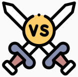
Estimation of ATE in Causal Survival: Comparison, Applications and Practical Recommendations

Julie Josse (Inria) & Bernard Sebastien (Sanofi R&D)

Charlotte VOINOT

September 3, 2024

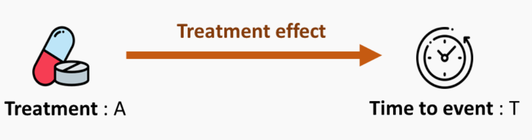
Causal Survival analysis: example of questions

Survival analysis  Causal inference

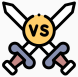
⇒ **Effect of a policy/intervention/treatment A on an time to event outcome T**

What is the impact of an oncology medicine on long term mortality ?

What is the impact of a medecine all along the time (it can be beneficial in the short term but harmful in the long term) ?



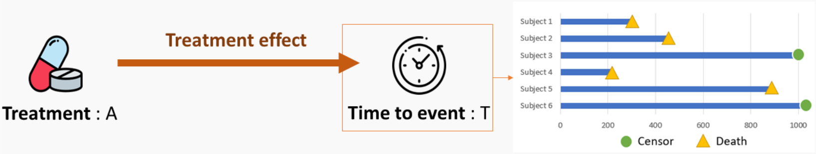
Causal Survival analysis: example of questions

Survival analysis  Causal inference

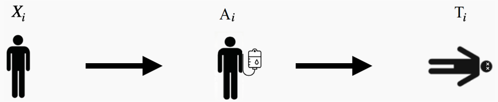
$\Rightarrow$ **Effect of a policy/intervention/treatment A on an time to event outcome T**

 What is the impact of an oncology medicine on long term mortality ?

 What is the impact of a medecine all along the time (it can be beneficial in the short term but harmful in the long term) ?



Potential outcomes



$$\Rightarrow n \textbf{ i.i.d. } (\underbrace{X_i}_{\textbf{covariates}}, \overbrace{A_i}^{\textbf{treatment}}, \underbrace{T_i}_{\textit{outcome}}) \in \mathbb{R}^d \times \{0,1\} \times \mathbb{R} \times \mathbb{R}$$

Let's say that in our example X_1 = sex and X_2 = age.

Covariates		**Treatment**	**Censoring**	**Status**	**Outcomes**		
X_1	X_2	**A**	**C**	Δ	**T(0)**	**T(1)**	$\tilde{T}$
1	24	1	?	1	?	200	200
2	52	0	?	1	100	?	100
1	33	1	200	0	?	?	200

In grey, the observed data : $(X_i, A_i, \Delta_i, \widetilde{T}_i)$ with $\widetilde{T}_i = min(T_i, C_i)$

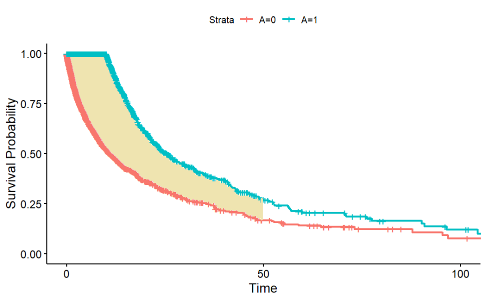
$\Rightarrow$ T is not directly observed

Causal effect in survival analysis

Difference in RMST : Average treatment effect in survival analysis

$$\hat{\theta}_{RMST}(\tau) = E[min(T(1), \tau) - min(T(0), \tau)]$$
$$= \int_0^\tau \left(\hat{S}_1(t) - \hat{S}_0(t)\right) dt$$

RMST can be defined as a measure of average survival from time 0 to time τ a **fixed time horizon**



$\hat{\theta}_{RMST}(\tau = 50) = 10$ means that on average the treatment increases the survival time by 10 days at 50 days.

Figure 1: Plot of stratified kaplan meier survival function and the representation of $\theta_{RMST}(\tau = 50)$ (in yellow)

Identifiability assumptions

S.T.U.V.A. $T = AT(1) + (1-A)T(0)$

RCT & Independent censoring

- **Random treatment assignment**
  $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Independent censoring**
  $C \perp\!\!\!\perp T(0), T(1), X, A$

Identifiability assumptions

S.T.U.V.A. $T = AT(1) + (1 - A)T(0)$

RCT & Independent censoring

- **Random treatment assignment**
  $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Independent censoring**
  $C \perp\!\!\!\perp T(0), T(1), X, A$

RCT & Dependent censoring

- **Random treatment assignment**
  $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Conditionally independent censoring**
  $C \perp\!\!\!\perp T(0), T(1) | X, A$
- **Positivity for censoring**
  $0 < P(C > t \mid X = x, A = a) < 1$

Identifiability assumptions

S.T.U.V.A. $T = AT(1) + (1-A)T(0)$

RCT & Independent censoring

- **Random treatment assignment**
  $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Independent censoring**
  $C \perp\!\!\!\perp T(0), T(1), X, A$

RCT & Dependent censoring

- **Random treatment assignment**
  $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Conditionally independent censoring**
  $C \perp\!\!\!\perp T(0), T(1) | X, A$
- **Positivity for censoring**
  $0 < P(C > t \mid X = x, A = a) < 1$

