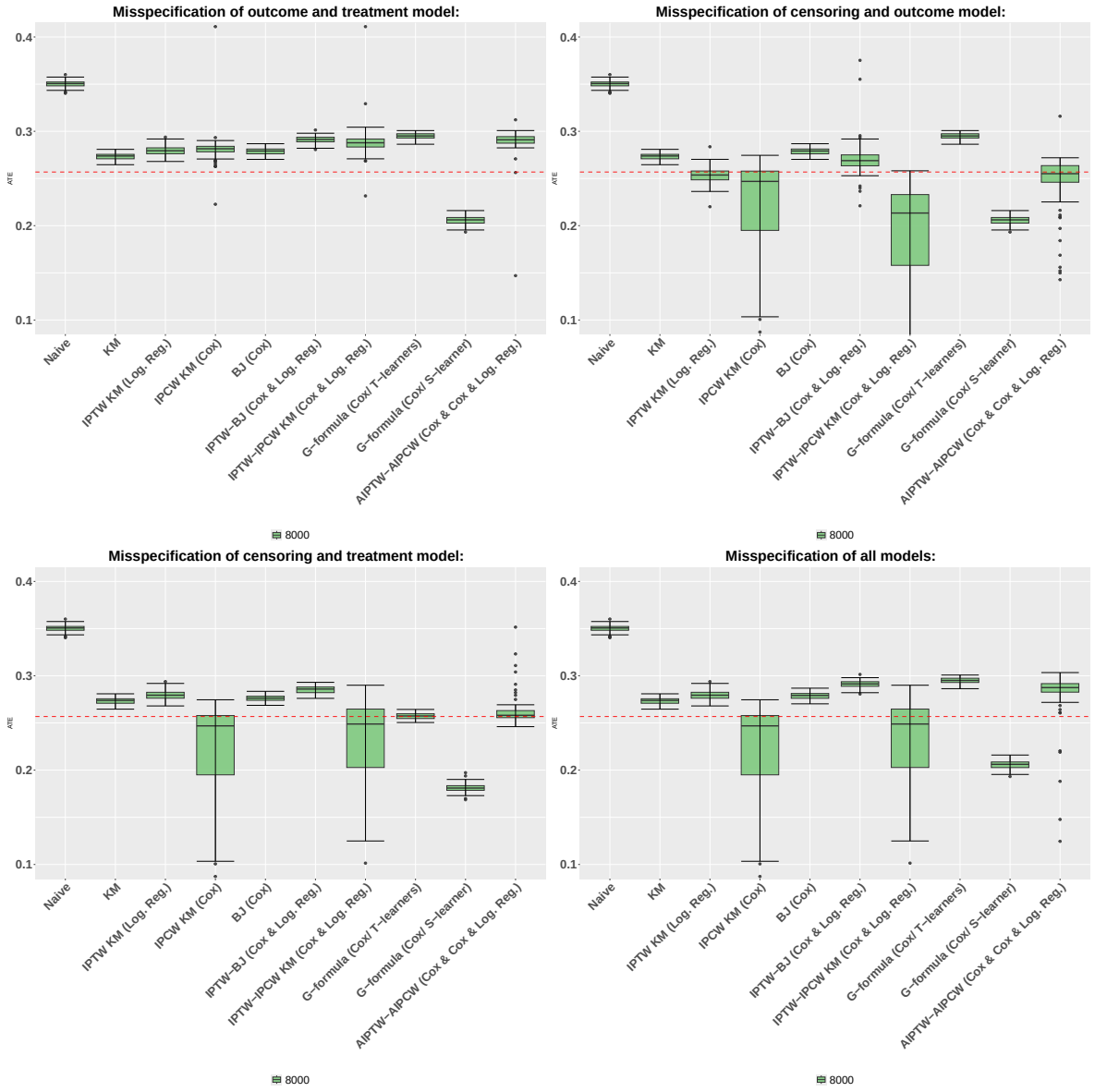


Figure 14: Estimation results of the ATE for an observational study with dependent censoring in case of a two or more misspecifications.

to a deeper understanding of how these estimators can offer more stable and interpretable estimates of treatment effects, particularly in complex real-world datasets.

