Variable Selection in Causal Survival

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

Charlotte VOINOT
PhD student at Sanofi R&D and Inria

Journées de Statistique

June 5, 2025

Plan

Introduction to Causal Inference

Variable selection: a fundamental issue

Towards causal survival: new challenges

Preliminary results on variable selection and perspectives

Causal Inference: Framework and Estimands

Main Goal

Deriving the causal treatment effect

Covariates		Treatment	Potential Outcomes		Observed Outcome
X_1	X_2	A	$Y(0)$	$Y(1)$	Y
1	24	1	?	200	200
2	52	0	100	?	100

$\Rightarrow$ Potential outcome framework (Rubin 1974)

Average Treatment Effect (ATE)

$$\tau_{\text{ATE}} = \mathbb{E}[Y(1) - Y(0)]$$

$\Rightarrow$ Depends on the population.

The ATE is the difference in average outcomes between a world where everyone is treated and a world where no one is.

Plan

Simple causal graph of observational study

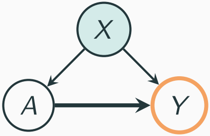
Causal effect can be derived in adjusting on minimum adjustment set.

Minimum adjustment set

The **confounders** are considered as the minimum adjustment.

It derives from **Uncounfondedness** assumption:

$A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X;$



How to find the best adjustment set ?

Simple causal graph of observational study

Causal effect can be derived in adjusting on minimum adjustment set.

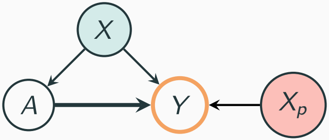
Minimum adjustment set

The **confounders** are considered as the minimum adjustment.

It derives from **Uncounfondedness** assumption:

$A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$;

But it can exist other variables such as **outcome related variables** (also known as risk factors or precision variables).



How to find the best adjustment set ?

Simple causal graph of observational study

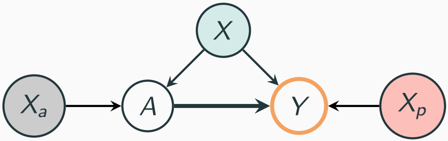
Causal effect can be derived in adjusting on minimum adjustment set.

Minimum adjustment set

The **confounders** are considered as the minimum adjustment.
It derives from **Uncounfondedness** assumption:
$A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$;

But it can exist other variables such as **outcome related variables** (also known as risk factors or precision variables).

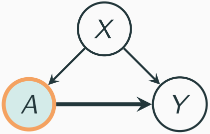
Or **variables predictive for the treatment only** (Instrumental variables).



How to find the best adjustment set ?

Recommendations for IPTW (treatment model)

Goal: *Achieve a reliable estimate of the causal effect.*



Identifiability

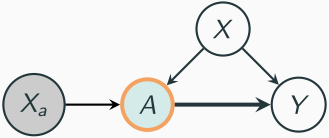
$$\mathbb{E}[Y_i(1) - Y_i(0)]$$
$$= \mathbb{E}[\mathbb{E}[Y_i(1) \mid X_i] - \mathbb{E}[Y_i(0) \mid X_i]]$$
$$= \mathbb{E}\left[\frac{\mathbb{E}[A_i \mid X_i] \cdot \mathbb{E}[Y_i(1) \mid X_i]}{e(X_i)} - \frac{\mathbb{E}[1 - A_i \mid X_i] \cdot \mathbb{E}[Y_i(0) \mid X_i]}{1 - e(X_i)}\right]$$
$$= \mathbb{E}\left[\frac{\mathbb{E}[A_i Y_i(1) \mid X_i]}{e(X_i)} - \frac{\mathbb{E}[(1 - A_i) Y_i(0) \mid X_i]}{1 - e(X_i)}\right]$$
$$= \mathbb{E}\left[\frac{A_i Y_i}{e(X_i)} - \frac{(1 - A_i) Y_i}{1 - e(X_i)}\right]$$

with $e(X) = \mathbb{P}(A = 1|X)$

- **Include confounders** X: required for identifiability.
- **Exclude instrumental variables** X_a: increases variance.
- **Include precision covariates** X_p: improves efficiency (under conditional exchangeability $A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid (X, X_p)$).