Obs & Independent censoring

- **Unconfoundedness** $A \perp\!\!\!\perp (T(0), T(1)) | X$
- **Positivity for treatment**
  $1 > P(A = a \mid X = x) > 0$
- **Independent censoring**
  $C \perp\!\!\!\perp T(0), T(1), X, A$

Identifiability assumptions

S.T.U.V.A. $T = AT(1) + (1 - A)T(0)$

RCT & Independent censoring

- **Random treatment assignment** $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Independent censoring** $C \perp\!\!\!\perp T(0), T(1), X, A$

Obs & Independent censoring

- **Unconfoundedness** $A \perp\!\!\!\perp (T(0), T(1))|X$
- **Positivity for treatment** $1 > P(A = a \mid X = x) > 0$
- **Independent censoring** $C \perp\!\!\!\perp T(0), T(1), X, A$

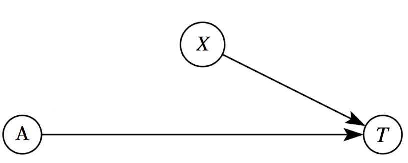
RCT & Dependent censoring

- **Random treatment assignment** $A \perp\!\!\!\perp (T(0), T(1), C, X)$
- **Conditionally independent censoring** $C \perp\!\!\!\perp T(0), T(1)|X, A$
- **Positivity for censoring** $0 < P(C > t \mid X = x, A = a) < 1$

Obs & Dependent censoring

- **Unconfoundedness** $A \perp\!\!\!\perp (T(0), T(1))|X$
- **Positivity for treatment** $1 > P(A = a \mid X = x) > 0$
- **Conditionally independent censoring** $C \perp\!\!\!\perp T(0), T(1)|X, A$
- **Positivity for censoring** $0 < P(C > t \mid X = x, A = a) < 1$

RCT & Independent censoring



Non-adjusted Kaplan-Meier estimator [Kaplan and Meier 1958]

It corresponds to a simple Kaplan meier estimator:

$$\hat{S}_{KM}(t \mid a) = \prod_{j=1, t_j <= t} \left(1 - \frac{\sum_i I\left\{ \tilde{T}_i = t_j, \Delta_i = 1, A_i = a \right\}}{\sum_k I\left\{ \tilde{T}_k \geq t_j, A_i = a \right\}} \right)$$

$$\hat{\theta}_{RMST}(\tau) = \int_0^\tau \left(\hat{S}_1(t) - \hat{S}_0(t) \right) dt$$

RCT & Dependent censoring

Non-adjusted Kaplan-Meier is biased in this context:

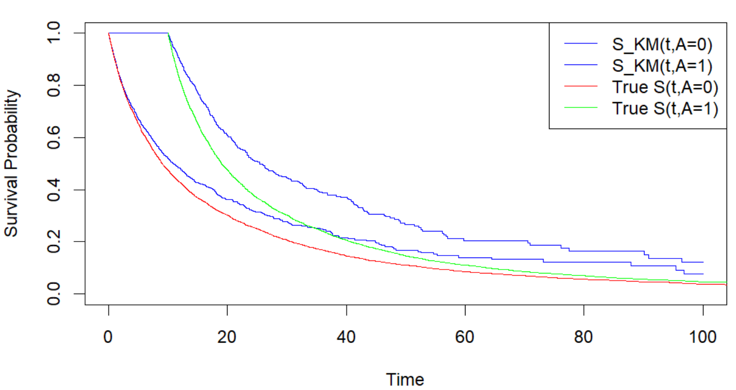


Figure 2: Plot of stratified kaplan meier survival function (A=1 and A=0) and the true survival function

⇒ The probability of survival is no longer consistent in using Non adjusted KM [Willems et al. 2018].

RCT & Dependent censoring

Notion of censoring unbiased transformation [Fan and Gijbels 1994]:

$$T^* = \Delta\phi_1(\mathbf{X}, A, \tilde{T}) + (1-\Delta)\phi_2(\mathbf{X}, A, \tilde{T})$$

The basic requirement is that $E(T^*|X, A) = E(T \wedge \tau|X, A)$.

- **IPC transformation** [Koul, Susarla, and Ryzin 1981]

$$T^*(\tau) = \frac{\widetilde{T} \wedge \tau * \Delta^\tau}{S_c(\widetilde{T} \wedge \tau|X, A)}$$

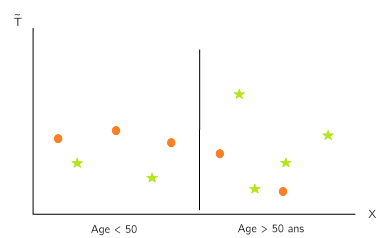
with $\hat{S}_c(t|X_i, A_i)$ is the probability of remain uncensored given the covariates and $\Delta^\tau = \mathbb{1}(\{\widetilde{T} > \tau\}) + \mathbb{1}(\{\widetilde{T} \leq \tau\}) \cdot \Delta$ is the censoring indicator of the restricted time (or restricted status).