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.

7 Appendix A: Proofs

7.1 Proofs of Section 2.1

Proof. (Proposition 2.1). Consistency is a trivial consequence of the law of large number and the identity 11. To show that $\hat{S}_{\mathrm{KM}}$ is unbiased, let us introduce $\mathscr{F}_k$ be the filtration generated by the set of variables

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

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

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

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

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

and then that

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

By induction, we easily find that

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

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

Proof. (Proposition 2.2). The asymptotic normality is a mere consequence of the joint asymptotic normality of $(N_k(a), D_k(a))_{t_k \leq t}$ with an application of the δ-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

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

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

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

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

which is what we wanted to show. □

7.2 Proofs of Section 2.2

Proof. (Proposition 2.5). Assumption 17 allows the tranformation to be well-defined. Furthermore, it holds

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

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

Proof. *(Proposition 2.6).* Similarly to the computations done in the proof of Proposition 2.5, it is easy to show that

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

and likewise that

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

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

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

yielding strong consistency. Asymptotic normality is straightforward. □

Proof. *(Proposition 2.7).* There holds

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

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 2.1).* 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 ϕ_0 is a function of $(\tilde{T},X,A)$ and that $\tilde{T}=C$ 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 ϕ_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 RHS is always non-negative, and is zero for the BJ tranformation. □

7.3 Proofs of Section 3.1

Proof. (Proposition 3.1). The fact that it is strongly consistent and asymptotically normal is again a simple application of the law of large number and of the δ-method. We indeed find that, for $t_k \leq \tau$

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

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

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

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

7.4 Proofs of Section 3.2

Proof. (Proposition 3.2). On the event $\{\Delta^\tau = 1, A = 1\}$, there holds $\tilde{T} \wedge \tau = T \wedge \tau = T(1) \wedge \tau$, whence we find that,

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

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

Proof. (Proposition 3.3). By consistency of $\widehat{G}(\cdot|X, A)$ and $\hat{e}$ and by continuity, it suffices to look at the asymptotic behavior of the oracle estimator

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

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

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

□

Proof. *(Proposition 3.4).* Asymptotic normality comes from a mere application of the δ-method, while strong consistency follows from the law of large number and the follozing computations. Like for the proof of Proposition 2.6, one find, by first conditionning wrt $X, A, T(a)$, that, for $t_k \leq \tau$,

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

and likewise that

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

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

Proof. *(Proposition 3.5).* We write

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

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

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

For the second term in the RHS, notice that on the event $\{\Delta^\tau = 0, A=1\}$, there holds $\tilde{T} = C < T(1)\wedge\tau$, 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 3.6).* By consistency of $\widehat{G}(\cdot|X,A)$ and $\hat{e}$ and by continuity, it suffices to look at the asymptotic behavior of the oracle estimator

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

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 3.7).* We can write that

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

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. □

8 Appendix B: Descriptive statistics

8.1 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:   980"
```

```
       X1                X2                X3                X4          
 Min.   :-2.3656   Min.   :-2.8782   Min.   :-3.9454   Min.   :-2.1623  
 1st Qu.: 0.3167   1st Qu.: 0.2922   1st Qu.:-1.6118   1st Qu.: 0.3072  
 Median : 1.0179   Median : 0.9393   Median :-0.9719   Median : 1.0386  
 Mean   : 1.0226   Mean   : 0.9497   Mean   :-0.9681   Mean   : 1.0275  
 3rd Qu.: 1.7362   3rd Qu.: 1.5907   3rd Qu.:-0.2807   3rd Qu.: 1.7090  
 Max.   : 4.1102   Max.   : 4.3237   Max.   : 2.4190   Max.   : 4.4077  
       C                 T1                T0              status      
 Min.   :  0.1045   Min.   : 10.01   Min.   :  0.0075   Min.   :0.0000  
 1st Qu.:  9.3477   1st Qu.: 12.68   1st Qu.:  2.6835   1st Qu.:0.0000  
 Median : 23.0082   Median : 17.80   Median :  7.7952   Median :1.0000  
 Mean   : 32.6855   Mean   : 32.93   Mean   : 22.9342   Mean   :0.6806  
 3rd Qu.: 46.1735   3rd Qu.: 31.97   3rd Qu.: 21.9735   3rd Qu.:1.0000  
 Max.   :270.7802   Max.   :511.37   Max.   :501.3667   Max.   :1.0000  
    T_tild          
 Min.   :  0.00751  
 1st Qu.:  2.07849  
 Median :  5.65401  
 Mean   : 10.46799  
 3rd Qu.: 13.00234  
 Max.   :142.90814  
```