Recommendations for IPTW (treatment model)

Goal: *Achieve a reliable estimate of the causal effect.*



Variance inflation[a][b]

$$\underbrace{\mathbf{E}\left[\frac{\sigma_T^2(X) + (\beta_T(X) - \beta_T)^2}{p(X)} + \frac{\sigma_C^2(X) + (\beta_C(X) - \beta_C)^2}{1 - p(X)}\right]}_{\mathbb{V}_{IPW} \text{ when using } X}$$

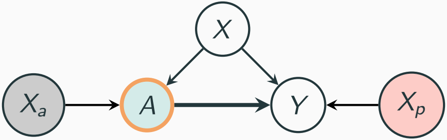
$$\leqslant \underbrace{\mathbf{E}\left[\frac{\sigma_T^2(X_1) + (\beta_T(X_1) - \beta_T)^2}{p(X_1)} + \frac{\sigma_C^2(X_1) + (\beta_C(X_1) - \beta_C)^2}{1 - p(X_1)}\right]}_{\mathbb{V}_{IPW} \text{ when using } X_1 = (X, X_a)}$$

- **Include confounders** X: required for identifiability.
- **Exclude instrumental variables** X_a: increases variance.
- **Include precision covariates** X_p: improves efficiency (under conditional exchangeability $A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid (X, X_p)$).

[a] Kangjie Zhou and Jinzhu Jia (2021). **Variance Reduction for Causal Inference.**
[b] Jay Bhattacharya and William B Vogt (2007). **Do instrumental variables belong in propensity scores?**

Recommendations for IPTW (treatment model)

Goal: *Achieve a reliable estimate of the causal effect.*



Variance reduction[a]

$$\Sigma_{IPW}^{V} = \Sigma_{IPW} - \underbrace{\left(\mathbf{H}_{\gamma} - \mathbf{E}_{\gamma\beta}\mathbf{E}_{\beta\beta}^{-1}\mathbf{H}_{\beta}\right)^{\mathrm{T}} \mathbf{H}_{\gamma\beta}^{-1} \left(\mathbf{H}_{\gamma} - \mathbf{E}_{\gamma\beta}\mathbf{E}_{\beta\beta}^{-1}\mathbf{H}_{\beta}\right)}_{\text{positive}}$$

with

- Σ_{IPW}^{V} the variance of the estimator including X and X_p in the **estimated** propensity score model.
- Σ_{IPW} the variance of the estimator including only X.

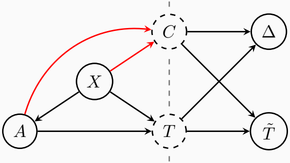
- **Include confounders** X: required for identifiability.
- **Exclude instrumental variables** X_a: increases variance.
- **Include precision covariates** X_p: improves efficiency (under conditional exchangeability $A \perp\!\!\!\perp \{Y(1), Y(0)\} \mid (X, X_p)$).

[a] Jared K. Lunceford and Marie Davidian (2004). **"Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study".** In: *Statistics in Medicine* 23.19, pp. 2937–2960.

Plan

Causal Survival analysis: Causal inference and Survival analysis

Causal inference — Survival analysis



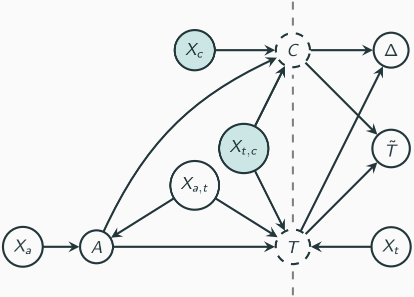
$$\Rightarrow n \textbf{ i.i.d. } (\underbrace{X_i}_{\textbf{covariates}}, \underbrace{A_i}_{\textbf{treatment}}, \underbrace{T_i}_{\textit{outcome}}) \in \mathbb{R}^d \times \{0,1\} \times \mathbb{R}^+$$

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

Potential outcome framework of Rubin 1974

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

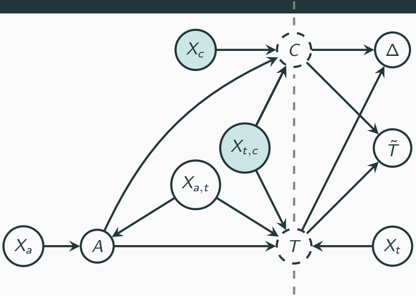
Causal Survival analysis: Causal inference and Survival analysis



