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
Bernard Sebastien[3] CMEI, Sanofi R&D
Julie Josse[4] Premedical, INRIA
INSERM
Université de Montpellier

Last modified: 2024-11-07

Abstract

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

Contents

1 Introduction **3**

1.1 Notations and definition of the estimand 4

2 Causal survival analysis with a Randomized Control Trial **5**

2.1 Independent censoring 6

2.1.1 Estimation with Kaplan-Meier 7

2.1.2 Estimation with Cox model 8

2.2 Conditionally independent censoring 10

2.2.1 The Inverse-Probability-of-Censoring Weighted transformation 11

2.2.2 The Buckley-James transformation 13

2.2.3 Augmented corrections 14

[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

3 Causal survival analysis with an observational study **15**

3.1 Independent censoring . . . 16

3.2 Conditional independent censoring . . . 17

3.2.1 IPTW-IPCW transformations . . . 17

3.2.2 IPTW-BJ transformations . . . 19

3.2.3 The *G*-formula . . . 19

3.2.4 Double augmented corrections . . . 20

4 Implementation **20**

4.1 Summary of the estimators . . . 20

4.2 Available packages . . . 23

4.3 Unadjusted Kaplan-Meier . . . 25

4.4 IPCW Kaplan-Meier . . . 27

4.5 Buckley-James based estimator . . . 30

4.6 IPTW Kaplan-Meier . . . 31

4.7 G-formula . . . 32

4.8 IPTW-IPCW Kaplan-Meier . . . 33

4.9 IPTW-BJ estimator . . . 34

4.10 AIPTW-AIPCW . . . 35

5 Simulations **39**

5.1 RCT . . . 39

5.1.1 Data Generating Process . . . 39

5.1.2 Estimation of the RMST . . . 42

5.2 Observational study with parametric data generating process . . . 46

5.2.1 Data Generating Process . . . 46

5.2.2 Estimation of the RMST . . . 49

5.3 Observational study with nonlinear relationships and conditionally independent censoring . . . 51

5.3.1 Data Generating Process . . . 51

5.3.2 Estimation of the RMST . . . 53

5.4 Mispecification of nuisance components . . . 53

5.4.1 Data Generating Process . . . 53

5.4.2 Estimation of the RMST . . . 56

5.5 Violation of positivity assumption for censoring . . . 58

6 Conclusion and perspective **59**

References **62**

7 Appendix A: Proofs **64**

7.1 Proofs of Section 2.1 64

7.2 Proofs of Section 2.2 66

7.3 Proofs of Section 3.1 67

7.4 Proofs of Section 3.2 67

8 Appendix B **68**

8.1 Descriptive statistics 68

8.1.1 RCT 68

8.1.2 Observational study with linear relationship 71

8.1.3 Observational study with non-linear relationship 73

8.1.4 Observational study with interaction 74

Session information **76**

[CB: il faudra penser à décommenter le usepackage orcidlink dans computo/partial/include-in-header pour le build final (c'est ce qui me fait bug la compilation de mon côté)]

[CB: todo pour moi-meme: repercuter les modifs sur Δ^τ partout et gerer les cross references pour les hypotheses dans tous les statements / preuve] [CB: jpense il faut parler des poids stabilisés quelque part ça a l'air important]

1 Introduction

Causal survival analysis is a growing field that integrates causal inference (D. B. Rubin 1974; Hernán and Robins 2010) with survival analysis 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. The most common form of censoring is right-censoring, where the event of interest has not occurred by the last observation, implying it may happen later.

In this field, the causal effect of a treatment is often measured using the Restricted Mean Survival Time (RMST), which offers an intuitive interpretation of the average survival time over a specific period. RMST addresses the limitations of interpreting the hazard ratio as a causal treatment effect (Martinussen 2022; Martinussen and Vansteelandt 2013; Martinussen, Vansteelandt, and Andersen 2020; Hernán 2010), which is often debated due to its potential non-causal nature. Additionally, unlike the hazard ratio, the RMST has the desirable property of being a collapsible measure, meaning that the population effect can be expressed as a weighted average of subgroup effects, with positive weights that sum to 1 (Huitfeldt, Stensrud, and Suzuki 2019; Greenland, Robbins, and Pearl 1999; Pearl 2000).

However, causal survival analysis is still a relatively new domain, and 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

Table 1: Example of survival data with covariates, treatment, the censoring time, the status of censoring and the potential outcomes, the partially observed outcome and the observed outcomes.

ID	Covariates			Treatment	Censoring	Status	Potential outcome		Partially observed 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

available methods, outlines the underlying assumptions, and provides an evaluation of estimator performance-—particularly in finite sample settings.

In this paper, we begin by presenting the necessary notations in Section 1.1. We then detail the identifiability assumptions and available estimators within the context of both randomized trials (Section 2) and observational data (Section 3). We give their statistical properties (consistency, asymptotic normality) and complete the proofs when there were missing. Finally, in Section 5, we conduct a numerical comparison of these estimators through simulations under various conditions, including independent and conditionally independent censoring, correct and incorrect model specifications, and violations of positivity assumptions. We conclude with practical recommendations on estimator selection, based on criteria such as convergence behavior, computational complexity, and efficiency.

1.1 Notations and definition of the estimand

[CB: faire une section 1.2 si on fait un section 1.1]

Let's consider (X_i, A_i, C_i, T_i) a sample of n i.i.d. realizations of the quadruplet (X, A, C, T) where $X \in \mathbb{R}^p$ denotes the baseline covariates, $A \in \{0, 1\}$ the binary treatment, $C \in \mathbb{R}^+$ the censoring time and $T \in \mathbb{R}^+$ the survival time.

We consider the potential outcome framework by D. B. Rubin (1974) and note $T(0)$ the survival time to the event of interest had the patient received control and $T(1)$ the survival time to the event of interest had the patient received treatment. 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)

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

However, due to censoring, the outcome T is not completely observed (Turkson, Ayiah-Mensah, and Nimoh 2021) and one only observes $\tilde{T} = T \wedge C = \min(T, C)$. When an observation is censored, the observed time is equal to the censoring time. The censoring considered is type II censoring, also known as right censoring. We further introduce the notation $\Delta = I\{T \leq C\}$ the status of censoring, where $I\{\cdot\}$ is the indicator function. The observed and non observed data is represented 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 (Chen and Tsiatis 2001).

Definition 1.1 (Causal effect: Difference between Restricted Mean Survival Time[5]).

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

with $T \wedge \tau = \min(T, \tau)$, the truncated partial outcome at τ, with τ a fixed time horizon.

Let us note $S^{(a)}(t) := \mathbb{P}(T(a) > t)$ for $a \in \{0, 1\}$ the survival curves, 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 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. The difference in RMST, illustrated in Figure 1, is a time-dependent measure that varies with the value of τ.

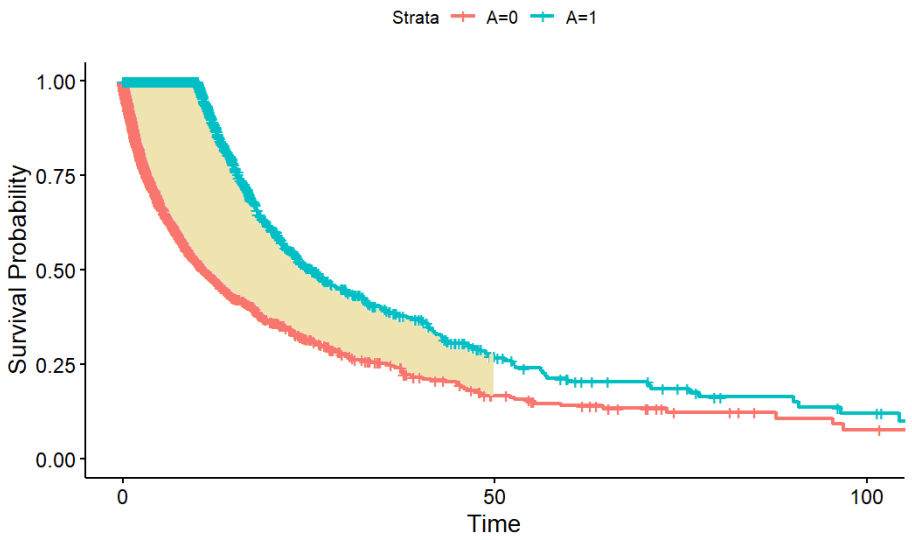


Figure 1: Plot of Kaplan-Meier survival curve for treated and control. The θ_{RMST} at $\tau = 50$ corresponds to the yellow shaded area between the two survival curves.

2 Causal survival analysis with a Randomized Control Trial

Randomized clinical 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).

[5]The causal effect can be also measured as the difference of the survival probability between treated and control (Ozenne et al. 2020).

Assumption. (Random treatment assignment) There holds:

$$A \perp\!\!\!\perp (T(0), T(1), X) \tag{3}$$

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

Assumption. (Trial positivity)

$$0 < \mathbb{P}(A = 1) < 1 \tag{4}$$

Under Equation 3 and Equation 4, classical causal identifiability equations can be written to express θ_{RMST} without potential outcomes.

$$\begin{aligned} \theta_{\text{RMST}} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\ &= \mathbb{E}[T(1) \wedge \tau | A = 1] - \mathbb{E}[T(0) \wedge \tau | A = 0] \quad \text{(Random treatment assignment)} \\ &= \mathbb{E}[T \wedge \tau | A = 1] - \mathbb{E}[T \wedge \tau | A = 0]. \quad \text{(SUTVA)} \end{aligned} \tag{5}$$

However, Equation 5 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 [CB: todo]).

2.1 Independent censoring

The most well-known assumption about censoring in survival analysis is that of independent censoring:

Assumption. (Independent censoring)

$$C \perp\!\!\!\perp (T(0), T(1), X, A) \tag{6}$$

Under Equation 6, subjects censored at time t are representative of all subjects who remain at risk at time t. It is as if the censored subjects were randomly selected from all subjects. 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

$$G(t) \geq \varepsilon \quad \text{for all} \quad t \in [0, \tau). \tag{7}$$

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. The two Assumptions 6 and 7 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 of which being the Kaplan-Meier product limit estimator.

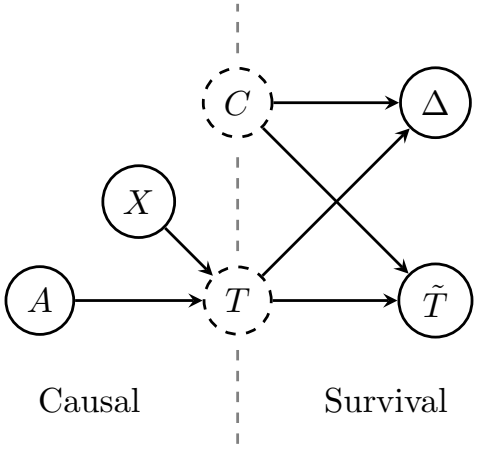


Figure 2: Causal graph in RCT survival data with independent censoring.

2.1.1 Estimation with Kaplan-Meier

To motivate the definition of the latter and explicit the identifiability identity, let us 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.

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

Using Assumptions 6 and 7, we find that

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

yielding the final identity

$$S(t|A=a) = \prod_{t_k \leq t} \left(1 - \frac{\mathbb{P}(\tilde{T}=t_k, \Delta=1, A=a)}{\mathbb{P}(\tilde{T} \geq t_k, A=a)}\right). \tag{8}$$

This last equation suggests in turn to introduce the quantities

$$D_k(a) := \sum_{i=1}^{n} I(\tilde{T}_i = t_k, \Delta_i = 1, A=a) \quad \text{and} \quad N_k(a) := \sum_{i=1}^{n} I(\tilde{T}_i \geq t_k, A=a), \tag{9}$$

which correspond respectively to the number of deaths (D_k) and of individuals at risk (N_k) 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 9, we let

$$\hat{S}_{\text{KM}}(t|A=a) := \prod_{t_k \leq t} \left(1 - \frac{D_k(a)}{N_k(a)}\right). \tag{10}$$

The assiociated RMST estimator is then simply defined as

$$\hat{\theta}_{\mathrm{KM}} = \int_0^\tau \hat{S}_{KM}(t|A=1) - \hat{S}_{KM}(t|A=0)dt. \tag{11}$$

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, 3, 4, 6 and 7, 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 7 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, 3, 4, 6 and 7, 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):

$$\widehat{\mathrm{Var}}\left(\hat{S}_{\mathrm{KM}}(t|A=a)\right) := \hat{S}_{\mathrm{KM}}(t|A=a)^2 \sum_{t_k \leq t} \frac{D_k(a)}{N_k(a)(N_k(a) - D_k(a))}. \tag{12}$$

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.

2.1.2 Estimation with Cox model

Building on Equation 5, we can also express the RMST as a function of the conditionnal response:

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

If we were provided with an estimator $\hat{\mu}(x, a)$ of

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

we could consider the following plug-in estimator of the RMST, also known as the G-formula:

$$\hat{\theta}_{\text{G-formula}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}\left(X_i, 1\right) - \hat{\mu}\left(X_i, 0\right). \tag{13}$$

The Cox proportional hazards model (Cox 1972) is one of the most widely used estimators for conditional response in survival analysis. This model assumes that the hazard function [CB: need to properly define the hazard function somewhere before] at time t is associated with baseline covariates X as follows:

$$\lambda(t|A=0,X) = \lambda_0(t)e^{\beta_0^T X},$$

where $\lambda_0(.)$ is an unspecified baseline hazard function, and β_0 is a $p\times 1$ vector of unknown regression parameters. Then, the conditional survival curve can be expressed as:

$$S(t \mid Z) = \mathrm{e}^{-\Lambda_0(t)\exp(\beta_0^T X)}$$

The estimator of the cumulative hazard function $\hat{\Lambda}_0(t)$ and $\hat{\beta}_0$ can be obtained by using the Maximum of Likelihood Estimator and the Breslow estimator (approximation of the Maximum of Likelihood in proportional hazards regression) (Breslow 1974).

Figure 3 illustrates the estimation of the difference in Restricted Mean Survival Time using G-formula.

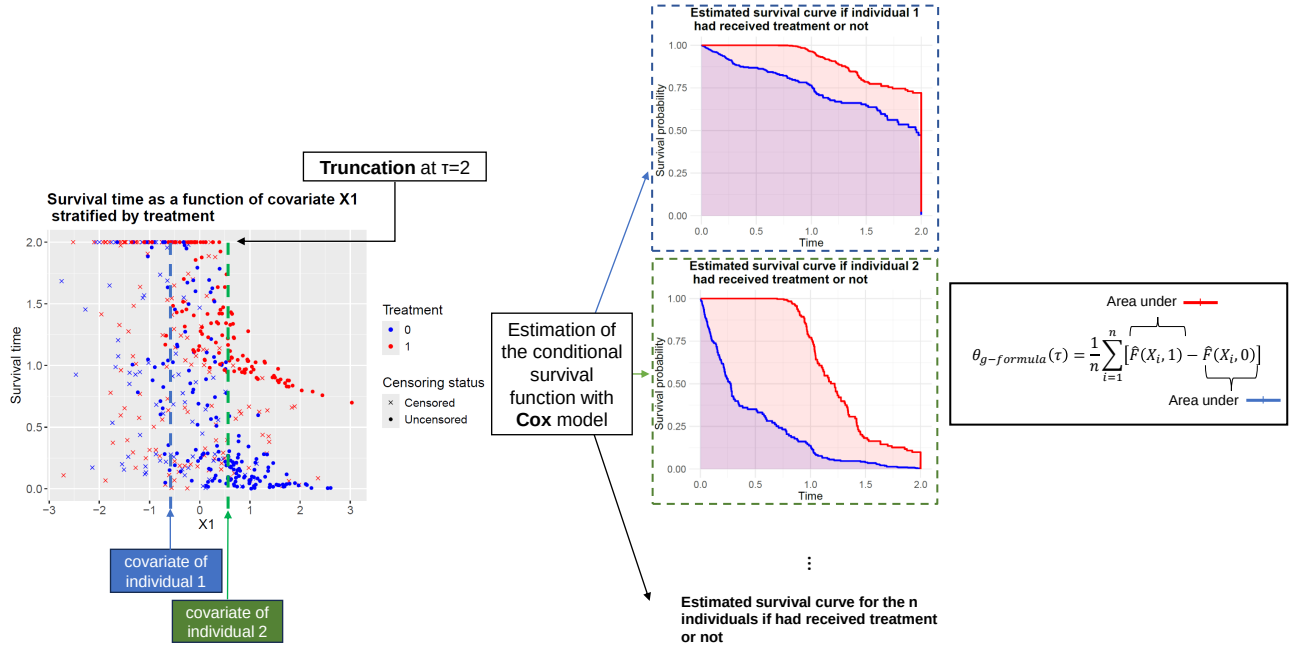


Figure 3: Illustration of the G-formula for estimating θ_{RMST} in an RCT when only one covariates X_1 influences the outcome

It is possible to use other methods to estimate the conditional survival function such as parametric survival models (for example a Weibull model) or non-parametric models such as survival forests (Ishwaran et al. 2008).

In causal inference, the goal is to estimate the parameter of interest here θ_{RMST}. To achieve this, it is necessary to estimate intermediate quantities, such as the conditional survival function in this case. These intermediate quantities are referred to as nuisance parameters. [CB: c'est à developper?]

[CB: ce serait bien de state des results dans cette sous-section]

2.2 Conditionally independent censoring

An alternative hypothesis in survival analysis that relaxes the assumption of independent censoring is conditionally independent censoring.

Assumption. (Conditionally independent censoring)

$$C \perp\!\!\!\perp (T(0), T(1))|X, A \tag{14}$$

Under Equation 14, 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

$$G(t|A, X) \geq \epsilon. \tag{15}$$

This assumption ensures that, for any time $t \leq \tau$, the probability of censoring within subgroups is not equal to 1, allowing for balance in the censoring mechanism [IM: do you mean non-empty risk sets at any time t? I'm not sure what is meant by balanced censoring mechanism]. If violated, it means that results are either only observed after time t, which can limit the analysis [IM: from the formulation at the beginning of the sentence it sounds like there is another possible consequence of the violation].