```
[1] "Descriptive statistics for group A=1:   1020"
```

```
       X1                 X2                 X3                 X4          
 Min.   :-2.4748   Min.   :-1.9933   Min.   :-4.4205   Min.   :-2.3556  
 1st Qu.: 0.2580   1st Qu.: 0.2772   1st Qu.:-1.6883   1st Qu.: 0.3147  
 Median : 0.9096   Median : 1.0073   Median :-1.0125   Median : 1.0227  
 Mean   : 0.9298   Mean   : 0.9869   Mean   :-0.9984   Mean   : 0.9969  
 3rd Qu.: 1.6135   3rd Qu.: 1.6621   3rd Qu.:-0.3168   3rd Qu.: 1.6969  
 Max.   : 4.1155   Max.   : 4.9003   Max.   : 2.4107   Max.   : 4.2795  
       C                  T1                 T0               status      
 Min.   :  0.0188   Min.   : 10.00   Min.   :  0.0036   Min.   :0.0000  
 1st Qu.:  9.3258   1st Qu.: 12.92   1st Qu.:  2.9202   1st Qu.:0.0000  
 Median : 22.7248   Median : 18.76   Median :  8.7567   Median :0.0000  
 Mean   : 33.5525   Mean   : 33.16   Mean   : 23.1638   Mean   :0.4853  
 3rd Qu.: 46.1063   3rd Qu.: 33.88   3rd Qu.: 23.8787   3rd Qu.:1.0000  
 Max.   :285.9917   Max.   :784.55   Max.   :774.5498   Max.   :1.0000  
    T_tild          
 Min.   :  0.0188  
 1st Qu.:  9.3258  
 Median : 12.9242  
 Mean   : 16.6361  
 3rd Qu.: 20.5556  
 Max.   :130.6259  
```

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:   1024"
```

```
       X1                 X2                 X3                 X4          
 Min.   :-2.2603   Min.   :-1.9757   Min.   :-4.4690   Min.   :-2.4477  
 1st Qu.: 0.3551   1st Qu.: 0.3269   1st Qu.:-1.6839   1st Qu.: 0.3016  
 Median : 0.9820   Median : 0.9787   Median :-0.9958   Median : 0.9389  
 Mean   : 0.9785   Mean   : 0.9766   Mean   :-0.9968   Mean   : 0.9531  
 3rd Qu.: 1.6186   3rd Qu.: 1.6356   3rd Qu.:-0.3445   3rd Qu.: 1.6510  
 Max.   : 4.4068   Max.   : 4.3839   Max.   : 2.2077   Max.   : 4.9331  
       C                   T1                T0               status      
 Min.   :  0.00345   Min.   : 10.01   Min.   :  0.007   Min.   :0.0000  
 1st Qu.:  2.54411   1st Qu.: 13.12   1st Qu.:  3.124   1st Qu.:0.0000  
 Median :  6.58437   Median : 19.38   Median :  9.380   Median :0.0000  
 Mean   : 13.67323   Mean   : 34.33   Mean   : 24.334   Mean   :0.4297  
 3rd Qu.: 16.37393   3rd Qu.: 34.71   3rd Qu.: 24.715   3rd Qu.:1.0000  
 Max.   :167.13919   Max.   :677.71   Max.   :667.713   Max.   :1.0000  
   status_tau          T_tild          
 Min.   :0.0000   Min.   :  0.00345  
```

```
 1st Qu.:0.0000   1st Qu.:  1.20895
 Median :0.0000   Median :  3.67413
 Mean   :0.4697   Mean   :  7.74005
 3rd Qu.:1.0000   3rd Qu.:  9.07585
 Max.   :1.0000   Max.   :137.56903
```

```
[1] "Descriptive statistics for group A=1:   976"
```

```
       X1                 X2                 X3                 X4
 Min.   :-2.3647   Min.   :-2.7951   Min.   :-3.7863   Min.   :-2.1308
 1st Qu.: 0.3285   1st Qu.: 0.4533   1st Qu.:-1.6386   1st Qu.: 0.2980
 Median : 0.9736   Median : 1.0693   Median :-1.0135   Median : 1.0010
 Mean   : 0.9504   Mean   : 1.0513   Mean   :-0.9787   Mean   : 0.9955
 3rd Qu.: 1.6065   3rd Qu.: 1.6961   3rd Qu.:-0.3198   3rd Qu.: 1.7377
 Max.   : 4.1157   Max.   : 3.9276   Max.   : 2.3075   Max.   : 3.9483
       C                  T1                 T0               status
 Min.   :  0.0064   Min.   : 10.02   Min.   :  0.0161   Min.   :0.0000
 1st Qu.:  3.0802   1st Qu.: 12.81   1st Qu.:  2.8125   1st Qu.:0.0000
 Median :  8.1820   Median : 18.62   Median :  8.6202   Median :0.0000
 Mean   : 18.3713   Mean   : 34.16   Mean   : 24.1625   Mean   :0.2326
 3rd Qu.: 21.4542   3rd Qu.: 33.87   3rd Qu.: 23.8701   3rd Qu.:0.0000
 Max.   :512.9616   Max.   :996.78   Max.   :986.7765   Max.   :1.0000
   status_tau         T_tild
 Min.   :0.0000   Min.   :  0.0064
 1st Qu.:0.0000   1st Qu.:  3.0802
 Median :0.0000   Median :  8.1820
 Mean   :0.2971   Mean   : 12.8878
 3rd Qu.:1.0000   3rd Qu.: 15.0043
 Max.   :1.0000   Max.   :512.9616
```