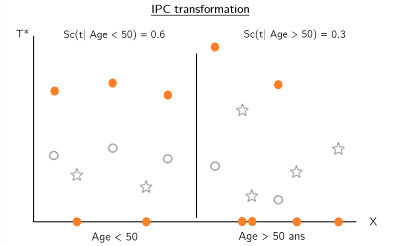
- **Buckley-James transformation** [Buckley and James 1979]

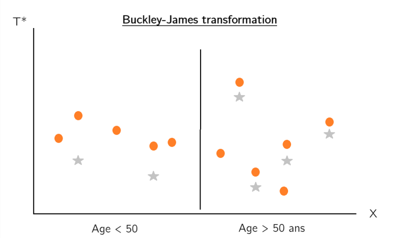
$$T^*(\tau) = \Delta^\tau * (\widetilde{T} \wedge \tau) + (1-\Delta^\tau) * \mathbb{E}[T \wedge \tau|X, A, T \wedge \tau > \widetilde{T} \wedge \tau]$$

RCT & Dependent censoring

Notion of censoring unbiased transformation [Fan and Gijbels 1994]:







RCT & Dependent censoring

IPCW adjusted Kaplan-Meier estimator [Robins and Finkelstein 2000]

$$\hat{S}_{IPCW}(t \mid A = a) = \prod_{j=1, t_j \leq t} \left(1 - \frac{\sum_i \hat{w}_i(t_j, X_i) \cdot I\left\{ \widetilde{T}_i = t_j, C_i \geq t_j, A_i = a \right\}}{\sum_k \hat{w}_k(t_j, X_k) \cdot I\left\{ \widetilde{T_k} \geq t_j, C_k \geq t_j, A_k = a \right\}} \right)$$

with $\hat{w}_i(t, X_i) = \frac{\Delta_i^\tau}{\hat{S}_c(t|X_i, A_i)}$: every uncensored observation is weighted by the inverse of the probability of remain uncensored given the covariates.

$$\hat{\theta}_{RMST} = \int_0^\tau \hat{S}_{IPCW}(t, A = 1) - \hat{S}_{IPCW}(t, A = 0) dt$$

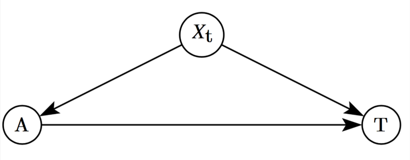
RCT & Dependent censoring

BJ based estimator

$$\hat{\theta}_{RMST} = \frac{1}{n_1} * \sum_{i=1}^{n_1} \left[\Delta_i^\tau * (\widetilde{T}_i \wedge \tau) + (1 - \Delta_i^\tau) * \hat{Q}_S(C_i | X, A) \mid A = 1 \right] - \frac{1}{n_0} * \sum_{j=1}^{n_0} \left[\Delta_j^\tau * (\widetilde{T}_j \wedge \tau) + (1 - \Delta_j^\tau) * \hat{Q}_S(C_j | X, A) \mid A = 0 \right]$$

- n_1 corresponds to the number of observations in the treated group
- n_2 corresponds to the number of observations in the control group
- $\hat{Q}_S(\widetilde{T} \wedge \tau \mid X, A) = \frac{1}{\hat{S}(\widetilde{T} \wedge \tau | X, A)} \int_{\widetilde{T} \wedge \tau}^{+\infty} \widetilde{T} \wedge \tau . d\hat{F}(\widetilde{T} \wedge \tau \mid X, A)$ the estimation function of the remaining survival function

Obs & Independent censoring



Risk of confounding bias

$\Rightarrow$ Need for balancing differences between the two groups.

IPTW adjusted Kaplan-Meier estimator [Xie and Liu 2005]

$$\hat{S}_{IPTW}(t \mid A=a) = \prod_{j=1, t_j \leq t} \left(1 - \frac{\sum_i \hat{w}_i(t_j, X_i) \cdot I\left\{ \widetilde{T}_i = t_j, C_i \geq t_j, A_i = a \right\}}{\sum_k \hat{w}_k(t_j, X_k) \cdot I\left\{ \widetilde{T_k} \geq t_j, C_k \geq t_j, A_k = a \right\}} \right)$$

with $\hat{w}_i(t, X_i) = \frac{A_i}{\hat{e}(X_i)} + \frac{1-A_i}{1-\hat{e}(X_i)}$

$$\hat{\theta}_{RMST} = \int_0^{\tau} \hat{S}_{IPTW}(t, A=1) - \hat{S}_{IPTW}(t, A=0) dt$$

Obs & Dependent censoring

Risk of censoring bias & confounding bias

$\Rightarrow$ Overcome by using IPC transformation

and IPT weighting

IPTW-IPCW adjusted Kaplan-Meier estimator

$$\hat{S}_{IPTW-IPCW}(t \mid A=a) = \prod_{j=1, t_j <= t} \left(1 - \frac{\sum_i \hat{w}_i(t, X_i) * I\{T_i = t_j, C_i \geq t_j, A_i = a\}}{\sum_i \hat{w}_i(t, X_i) * I\{T_i \geq t_j, C_i \geq t_j, A_i = a\}}\right)$$

with $\hat{w}_i(t, X_i) = \frac{\Delta_i^\tau}{\hat{S}_C\left(\widetilde{T} \wedge \tau | A_i, X_i\right)} * \left(\frac{A_i}{\hat{e}(X_i)} + \frac{1-A_i}{1-\hat{e}(X_i)}\right)$: every uncensored observation is weighted by the inverse of remain uncensored and by the inverse propensity score given the covariates.