In practice, adjusting the threshold time τ can help meet this assumption. For example, in a 5-year clinical study, if patients leave due to severe side effects or worsening health, censoring becomes dependent. In such cases, the likelihood of remaining uncensored for severely ill patients at 5 years is zero. To address this, τ can be adjusted so that participants have a chance to remain uncensored up to a revised threshold time.
Figure 4 represents the causal graph when the study is randomized with conditionally independent censoring.

Under Assumptions 14 and 15, the Kaplan-Meier estimator as defined in Definition 2.1 can fail to estimate survival probabilities due to conditionally independent censoring (Willems et al. 2018). One classical strategy to circumvent this effect is to use so-called *censoring unbiased transformation*.

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

$$T^* = \begin{cases} \phi_0(\tilde{T} \wedge \tau, X, A) & \text{if} \quad \Delta^\tau = 0, \\ \phi_1(\tilde{T} \wedge \tau, X, A) & \text{if} \quad \Delta^\tau = 1. \end{cases} \tag{16}$$

for two wisely chosen transformations ϕ_0 and ϕ_1, and where

$$\Delta^\tau := I\{T \wedge \tau \leq C\} = \Delta + (1 - \Delta) I(\tilde{T} \geq \tau) \tag{17}$$

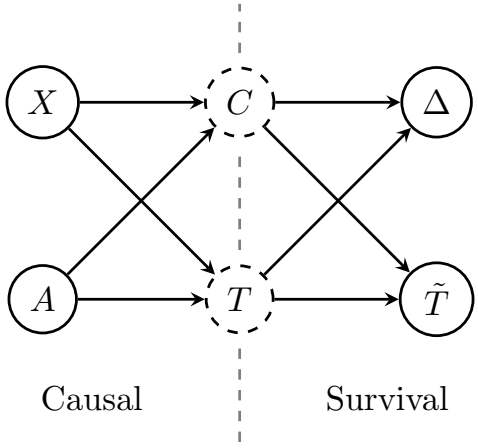


Figure 4: Causal graph in RCT survival data with dependent censoring.

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:

$$\mathbb{E}[T^*|A, X] = \mathbb{E}[T \wedge \tau|A, X] \quad \text{almost surely.} \tag{18}$$

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.

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.

2.2.1 The Inverse-Probability-of-Censoring Weighted transformation

The Inverse-Probability-of-Censoring Weighted (IPCW) transformation involves discarding censored observations and applying weights to uncensored data. More precisely, we define

$$T^*_{\text{IPCW}} = \frac{\Delta^\tau}{G(\tilde{T} \wedge \tau|X, A)}\tilde{T} \wedge \tau, \tag{19}$$

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 the same covariate group, correcting for conditionally independent censoring and reducing selection bias (Howe et al. 2016).

Proposition 2.3. *Under Assumptions 1, 3, 14 and 15, the IPCW transform 19 is a censoring unbiased transformation in the sense of Equation 18.*

The proof of Proposition 2.3 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. We refer to Section Section 2.1.2 for a development regarding the estimation of this quantity. Estimating conditional censoring or the conditional survival function can be approached similarly, as both involve

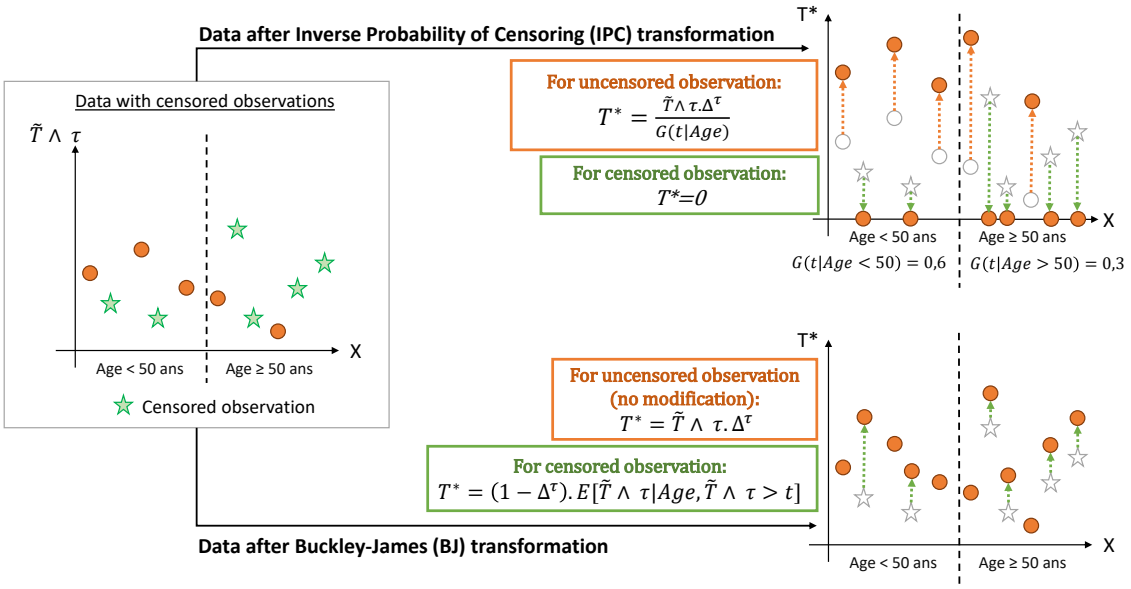


Figure 5: Illustration on Inverse-Probability-of-Censoring and Buckley-James transformation

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. Once an estimator $\widehat{G}(\cdot|A, X)$ of the later is provided, a very natural estimator of the RMST based on the IPCW transformation would be

$$\hat{\theta}_{\text{IPCW}} = \sum_{i=1}^{n} \left(\frac{A_i}{n_1} - \frac{1 - A_i}{n_0} \right) \frac{\Delta_i^\tau}{\widehat{G}(\tilde{T} \wedge \tau | A_i, X_i)} \tilde{T}_i,$$

where $n_a := \#\{i \in [n] \mid A_i = a\}$.

Proposition 2.4. *Under Assumptions1, 3, 14 and 15, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\hat{\theta}_{\text{IPCW}}$.*

The proof of Proposition 2.3 can be found in Section 7.2. 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 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

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

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

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:

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

The subsequent RMST estimator then take the simple form

$$\hat{\theta}_{\text{IPCW-KM}} = \int_0^\tau \hat{S}_{\text{IPCW}}(t|A=1) - \hat{S}_{\text{IPCW}}(t|A=0)dt. \tag{20}$$

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

Proposition 2.5. *Under Assumptions 1, 3, 14 and 15, 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 2.5 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 still achieve consistency and asymptotic normality if the estimator $\widehat{G}(\cdot|A,X)$ that we provide is consistent, as shown in Section 7.2 and stated below.

Proposition 2.6. *Under Assumptions 1, 3, 14 and 15, if $\widehat{G}$ is uniformly weakly (resp. strongly) consistent then so is $\hat{S}_{\text{IPCW}}(t|A=a)$.*

2.2.2 The Buckley-James transformation

One weakness of the IPCW transformation is that it discards all censored data. The Buckley-James (BJ) transformation 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:

$$T^*_{\text{BJ}} = \Delta^\tau(\tilde{T} \wedge \tau) + (1-\Delta^\tau)Q_S(\tilde{T} \wedge \tau|X,A), \tag{21}$$

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.

Proposition 2.7. *Under Assumptions 1, 3, 14 and 15, the BJ transform 21 is a censoring unbiased transformation in the sense of Equation 18.*

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 refer to Sec [CB: todo] 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 is

$$\hat{\theta}_{\mathrm{BJ}} = \sum_{i=1}^{n} \left(\frac{A_i}{n_1} - \frac{1-A_i}{n_0} \right) \left\{ \Delta_i^\tau (\tilde{T}_i \wedge \tau) + (1-\Delta_i^\tau) \widehat{Q}_S(\tilde{T}_i \wedge \tau | X_i, A_i) \right\}. \tag{22}$$

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

Proposition 2.8. *Under Assumptions 1, 3, 14 and 15, if $\widehat{Q}_S$ is uniformly weakly (resp. strongly) consistent then so is $\hat{\theta}_{\mathrm{BJ}}$.*

–>

The proof of this result can be found in Section 7.2. 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 16, 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),

$$T^*_{\mathrm{DR}} = \frac{\Delta^\tau \tilde{T} \wedge \tau}{G(\tilde{T} \wedge \tau | X, A)} + \frac{(1-\Delta^\tau) Q_S(\tilde{T} \wedge \tau | X, A)}{G(\tilde{T} \wedge \tau | X, A)} - \int_0^{\tilde{T} \wedge \tau} \frac{Q_S(t|X,A)}{G(t|X,A)^2} \, \mathrm{d}G(t|X,A). \tag{23}$$

This transformations depends on the knowledge of both conditional survival functions G and S and will be thus sometimes denoted $T^*_{\mathrm{DR}}(G,S)$ to stress this dependency. This transformations is not only a censoring unbiased transform in the cens of Equation 18, but is also doubly robust in the following sense.

Proposition 2.9. *We let J, H be two conditional survival functions. Under Assumptions 1, 3, 14 and 15, if F also satisfies Assumption 15, and if $F(\cdot|X,A)$ is absolutely continuous wrt $G(\cdot|X,A)$, then the transformation $T^*_{\mathrm{DR}} = T^*_{\mathrm{DR}}(F,R)$ satisfy*

$$\mathbb{E}[T^*_{\mathrm{DR}}|X,A] = \mathbb{E}[T \wedge \tau | X, A] \quad \textit{if} \quad F = G \quad \textit{or} \quad R = S.$$

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 with an observational study

The previous estimators are suited for randomized controlled trial (RCT) settings but not for more complex contexts like 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. As a result, the treatment effect is confounded by variables which influence both the time-to-event outcome T and the treatment allocation A. The assumption of randomized treatment assignment, as presented in Equation 3 (see Section 2), is not satisfied in observational studies. To enable the identifiability of the causal estimand, additional assumptions regarding treatment allocation are needed. These assumptions are standard in causal inference methods with observational data and can be extended to identify θ_{RMST}:

Assumption. (Conditional exchangeability / Unconfoundedness) It holds

$$A \perp\!\!\!\perp (T(0), T(1))|X \tag{24}$$

Under Equation 24, the treatment assignment is randomly assigned conditionally on the covariates X. It is as if the treatment for all subjects were randomly selected inside each subgroup. Exactly like Equation 14, this assumption assumes that there are no unmeasured confounders as unobserved confounders make it impossible to distinguish correlation from causality.

Assumption. (Positivity / Overlap for treatment) For $a \in \{0, 1\}$, it holds

$$0 < P(A = a \mid X) < 1 \quad \text{almost surely.} \tag{25}$$

The Equation 25 assumption requires adequate overlap in covariate distributions between treatment groups, meaning every observation must have a non-zero probability of being treated.

In addition to confounding bias, censoring bias must also be addressed, as discussed in Section 2. The censoring mechanism assumptions from Equation 14 and Equation 6 remain applicable. The next section will present the identifiability formula and corresponding estimator when censoring is independent in observational studies.

Because Assumption 1 does not hold anymore, neither does the previous idenfiability Equation 5. Letting the propensity score $e(X) := \mathbb{P}(A = 1|X)$ (which is in $(0, 1)$ almost surely, thanks to Assumption 25), we can write

$$\begin{aligned}
\theta_{\text{RMST}} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\
&= \mathbb{E}\left[\mathbb{E}[T(1) \wedge \tau|X] - \mathbb{E}[T(0) \wedge \tau|X]\right] \\
&= \mathbb{E}\left[\mathbb{E}[T(1) \wedge \tau|X, A = 1] - \mathbb{E}[T(0) \wedge \tau|X, A = 0]\right] \quad \text{(unconfoundness)} \\
&= \mathbb{E}\left[\mathbb{E}[T \wedge \tau|X, A = 1] - \mathbb{E}[T \wedge \tau|X, A = 0]\right]. \quad \text{(SUTVA)}
\end{aligned} \tag{26}$$

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

$$S^{(a)}(t) = \mathbb{P}(T(a) > t) = \mathbb{E}\left[\mathbb{P}(T > t|X, A = a)\right]. \tag{27}$$

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.

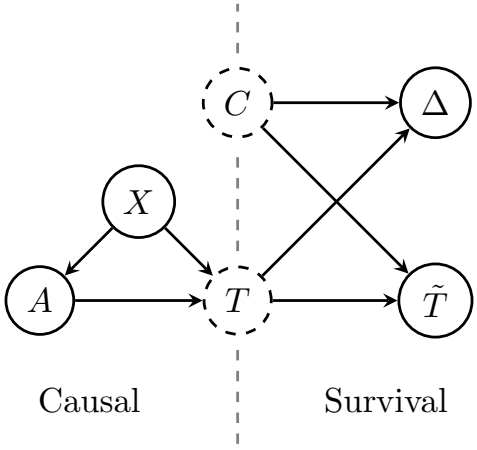


Figure 6: Causal graph in observational survival data with independent censoring (A is the treatment, X the confounding variable and T is the survival time outcome and C is the censoring).

3.1 Independent censoring

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

Under Assumption 6, we saw in Section 2.1 that the Kaplan-Meier estimator could conveniently handle censoring. Building on Equation 27, we can write

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

which suggests to adapt the classical KM estimator by reweighting it by the propensity score, as detailed below. The use of propensity score in causal inference has been introduced by Rosenbaum and Rubin (1983) and extended to survival analysis by Xie and Liu (2005). Propensity scores are often estimated using parametric models such as logistic regression but non parametric methods such as probability forests can be used as well.

Definition 3.1. (IPTW Kaplan-Meier estimator) We let $\hat{e}(\cdot)$ be an estimator of the propensity score $e(\cdot)$. We introduce

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

be the numbers of deaths and of individual at risk at time t_k, reweighted by the propensity score. The IPTW version of the KM estimator is then defined as:

$$\hat{S}_{\text{IPTW}}(t|A=a) = \prod_{t_k \leq t}\left(1 - \frac{D_k^{\text{IPTW}}(a)}{N_k^{\text{IPTW}}(a)}\right). \tag{28}$$

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, 24, 25, 6 and 7 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 $e(\cdot)$, 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* $S_{\text{IPTW}}(t|A=a)$ *is a strongly (resp. weakly) consistent estimator of* $S^{(a)}(t)$.

The resulting RMST estimator simply takes the formn:

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

We are not aware of any formal results concerning the bias and the asymptotic variance of the oracle estimator $S^*_{\text{IPTW}}(t|A=a)$.We refer to Xie and Liu (2005) for heuristics concerning these questions. [CB: c'est mieux?]

3.2 Conditional independent censoring

Under Equation 24 (uncounfoundedness) and Equation 14 (conditional independent censoring), the causal effect is affected both by confounding variables (confounding bias) and by conditional censoring. The associated causal graph is depicted in Figure 7. A very natural approach is to reweight the tranformations previously considered to handle the conditional censoring by the propensity score to disentangle both effects at the same time.

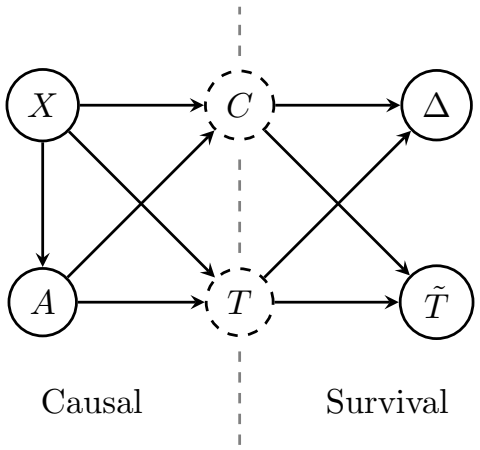


Figure 7: Causal graph in observational survival data with dependent censoring.

3.2.1 IPTW-IPCW transformations

Inspired by the IPW estimators from causal inference (Hirano, Imbens, and Ridder 2003), one can also reweight the observed time by the inverse of the propensity score in a attempt to remove the confounding effect of the covariates. Building on the IPCW tranformation introduced in Section 2.2.1, we would obtain the so-called IPTW-IPCW tranformation:

$$T^*_{\text{IPTW-IPCW}} = \left(\frac{A}{e(X)} + \frac{1-A}{1-e(X)}\right) \frac{\Delta^\tau}{G(\tilde{T} \wedge \tau | X, A)} \tilde{T} \wedge \tau. \tag{30}$$

Proposition 3.2. *Under Assumptions 1, 24, 25, 14 and 15, the IPTW-IPCW transform 30 is a censoring unbiased transformation in the following sense: for $a \in \{0,1\}$, it holds*

$$\mathbb{E}[I\{A=a\}T^*_{\text{IPTW-IPCW}}|X] = \mathbb{E}[T(a) \wedge \tau|X].$$

The proof of Proposition 3.2 can be found in Section 7.4. This transform now depends on two nuisance parameters, namely the conditional survival function of the censoring, and the propensity score. Once estimators of these quantities are provided, one could look at

$$\hat{\theta}_{\text{IPTW-IPCW}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A}{\hat{e}(X)} - \frac{1-A}{1-\hat{e}(X)}\right)\frac{\Delta_i^\tau}{\widehat{G}(\tilde{T}\wedge\tau|A_i,X_i)}\tilde{T}_i \wedge \tau, \tag{31}$$

Proposition 3.3. *Under Assumptions 1, 24, 25, 14 and 15, and if $\widehat{G}(\cdot|X,A)$ and $\hat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then estimator 31 is a weakly (resp. strongly) consistent estimator of the 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 tranform 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

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

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

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:

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

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

$$\hat{\theta}_{\text{IPTW-IPCW}} = \int_0^\tau \hat{S}_{\text{IPTW-IPCW}}(t|A=1) - \hat{S}_{\text{IPTW-IPCW}}(t|A=0)dt. \tag{32}$$

This estimator is consistent, provided that $\hat{e}$ and $\widehat{G}$ are too.

Proposition 3.4. *Under Assumptions 1, 24, 25, 14 and 15, 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.