It can exist additional variables than propose in the previous causal graph:

- Precision variables X_t;
- Instrumental variables X_a;
- **Censoring related variables** X_c;
- Confounders $X_{a,t}$;
- **Dependent censoring variables** $X_{t,c}$;

Causal Survival analysis: Causal inference and Survival analysis



It can exist additional variables than propose in the previous causal graph:

- Precision variables X_t;
- Instrumental variables X_a;
- **Censoring related variables** X_c;
- Confounders $X_{a,t}$;
- **Dependent censoring variables** $X_{t,c}$;

$\Rightarrow$ What is the impact of these variables on the variance of the estimators ?

Causal effect in survival analysis

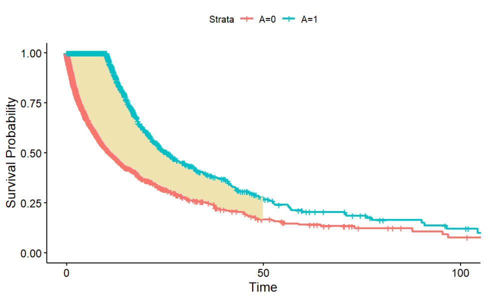
Restricted Mean Survival Time (RMST)

$$\mathrm{RMST}(\tau) = \mathbb{E}[min(T, \tau)] = \int_0^\tau \hat{S}(t)dt \text{ with } S(t) = \mathbb{P}(T > t)$$

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

Average treatment effect in survival analysis

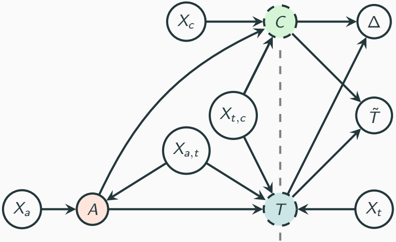
$$\hat{\theta}_{RMST}(\tau) = \mathrm{RMST}_1(\tau) - \mathrm{RMST}_0(\tau)$$



$\hat{\theta}_{RMST}(\tau = 50) = 10$ 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)

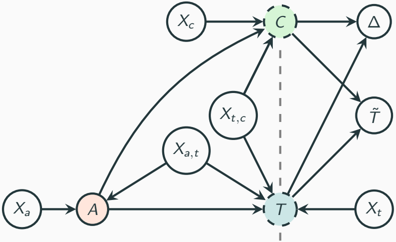
Estimators in observational study and dependent censoring



Estimators based on Censoring unbiased transformations + IPW

- Transformation for dependent censoring:
  - **IPCW**: Inverse probability of censoring weighting ⇒ **Estimation of censoring model** (minimal set = $X_{t,c}$).
  - **BJ**: Buckley-James transformation ⇒ **Estimation of outcome model** (minimal set = $X_{a,t}$).
- Weighting for confounding bias: **IPTW** Inverse probability weighting for treatment ⇒ **Estimation of treatment model** (minimal set= $X_{a,t}$)

Estimators in observational study and dependent censoring



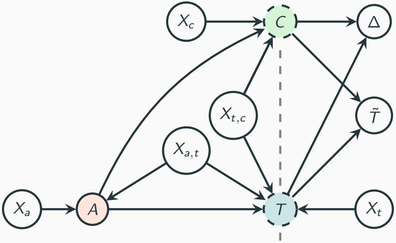
Estimators based on Censoring unbiased transformations + IPW

- Transformation for dependent censoring:
  - **IPCW**: Inverse probability of censoring weighting $\Rightarrow$ **Estimation of censoring model** (minimal set $= X_{t,c}$).
  - **BJ**: Buckley-James transformation $\Rightarrow$ **Estimation of outcome model** (minimal set $= X_{a,t}$).
- Weighting for confounding bias: **IPTW** Inverse probability weighting for treatment $\Rightarrow$ **Estimation of treatment model** (minimal set$= X_{a,t}$)

Estimators based on survival functions

- **Weighted Kaplan-Meier by IPCW and IPTW** $\Rightarrow$ **Estimation of censoring model** (minimal set $= X_{t,c}$) + **Estimation of treatment model** (minimal set$= X_{a,t}$)
- **G-formula** $\Rightarrow$ **Estimation of outcome model** (minimal set $= X_{a,t}$).