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.

8.2 Observational study with linear relationship

```
# Observational data with no informative censoring
data_obs1 <- simulate_data_obs(n = 2000, tau = 25, scenario = "Obs1")

# Observational data simulation with dependent censoring
data_obs2 <- simulate_data_obs(n = 2000, tau = 25, scenario = "Obs2",
                               coefC = 0.03, parsC = c(0.7,0.3,-0.25,-0.1))
```

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:   1123"
```

```
       X1                 X2                 X3                 X4          
 Min.   :-1.6361   Min.   :-2.4576   Min.   :-3.05982   Min.   :-1.8151  
 1st Qu.: 0.5548   1st Qu.: 0.4975   1st Qu.:-1.13132   1st Qu.: 0.5994  
 Median : 1.2563   Median : 1.1748   Median :-0.57831   Median : 1.2098  
 Mean   : 1.2379   Mean   : 1.1873   Mean   :-0.52982   Mean   : 1.2495  
 3rd Qu.: 1.8787   3rd Qu.: 1.8318   3rd Qu.: 0.05711   3rd Qu.: 1.9247  
 Max.   : 4.8759   Max.   : 4.6917   Max.   : 1.94260   Max.   : 4.2679  
       C                  T1                 T0                status      
 Min.   :  0.0124   Min.   :  10.00   Min.   :   0.0008   Min.   :0.0000  
 1st Qu.:  9.2665   1st Qu.:  12.42   1st Qu.:   2.4160   1st Qu.:0.0000  
 Median : 23.6780   Median :  17.64   Median :   7.6412   Median :1.0000  
 Mean   : 33.6367   Mean   :  31.75   Mean   :  21.7456   Mean   :0.6972  
 3rd Qu.: 46.3696   3rd Qu.:  32.23   3rd Qu.:  22.2270   3rd Qu.:1.0000  
 Max.   :386.3407   Max.   :1236.60   Max.   :1226.6041   Max.   :1.0000  
    T_tild          
 Min.   :  0.00084  
 1st Qu.:  1.87894  
 Median :  5.33110  
 Mean   : 10.34272  
 3rd Qu.: 12.52973  
 Max.   :169.53599  

[1] "Descriptive statistics for group A=1:   877"