Corollary 3.2. *Under Assumptions 1, 24, 25, 14 and 15, 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 $\hat{S}_{\text{IPTW}-\text{IPCW}}(t|A = a)$ 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

We can also apply IPTW to the Buckley-James transformation introduced in Section 2.2.2. We obtain

$$T^*_{\text{IPTW}-\text{BJ}} = \left(\frac{A}{e(X)} + \frac{1-A}{1-e(X)}\right)\left(\Delta^\tau \tilde{T} \wedge \tau + (1-\Delta^\tau)Q_S(\tilde{T} \wedge \tau | A, X)\right). \tag{33}$$

Proposition 3.5. *Under Assumptions 1, 24, 25, 14 and 15, the IPTW-BJ transform 33 is a censoring unbiased transformation in the following sense: for $a \in \{0, 1\}$, it holds*

$$\mathbb{E}[I\{A = a\}T^*_{\text{IPTW}-\text{BJ}}|X] = \mathbb{E}[T(a) \wedge \tau | X].$$

The proof of Proposition 3.5 can be found in Section 7.4. This transform again depends on two nuisance parameters which, once estimated, yields the following estimator of the ATE:

$$\hat{\theta}_{\text{IPTW}-\text{BJ}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A}{\hat{e}(X)} - \frac{1-A}{1-\hat{e}(X)}\right)\left(\Delta^\tau \tilde{T} \wedge \tau + (1-\Delta^\tau)\widehat{Q}_S(\cdot|A, X)\right), \tag{34}$$

Proposition 3.6. *Under Assumptions 1, 24, 25, 14 and 15, and if $\widehat{Q}_S(\cdot|X, A)$ and $\hat{e}(\cdot)$ are uniformly weakly (resp. strongly) consistent estimators, then estimator 34 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 The *G*-formula

Like in Section 2.1.2, one can estimate the RMST by leveraging Equation 26 with the *G*-formula Equation 13.

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

Chen and Tsiatis (2001) study this estimator in the particular case where $\hat{\mu}_a$ are estimated using Cox models. They are able to show that the resulting estimator is strongly consistence and asymptotically normal, and they give an explicit formulation of the asymptotic variance as a function of the parameters of the Cox model. Foster, Taylor, and Ruberg (2011) and Künzel et al. (2019) empirically study this estimator using Survival Forest, with the former employing it as a T-learner (referred to as 'Virtual Twins') and the latter as an S-learner.

3.2.4 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

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

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. It is easy to show that T^*_{QR} is quadruply robust in the following sense.

Proposition 3.7. *Let G, H be two conditional survival function, p be a propensity score, and ν be a conditional response. Then, under the same assumtion on F, R as in Proposition 2.9, and under Assumptions 1, 24, 25, 14 and 15, the transformations $T^*_{\text{QR}} = T^*_{\text{QR}}(F, R, p, \nu)$ satisfies, fro $a \in 0, 1$,*

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

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):

$$\hat{\theta}_{\text{AIPTW-AIPCW}} := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{\hat{e}(X_i)} - \frac{1-A_i}{1-\hat{e}(X_i)}\right)(T^*_{\text{DR}}(\widehat{G}, \hat{S})_i - \hat{\mu}(X_i, A_i)) + \hat{\mu}(X_i, 1) - \hat{\mu}(X_i, 0). \quad (35)$$

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

4 Implementation

In this section, we first review the notations and summarize the various estimators and their properties. We then present the packages available for directly computing θ_{RMST}, specifying the particular settings under which each is applicable, whether for randomized controlled trial (RCT) data, observational data, or certain censoring assumptions. Finally, we 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.

4.1 Summary of the estimators

Table 2 summarizes of the notation used in the previous sections:

Table 2: Summary of the notations.

Symbol	Description
X	Covariates

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 mis-specification has no impact. Underlined estimators are those implemented in available packages

Estimator	Context of application	Outcome model	Censoring model	Treatment model	Resistant to nuisance mis-specification
Unadjusted KM	RCT & Indep. cens.				
IPCW-KM	RCT & Dep. cens.		(G)		No
BJ		(Q_S)			No
IPTW-KM	Obs. & Indep. cens.			(e)	No
IPCW-IPTW-KM	Obs & Dep. cens.		(G)	(e)	No
G-formula		(μ)			No
IPTW-BJ		(Q_S)		(e)	No
AIPTW-AIPCW		(Q_S, μ)	(G)	(e)	Resistant if outcome model well specified or if censoring and treatment model well specified.

Symbol	Description
A	Treatment indicator ($A = 1$ for treatment, $A = 0$ for control)
T	Survival time
$T(1), T(0)$	Potential survival time respectively with and without treatment
$S^{(1)}, S^{(0)}$	Potential survival curve $\left(S^{(a)}(t) = \mathbb{P}(T(a) > t)\right)_{a\in\{0,1\}}$ of the potential survival time
C	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 $\mathbb{I}\{\tilde{T} \geq \tau\} + \mathbb{I}\{\tilde{T} < \tau\} \cdot \Delta$
$\{t_1, t_2, \dots, t_K\}$	Support of the survival time and censoring time distribution
$e(x)$	Propensity score $\mathbb{E}[A\|X = x]$
$\mu(x, a)$	$\mathbb{E}[T \wedge \tau \mid X = x, A = a]$
$S(t\|a, x)$	Conditional survival function, $\mathbb{P}(T > t\|X = x, A = a)$.
$G(t\|a, x)$	left-limit of the conditional survival function of the censoring $\mathbb{P}(C \geq t\|X = x, A = a)$
$Q_S(t\|x, a)$	$\mathbb{E}[T \wedge \tau \mid X = x, A = a, T \wedge \tau > t]$

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 mis-specification of the nuisance parameters. [CB: y'a pas le double robust simple de Rubin et van der Laan?]

In the next sections, the following packages will be used. [CB: a mettre en section package? on comprend pas trop ce que ça vient faire la]

```r
library(survival) # Implementation of Kaplan-Meier

library(survRM2)
library(RISCA)
library(grf) # causal_survival_forest function and also
# survival_forest and probability_forest


library(MASS) # mvrnorm function for simulation
library(rms) # cph function and predict for cph object

library(dplyr)
library(ggplot2)
library(gridExtra) # to plot multiple graph on a page
```

We also use utilitary functions that will be used through the different implementation:

- `estimate_propensity_score`: function to estimate propensity scores $e(X)$ using either parametric (i.e. logistic regression with the argument `"glm"`) or non-parametric methods (i.e. probability forest with the argument `"probability forest"` based on the `probability_forest` function from the grf(Tibshirani et al. 2017) package). This latter can include cross-fitting (n.folds>1).
- `estimate_survival_function`: function to estimate the conditional survival model, which supports either Cox models (argument `type_of_model ="cox"`) or survival forests (argument `type_of_model ="survival forest"`) which uses the function `survival_forest` from the grf(Tibshirani et al. 2017) package). This latter can also include cross-fitting (n.folds>1). The estimation can be done as 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 which uses the previous `estimate_survival_function`.
- `Q_Y`: function to find the remaining survival function from `Q_t_hat` at the specific time to event. [CB: c'est a dire? c'est pas un simple appel de Q t hat?]
- `integral_rectangles`: function to estimate the integral of a decreasing step function using the rectangle method with the function's x and y coordinates.
- `expected_survival`: function to estimate the integral of a continuous survival function using the trapezoidal method.
- `integrate`: function to estimate the integral at specific time points `Y.grid` of a given `integrand` which takes its values at `times`.

4.2 Available packages

Currently, there are few sustained implementations available for estimating RMST in the presence of right censoring. Notable exceptions [CB: exception?] include the packages survRM2 ["survRM2: Comparing Restricted Mean Survival Time" (2015)][CB:mauvais format biblio], grf (Tibshirani et al. 2017) and RISCA (Foucher, Le Borgne, and Chatton 2019).

[CB: dans toute cette section je pointerais plutot vers les equations definissant estimateurs que vers les assumptions, sinon on a l'impression que qu'on ne peut pas appeler ces fonctions en dehors de ces hypotheses ce qui est faux.]

SurvRM2

Under Equation 3 (random treatment assignment) and Equation 6 (independent censoring), the difference in RMST with Unadjusted Kaplan-Meier $\hat{\theta}_{KM}$ (Equation 11) 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.

```
library(survRM2)
RMST_survRM2 <- function(data, tau) {
  ATE_pack <- rmst2(data$T_obs, data$status, arm = data$A, tau = tau)
  RMST <- ATE_pack[[5]][1]
  return(RMST)
}
```

[IM: c'est pas claire pour moi si cette fonction $RMST_survRM2$ calcule directement $theta_rmst$ ou les 2 rmst qui correspondent aux deux groupes de traitement. Si elle donne $theta_rmst$, je l'appelerai peut-etre $theta_rmst_survrm2$] [CV : à faire à la fin, à homogénéiser partout]

RISCA

The RISCA package provides several methods for estimating θ_{RMST}. Under the assumptions of random treatment assignment (Equation 3) and independent censoring (Equation 6), the difference in RMST with Unadjusted Kaplan-Meier $\hat{\theta}_{KM}$ (Equation 11) 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.

Under the assumptions of unconfoundedness (Equation 24), treatment positivity (Equation 25), and independent censoring (Equation 6), IPTW Kaplan-Meier (Equation 28) 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 2.1.2 and Section 3.2.3, 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.

```
# Function to estimate RMST using single learner G-formula with Cox model
RISCA_gf <- function(data,
```

```
                    tau,
                    X.names.outcome) {

  # Define the outcome formula for the Cox model
  outcome <- paste(c('Surv(', "T_obs", ',', "status", ')'), collapse = "")
  # Single learner : the treatment arm is a predictor
  formula <- as.formula(paste(outcome, paste(c(X.names.outcome, 'A'),
                                             collapse = " + "), sep = " ~ "))

  # Fit the Cox proportional hazards model
  cox.cdt <- coxph(formula, data = data, x = TRUE)
  summary(cox.cdt)

  # Compute the effect of the treatment (ATE) using the G-formula
  gc.ate <- gc.survival(
    object = cox.cdt,
    data = data,
    group = "A",
    times = "T_obs",
    failures = "status",
    max.time = tau,
    iterations = 100,
    effect = "ATE",
    n.cluster = 1
  )

  # Extract the ATE
  ATE_RISCA_gf <- gc.ate$delta[[1]]
  return(ATE_RISCA_gf)
}
```

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.

```
# Function to estimate RMST using Causal Survival Random Forest (CSRF)
CSRF <- function(data, X.names, tau) {
  # Fit a causal survival forest
  cf <- causal_survival_forest(X = as.matrix(data[, X.names]), Y = as.matrix(data$T_obs), W = as.m

  # Predict using the fitted forest
  cf.predict <- predict(cf)

  # Estimate the average treatment effect (ATE)
  ATE_csf <- average_treatment_effect(cf)
```

```
  # Return the estimated ATE
  return(ATE_csf[[1]])
}
```

[CB: mettre toutes les sous-section suivantes dans une seule sous-section et organiser en paragraphe]

4.3 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 11 can then be calculated with the `RMST_1` function. Here, the integral is computed using the `integral_rectangles` utility function, available in the `utilitary.R` file. [CB: je ne mentionnerais pas du tout ces fonctions utilitaires dans le corps du texte]

```
# Kaplan-Meier estimator handmade implementation
# The database 'data' must be in the same form as that shown in
# notation (Table 1) and with the same variable name (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 data frame 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):
```

```r
# Two possibilities of 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) 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
  fit0 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A0,])  # Groupe A = 0
  fit1 <- survfit(Surv(T_obs, status) ~ 1, data = data[data$A == A1,])  # Groupe A = 1

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

  # Extraire les RMST des résultats
  rmst0 <- summary_fit0$table["rmean"][[1]]
  rmst1 <- summary_fit1$table["rmean"][[1]]

  # Calculer la différence des RMST entre les deux groupes
  difference_rmst <- rmst1 - rmst0
return(difference_rmst)
}
```

4.4 IPCW Kaplan-Meier

We provide a customized function, `adjusted.KM`, to facilitate the understanding of the IPCW Kaplan-Meier approach. The difference in RMST, estimated with $\hat{\theta}_{\text{IPCW}}$ as in Equation 20, can then be calculated using the `IPCW_Kaplan_meier` function. 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
```

```r
  # 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 observe T_obs_tau from the all
  # 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 difference in RMST
  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 pour A = 0
```

```
  summary_fit1 <- summary(fit1, rmean = tau)  # RMST pour A = 1

  # Extraire les RMST des résultats
  rmst0 <- summary_fit0$table["rmean"][[1]]
  rmst1 <- summary_fit1$table["rmean"][[1]]

  # Calculer la différence des RMST entre les deux groupes
  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.

4.5 Buckley-James based estimator

The function `BJ` estimates θ_{RMST} by implementing the Buckley-James estimator as in Equation 22. It uses two functions available in the `utilitary.R` file, namely `Q_t_hat` and `Q_Y`.

```
# 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 (Restricted Mean Survival Time) 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)
  ))
}
```

4.6 IPTW Kaplan-Meier

The function `IPTW_Kaplan_meier` implements the IPTW-KM estimator in Equation 29. It uses the `estimate_propensity_score` function from the `utilitary.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))
}
```

4.7 G-formula

[CV : J'ai reduit le code, ça va modifier le texte :][CV: attention j'ai changé aussi légèrement estimate_survival_function] 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.

Our nuisance models include both survival forests and Cox regression, [CB: a supprimer IMO ->] with cross-fitting applied to survival forests when `n.folds` > 1. Integration is handled by the trapezoidal rule (detailed in appendix **?@sec-trapez**), implemented through the `Expected_survival` utility function in `utilitary.R`. The function `cph` from rms function from the rms package is preferred because it allows for predictions on a specified time scale, a feature that the `coxph` from survival does not support. [CB: <-]

```
# 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 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)
}

# 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)
}
```

4.8 IPTW-IPCW Kaplan-Meier

The `IPTW_IPCW_Kaplan_meier` function implements the IPTW-IPCW Kaplan Meier estimator from Equation 32. 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` (detailed and implemented in Section 4.4) which computes an adjusted Kaplan Meier estimator using the appropriate weight.

```
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
```

```r
  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$T_obs_tau,
                   failures = data$status_tau,
                   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))
}
```

4.9 IPTW-BJ estimator

The `IPTW_BJ` implements the IPTW-BJ estimator in Equation 34. It uses the utilitary functions, from the `utilitary.R` file, `estimate_propensity_score` to estimate the nuisance parameters.

```r
# In using the min
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 and other relevant metrics
  return(RMST)
}
```

4.10 AIPTW-AIPCW

The `AIPTW_AIPCW` function implements the AIPTW_AIPCW estimator Equation 35 using the utilitary function from the `utilitary.R` file `estimate_propensity_score` 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)]


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
  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))

  # 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))
}
```

5 Simulations

We compare the behaviors and performances of the estimators using simulations. [CB: a supprimer ou a étoffer] [CB: remarque general: ce serait pas mal de continuer le labelling Scenario 1/2/3/... etc pour faciliter la lecture. Il est present en section 5.1 mais plus dans la suite.]

5.1 RCT

5.1.1 Data Generating Process

We generate RCTs with independent censoring (Scenario 1) and conditionally independent censoring (Scenario 2). The survival time and the censoring time (when there is dependency between the censoring time and the covariates) is simulated using the cumulative hazard inversion method for exponential models (Leemis, Shih, and Reynertson 1990; Bender, Augustin, and Blettner 2005). More specifically, n samples $(X_i, A_i, C, T_i(0), T_i(1))$ are generated as follows: [CB: je reformulerais en terme de S, G plutot qu avec les fonctions de hasard instantannee]

- $X \sim \mathcal{N}\big(\mu = [1, 1, -1, 1]^\top, \Sigma = I_4\big)$.
- $e(X) = P(A = 1|X) = 0.5$.
- $\lambda(0)(X) = 0.01 \cdot \exp\{0.5X_1 + 0.5X_2 - 0.5X_3 + 0.5X_4\}$ hazard for the event time $T(0)$.
- The hazard for the censoring time C:
  - For Scenario 1: $\lambda_c = 0.03$ does not depend on covariates.
  - For Scenario 2: $\lambda_c(X) = 0.03 \cdot \exp\{0.7X_1 + 0.3X_2 - 0.25X_3 - 0.1X_4 - 0.2A\}$.
- $T(1) = T(0) + 10$.
- the survival time is $T = AT(1) + (1 - A)T(0)$.
- The observed time is $\tilde{T} = \min(T, C)$.
- The status is $\Delta = 1(T \leq C)$.
- The threshold time τ is set to 25.