$$\hat{\theta}_{RMST} = \int_0^\tau \hat{S}_{IPTW-IPCW}(t, A=1) - \hat{S}_{IPTW-IPCW}(t, A=0) dt$$

Obs & Dependent censoring

Risk of censoring bias & confounding bias
⇒ In using Buckley-James transformation
and IPT weighting

IPTW-BJ estimator

$$\hat{\theta}_{\mathrm{IPTW-BJ}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\Delta^{\tau}\widetilde{T}\wedge\tau + (1-\Delta^{\tau})\hat{Q}_S(\widetilde{T}\wedge\tau|X,A)\right)\left(\frac{A_i}{\hat{e}(X_i)} - \frac{1-A_i}{1-\hat{e}(X_i)}\right)$$

Obs & Dependent censoring

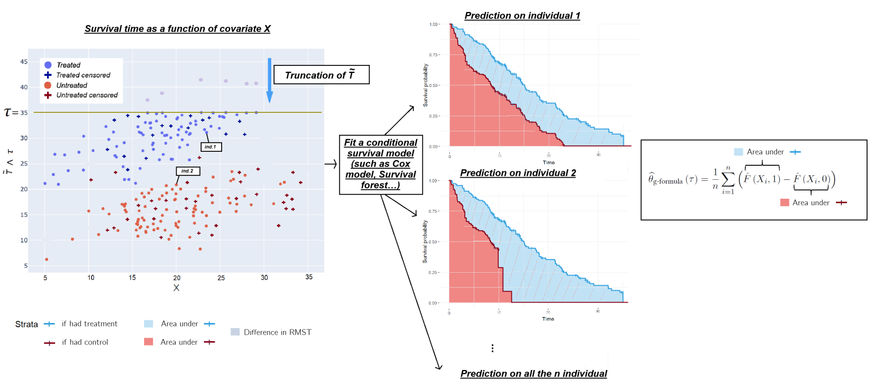
G-formula estimator

$$\widehat{\theta}_{\text{g-formula}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{F}\left(X_i, 1\right) - \hat{F}\left(X_i, 0\right)\right)$$

with $\hat{F}(x, a) \triangleq \mathbb{E}[T \wedge \tau \mid X = x, A = a] = \int_0^\tau S(t|X = x, A = a)$ the integral of the conditional survival function truncated at τ.

- G-formula compute the θ_{RMST} for each individual based on its covariates $\rightarrow$ can be estimated by Cox model, survival forest or parametric model (such as Weibull model, exponential model...).
- Then, it does the mean of the conditional θ_{RMST} on all individuals.

Obs & Dependent censoring



Two possibilities for estimating $\hat{S}(t|x,a)$:

- S-learner: **Fit one model** on all data with covariate ajdustment on X and A.
- T-learner: **Fit two models** (stratified analysis) on data with A=1 and with A=0 and adjustment on X.

Obs & Dependent censoring

Augmented version of IPTW-IPCW [Ozenne et al. 2020], a mix of :

$\rightarrow$ AIPTW (the equivalent of AIPW in causal inference)

$\rightarrow$ An other unbiased censoring transformation (AIPCW)

Augmented version of IPTW-IPCW [Ozenne et al. 2020]:

- **AIPTW** (Augmented Inverse Probability of Treatment Weighting [James M. Robins and Zhao 1994; James M. Robins and Zhao 1995; Chernozhukov et al. 2016]):

We consider **complete observation**, as the transformation will estimate it.

$$
\begin{aligned}
\theta_{AIPTW} &= \mathbb{E}[\mathbb{E}(T \wedge \tau | X, A = 1)] - \mathbb{E}[\mathbb{E}(T \wedge \tau | X, A = 0)] \\
&= \mathbb{E}\left[\frac{A.T \wedge \tau}{\hat{e}(X)} + \hat{F}(X, A = 1) * \frac{\hat{e}(X) - A}{\hat{e}(X)}\right] - \\
&\quad \mathbb{E}\left[\frac{(1 - A) * T \wedge \tau}{1 - \hat{e}(X)} + \hat{F}(X, A = 0) * \frac{(1 - \hat{e}(X)) - A}{1 - \hat{e}(X)}\right]
\end{aligned}
$$

Augmented version of IPTW-IPCW [Ozenne et al. 2020]:

- **AIPCW** (Augmented Inverse Probability Censoring Weighting [Rubin and Laan 2007]):

$$T^\star(O) = \frac{\widetilde{T} \wedge \tau \Delta^\tau}{S_c(\widetilde{T} \wedge \tau \mid X, A)} + \frac{Q_S(\widetilde{T} \wedge \tau \mid X, A)(1 - \Delta^\tau)}{S_c(\widetilde{T} \wedge \tau \mid X, A)} - \int_{-\infty}^{\widetilde{T} \wedge \tau} \frac{Q_S(\widetilde{T} \wedge \tau \mid X, A)}{S_c^2(\widetilde{T} \wedge \tau \mid X, A)} \, d(1 - S_c(\widetilde{T} \wedge \tau \mid X, A))$$

- The first term: IPC transformation
- The second term: BJ transformation
- The third one is an augmentation term