       X1                 X2                 X3                 X4          
 Min.   :-2.2932   Min.   :-2.2910   Min.   :-4.4865   Min.   :-2.3318  
 1st Qu.: 0.1193   1st Qu.: 0.1227   1st Qu.:-2.1535   1st Qu.: 0.1581  
 Median : 0.7948   Median : 0.7279   Median :-1.6229   Median : 0.7520  
 Mean   : 0.7925   Mean   : 0.7440   Mean   :-1.6170   Mean   : 0.7552  
 3rd Qu.: 1.4620   3rd Qu.: 1.3776   3rd Qu.:-1.0601   3rd Qu.: 1.3920  
 Max.   : 4.5127   Max.   : 3.9116   Max.   : 0.5652   Max.   : 4.2323  
       C                  T1                 T0                status      
 Min.   :  0.07718   Min.   :  10.02   Min.   :   0.0173   Min.   :0.0000  
 1st Qu.:  8.58655   1st Qu.:  13.07   1st Qu.:   3.0662   1st Qu.:0.0000  
 Median : 22.04855   Median :  18.72   Median :   8.7195   Median :0.0000  
 Mean   : 32.48464   Mean   :  33.01   Mean   :  23.0090   Mean   :0.4789  
 3rd Qu.: 44.91589   3rd Qu.:  32.30   3rd Qu.:  22.2992   3rd Qu.:1.0000  
 Max.   :288.59981   Max.   :1193.65   Max.   :1183.6468   Max.   :1.0000  
    T_tild          
 Min.   :  0.07718  
 1st Qu.:  8.58655  
 Median : 12.95647  
 Mean   : 16.40195  
 3rd Qu.: 19.88477  
 Max.   :154.77685  
```

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:   1115"
```

```
       X1                 X2                 X3                 X4          
 Min.   :-2.0569   Min.   :-1.7694   Min.   :-3.1452   Min.   :-2.828  
 1st Qu.: 0.5876   1st Qu.: 0.5641   1st Qu.:-1.0435   1st Qu.: 0.557  
 Median : 1.1991   Median : 1.2337   Median :-0.5226   Median : 1.237  
 Mean   : 1.1876   Mean   : 1.2222   Mean   :-0.4871   Mean   : 1.208  
 3rd Qu.: 1.8542   3rd Qu.: 1.8908   3rd Qu.: 0.1015   3rd Qu.: 1.879  
 Max.   : 4.2152   Max.   : 4.5192   Max.   : 2.2314   Max.   : 3.773  
       C                   T1                T0                status      
 Min.   :  0.00884   Min.   : 10.01   Min.   :  0.0059   Min.   :0.000  
 1st Qu.:  2.36853   1st Qu.: 12.49   1st Qu.:  2.4911   1st Qu.:0.000  
 Median :  6.60106   Median : 18.06   Median :  8.0611   Median :0.000  
 Mean   : 14.31117   Mean   : 29.91   Mean   : 19.9060   Mean   :0.452  
 3rd Qu.: 15.49411   3rd Qu.: 31.65   3rd Qu.: 21.6536   3rd Qu.:1.000  
 Max.   :311.13441   Max.   :542.81   Max.   :532.8064   Max.   :1.000  
   status_tau          T_obs                 e           
 Min.   :0.0000   Min.   :  0.00593   Min.   :0.0000123  
 1st Qu.:0.0000   1st Qu.:  1.16135   1st Qu.:0.0223584  
 Median :0.0000   Median :  3.25955   Median :0.0986912  
 Mean   :0.4906   Mean   :  7.41934   Mean   :0.1968066  
 3rd Qu.:1.0000   3rd Qu.:  8.22174   3rd Qu.:0.3014859  
 Max.   :1.0000   Max.   :184.07543   Max.   :0.9641430  
```

```
[1] "Descriptive statistics for group A=1:   885"
```

```
       X1                  X2                  X3                 X4          
 Min.   :-1.92165   Min.   :-2.04608   Min.   :-4.4119   Min.   :-2.87570  
 1st Qu.: 0.05963   1st Qu.: 0.05623   1st Qu.:-2.1584   1st Qu.: 0.03808  
 Median : 0.74540   Median : 0.77309   Median :-1.5817   Median : 0.71762  
 Mean   : 0.72415   Mean   : 0.73933   Mean   :-1.6208   Mean   : 0.70225  
 3rd Qu.: 1.41800   3rd Qu.: 1.39827   3rd Qu.:-1.0984   3rd Qu.: 1.32295  
 Max.   : 3.48278   Max.   : 3.63049   Max.   : 0.8166   Max.   : 3.24713  
       C                   T1                T0                status      
 Min.   :  0.01812   Min.   : 10.01   Min.   :  0.0149   Min.   :0.0000  
 1st Qu.:  2.74555   1st Qu.: 13.33   1st Qu.:  3.3285   1st Qu.:0.0000  
 Median :  7.60688   Median : 18.19   Median :  8.1880   Median :0.0000  
 Mean   : 15.27302   Mean   : 29.97   Mean   : 19.9722   Mean   :0.2034  
 3rd Qu.: 17.42381   3rd Qu.: 31.30   3rd Qu.: 21.3038   3rd Qu.:0.0000  
 Max.   :209.88370   Max.   :664.89   Max.   :654.8862   Max.   :1.0000  
   status_tau          T_obs                 e          
 Min.   :0.0000   Min.   :  0.01812   Min.   :0.0125  
 1st Qu.:0.0000   1st Qu.:  2.74555   1st Qu.:0.5782  
 Median :0.0000   Median :  7.60688   Median :0.8458  
```

```
 Mean   :0.2554   Mean   : 10.70094   Mean   :0.7479
 3rd Qu.:1.0000   3rd Qu.: 13.39265   3rd Qu.:0.9622
 Max.   :1.0000   Max.   :180.73773   Max.   :0.9999
```