The observed samples are $(X_i, A_i, \Delta_i, \widetilde{T}_i)$.

```
############ 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,
```

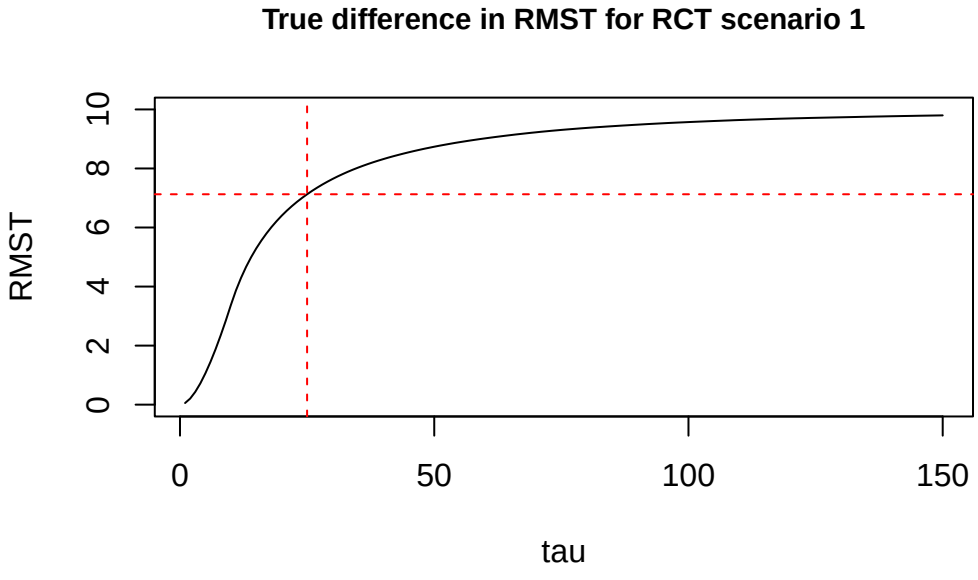
```
                                parsC = c(0.7, 0.3, -0.25, -0.1),
                                parsC_A = c(-0.2),
                                scenario = "RCT2",
                                mis_specification="none") {

# 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)

  # 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)
}
```

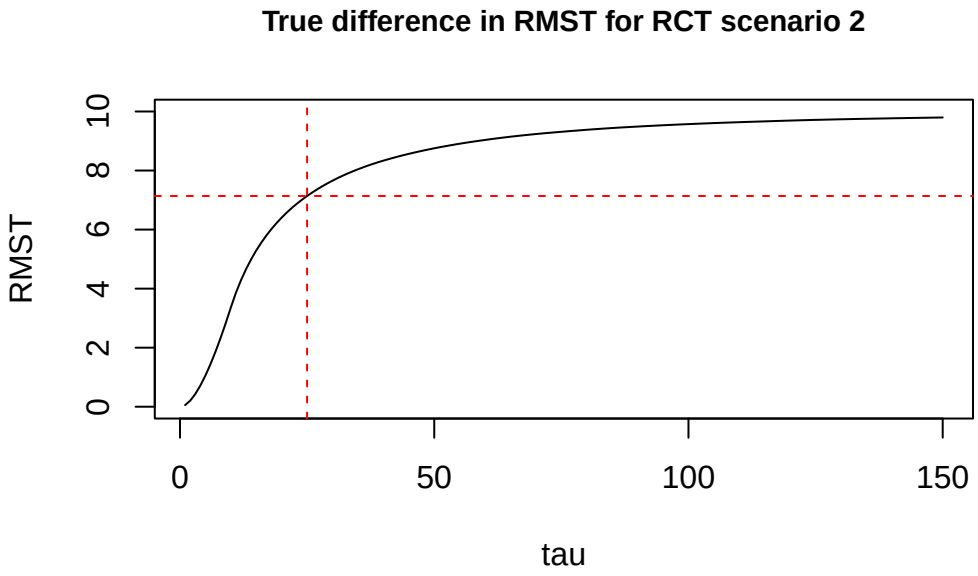
The descriptive statistics of the two datasets are displayed in Annex (Section 8.1.1).

The following implementation provides the value of θ_{RMST} for a given time horizon τ. Note that this value is the same for Scenario 1 and Scenario 2. The graph of the difference in RMST as a function of τ for both Scenarii is displayed below. [CB: je ne m'y connais pas en R donc je vous laisse modifier le code pour qu'il n'y qu'un graphe + changer le titre]

```
# 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 at time 25 is 7.1"
```




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

5.1.2 Estimation of the RMST

For each setting, we estimate the difference in RMST using the following methods. In parentheses, we indicate how the nuisance components are estimated. Default tuning parameters are applied for forest-based methods, and cross-fitting is performed with five folds. Where available, we use package implementations. [CB: on comprends pas trop ce que ça vient faire la puisqu'on utilise les memes estimateurs dans le cas obs. Je mettrais ça soit en fin de partie implementation, soit en preambule de la partie simulation][CB: de plus j'enleverai NAIVE car on en a jamais parler avant]

[IM : rajouter le nom des fonctions utilisées + argument][CB: oui pk pas]

- **Naive**: The Naive estimator excludes censored observations and directly applies the θ_{RMST} formula to uncensored data.
- **SurvRM2 - KM**: Using the SurvRM2 package for Kaplan-Meier estimation.
- **IPTW KM (Logistic Regression)**: Inverse Probability of Treatment Weighting (IPTW) with Kaplan-Meier, using logistic regression for treatment assignment.
- **RISCA - IPTW KM**: IPTW with Kaplan-Meier implemented through the RISCA package.
- **IPCW KM (Cox)**: Inverse Probability of Censoring Weighting (IPCW) with Kaplan-Meier, using a Cox model for censoring.
- **IPTW-IPCW KM (Cox & Logistic Regression)**: Combination of IPTW and IPCW with Kaplan-Meier, using a Cox model for censoring and logistic regression for treatment assignment.
- **BJ (Cox)**: Buckley-James estimator with a Cox model.
- **IPTW-BJ (Cox & Logistic Regression)**: IPTW applied to the Buckley-James estimator, using a Cox model for censoring and logistic regression for treatment assignment.

- **G_formula (Cox / T-learners)**: G-formula with a Cox model using T-learners.
- **G_formula (Cox / S-learner)**: G-formula with a Cox model using an S-learner.
- **RISCA - G_formula (S-learner)**: G-formula implemented through the RISCA package using an S-learner.
- **AIPTW-AIPCW (Cox & Logistic Regression)**: Augmented IPTW and IPCW estimator using Cox regression for censoring and logistic regression for treatment assignment.
- **grf - Causal Survival Forest**: Generalized random forest-based causal survival estimator.
- **IPTW KM (Forest)**: IPTW applied to Kaplan-Meier estimation, using a forest model for treatment assignment.
- **IPCW KM (Forest)**: IPCW applied to Kaplan-Meier estimation, using a forest model for censoring.
- **BJ (Forest)**: Buckley-James estimator using a forest model.
- **IPTW-BJ (Forest)**: IPTW applied to the Buckley-James estimator, using a forest model for treatment assignment and censoring.
- **IPTW-IPCW KM (Forest)**: Combination of IPTW and IPCW with Kaplan-Meier, using forest models for both treatment assignment and censoring.
- **G_formula (Forest / T-learners)**: G-formula with forest models using T-learners.
- **G_formula (Forest / S-learner)**: G-formula with forest models using an S-learner.
- **AIPTW-AIPCW (Forest)**: Augmented IPTW and IPCW estimator using forest models for both treatment assignment and censoring.

[IM: je sais pas si c'est facile a faire mais dans les figures peut-etre dans l'axe des X les noms des estimateurs pourraient avoir differentes couleurs en fonction de la categorie d'estimateur ou du type de nuisance parameter model?]

Figure 8 shows the distribution of the difference in RMST for 100 simulations in Scenario 1 and different sample sizes: 500, 1000, 2000, 4000. We chose to set $\tau = 25$. The corresponding 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)

# 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))
```

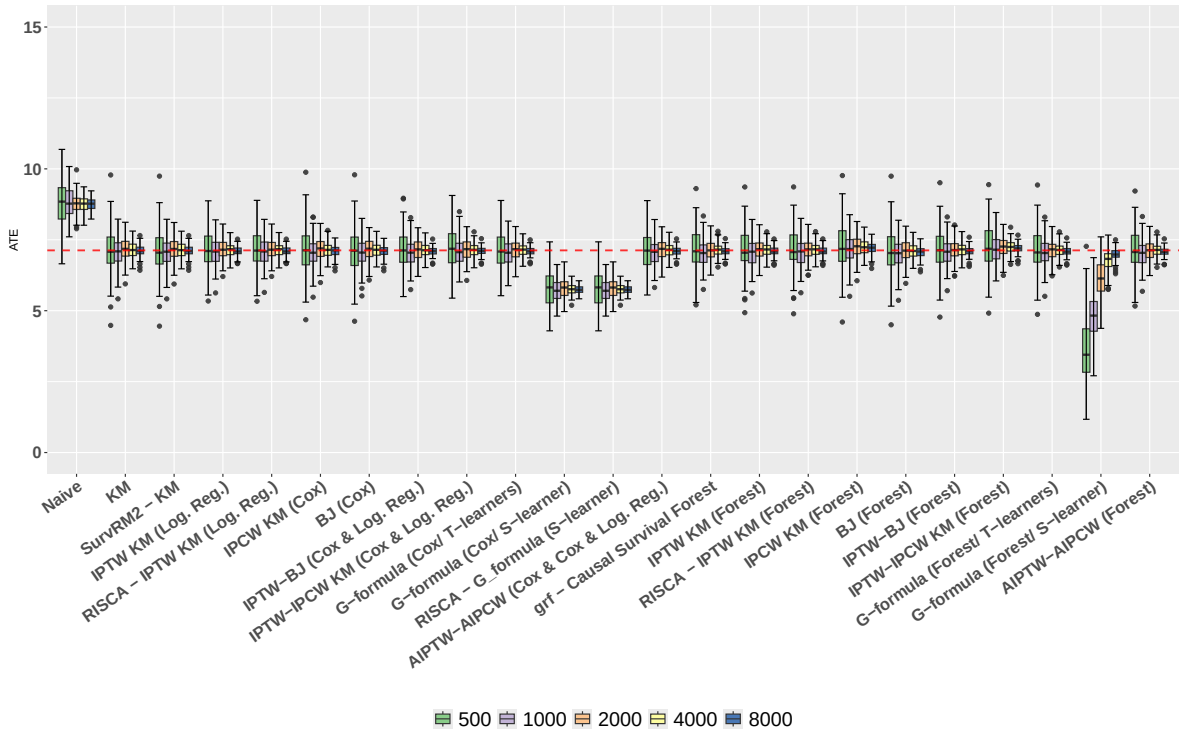


Figure 8: Results of the ATE for the simulation of a RCT with independent censoring.

In this setting, and in accordance to the theory, the simplest estimator (unadjusted KM) performs just as well as the others,and presents an extremely small bias (as derived in Section 2.1.1).

The naive estimator is biased, as expected, and the bias in both the G-formula (RISCA) and the manual G-formula S-learner implementation arises from a violation of the proportional hazards assumption, as the treatment effect is additive ($T(1) = T(0) + 10$). Since survival times are generated with an exponential distribution [CV: to check la formulation], a Cox model adjustment is well-suited for survival time estimation and can be seen as a correct specification of nuisance parameters. Consequently, the G-formula (Cox/ T-Learners) is highly appropriate, yielding a small-bias estimator across all sample sizes.

[CB: tentative de reformulation –>] Because the treatment effect is additive, it violates the assumption that T would follow a Cox model in the variables (X, A), whence the very poor performance of both G-formula (Cox/S-learners). 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. [CB: <–]

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, the 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 forests based methods. Regarding G-formula (S-learner/ Forest), it may converge but requires very large sample size.

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-

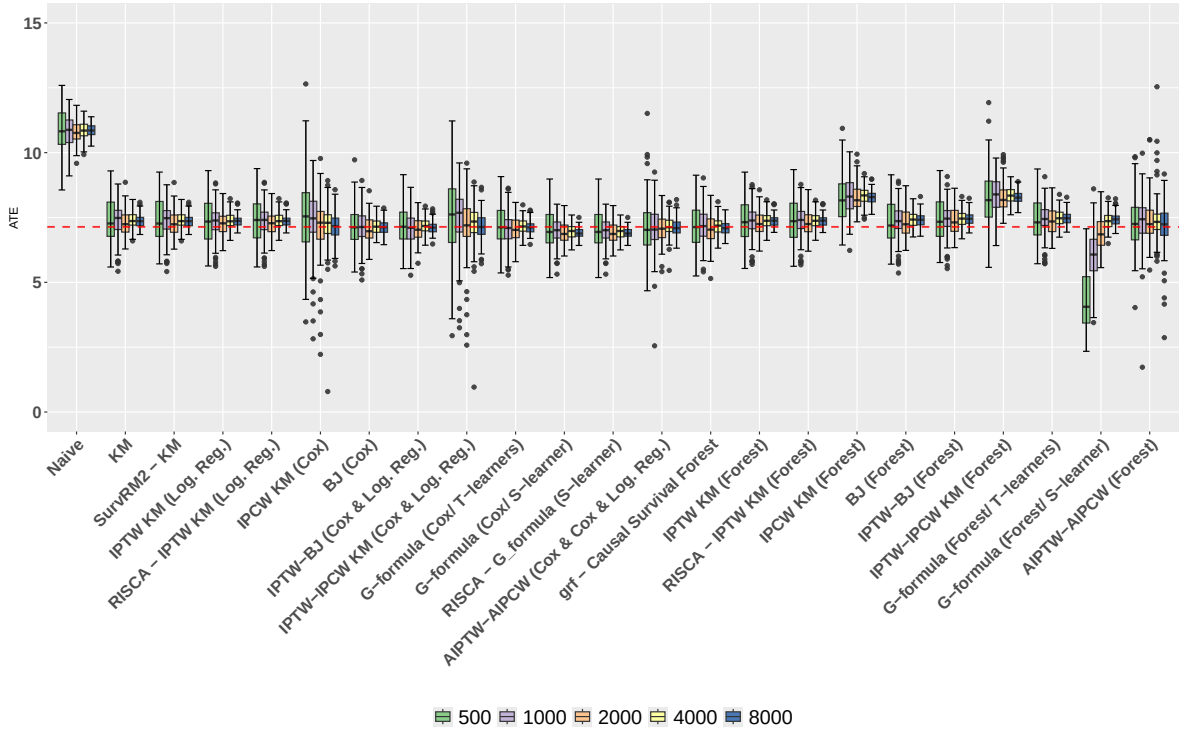


Figure 9: Estimation results of the ATE for the simulation of a RCT with dependent censoring.

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) and IPTW-IPCW KM (Cox & Log.Reg.) are slightly biased up to 4,000 observations and very unstable due to extreme censoring probabilities. In contrast, the Buckley-James estimator BJ (Cox), converges faster and exhibit a small bias 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) highlighting the importance of the choice of nuisance parameter estimation. [CB: comprend pas] Notably, all estimators exhibit higher variability compared to Scenario 1.

5.2 Observational study with parametric data generating process

5.2.1 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, n samples $(X_i, A_i, C, T_i(0), T_i(1))$ are generated as in Section 5.1, except:

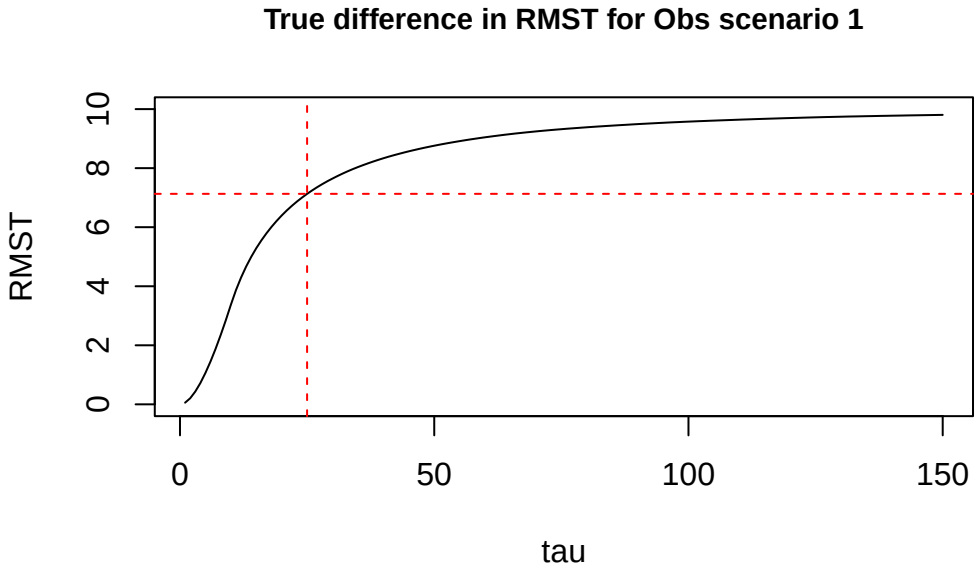
- $\text{logit}\{e(X)\} = -1X_1 - 1X_2 - 2.5X_3 - 1X_4$ for the treatment allocation (A).
- The hazard for the censoring time C:
  - For scenario 2: $\lambda_c(X) = 0.03 \cdot \exp\{0.7X_1 + 0.3X_2 - 0.25X_3 - 0.1X_4\}$.

The descriptive statistics for the two observational data with independent (`Obs1`) and conditionally independent censoring (`Obs2`) are displayed in Appendix (Section 8.1.2). Note that we did not modify the survival distribution, the target difference in RMST is thus the same.

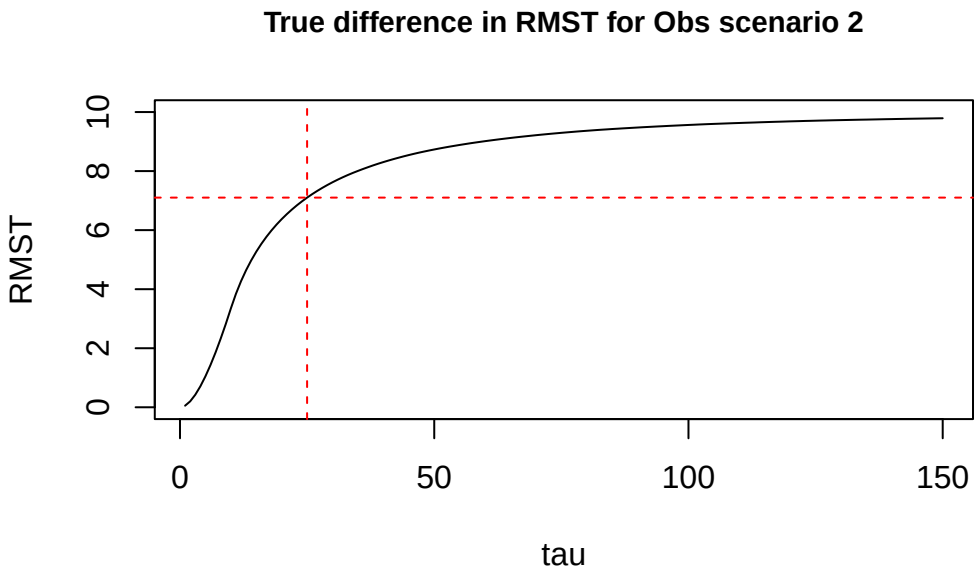
[CB: euh pourquoi on change C? revoir ce passage pour parler des Scenarii 3 et 4] [CB: il faut enlever le graphe du RMST car il ne change pas]