Obs & Dependent censoring

Augmented estimator: AIPTW-AIPCW

$$\hat{\theta}_{\text{AIPTW-AIPCW}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{\hat{e}\left(X_i\right)} - \frac{1-A_i}{1-\hat{e}\left(X_i\right)}\right)\hat{T}^*_{\text{DR}} + \hat{F}\left(X_i, A=1\right)\left(1-\frac{A_i}{\hat{e}\left(X_i\right)}\right) - \hat{F}\left(X_i, A=0\right)\left(1-\frac{1-A_i}{1-\hat{e}\left(X_i\right)}\right)$$

with $\hat{T}^*_{\text{DR}} = \frac{\widetilde{T}_i \wedge \tau \cdot \Delta^{\tau}}{\hat{S}_C\left(\widetilde{T}_i \wedge \tau | X_i\right)} + \frac{Q_{\hat{S}}(\widetilde{T}_i \wedge \tau | X, A)\cdot(1-\Delta^{\tau})}{\hat{S}_C(\widetilde{T}_i \wedge \tau | X_i)} - \int_0^{\widetilde{T}_i \wedge \tau} \frac{Q_{\hat{S}}(c|X_i, A_i)}{\hat{S}^2_C(c|X_i, A_i)} d\hat{S}_C(c \mid X_i, A_i)$ and

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

$\Rightarrow$ 3 nuisance parameters to compute :

- Censoring model : $C \sim A + X$
- Propensity score model : $A \sim X$
- Conditional survival : $T \sim A + X$

Obs & Dependent censoring

Estimator	mis. outcome model	mis. censoring model	mis. treatment model	mis. outcome and censoring	mis. outcome and treatment	mis. censoring and treatment
Unadjusted KM						
IPCW-KM		☒				
BJ	☒					
IPTW-KM			☒			
IPTW-IPCW		☒	☒			
G-formula	☒					
IPTW-BJ	☒		☒			
AIPTW-AIPCW	✓	✓	✓	☒	☒	✓

Table 1: Consistency of estimator under model mis-specification. When all the nuisances models are mis-specified none of the estimators is consistent. ✓ indicates consistency of the estimator, ☒ indicates non consistency and empty box means there is no need to compute it.

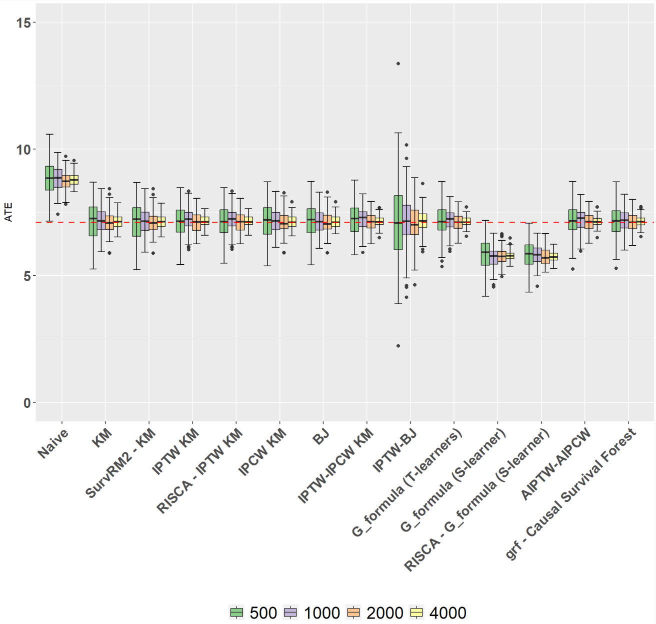
Simulations: Estimation of nuisance models

- 2 RCT (Ind. censoring & Dep. censoring), 2 Obs (Ind. censoring & Dep. censoring) in a context of **parametric simulation**.
- 1 Obs (Dep. censoring) in a context of **non-parametric simulation**.

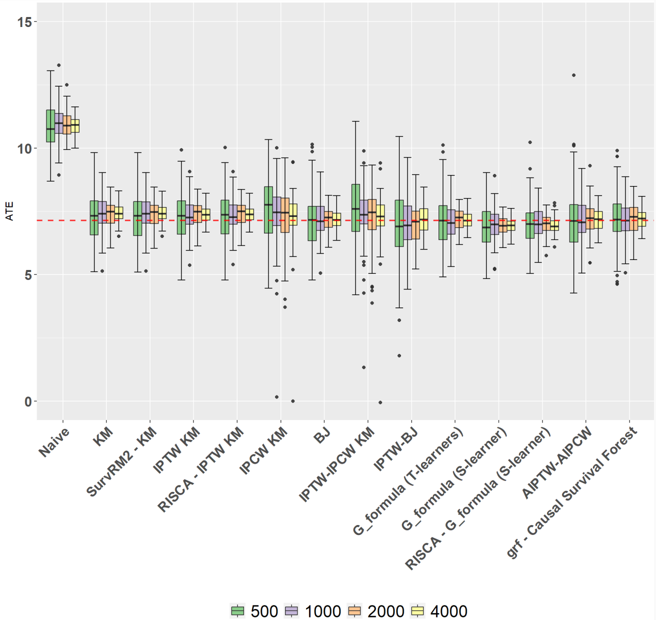
	Parametric simulation	**Non-parametric simulation**
Survival model	Cox model	Survival forest
Censoring model	Cox model	Survival forest
Propensity model	Logistic regression	Probability forest