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.

8.3 Observational study with interaction

```
mis <- simulate_data_mis(n=2000,tau=0.5)
summary(mis)
```

```
       X1                 X2                 X3                  X4          
 Min.   :-3.2335   Min.   :-3.0996   Min.   :-2.495303   Min.   :-2.7528  
 1st Qu.:-0.1413   1st Qu.:-0.1353   1st Qu.: 0.001702   1st Qu.:-0.1802  
 Median : 0.4971   Median : 0.5219   Median : 0.702774   Median : 0.5066  
 Mean   : 0.5057   Mean   : 0.4985   Mean   : 0.685353   Mean   : 0.4968  
 3rd Qu.: 1.1681   3rd Qu.: 1.2085   3rd Qu.: 1.327910   3rd Qu.: 1.1519  
 Max.   : 3.7553   Max.   : 3.6116   Max.   : 3.948293   Max.   : 3.7800  
      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.: 0.03598   1st Qu.: 1.036  
 Median :0.5   Median :1.0000   Median : 0.19056   Median : 1.191  
 Mean   :0.5   Mean   :0.5855   Mean   : 0.92662   Mean   : 1.927  
 3rd Qu.:0.5   3rd Qu.:1.0000   3rd Qu.: 0.66966   3rd Qu.: 1.670  
 Max.   :0.5   Max.   :1.0000   Max.   :47.08203   Max.   :48.082  
       C                T_obs              T_obs_tau              status      
 Min.   :   0.000   Min.   : 0.000001   Min.   :0.0000008   Min.   :0.000  
 1st Qu.:   0.139   1st Qu.: 0.048994   1st Qu.:0.0489938   1st Qu.:0.000  
 Median :   0.525   Median : 0.230527   Median :0.2305269   Median :0.000  
 Mean   :   5.869   Mean   : 0.532786   Mean   :0.2656433   Mean   :0.449  
 3rd Qu.:   1.694   3rd Qu.: 0.848952   3rd Qu.:0.5000000   3rd Qu.:1.000  
 Max.   :5220.676   Max.   :25.222266   Max.   :0.5000000   Max.   :1.000  
   status_tau      censor.status          e           
 Min.   :0.0000   Min.   :0.000   Min.   :0.0004177  
 1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.4084926  
 Median :1.0000   Median :1.000   Median :0.5835515  
 Mean   :0.6145   Mean   :0.551   Mean   :0.5829850  
 3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.7941064  
 Max.   :1.0000   Max.   :1.000   Max.   :0.9999594  
```