```
# 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)
}
```




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




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

5.2.2 Estimation of the RMST

Figure 10 below shows the distribution of the estimators of θ_{RMST} for the observational study with independent censoring.

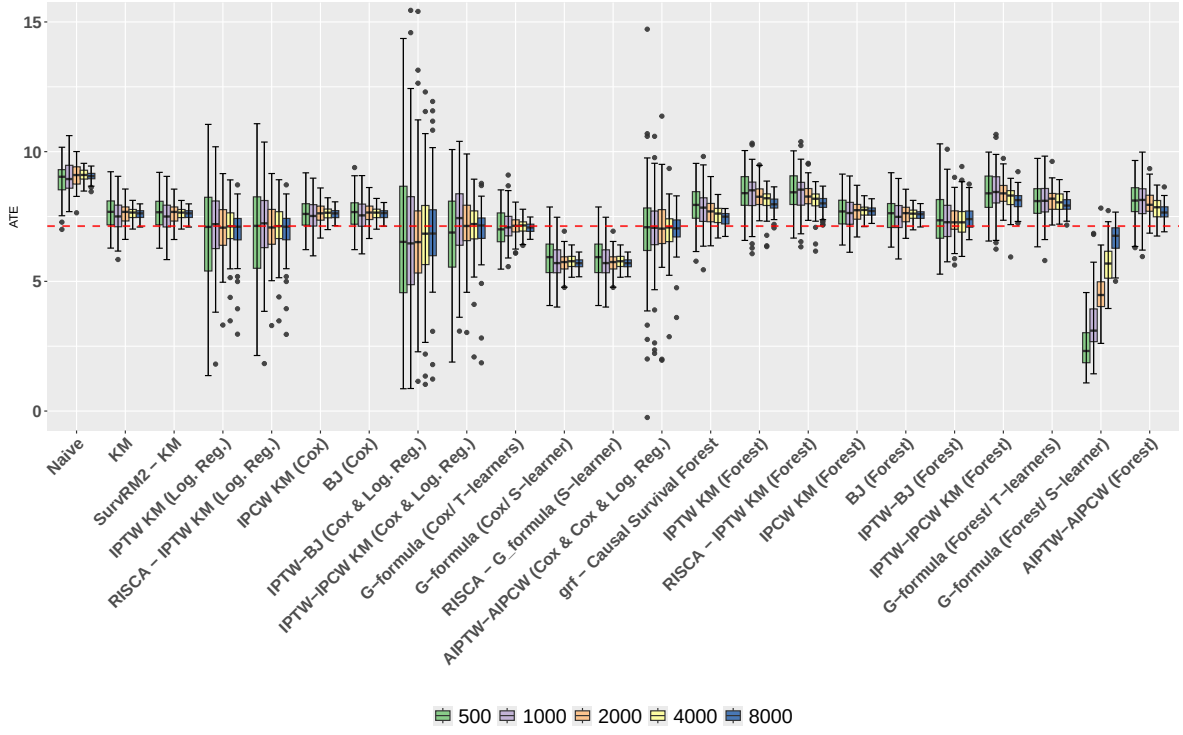


Figure 10: Estimation results of the ATE for the simulation of an observational study with independent censoring.

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 unecessary nuisance components. IPTW BJ (Cox & Log.Reg) is 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.

Figure 11 below show 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 $\tau = 25$.

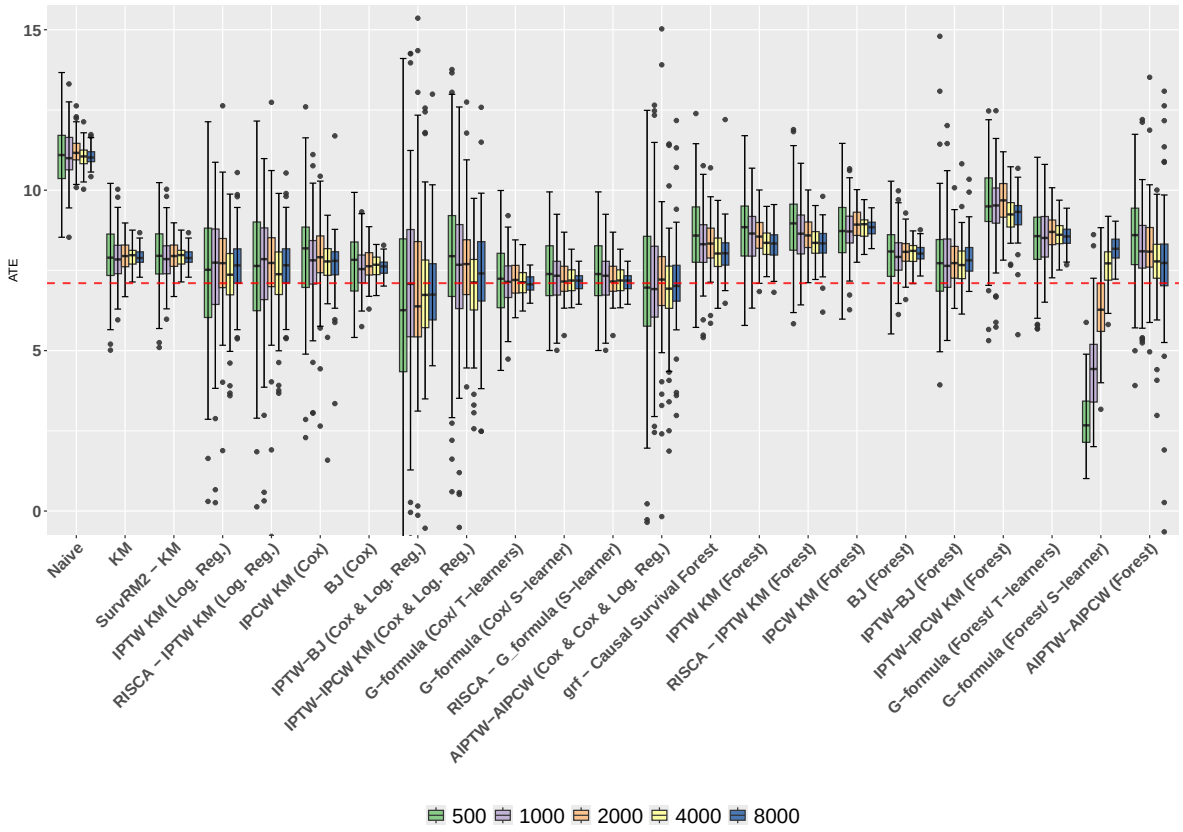


Figure 11: Estimation results of the ATE for the simulation of an observational study with dependent censoring.

In the simulation of an observational study with conditionally independent censoring, estimators that do not account for both censoring and confounding bias, such as KM, IPCW KM, IPTW KM, and their package equivalents, are biased. Notably, the IPTW-IPCW KM (Cox & Log.Reg) and IPTW-BJ (Cox & Log.Reg) estimators remain biased, even with 4,000 observations.

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 [JJ: mettre la ref de la section du resultat theorique sur la plus petite variance de gformula][CV: on a rien mis dans la partie théorique, je trouve pas] .

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 Observational study with nonlinear relationships and conditionally independent censoring

5.3.1 Data Generating Process

We conduct a simulation in which treatment allocation, as well as the survival and censoring models, are complex and cannot be accurately captured by a simple parametric or semi-parametric model. Instead, these complexities are better estimated using probability forests for the propensity model and survival forests for the conditional survival and censoring models. Specifically, we generate n samples $(X_i, A_i, C, T_i(0), T_i(1))$ in a setup similar to Scenario 4 in Cui et al. (2023):

- $X \sim \mathcal{N}\big(\mu = [1,1,1]^\top, \Sigma = I_3\big)$.
- T is generated from a Poisson distribution with mean $X_2 + X_3 + max(0; X_1 - 0,3)A$.
- C from a Poisson distribution with mean $1 + log(1 + exp(X_3))$.
- The propensity score is $e(x) = [(1 + exp(-X_1))(1 + exp(-X_2))]^{-1}$

Note that for subjects with $X_1 < 0,3$, treatment does not affect survival time. The horizon time τ is fixed at 2.

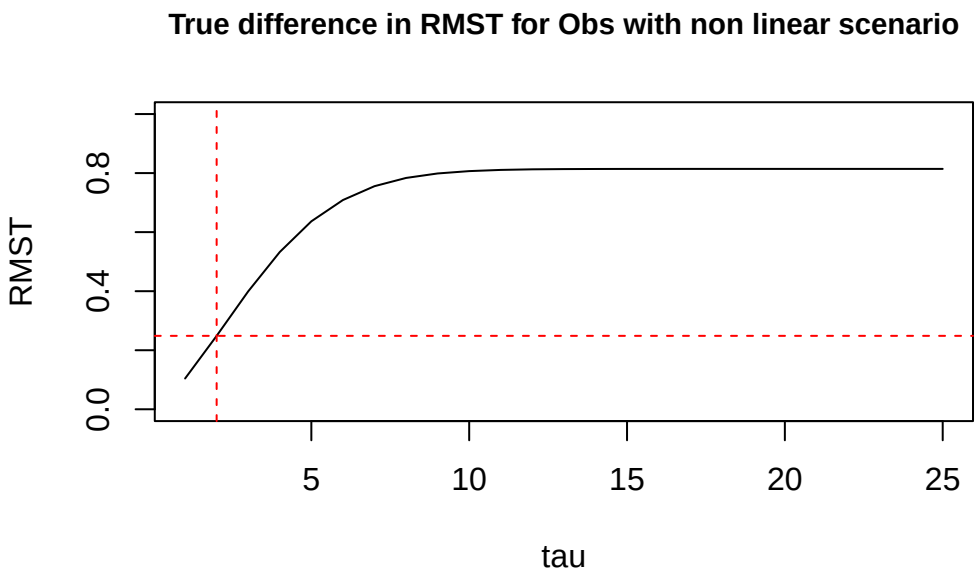
```
simulate_data_complex <- function(n = 2000, tau, parsC = c(0,0,1)){
  # Load necessary library
  library(MASS)

  # Generate covariates
  X <- mvrnorm(n, mu = c(1, 1, 1), Sigma = diag(3))
  X <- as.data.frame(X)
  colnames(X) <- c("X1", "X2", "X3")

  # Convert data frame to matrix for matrix operations
  X_treatment <- as.matrix(X)

  # Generate treatment
  e <- 1 / ((1 + exp(-X_treatment[, "X1"])) * (1 + exp(-X_treatment[, "X2"])))
  A <- sapply(e, function(p) rbinom(1, size = 1, prob = p))

  # Generate potential outcomes
  lambda_1 <- X_treatment[, "X2"] + X_treatment[, "X3"] +
              pmax(0, X_treatment[, "X1"] - 0.3) * 1
  lambda_0 <- X_treatment[, "X2"] + X_treatment[, "X3"]

  T1 <- rpois(n, lambda_1)
  T0 <- rpois(n, lambda_0)
```

```
  T1[is.na(T1)] <- 0
  T0[is.na(T0)] <- 0

  T_true <- T1 * A + T0 * (1 - A)

  # Generate censoring time
  lambda_C <- 1 + log(1 + exp(parsC[1]*X_treatment[, "X1"] +
                              parsC[2]*X_treatment[, "X2"] +
                              parsC[3]*X_treatment[, "X3"]))
  C <- rpois(n, lambda_C)

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

  # Create the final data frame
  DATA_target_population <- data.frame(
    X1 = X$X1,
    X2 = X$X2,
    X3 = X$X3,
    tau = tau,
    A = A,
    T1 = T1,
    T0 = T0,
    T_true = T_true,
    C = C,
    T_obs = T_obs,
    T_obs_tau = T_obs_tau,
    status = status,
    censor_status = censor_status,
    status_tau = status_tau,
    e = e
  )

  return(DATA_target_population)
}
```

The descriptive statistics for this data is given in Appendix (Section 8.1.3).

```
Warning in rpois(n, lambda_1): NAs produced
```

```
Warning in rpois(n, lambda_0): NAs produced
```




```
[1] "The ground truth for Observation with non linear scenario at time 2 is 0.25"
```

5.3.2 Estimation of the RMST

Figure 12 below show the distribution of the θ_{RMST} estimates for the observational study with conditionally independent censoring in the context of this non-parametric simulation. The red dashed line represents the true θ_{RMST} for $\tau = 2$.

This outcome is quite surprising, as one might expect forest-based methods to outperform others given the data-generating process (DGP). Yet, all methods face significant challenges except for IPTW-BJ (Cox & Log.Reg.). IPTW-BJ (Forest) and Causal Survival Forest also show convergence at 8000 observations. AIPTW-AIPCW (Forest) here exhibits a small biais. G-formula (Forest/T-learners) will also converge but its rate is slower. Lastly, and again unexpectedly, the G-formula (Cox/S-learner) provides satisfactory results despite the proportional hazards assumption being violated. [CV: Je trouve qu'on voit quand même un biais non ?]

5.4 Mispecification of nuisance components

5.4.1 Data Generating Process

We generate an observational study with covariate interactions and conditionally independent censoring. The objective is to assess the impact of misspecifying nuisance components; specifically, we will use models that omit interactions to estimate these components. This approach enables us to evaluate the robustness properties of various estimators.

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

- $X \sim \mathcal{N}\big(\mu = [1, 1, -1, 1, -2, -5]^\top, \Sigma = I_6\big)$.
- T is generated from an exponential distribution with mean $X_1^2 + X_2^2 + X_1 * X_2$.
- C from an exponential distribution with mean $0.1 * X_1^2 + 0.1 * X_3^2 + X_1 * X_3$.
- The propensity score is $e(x) = \frac{1}{1+exp(0.05*X_1 - 0.01*X_2 + 1*X_3 - 0.1X_4)}$.

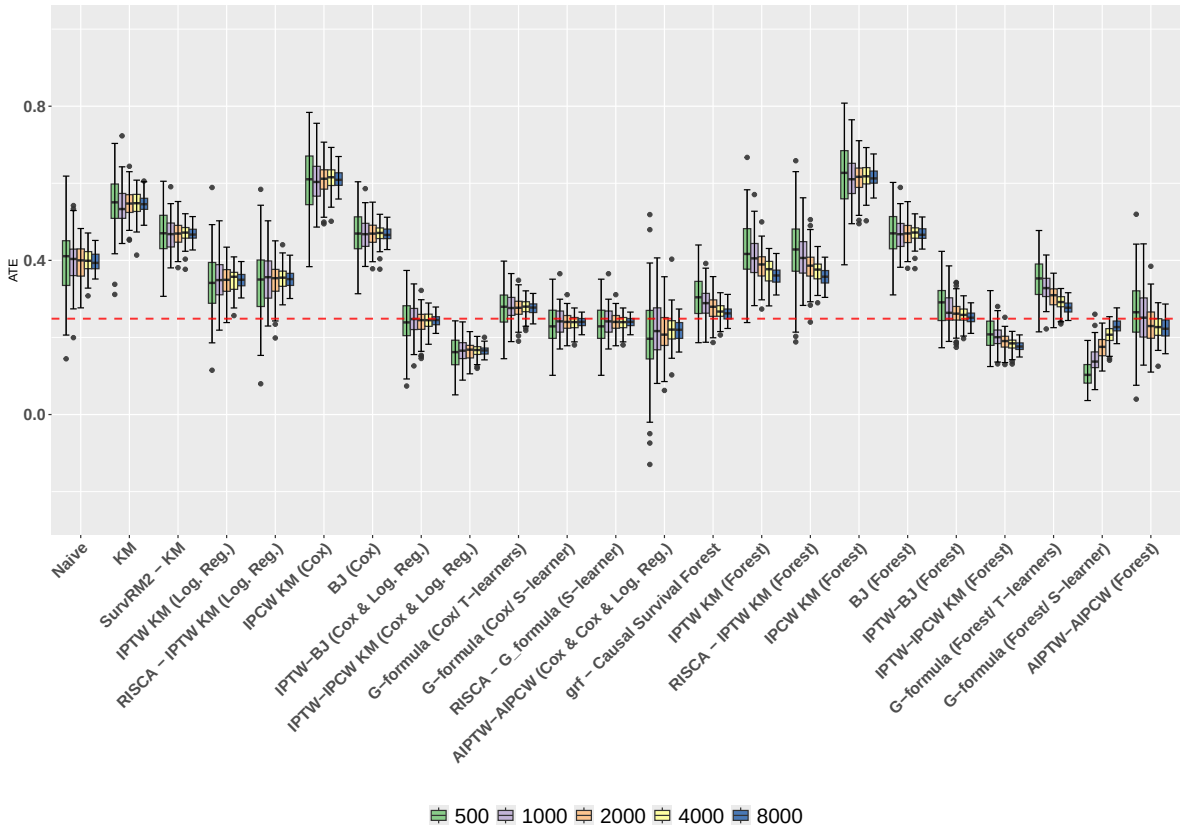
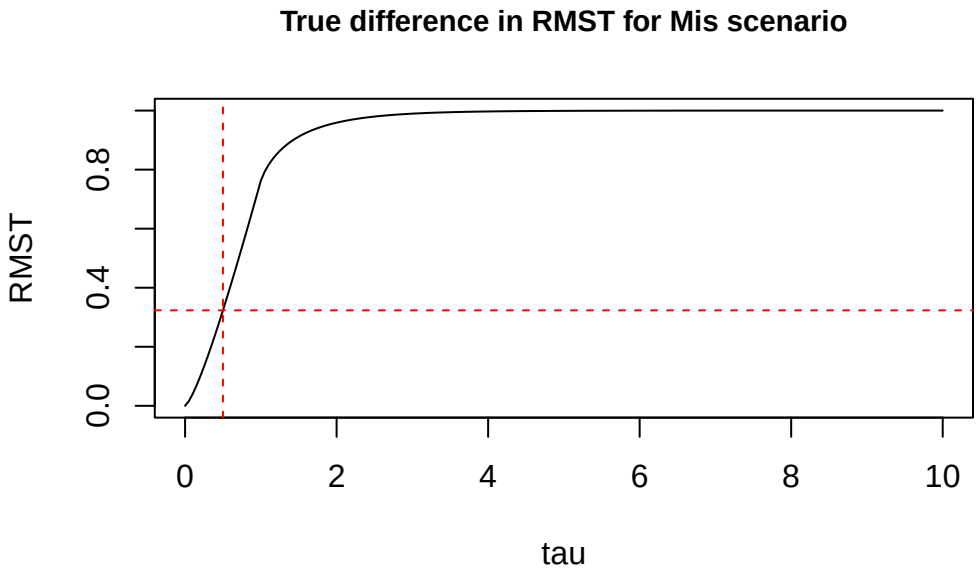


Figure 12: Estimation results of the ATE for an observational study with dependent censoring and complex relationships.