Table 2: Model used for nuisance parameter estimation in the different simulation

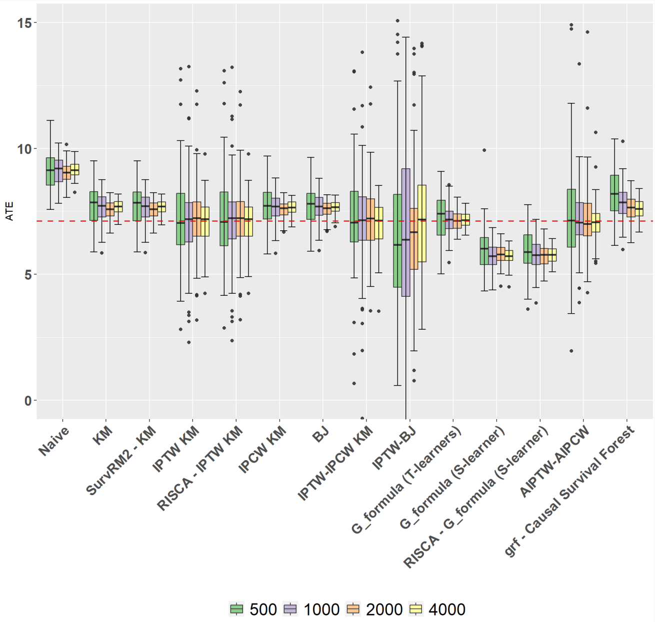
Parametric simulation: RCT & ind. censoring



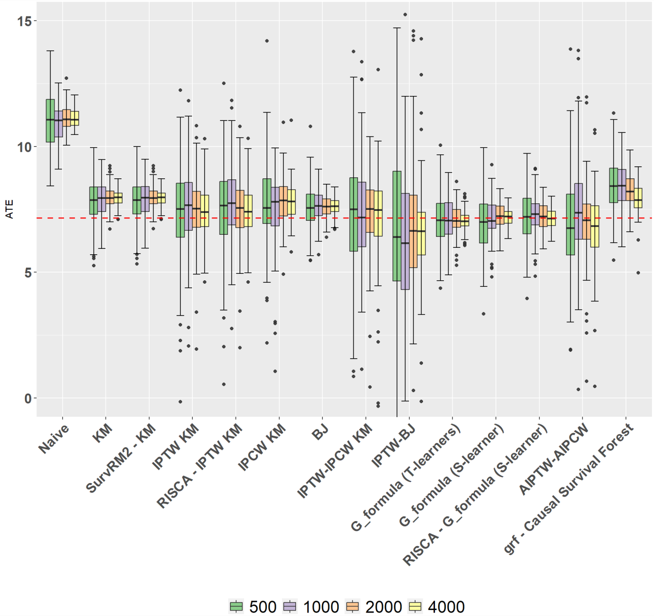
Parametric simulation: RCT & dep. cens



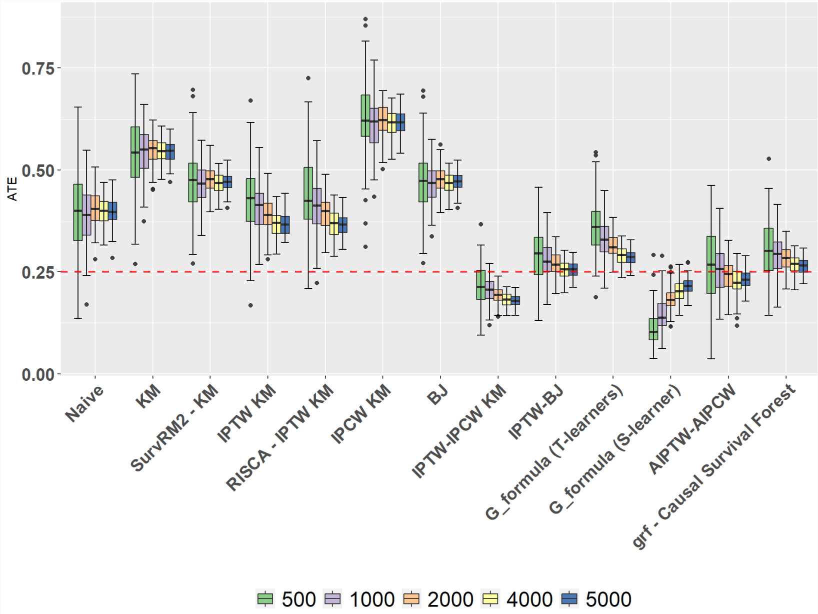
Parametric simulation: Obs & ind. censoring



Parametric simulation: Obs & dep.censoring



Non parametric simulation: Obs & dep.censoring



Conclusions:

- RMST is one extension of causal inference in survival analysis.
- The choice of the time horizon τ impacts the assumption of censoring positivity.

PARAMETRIC SIMULATION

- Robust estimators such as AIPTW-AIPCW is outperformed by G-formula estimator.
- BUT G-formula is not robust to mis-specification.
- AND in parameteric setting, G-formula relies on Cox model (proportional hazard assumption).

NON PARAMETRIC SIMULATION

- IPTW-BJ, AIPTW-AIPCW and Causal Survival Forest outperform the other in non-parametric setting.
- BUT IPTW-BJ is not robust to mis-specification.
- AIPTW-AIPCW and Causal Survival Forest need more data due to their complexity.