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:   829"
```

```
       X1                  X2                  X3                  C
 Min.   :-3.2335   Min.   :-2.96161   Min.   :-2.1972   Min.   :   0.000
 1st Qu.:-0.1493   1st Qu.: 0.06245   1st Qu.:-0.1584   1st Qu.:   0.153
 Median : 0.5185   Median : 0.68032   Median : 0.4637   Median :   0.556
 Mean   : 0.5322   Mean   : 0.68568   Mean   : 0.4351   Mean   :   9.947
 3rd Qu.: 1.2051   3rd Qu.: 1.41138   3rd Qu.: 1.0089   3rd Qu.:   1.600
 Max.   : 3.7553   Max.   : 3.50320   Max.   : 3.6948   Max.   :5220.676
       T1                T0                 status            T_obs
 Min.   : 1.000   Min.   : 0.00000   Min.   :0.0000   Min.   : 0.000001
 1st Qu.: 1.027   1st Qu.: 0.02678   1st Qu.:0.0000   1st Qu.: 0.019270
 Median : 1.136   Median : 0.13610   Median :1.0000   Median : 0.078731
 Mean   : 1.715   Mean   : 0.71527   Mean   :0.6912   Mean   : 0.254981
 3rd Qu.: 1.545   3rd Qu.: 0.54460   3rd Qu.:1.0000   3rd Qu.: 0.290261
 Max.   :33.653   Max.   :32.65311   Max.   :1.0000   Max.   :10.710763
   status_tau            e
 Min.   :0.0000   Min.   :0.0004177
 1st Qu.:1.0000   1st Qu.:0.2389636
 Median :1.0000   Median :0.4341506
 Mean   :0.7684   Mean   :0.4147071
 3rd Qu.:1.0000   3rd Qu.:0.5668892
 Max.   :1.0000   Max.   :0.9841576

[1] "Descriptive statistics for group A=1:   1171"

       X1                  X2                  X3                  C
 Min.   :-2.5575   Min.   :-3.0996   Min.   :-2.4953   Min.   :  0.00012
 1st Qu.:-0.1339   1st Qu.:-0.2678   1st Qu.: 0.1806   1st Qu.:  0.13043
 Median : 0.4792   Median : 0.3589   Median : 0.8964   Median :  0.51127
 Mean   : 0.4869   Mean   : 0.3659   Mean   : 0.8625   Mean   :  2.98163
 3rd Qu.: 1.1427   3rd Qu.: 1.0111   3rd Qu.: 1.6001   3rd Qu.:  1.75083
 Max.   : 3.7111   Max.   : 3.6116   Max.   : 3.9483   Max.   :254.44133
       T1                T0                 status            T_obs
 Min.   : 1.000   Min.   : 0.00000   Min.   :0.0000   Min.   : 0.000123
 1st Qu.: 1.044   1st Qu.: 0.04424   1st Qu.:0.0000   1st Qu.: 0.130434
 Median : 1.232   Median : 0.23242   Median :0.0000   Median : 0.511272
 Mean   : 2.076   Mean   : 1.07624   Mean   :0.2775   Mean   : 0.729455
 3rd Qu.: 1.770   3rd Qu.: 0.77015   3rd Qu.:1.0000   3rd Qu.: 1.051934
 Max.   :48.082   Max.   :47.08203   Max.   :1.0000   Max.   :25.222266
   status_tau            e
 Min.   :0.0000   Min.   :0.01723
 1st Qu.:0.0000   1st Qu.:0.55090
 Median :1.0000   Median :0.73045
 Mean   :0.5056   Mean   :0.70212
 3rd Qu.:1.0000   3rd Qu.:0.88633
 Max.   :1.0000   Max.   :0.99996
```

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 θ_{RMST} at time τ.

Session information

```
R version 4.4.1 (2024-06-14)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 20.04.6 LTS

Matrix products: default
BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so;  LAPACK version 3.