```r
# DGP for mis-specification
simulate_data_mis <- function(n,
                              mu = c(0.5, 0.5, 0.7, 0.5, -0.1, -0.1),
                              sigma =  matrix(c(1, 0, 0, 0, 0, 0,
                                                0, 1, 0, 0, 0, 0,
                                                0, 0, 1, 0, 0, 0,
                                                0, 0, 0, 1, 0, 0,
                                                0, 0, 0, 0, 1, 0,
                                                0, 0, 0, 0, 0, 1),
                                              nrow = 6, byrow = TRUE),
                              colnames_cov = c("X1", "X2", "X3", "X4", "X5",
                                               "X6"),
                              parsA =  c(0.05, -0.01, 1, -0.1, 0, 0),
                              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"] + parsA[2]*X_treatment[, "X2"] +
    parsA[3]*X_treatment[, "X3"] + parsA[4]*X_treatment[, "X4"] +
    parsA[5]*X_treatment[, "X5"] + parsA[6]*X_treatment[, "X6"]

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

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


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

```r
  # 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 8.1.4).




```
[1] "The ground truth for mis scenario at time 0.45 is 0.32"
```

5.4.2 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. Next, we introduce misspecification individually for the treatment model, censoring model, and outcome model (Figure 13). 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). [CV: mis-specification ou misspecification ?]

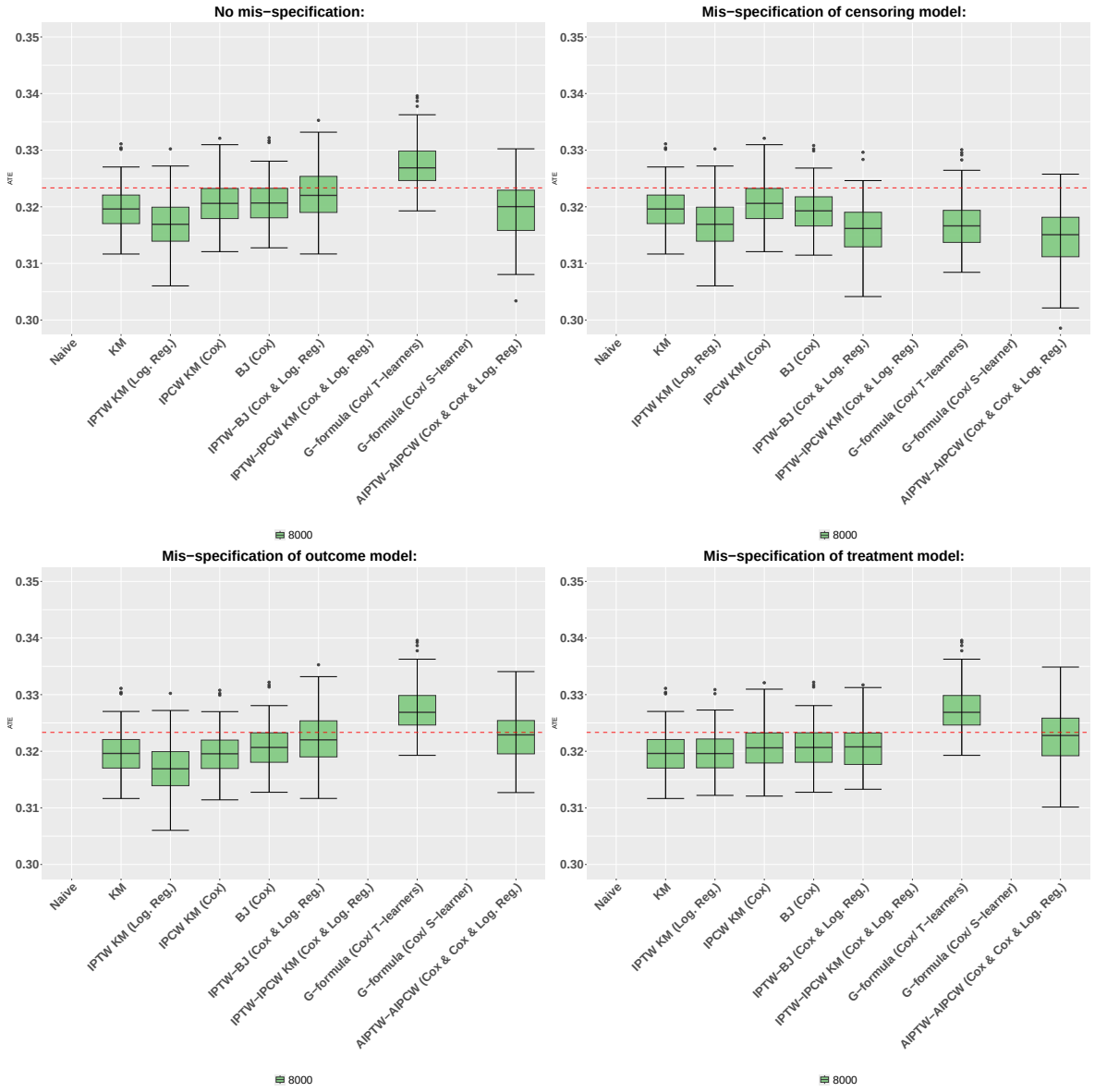


Figure 13: Estimation results of the ATE for an observational study with dependent censoring in case of a single mis-specification.

When there is no mispecification, IPTW-BJ (Cox & Log.Reg) is unbiased. IPTW-IPCW KM (Cox & Log.Reg) exhibits a large bias and is not displayed. Surprisingly, G-formula (Cox/ T-learners) and AIPTW-AIPCW (Cox & Cox & Reg.Log) are slightly biased.

When one nuisance model is mis-specified, [JJ: commenter des qu'on a la figure]

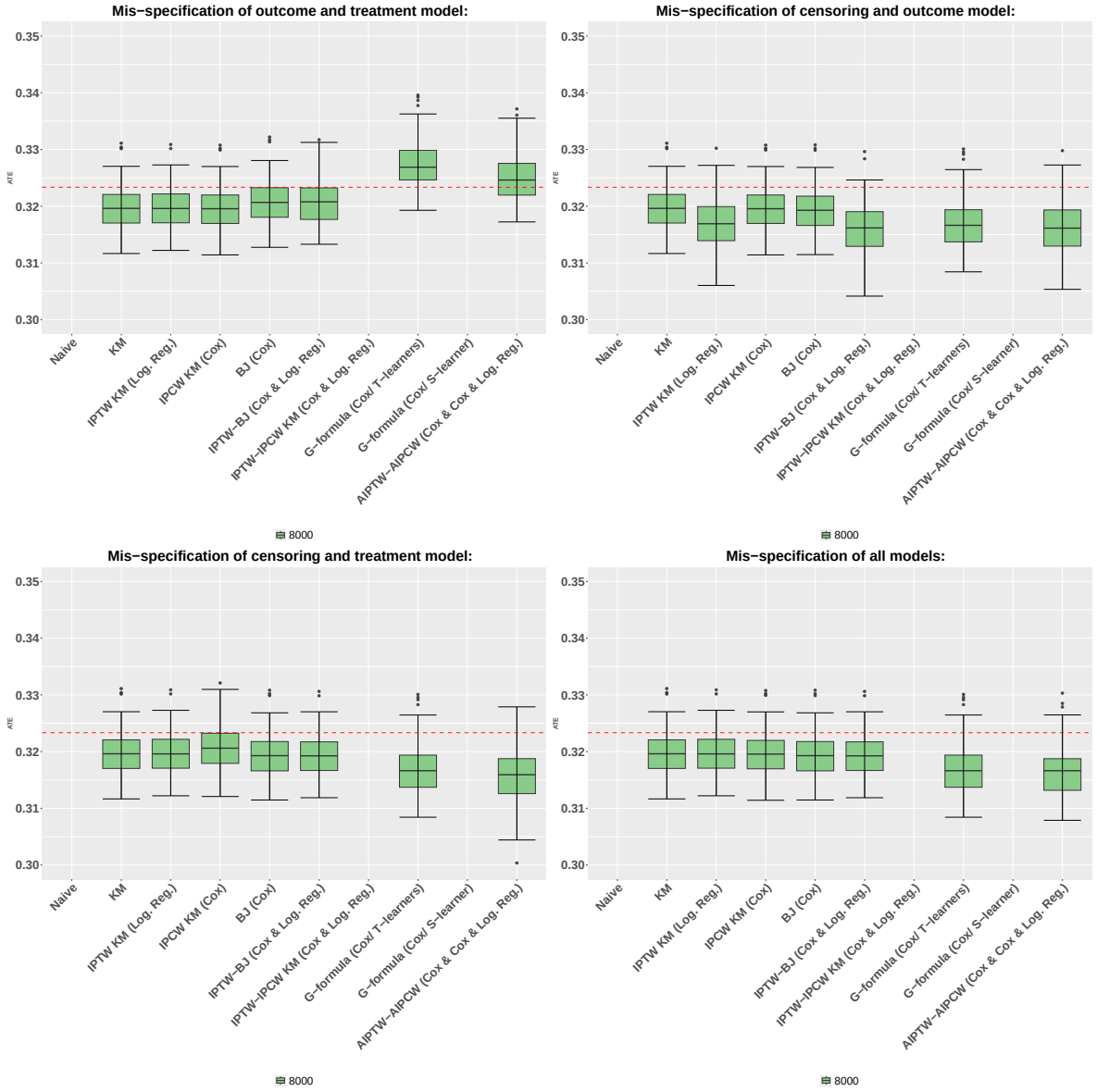


Figure 14: Estimation results of the ATE for an observational study with dependent censoring in case of a two or more mis-specifications.

Figure 14 shows that, as expected, when all nuisance models are misspecified, all estimators exhibit bias. AIPTW-AIPCW seems to converge only in cases where either the outcome and treatment models, or the outcome and censoring models, are misspecified, which deviates from initial expectations. It was anticipated that AIPTW-AIPCW would converge solely when both the censoring and treatment models were misspecified. [JJ: modifier au vu du plot]

5.5 Violation of positivity assumption for censoring

[CB: je suis pour enlever cette section car 1/ je pense que c'est assez clair que rien ne marche sans

cette hypothese 2/ on ne checke pas les autres hypotheses donc ça m'a l'air assez arbitraire de se concentrer sur la positivité de la censure et 3/ les lois proposées dans cette section ne violent pas cette hypothèse] In this section, we aim to demonstrate the impact of violating the positivity assumption for censoring on estimator convergence. To enable this violation, we use the same simulation setup as for randomized controlled trials (RCTs) with conditionally independent censoring in Section 5.1 (Figure 9), as well as for observational studies with conditionally independent censoring in Section 5.2 (Figure 11), but with a stronger dependency on covariates.

- For RCTs: $\lambda_c(X) = 0.002 \cdot \exp\{2X_1 - 4X_2 - 5X_3 + 0.2X_4 - 0.2A\}$.
- For observational studies: $\lambda_c(X) = 0.002 \cdot \exp\{2X_1 - 4X_2 - 5X_3 + 2X_4\}$.

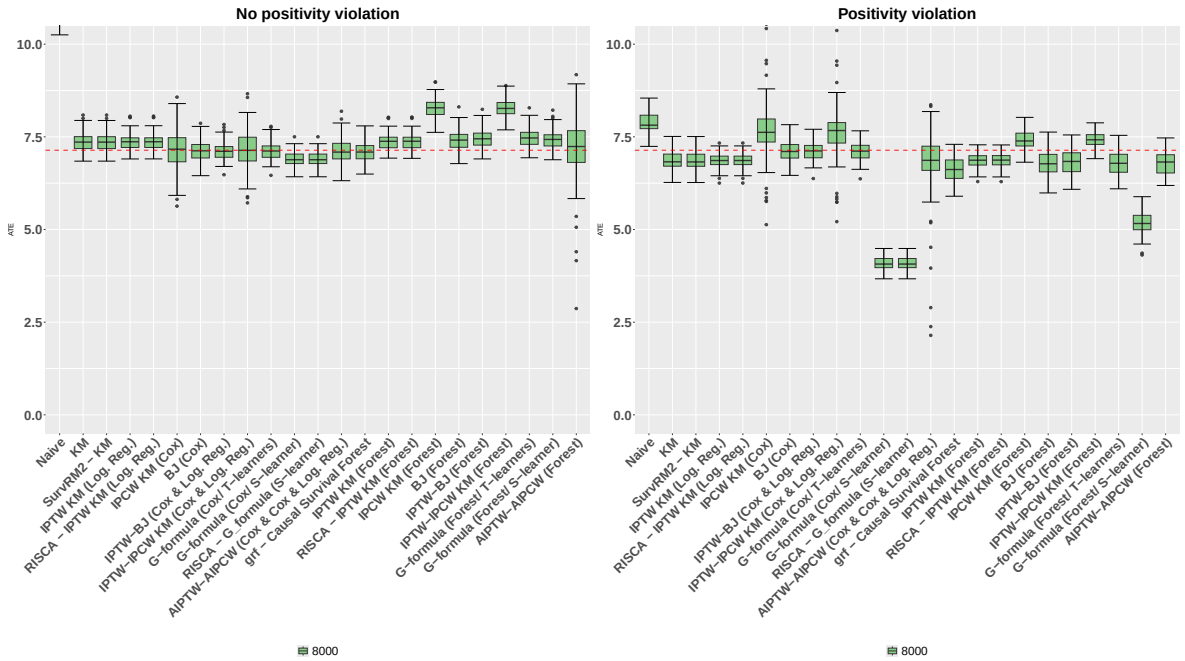


Figure 15: Estimation results of the ATE for the simulation of a RCT with dependent censoring (parametric relationship) when positivity for censoring is violated.

When the censoring positivity assumption is violated in this RCT (Figure 15), all estimators that adjust for censoring using inverse probability weighting (IPCW) methods—such as IPCW KM, IPTW-IPCW KM, and AIPTW-AIPCW—become biased. The causal survival forest shows only a slight bias, while, surprisingly, the G-formula (Cox/T-learners), BJ (Cox), and IPTW-BJ (Cox & Logistic Regression) remain unbiased.

The conclusion in the observational case (Figure 16) are similar but despite beeing unbiased IPTW-BJ (Cox) and the G-formula (Cox/T-learners) pay a price in terms of variance. This apparent unsencitivity to the violation of the positivity assumption for such estimators is valid as soon as the data generating process for the survival time is a Cox model. In a non-parametric setting for DGPs, presented in the boxplots below, all estimators suffers from this violation.

6 Conclusion and perspective

[IM: si je me souviens bien, beaucoup d'estimateurs en analyse de survie ont des problemes quand il y a beaucoup de 'ties' dans les donnees. je rajouterais ici peut-etre une phrase qui en parle et peut-

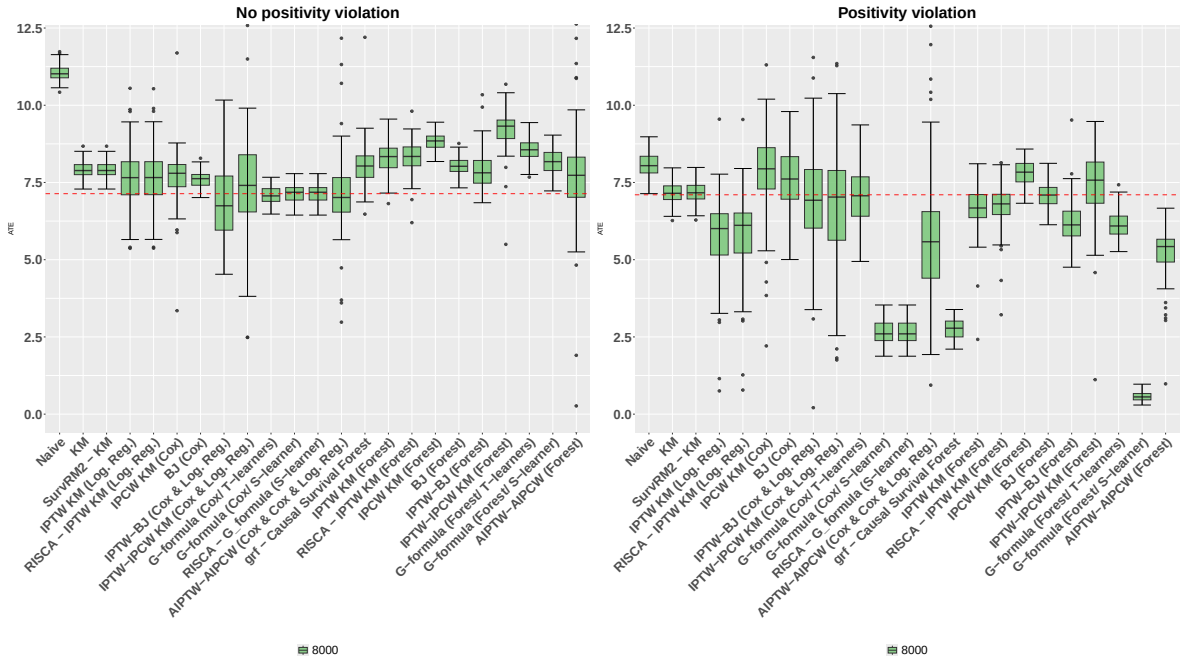


Figure 16: Estimation results of the ATE for the simulation of an observational study with dependent censoring (parametric relationship) when positivity for censoring is violated.

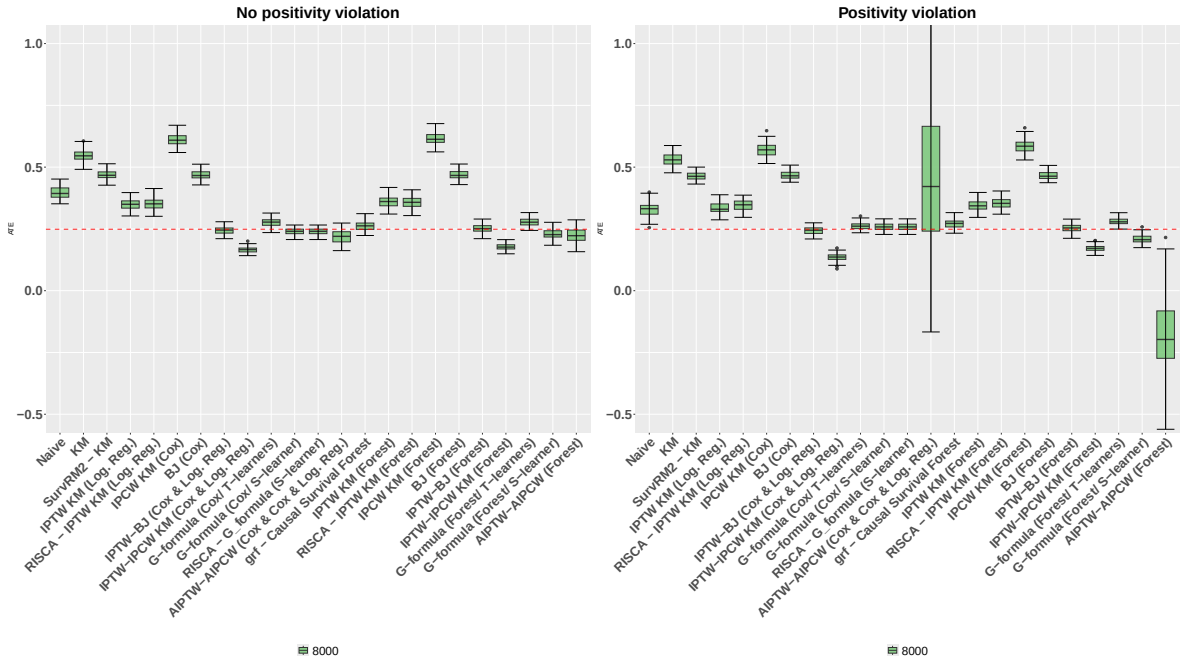


Figure 17: Estimation results of the ATE for the simulation of an observational study with dependent censoring (non-linear relationship) when positivity for censoring is violated.

etre meme un peu plus de details sur quels estimateurs soufferaient de la presence de beaucoup de ties et lesquels moins]

[IM: peut-etre ca vaut la peine de reprendre ici l'argument ATE_RMST vs HR et de dire que les estimateurs ATE_RMST existent et ont ete etudies et que l'analyse et le code fourni avec cet article permettront d'etudier davantage l'avantage de ATE_RMST pour l'analyse causale en analyse de survie?]

[IM: et une question naive - sans avoir regarde en detail les preuves en annexe - je me demandais si ce qui compte surtout pour la convergence des differents estimateurs c'est la sample size ou le nombre d'evenements observes dans chaque groupe. est-ce qu'il serait possible de rajouter dans les experiences le nombre d'evenements (moyen) pour chaque sample size et chaque groupe de traitement?]

In this study, we provide implementations for a comprehensive set of estimators to assess treatment effects on survival time using restricted mean survival time, filling a notable gap in available resources. Although demand for causal survival analysis tools is growing, few packages offer robust, diverse estimator options. Our code aims to support further research and practical applications in causal survival analysis, where software options remain limited.

For users, we recommend different estimators based on the context of their data and modeling assumptions.

The key insights from this study are as follows: In an **RCT setting** with independent censoring, any causal survival estimator, combined with any method for estimating nuisance parameters, accurately estimates θ_{RMST}, with the exception of the G-formula S-learner, which relies on specific assumptions. However, when censoring is dependent, estimators that fail to account for this are, unsurprisingly, biased. That said, it is reasonable to assume that users may have sufficient information to determine whether censoring is dependent. As a precaution, it is generally advisable to use estimators that accommodate dependent censoring, as the cost of estimating potentially unnecessary nuisance parameters is low (slight increase in variance). In scenarios with dependent censoring, the choice of (semi)-parametric or non-parametric methods for estimating nuisance parameters has a more substantial impact. Notably, causal survival forests and AIPTW-AIPCW forests perform well even when data are generated from simpler models, while other methods struggle. In these cases, the strong performance of BJ (Cox) and the G-formula (Cox/T-learners) is especially noteworthy. In an **observational data setting**, estimators that fail to account for confounders are, predictably, biased. However, users are generally aware of whether they are working within an RCT or an observational framework. In the case of independent censoring, the choice of nuisance parameter estimation becomes critical: Causal Survival Forest, AIPTW-AIPCW (Forest), and G-formula (Forest/T-learners) exhibit slow convergence when data are generated with Cox or logistic models. In this context, only the G-formula (Cox/T-learners) estimator proves to be truly effective. These observations become even more pronounced when censoring is dependent.

[JJ: finir commentaire récapitulatif des resulats par exemple sur les mispecifications]

In parametric settings, violations of positivity tend to have minimal impact for G-formula and IPTW-BJ estimators, as the extrapolation needed to handle these violations is feasible but risky. However, in non-parametric contexts, violating the positivity assumption can lead to erroneous extrapolations, making these estimators unsuitable for use.

Based on the simulations and theoretical results, it would be practical to consider moving away from the IPCW and IPTW-IPCW estimators, as they often exhibit excessive variability. Instead, we recommend implementing Causal Survival Forest, G-formula (T-learners), 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.

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.

[JJ: Donner d'autres perspectives: BJ semble plus interessant quand meme]

A key limitation of our simulations is the use of larger datasets with relatively simple relationships, which may not reflect the complexity of real-world scenarios. Most survival analysis datasets are smaller and more intricate, so the stability of certain estimators observed here may not fully generalize. It would be valuable to test these methods on real-world datasets to better assess their performance in practical applications.

Looking ahead, one promising avenue for improving these estimators is through optimizing variable selection for the conditional censoring, conditional survival, and treatment models. It has been shown that adding precision variables in causal inference enhances the variance of G-formula-like estimators. A potential area of exploration would be to investigate whether similar improvements can be made in causal survival analysis by refining the selection of covariates that influence censoring and survival outcomes.

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.

Bender, Ralf, Thomas Augustin, and Maria Blettner. 2005. "Generating Survival Times to Simulate Cox Proportional Hazards Models." *Statistics in Medicine* 24 (11): 1713–23. https://doi.org/https://doi.org/10.1002/sim.2059.

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/https://doi.org/10.1111/j.0006-341X.2001.01030.x.

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/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.

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/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.

Greenland, S., J. M. Robbins, and J. Pearl. 1999. "Confounding and Collapsibility in Causal Inference." *Statistical Science* 14: 29–46.

Hernán, Miguel A. 2010. "The Hazards of Hazard Ratios." *Epidemiology* 21 (1): 13–15.

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.

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.

Leemis, Lawrence M., Li-Hsing Shih, and Kurt Reynertson. 1990. "Variate Generation for Accelerated Life and Proportional Hazards Models with Time Dependent Covariates." *Statistics & Probability Letters* 10 (4): 335–39. https://doi.org/https://doi.org/10.1016/0167-7152(90)90052-9.

Martinussen, Torben. 2022. "Causality and the Cox Regression Model." Journal Article. *Annual Review of Statistics and Its Application* 9 (Volume 9, 2022): 249–59. https://doi.org/https://doi.org/10.1146/annurev-statistics-040320-114441.

Martinussen, Torben, and Stijn Vansteelandt. 2013. "On Collapsibility and Confounding Bias in Cox and Aalen Regression Models." *Lifetime Data Anal.* 19 (3): 279–96.

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/https://doi.org/10.1002/bimj.201800298.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

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.

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

"survRM2: Comparing Restricted Mean Survival Time." 2015. *CRAN: Contributed Packages*. The R Foundation. https://doi.org/10.32614/cran.package.survrm2.

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.

Willems, SJW, A Schat, MS van Noorden, and M Fiocco. 2018. "Correcting for Dependent Censoring in Routine Outcome Monitoring Data by Applying the Inverse Probability Censoring Weighted Estimator." *Statistical Methods in Medical Research* 27 (2): 323–35. https://doi.org/10.1177/0962280216628900.

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/https://doi.org/10.1002/sim.2174.

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 8. To show that $\hat{S}_{\mathrm{KM}}$ is unbiased, let us introduce $\mathscr{F}_k$ be the filtration generated by the set of variables

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

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 $k \geq 2$

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

where we used that $T_i(a)$ is idependant from A_i by Assumption 3. We then easily derive from this that

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

and then that

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

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:

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

So that we deduce that

$$\begin{aligned}n\mathrm{Var}\hat{S}_{\mathrm{KM}}(t|A = a) &= n\left(\mathbb{E}S_{\mathrm{KM}}(t|A = a)^2 - S^{(a)}(t)^2\right) \\ &= nS^{(a)}(t)^2\left(\mathbb{E}\left[\prod_{t_k \leq t}\left(1 + \frac{1 - s_k(a)}{s_k(a)N_k(a)} I\{N_k(a) > 0\} + O(I\{N_k(a) = 0\})\right)\right] - 1\right).\end{aligned}$$

Now using that $N_k(a) = nr_k(a) + \sqrt{n}O_{\mathbb{P}}(1)$ (with the $O_{\mathbb{P}}(1)$ having uniformly bounded variance), we find that

$$\begin{aligned}\prod_{t_k \leq t}\left(1 + \frac{1 - s_k(a)}{s_k(a)N_k(a)} I\{N_k(a) > 0\} + O(I\{N_k(a) = 0\})\right) &= \prod_{t_k \leq t}\left(1 + \frac{1}{n}\frac{1 - s_k(a)}{s_k(a)r_k(a)} + n^{-3/2}O_{\mathbb{P}}(1)\right) \\ &= 1 + \frac{1}{n}\sum_{t_k \leq t}\frac{1 - s_k(a)}{s_k(a)r_k(a)} + n^{-3/2}O_{\mathbb{P}}(1),\end{aligned}$$

which in turn yields (because the $O_{\mathbb{P}}(1)$ has still uniformly bounded variance):

$$n\mathrm{Var}\hat{S}_{\mathrm{KM}}(t|A = a) = V_{\mathrm{KM}}(t|A = a) + o(1),$$

which is what we wanted to show. □

7.2 Proofs of Section 2.2

Proof. (Proposition 2.3). Assumption 15 allows the tranformation to be well-defined. Furthermore, it holds

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

We used in the second equality that on the event $\{\Delta^\tau=1\}$, we have $\tilde{T}\wedge\tau = T\wedge\tau$. We used Assumption 1 in the second and last inequality to affirm that $T=T(a)$ on the event $A=a$. Finally, we used in the fourth equality that $G(T(a)\wedge\tau|A,X) = E[I\{T(a)\wedge\tau\le C\}|X,T(a),A]$ thanks to Assumption 14. □

Proof. (Proposition 2.5). Similarly to the computations done in the proof of Proposition 2.3, it is easy to show that

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

and likewise that

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

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

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

yielding strong consistency. Asymptotic normality is straightforward. □

Proof. (Proposition 2.7). There holds

$$\mathbb{E}[T_{\text{BJ}}|X,A] = \mathbb{E}\left[\Delta^\tau T + (1-\Delta^\tau)\frac{\mathbb{E}[T\wedge\tau\times I\{T\wedge\tau>C\}|C,A,X]}{\mathbb{P}(T>C|C,A,X)}\middle|X,A\right] \tag{36}$$

$$= \mathbb{E}[\Delta^\tau T\wedge\tau|X,A] + \mathbb{E}\left[I\{T\wedge\tau>C\}\frac{\mathbb{E}[T\wedge\tau\times I\{T\wedge\tau>C\}|C,A,X]}{\mathbb{E}[I\{T\wedge\tau\ge C\}|C,A,X]}\middle|X,A\right]. \tag{37}$$

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

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

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 16. 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 $\tilde{T} \wedge \tau = C \wedge \tau$ on $\{\Delta^\tau = 1\}$, the second term can be rewritten in the following way. We let R be a generic quantity that does not depend on ϕ_0.

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

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

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

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

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 2.5, one find, by first conditionning wrt $X, A, T(a)$, that

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

and likewise that

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

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

Proof. *(Proposition 3.5).* We write

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

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

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

For the second term in the RHS, notice that on the event $\{\Delta^\tau=0,A=1\}$, there holds $\tilde{T}=C\leq T(1)\wedge\tau$, so that

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

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 RMST. □

8 Appendix B

8.1 Descriptive statistics

8.1.1 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:   988"
```

```
       X1                  X2                  X3                  X4
 Min.   :-2.4644   Min.   :-1.946   Min.   :-4.3774   Min.   :-2.0386
 1st Qu.: 0.3555   1st Qu.: 0.332   1st Qu.:-1.6887   1st Qu.: 0.3742
 Median : 0.9725   Median : 1.047   Median :-1.0536   Median : 1.0192
 Mean   : 0.9961   Mean   : 1.034   Mean   :-1.0320   Mean   : 1.0178
 3rd Qu.: 1.6515   3rd Qu.: 1.761   3rd Qu.:-0.3837   3rd Qu.: 1.6777
 Max.   : 4.1133   Max.   : 4.484   Max.   : 2.7331   Max.   : 3.8889
       C                   T1                 T0               status
 Min.   :  0.00811   Min.   : 10.00   Min.   :  0.0023   Min.   :0.0000
 1st Qu.:  9.93841   1st Qu.: 12.36   1st Qu.:  2.3619   1st Qu.:0.0000
 Median : 23.32463   Median : 18.24   Median :  8.2433   Median :1.0000
 Mean   : 33.24563   Mean   : 29.36   Mean   : 19.3599   Mean   :0.6822
 3rd Qu.: 44.35759   3rd Qu.: 31.07   3rd Qu.: 21.0735   3rd Qu.:1.0000
 Max.   :220.78888   Max.   :431.24   Max.   :421.2415   Max.   :1.0000
     T_tild
 Min.   :  0.00231
 1st Qu.:  1.93105
 Median :  5.87472
 Mean   : 10.99796
 3rd Qu.: 13.58738
 Max.   :147.95345

[1] "Descriptive statistics for group A=1:   1012"

       X1                  X2                  X3                  X4
 Min.   :-2.5003   Min.   :-2.2885   Min.   :-3.9924   Min.   :-2.8426
 1st Qu.: 0.3649   1st Qu.: 0.3237   1st Qu.:-1.7140   1st Qu.: 0.3892
 Median : 1.0151   Median : 1.0076   Median :-1.0097   Median : 1.0374
 Mean   : 0.9794   Mean   : 1.0122   Mean   :-1.0177   Mean   : 1.0119
 3rd Qu.: 1.6083   3rd Qu.: 1.6618   3rd Qu.:-0.3398   3rd Qu.: 1.6164
 Max.   : 3.9174   Max.   : 4.1442   Max.   : 2.3581   Max.   : 3.9375
       C                   T1                 T0              status
 Min.   :  0.08514   Min.   : 10.00   Min.   :  0.002   Min.   :0.000
 1st Qu.:  8.96279   1st Qu.: 12.75   1st Qu.:  2.747   1st Qu.:0.000
 Median : 21.15034   Median : 18.67   Median :  8.667   Median :0.000
 Mean   : 32.58060   Mean   : 31.10   Mean   : 21.097   Mean   :0.496
 3rd Qu.: 44.32894   3rd Qu.: 30.28   3rd Qu.: 20.282   3rd Qu.:1.000
 Max.   :255.20092   Max.   :551.81   Max.   :541.807   Max.   :1.000
     T_tild
 Min.   :  0.08514
 1st Qu.:  8.96279
 Median : 12.81409
 Mean   : 16.00726
 3rd Qu.: 19.41034
 Max.   :212.97525
```