References

Buckley and James (Dec. 1979). **"Linear regression with censored data".** In: *Biometrika* 66.3, pp. 429–436. ISSN: 0006-3444. DOI: 10.1093/biomet/66.3.429. eprint: https://academic.oup.com/biomet/article-pdf/66/3/429/631183/66-3-429.pdf. URL: https://doi.org/10.1093/biomet/66.3.429.

Chernozhukov, Victor et al. (2016). **Double/Debiased Machine Learning for Treatment and Causal Parameters.** eprint: arXiv:1608.00060.

Fan, Jianqing and Irene Gijbels (1994). **Local polynomial modelling and its applications.** URL: https://api.semanticscholar.org/CorpusID:118744555.

James M. Robins, Andrea Rotnitzky and Lue Ping Zhao (1994). **“Estimation of Regression Coefficients When Some Regressors are not Always Observed”.** In: *Journal of the American Statistical Association* 89.427, pp. 846–866. DOI: 10.1080/01621459.1994.10476818. eprint: https://doi.org/10.1080/01621459.1994.10476818. URL: https://doi.org/10.1080/01621459.1994.10476818.

— (1995). **“Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data”.** In: *Journal of the American Statistical Association* 90.429, pp. 106–121. DOI: 10.1080/01621459.1995.10476493. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1995.10476493. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476493.

Kaplan, E. L. and Paul Meier (1958). **"Nonparametric Estimation from Incomplete Observations".** In: *Journal of the American Statistical Association* 53.282, pp. 457–481. DOI: 10.1080/01621459.1958.10501452. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501452. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452.

Koul, H., V. Susarla, and J. Van Ryzin (1981). **"Regression Analysis with Randomly Right-Censored Data".** In: *The Annals of Statistics* 9.6, pp. 1276–1288. DOI: 10.1214/aos/1176345644. URL: https://doi.org/10.1214/aos/1176345644.

Ozenne, Brice Maxime Hugues et al. (2020). **"On the estimation of average treatment effects with right-censored time to event outcome and competing risks".** In: *Biometrical Journal* 62.3, pp. 751–763. DOI: https://doi.org/10.1002/bimj.201800298. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201800298. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201800298.

Robins, James M. and Dianne M. Finkelstein (2000). **"Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests".** In: *Biometrics*.

Rubin, Daniel and Mark J. van der Laan (2007). In: *The International Journal of Biostatistics* 3.1. DOI: doi:10.2202/1557-4679.1052. URL: https://doi.org/10.2202/1557-4679.1052.

Willems, SJW et al. (2018). **“Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator”.** In: *Statistical Methods in Medical Research*.

Xie, Jun and Chaofeng Liu (2005). **“Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data”.** In: *Statistics in Medicine*.

Appendix

RMST and survival probability

Why $\theta_{RMST}(\tau) = E[min(T(1), \tau) - min(T(0), \tau)]$ and $\theta_{RMST}(\tau) = \int_0^\tau (S_1(t) - S_0(t))\, dt$ are equal ?

$E(T \wedge \tau) = E(\int_0^{T\wedge\tau} 1dt) = E(\int_0^\tau I\{T > t\}dt) = \int_0^\tau E(I\{T > t\})dt = \int_0^\tau S(t)dt$

Non adjusted Kaplan meier

$$
\begin{aligned}
\theta_{RMST} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\
&= \int_0^\tau \mathbb{E}[I\{T(1) > t\} - I\{T(0) > t\}]dt && \text{(By definition)} \\
&= \int_0^\tau \mathbb{E}[I\{T(1) > t\}] - \mathbb{E}[I\{T(0) > t\}]dt && \text{(By linearity)} \\
&= \int_0^\tau \mathbb{E}[I\{T(1) > t|A = 1\}] - \mathbb{E}[I\{T(0) > t|A = 0\}]dt && \text{(Random treatment assignm} \\
&= \int_0^\tau \mathbb{E}[I\{T > t|A = 1\}] - \mathbb{E}[I\{T > t|A = 0\}]dt && \text{(By consistency)} \\
&= \int_0^\tau \mathbb{P}(T > t|A = 1) - \mathbb{P}(T > t|A = 0)dt \\
&= \int_0^\tau S(t|A = 1) - S(t|A = 0)dt
\end{aligned}
$$