locale:
 [1] LC_CTYPE=C.UTF-8       LC_NUMERIC=C           LC_TIME=C.UTF-8
 [4] LC_COLLATE=C.UTF-8     LC_MONETARY=C.UTF-8    LC_MESSAGES=C.UTF-8
 [7] LC_PAPER=C.UTF-8       LC_NAME=C              LC_ADDRESS=C
[10] LC_TELEPHONE=C         LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C

time zone: UTC
tzcode source: system (glibc)

attached base packages:
[1] splines   stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] gridExtra_2.3     rms_6.8-2         Hmisc_5.1-3       MASS_7.3-60.2
 [5] grf_2.3.2         RISCA_1.0.5       mosaicData_0.20.4 ggformula_0.12.0
 [9] dplyr_1.1.4       Matrix_1.7-0      ggplot2_3.5.1     lattice_0.22-6
[13] tune_1.2.1        reticulate_1.39.0 relsurv_2.2-9     date_1.2-42
[17] survRM2_1.0-4     survival_3.6-4

loaded via a namespace (and not attached):
  [1] RColorBrewer_1.1-3   rstudioapi_0.16.0    jsonlite_1.8.8
  [4] shape_1.4.6.1        magrittr_2.0.3       TH.data_1.1-2
  [7] farver_2.1.2         rmarkdown_2.28       vctrs_0.6.5
 [10] base64enc_0.1-3      tinytex_0.53         htmltools_0.5.8.1
 [13] forcats_1.0.0        dials_1.3.0          polspline_1.1.25
 [16] haven_2.5.4          Formula_1.2-5        pROC_1.18.5
 [19] caret_6.0-94         parallelly_1.38.0    htmlwidgets_1.6.4
 [22] plyr_1.8.9           sandwich_3.1-0       zoo_1.8-12
 [25] lubridate_1.9.3      gam_1.22-4           lifecycle_1.0.4
 [28] iterators_1.0.14     pkgconfig_2.0.3      R6_2.5.1
 [31] fastmap_1.2.0        future_1.34.0        digest_0.6.37
 [34] colorspace_2.1-1     furrr_0.3.1          mosaic_1.9.1
 [37] labeling_0.4.3       fansi_1.0.6          yardstick_1.3.1
 [40] timechange_0.3.0     nnls_1.5             compiler_4.4.1
 [43] withr_3.0.1          doParallel_1.0.17    htmlTable_2.4.3
 [46] backports_1.5.0      SuperLearner_2.0-29  lava_1.8.0
 [49] quantreg_5.98        ModelMetrics_1.2.2.2 tools_4.4.1
 [52] foreign_0.8-86       future.apply_1.11.2  nnet_7.3-19
 [55] glue_1.7.0           nlme_3.1-164         grid_4.4.1
 [58] checkmate_2.3.2      cluster_2.1.6        reshape2_1.4.4
```

```
 [61] generics_0.1.3       recipes_1.1.0        gtable_0.3.5
 [64] labelled_2.13.0      class_7.3-22         tidyr_1.3.1
 [67] data.table_1.16.0    hms_1.1.3            rsample_1.2.1
 [70] utf8_1.2.4           foreach_1.5.2        pillar_1.9.0
 [73] stringr_1.5.1        lhs_1.2.0            SparseM_1.84-2
 [76] tidyselect_1.2.1     knitr_1.48           stats4_4.4.1
 [79] xfun_0.47            statmod_1.5.0        hardhat_1.4.0
 [82] mosaicCore_0.9.4.0   timeDate_4032.109    stringi_1.8.4
 [85] DiceDesign_1.10      yaml_2.3.10          workflows_1.1.4
 [88] evaluate_0.24.0      codetools_0.2-20     kernlab_0.9-33
 [91] tibble_3.2.1         cli_3.6.3            rpart_4.1.23
 [94] munsell_0.5.1        Rcpp_1.0.13          globals_0.16.3
 [97] png_0.1-8            parallel_4.4.1       MatrixModels_0.5-3
[100] gower_1.0.1          parsnip_1.2.1        cubature_2.1.1
[103] GPfit_1.0-8          listenv_0.9.1        glmnet_4.1-8
[106] mvtnorm_1.3-1        ipred_0.9-15         scales_1.3.0
[109] prodlim_2024.06.25   ggridges_0.5.6       purrr_1.0.2
[112] rlang_1.1.4          multcomp_1.4-26
```