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:   1026"
```

```
       X1                  X2                  X3                  X4          
 Min.   :-2.2716   Min.   :-2.8883   Min.   :-4.0026   Min.   :-2.0497  
 1st Qu.: 0.3372   1st Qu.: 0.3769   1st Qu.:-1.6492   1st Qu.: 0.2268  
 Median : 0.9547   Median : 1.0265   Median :-0.9512   Median : 0.9378  
 Mean   : 0.9647   Mean   : 1.0255   Mean   :-0.9746   Mean   : 0.9250  
 3rd Qu.: 1.6365   3rd Qu.: 1.6419   3rd Qu.:-0.2882   3rd Qu.: 1.6178  
 Max.   : 3.9996   Max.   : 5.0666   Max.   : 2.1646   Max.   : 3.6106  
       C                   T1                  T0                status      
 Min.   :  0.00192   Min.   : 10.01   Min.   :  0.0098   Min.   :0.000  
 1st Qu.:  2.46808   1st Qu.: 13.36   1st Qu.:  3.3606   1st Qu.:0.000  
 Median :  7.05391   Median : 19.75   Median :  9.7488   Median :0.000  
 Mean   : 15.14802   Mean   : 31.66   Mean   : 21.6555   Mean   :0.423  
 3rd Qu.: 16.76798   3rd Qu.: 34.70   3rd Qu.: 24.7032   3rd Qu.:1.000  
 Max.   :282.42840   Max.   :529.91   Max.   :519.9052   Max.   :1.000  
   status_tau         T_tild          
 Min.   :0.000   Min.   :  0.00192  
 1st Qu.:0.000   1st Qu.:  1.32283  
 Median :0.000   Median :  4.12980  
 Mean   :0.462   Mean   :  8.30952  
 3rd Qu.:1.000   3rd Qu.:  9.80172  
 Max.   :1.000   Max.   :237.66249  
```