IPCW Kaplan meier

$$
\begin{aligned}
\theta_{RMST} &= \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau] \\
&= \int_0^\tau \mathbb{E}[I\{T(1) > t\}] - \mathbb{E}[I\{T(0) > t\}] dt \\
&= \int_0^\tau \mathbb{E}\left[\mathbb{E}[I\{T > t\}|A = 1, X]\right] - \mathbb{E}\left[\mathbb{E}[I\{T > t\}|A = 0, X]\right] \\
&= \int_0^\tau \mathbb{E}\left[\frac{\mathbb{E}[I\{T > t|A = 1, X\}] * \mathbb{E}[I\{T \wedge \tau < C|A = 1, X\}]}{S_c(T(1) \wedge \tau|X, A = 1)}\right] - \\
&\mathbb{E}\left[\frac{\mathbb{E}[I\{T > t|A = 0, X\}] * \mathbb{E}[I\{T \wedge \tau < C|A = 0, X\}]}{S_c(T(0) \wedge \tau|X, A = 0)}\right] dt \\
&= \int_0^\tau \mathbb{E}\left[\frac{\mathbb{E}[I\{T > t|A = 1, X\} * I\{T \wedge \tau < C|A = 1, X\}]}{S_c(T \wedge \tau|X, A = 1)}\right] - \\
&\mathbb{E}\left[\frac{\mathbb{E}[I\{T > t|A = 0, X\} * I\{T \wedge \tau < C|A = 0, X\}]}{S_c(T \wedge \tau|X, A = 0)}\right] dt \\
&= \int_0^\tau \mathbb{E}\left[\frac{I\{T > t|A = 1\} * \Delta^\tau}{S_c(T \wedge \tau|X, A = 1)}\right] - \mathbb{E}\left[\frac{I\{T > t|A = 0\} * \Delta^\tau}{S_c(T \wedge \tau|X, A = 0)}\right] dt
\end{aligned}
$$

IPTW Kaplan meier

$$\theta = \mathbb{E}[T(1) \wedge \tau - T(0) \wedge \tau]$$

$$= \int_0^\tau \mathbb{E}[I\{T(1) > t\}] - \mathbb{E}[I\{T(0) > t\}]dt$$

(By linearity)

$$= \int_0^\tau \mathbb{E}\left[\mathbb{E}[I\{T(1) > t\}|X]\right] - \mathbb{E}\left[\mathbb{E}[I\{T(0) > t\}|X]\right]$$

(Law of total probability and Consistency)

$$= \int_0^\tau \mathbb{E}\left[\frac{\mathbb{E}[I\{T(1) > t|X\}] * \color{blue}{\mathbb{E}[A|X\}]}}{\color{blue}{e(X)}}\right] - \mathbb{E}\left[\frac{\mathbb{E}[I\{T(0) > t|X\}] * \color{red}{\mathbb{E}[1 - A|X\}]}}{\color{red}{1 - e(X)}}\right] dt$$

(In color, the terms are equal)

$$= \int_0^\tau \mathbb{E}\left[\frac{\mathbb{E}[I\{T(1) > t\} * A|X\}}{e(X)}\right] - \mathbb{E}\left[\frac{\mathbb{E}[I\{T(0) > t\} * (1 - A)|X]]}{1 - e(X)}\right] dt$$

(By unconfoundedness)

$$= \int_0^\tau \frac{\mathbb{E}[I\{T(1) > t\} * A]}{e(X)} - \frac{\mathbb{E}[I\{T(0) > t\} * (1 - A)]}{1 - e(X)} dt$$

(Law of total probability)

$$= \int_0^\tau \frac{\mathbb{E}[I\{T(1) > t|A = 1\} * A]}{e(X)} - \frac{\mathbb{E}[I\{T(0) > t|A = 0\} * (1 - A)]}{1 - e(X)} dt$$

$$= \int_0^\tau \frac{\mathbb{E}[I\{T > t|A = 1\} * A]}{e(X)} - \frac{\mathbb{E}[I\{T > t|A = 0\} * (1 - A)]}{1 - e(X)} dt$$

(By consistency)

$$= \int_0^\tau p(T \geq t|A = 1) * \left(\frac{A}{e(X)}\right) - p(T \geq t|A = 0) * \left(\frac{1 - A}{1 - e(X)}\right) dt$$

RCTs simulations

For the simulation, 2000 samples $(X_i, A_i, C, T_i(0), T_i(1))$ are generated in the following way:

- $X \sim \mathcal{N}\left(\mu = [1, 1, -1, 1]^\top, \Sigma = I_4\right)$

- $e(X) = 0.5$ (constant) for the propensity score (A)

- $\lambda(0)(X) = 0.01 \cdot \exp\{0.5X_1 + 0.5X_2 - 0.5X_3 + 0.5X_4\}$ hazard for the event time $T(0)$

- The hazard for the censoring time C:

  - Scenario 1: $\lambda_c = 0.03$.
  - 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 event time is $T = AT(1) + (1 - A)T(0)$

- The observed time is $\widetilde{T} = \min(T, C)$

- The status is $\Delta = 1(T \leq C)$

Obs simulations

For the simulation, 2000 samples $(X_i, A_i, C, T_i(0), T_i(1))$ are generated in the following way:

- $X \sim \mathcal{N}\left(\mu = [1, 1, -1, 1]^\top, \Sigma = I_4\right)$

- $\text{logit}\{e(X)\} = -1X_1 - 1X_2 + 2.5X_3 - 1X_4$ for the propensity score (A)

- $\lambda(0)(X) = 0.01 \cdot \exp\{0.5X_1 + 0.5X_2 - 0.5X_3 + 0.5X_4\}$ hazard for the event time $T(0)$

- The hazard for the censoring time C:

  - Scenario 1:$\lambda_c = 0.09$.
  - Scenario 2:$\lambda_c(X) = 0.03 \cdot \exp\{0.7X_1 + 0.3X_2 - 0.25X_3 - 0.1X_4\}$.

- $T(1) = T(0) + 10$

- the event time is $T = AT(1) + (1 - A)T(0)$

- The observed time is $\widetilde{T} = \min(T, C)$

- The status is $\Delta = 1(T \leq C)$