```
[1] "Descriptive statistics for group A=1:   974"
```

```
       X1                  X2                  X3                  X4          
 Min.   :-2.3569   Min.   :-2.9743   Min.   :-4.5163   Min.   :-1.9884  
 1st Qu.: 0.2814   1st Qu.: 0.3101   1st Qu.:-1.7605   1st Qu.: 0.3443  
 Median : 0.9748   Median : 1.0563   Median :-1.0139   Median : 0.9650  
 Mean   : 0.9662   Mean   : 1.0247   Mean   :-1.0407   Mean   : 0.9764  
 3rd Qu.: 1.5997   3rd Qu.: 1.7343   3rd Qu.:-0.3446   3rd Qu.: 1.6170  
 Max.   : 3.9569   Max.   : 4.0610   Max.   : 2.4304   Max.   : 4.1065  
       C                   T1                  T0                status      
 Min.   :  0.0258   Min.   : 10.00   Min.   :  0.0027   Min.   :0.0000  
 1st Qu.:  3.4013   1st Qu.: 12.61   1st Qu.:  2.6119   1st Qu.:0.0000  
 Median :  8.9837   Median : 18.38   Median :  8.3818   Median :0.0000  
 Mean   : 17.8350   Mean   : 31.21   Mean   : 21.2144   Mean   :0.2382  
 3rd Qu.: 19.7174   3rd Qu.: 33.05   3rd Qu.: 23.0468   3rd Qu.:0.0000  
 Max.   :430.0422   Max.   :473.98   Max.   :463.9799   Max.   :1.0000  
   status_tau          T_tild          
 Min.   :0.0000   Min.   :  0.02581  
 1st Qu.:0.0000   1st Qu.:  3.40130  
 Median :0.0000   Median :  8.98371  
 Mean   :0.2988   Mean   : 12.34652  
 3rd Qu.:1.0000   3rd Qu.: 14.84748  
 Max.   :1.0000   Max.   :190.03229  
```

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.

8.1.2 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:   1082"
```

```
       X1                X2                X3                X4
 Min.   :-2.394   Min.   :-2.2164   Min.   :-2.7210   Min.   :-1.8268
 1st Qu.: 0.548   1st Qu.: 0.5455   1st Qu.:-1.0444   1st Qu.: 0.6914
 Median : 1.239   Median : 1.2096   Median :-0.5087   Median : 1.3167
 Mean   : 1.240   Mean   : 1.2034   Mean   :-0.5057   Mean   : 1.2835
 3rd Qu.: 1.919   3rd Qu.: 1.8377   3rd Qu.: 0.0437   3rd Qu.: 1.9182
 Max.   : 4.399   Max.   : 4.7893   Max.   : 2.1439   Max.   : 4.5121
       C                  T1                T0               status
 Min.   :  0.0205   Min.   : 10.00   Min.   :  0.0023   Min.   :0.000
 1st Qu.: 10.1677   1st Qu.: 12.47   1st Qu.:  2.4712   1st Qu.:0.000
 Median : 24.0391   Median : 17.38   Median :  7.3772   Median :1.000
 Mean   : 33.1008   Mean   : 29.16   Mean   : 19.1581   Mean   :0.719
 3rd Qu.: 46.8047   3rd Qu.: 29.90   3rd Qu.: 19.9019   3rd Qu.:1.000
 Max.   :231.2026   Max.   :377.59   Max.   :367.5902   Max.   :1.000
    T_tild
 Min.   :  0.00225
 1st Qu.:  2.01065
 Median :  5.60056
 Mean   : 10.28857
 3rd Qu.: 12.92599
 Max.   :134.52845
```

```
[1] "Descriptive statistics for group A=1:   918"
```

```
       X1                X2                X3                X4
 Min.   :-2.5618   Min.   :-2.32950   Min.   :-4.489   Min.   :-2.10534
 1st Qu.: 0.1239   1st Qu.: 0.03293   1st Qu.:-2.199   1st Qu.: 0.09466
 Median : 0.7706   Median : 0.70441   Median :-1.606   Median : 0.72845
 Mean   : 0.7529   Mean   : 0.72167   Mean   :-1.645   Mean   : 0.73938
```

```
 3rd Qu.: 1.3872   3rd Qu.: 1.37002   3rd Qu.:-1.100   3rd Qu.: 1.38141
 Max.   : 4.4020   Max.   : 4.07447   Max.   : 1.188   Max.   : 3.78355
       C                  T1              T0              status
 Min.   :  0.01654   Min.   : 10.03   Min.   :  0.0283   Min.   :0.0000
 1st Qu.:  8.72284   1st Qu.: 13.12   1st Qu.:  3.1200   1st Qu.:0.0000
 Median : 20.89575   Median : 19.70   Median :  9.7046   Median :0.0000
 Mean   : 32.75458   Mean   : 36.19   Mean   : 26.1887   Mean   :0.4641
 3rd Qu.: 44.97785   3rd Qu.: 37.59   3rd Qu.: 27.5912   3rd Qu.:1.0000
 Max.   :235.55021   Max.   :503.33   Max.   :493.3260   Max.   :1.0000
    T_tild
 Min.   :  0.01654
 1st Qu.:  8.72284
 Median : 12.83592
 Mean   : 16.85740
 3rd Qu.: 20.36177
 Max.   :110.95169
```

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:   1095"

       X1                X2                X3                X4
 Min.   :-1.8652   Min.   :-1.6066   Min.   :-3.68385   Min.   :-1.7068
 1st Qu.: 0.6082   1st Qu.: 0.6446   1st Qu.:-1.04672   1st Qu.: 0.5201
 Median : 1.2159   Median : 1.2568   Median :-0.52517   Median : 1.1807
 Mean   : 1.2149   Mean   : 1.2555   Mean   :-0.50143   Mean   : 1.2110
 3rd Qu.: 1.8489   3rd Qu.: 1.8776   3rd Qu.: 0.04487   3rd Qu.: 1.9267
 Max.   : 4.7258   Max.   : 4.5065   Max.   : 2.25885   Max.   : 4.3384
       C                  T1              T0              status
 Min.   :  0.01021   Min.   : 10.00   Min.   :  0.0044   Min.   :0.0000
 1st Qu.:  2.17092   1st Qu.: 12.55   1st Qu.:  2.5480   1st Qu.:0.0000
 Median :  6.35684   Median : 17.56   Median :  7.5562   Median :0.0000
 Mean   : 13.36958   Mean   : 30.36   Mean   : 20.3602   Mean   :0.4393
 3rd Qu.: 15.32565   3rd Qu.: 30.17   3rd Qu.: 20.1717   3rd Qu.:1.0000
 Max.   :275.26050   Max.   :527.50   Max.   :517.5009   Max.   :1.0000
   status_tau         T_obs                e
 Min.   :0.0000   Min.   : 0.00437   Min.   :0.0000038
 1st Qu.:0.0000   1st Qu.: 1.11016   1st Qu.:0.0215818
 Median :0.0000   Median : 3.08897   Median :0.0977545
 Mean   :0.4694   Mean   : 6.60660   Mean   :0.1957585
 3rd Qu.:1.0000   3rd Qu.: 7.79404   3rd Qu.:0.3093209
 Max.   :1.0000   Max.   :90.75745   Max.   :0.9861897

[1] "Descriptive statistics for group A=1:   905"
```

```
       X1                  X2                  X3                 X4
 Min.   :-2.27195   Min.   :-2.43858   Min.   :-4.2450   Min.   :-2.3197
 1st Qu.: 0.06301   1st Qu.: 0.09218   1st Qu.:-2.1836   1st Qu.: 0.1054
 Median : 0.69634   Median : 0.79729   Median :-1.6027   Median : 0.7415
 Mean   : 0.70785   Mean   : 0.76903   Mean   :-1.6253   Mean   : 0.7351
 3rd Qu.: 1.39007   3rd Qu.: 1.45291   3rd Qu.:-1.0979   3rd Qu.: 1.3334
 Max.   : 4.17839   Max.   : 4.54814   Max.   : 0.6551   Max.   : 3.5652
       C                   T1                 T0               status
 Min.   :   0.0022   Min.   : 10.01   Min.   :  0.0077   Min.   :0.0000
 1st Qu.:   2.2449   1st Qu.: 13.02   1st Qu.:  3.0231   1st Qu.:0.0000
 Median :   6.6096   Median : 19.05   Median :  9.0495   Median :0.0000
 Mean   :  16.1463   Mean   : 33.94   Mean   : 23.9442   Mean   :0.1823
 3rd Qu.:  16.6410   3rd Qu.: 35.56   3rd Qu.: 25.5592   3rd Qu.:0.0000
 Max.   :1145.3422   Max.   :437.37   Max.   :427.3749   Max.   :1.0000
   status_tau          T_obs                 e
 Min.   :0.0000   Min.   :  0.00224   Min.   :0.01847
 1st Qu.:0.0000   1st Qu.:  2.24486   1st Qu.:0.59064
 Median :0.0000   Median :  6.60957   Median :0.84684
 Mean   :0.2354   Mean   : 10.64328   Mean   :0.74700
 3rd Qu.:0.0000   3rd Qu.: 13.56806   3rd Qu.:0.96053
 Max.   :1.0000   Max.   :112.88523   Max.   :0.99998
```

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.

8.1.3 Observational study with non-linear relationship

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

```
[1] "Descriptive statistics for group A=0:   1056"
```

```
       X1                  X2                  X3                C
 Min.   :-2.54709   Min.   :-2.91568   Min.   :-2.5305   Min.   : 0.000
 1st Qu.: 0.05991   1st Qu.: 0.04108   1st Qu.: 0.3415   1st Qu.: 1.000
 Median : 0.74140   Median : 0.69345   Median : 0.9930   Median : 2.000
 Mean   : 0.72414   Mean   : 0.72510   Mean   : 0.9827   Mean   : 2.439
 3rd Qu.: 1.38042   3rd Qu.: 1.38957   3rd Qu.: 1.6569   3rd Qu.: 3.250
 Max.   : 3.85426   Max.   : 4.03273   Max.   : 4.0144   Max.   :10.000
       T1               T0              status           T_obs
 Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   :0.000
 1st Qu.: 1.000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.000
 Median : 2.000   Median : 1.000   Median :1.0000   Median :1.000
 Mean   : 2.509   Mean   : 1.723   Mean   :0.7358   Mean   :1.189
 3rd Qu.: 4.000   3rd Qu.: 3.000   3rd Qu.:1.0000   3rd Qu.:2.000
 Max.   :13.000   Max.   :11.000   Max.   :1.0000   Max.   :6.000
```

```
   status_tau           e          
 Min.   :0.0000   Min.   :0.02114  
 1st Qu.:1.0000   1st Qu.:0.28005  
 Median :1.0000   Median :0.40927  
 Mean   :0.7955   Mean   :0.41366  
 3rd Qu.:1.0000   3rd Qu.:0.54331  
 Max.   :1.0000   Max.   :0.90419  
```

```
[1] "Descriptive statistics for group A=1:   944"
```

```
       X1                X2                X3                C          
 Min.   :-1.3471   Min.   :-1.6418   Min.   :-1.9722   Min.   :0.000  
 1st Qu.: 0.6845   1st Qu.: 0.6773   1st Qu.: 0.3601   1st Qu.:1.000  
 Median : 1.3020   Median : 1.2939   Median : 1.0152   Median :2.000  
 Mean   : 1.3004   Mean   : 1.2820   Mean   : 1.0107   Mean   :2.336  
 3rd Qu.: 1.9212   3rd Qu.: 1.9072   3rd Qu.: 1.6895   3rd Qu.:3.000  
 Max.   : 4.1689   Max.   : 3.9989   Max.   : 4.1720   Max.   :9.000  
       T1                T0              status            T_obs       
 Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   :0.000  
 1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.:0.0000   1st Qu.:1.000  
 Median : 3.000   Median : 2.000   Median :0.0000   Median :2.000  
 Mean   : 3.428   Mean   : 2.219   Mean   :0.4269   Mean   :1.779  
 3rd Qu.: 5.000   3rd Qu.: 3.000   3rd Qu.:1.0000   3rd Qu.:3.000  
 Max.   :14.000   Max.   :12.000   Max.   :1.0000   Max.   :7.000  
   status_tau           e          
 Min.   :0.0000   Min.   :0.1299  
 1st Qu.:0.0000   1st Qu.:0.4469  
 Median :1.0000   Median :0.5742  
 Mean   :0.5847   Mean   :0.5625  
 3rd Qu.:1.0000   3rd Qu.:0.6914  
 Max.   :1.0000   Max.   :0.9048  
```

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.

8.1.4 Observational study with interaction

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

```
       X1                X2                X3                 X4          
 Min.   :-2.8040   Min.   :-3.3269   Min.   :-2.40894   Min.   :-2.7028  
 1st Qu.:-0.2132   1st Qu.:-0.1704   1st Qu.: 0.06688   1st Qu.:-0.1792  
 Median : 0.4768   Median : 0.5100   Median : 0.76134   Median : 0.4917  
 Mean   : 0.4727   Mean   : 0.5058   Mean   : 0.73480   Mean   : 0.4938  
 3rd Qu.: 1.1413   3rd Qu.: 1.2000   3rd Qu.: 1.39093   3rd Qu.: 1.1727  
 Max.   : 4.3012   Max.   : 4.1397   Max.   : 4.11606   Max.   : 3.5114  
       X5                X6               tau              A          
```

```
Min.   :-3.7002   Min.   :-3.5151   Min.   :0.5   Min.   :0.000
1st Qu.:-0.8024   1st Qu.:-0.7782   1st Qu.:0.5   1st Qu.:0.000
Median :-0.0769   Median :-0.1050   Median :0.5   Median :1.000
Mean   :-0.1046   Mean   :-0.1122   Mean   :0.5   Mean   :0.627
3rd Qu.: 0.5898   3rd Qu.: 0.5706   3rd Qu.:0.5   3rd Qu.:1.000
Max.   : 3.2012   Max.   : 3.6200   Max.   :0.5   Max.   :1.000
      T0                T1               C                T_obs
Min.   :0.00000   Min.   :1.000   Min.   : 0.0000   Min.   :0.0000
1st Qu.:0.01060   1st Qu.:1.011   1st Qu.: 0.1057   1st Qu.:0.0354
Median :0.08564   Median :1.086   Median : 0.4024   Median :0.1792
Mean   :0.30152   Mean   :1.302   Mean   : 1.0571   Mean   :0.3884
3rd Qu.:0.35930   3rd Qu.:1.359   3rd Qu.: 1.0877   3rd Qu.:0.6174
Max.   :4.20179   Max.   :5.202   Max.   :49.0062   Max.   :3.8458
  T_obs_tau          status         status_tau     censor.status
Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.000
1st Qu.:0.0354   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000
Median :0.1792   Median :0.000   Median :1.0000   Median :1.000
Mean   :0.2368   Mean   :0.413   Mean   :0.5625   Mean   :0.587
3rd Qu.:0.5000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.000
Max.   :0.5000   Max.   :1.000   Max.   :1.0000   Max.   :1.000
      e
Min.   :0.07648
1st Qu.:0.50751
Median :0.67394
Mean   :0.64165
3rd Qu.:0.79933
Max.   :0.98338
```

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:   746"
```

```
      X1                X2                X3                C
Min.   :-2.8040   Min.   :-2.5874   Min.   :-2.4089   Min.   : 0.00011
1st Qu.:-0.2646   1st Qu.:-0.2258   1st Qu.:-0.4447   1st Qu.: 0.16511
Median : 0.4530   Median : 0.5224   Median : 0.1988   Median : 0.49179
Mean   : 0.4590   Mean   : 0.5005   Mean   : 0.1970   Mean   : 1.09255
3rd Qu.: 1.1193   3rd Qu.: 1.2010   3rd Qu.: 0.7974   3rd Qu.: 1.23762
Max.   : 4.0008   Max.   : 3.6887   Max.   : 3.0511   Max.   :49.00623
      T1               T0                status           T_obs
Min.   :1.000   Min.   :0.000000   Min.   :0.0000   Min.   :0.000000
1st Qu.:1.009   1st Qu.:0.008991   1st Qu.:1.0000   1st Qu.:0.007505
Median :1.089   Median :0.088579   Median :1.0000   Median :0.063862
Mean   :1.285   Mean   :0.284738   Mean   :0.7614   Mean   :0.164545
3rd Qu.:1.347   3rd Qu.:0.347138   3rd Qu.:1.0000   3rd Qu.:0.208099
Max.   :5.202   Max.   :4.201787   Max.   :1.0000   Max.   :2.238672
  status_tau            e
Min.   :0.0000   Min.   :0.07648
```

```
 1st Qu.:1.0000   1st Qu.:0.38644
 Median :1.0000   Median :0.54589
 Mean   :0.8029   Mean   :0.53447
 3rd Qu.:1.0000   3rd Qu.:0.68563
 Max.   :1.0000   Max.   :0.94391
```

```
[1] "Descriptive statistics for group A=1:   1254"
```

```
       X1                 X2                 X3                 C
 Min.   :-2.6470   Min.   :-3.3269   Min.   :-1.7054   Min.   : 0.00000
 1st Qu.:-0.1729   1st Qu.:-0.1252   1st Qu.: 0.4503   1st Qu.: 0.07898
 Median : 0.4905   Median : 0.5070   Median : 1.0455   Median : 0.33052
 Mean   : 0.4809   Mean   : 0.5089   Mean   : 1.0547   Mean   : 1.03601
 3rd Qu.: 1.1427   3rd Qu.: 1.1995   3rd Qu.: 1.6320   3rd Qu.: 1.03043
 Max.   : 4.3012   Max.   : 4.1397   Max.   : 4.1161   Max.   :41.67485
       T1              T0               status           T_obs
 Min.   :1.000   Min.   :0.00000   Min.   :0.0000   Min.   :0.000004
 1st Qu.:1.011   1st Qu.:0.01123   1st Qu.:0.0000   1st Qu.:0.078982
 Median :1.084   Median :0.08428   Median :0.0000   Median :0.330524
 Mean   :1.312   Mean   :0.31151   Mean   :0.2057   Mean   :0.521636
 3rd Qu.:1.374   3rd Qu.:0.37373   3rd Qu.:0.0000   3rd Qu.:1.000033
 Max.   :4.752   Max.   :3.75177   Max.   :1.0000   Max.   :3.845812
   status_tau            e
 Min.   :0.0000   Min.   :0.1388
 1st Qu.:0.0000   1st Qu.:0.6029
 Median :0.0000   Median :0.7337
 Mean   :0.4195   Mean   :0.7054
 3rd Qu.:1.0000   3rd Qu.:0.8342
 Max.   :1.0000   Max.   :0.9834
```

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
```