Estimators in observational study and dependent censoring



Estimators based on Censoring unbiased transformations + IPW

- Transformation for dependent censoring:
  - **IPCW**: Inverse probability of censoring weighting ⇒ **Estimation of censoring model** (minimal set = $X_{t,c}$).
  - **BJ**: Buckley-James transformation ⇒ **Estimation of outcome model** (minimal set = $X_{a,t}$).
- Weighting for confounding bias: **IPTW** Inverse probability weighting for treatment ⇒ **Estimation of treatment model** (minimal set= $X_{a,t}$)

Estimators based on survival functions

- **Weighted Kaplan-Meier by IPCW and IPTW** ⇒ **Estimation of censoring model** (minimal set = $X_{t,c}$) + **Estimation of treatment model** (minimal set= $X_{a,t}$)
- **G-formula** ⇒ **Estimation of outcome model** (minimal set = $X_{a,t}$).

Doubly robust estimator

Focus on AIPTW-AIPCW (Doubly Robust Estimator)

Augmented estimator: AIPTW-AIPCW [Ozenne et al. 2020]

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

$$\widehat{\theta}_{\mathrm{AIPTW-AIPCW}} := \frac{1}{n}\sum_{i=1}^{n} \Delta^*_{\mathrm{QR}}(\widehat{G}, \widehat{S}, \widehat{\mu}, \widehat{e}).$$

$\Rightarrow$ 3 nuisance parameters to compute :

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

Called doubly robust because the estimator is consistent if among the three quantities, conditional survival **or** propensity and censoring model are consistently estimated.

Plan

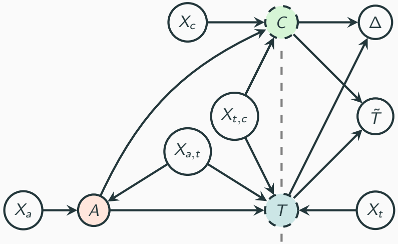
Introduction to Causal Inference

Variable selection: a fundamental issue

Towards causal survival: new challenges

Preliminary results on variable selection and perspectives

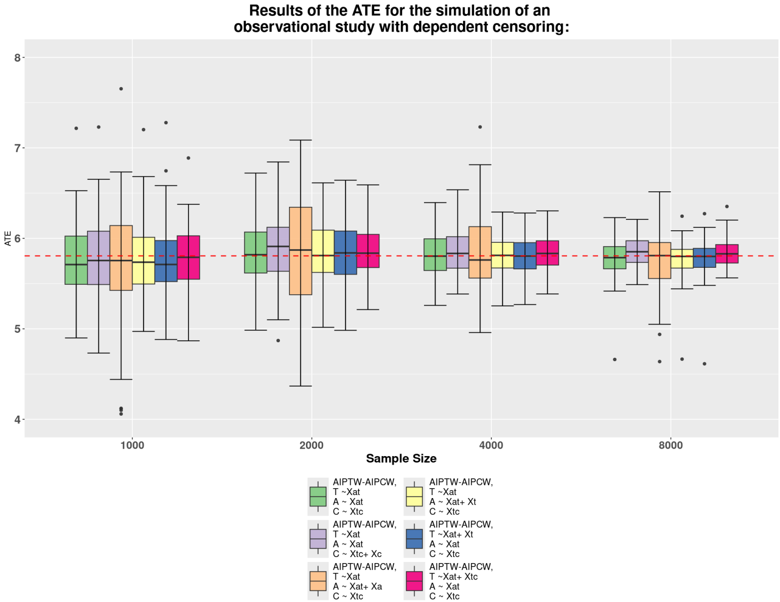
Focus on AIPTW-AIPCW (Doubly Robust Estimator)



Estimator	Outcome model	Censoring model	Treatment model
Minimal sufficient set	$T \sim X_{a,t}$	$C \sim X_{t,c}$	$A \sim X_{a,t}$
Extended set with X_c	$T \sim X_{a,t}$	$C \sim X_{t,c} + X_c$	$A \sim X_{a,t}$
Extended set with X_a	$T \sim X_{a,t}$	$C \sim X_{t,c}$	$A \sim X_{a,t} + X_a$
Extended set with X_t	$T \sim X_{a,t} + X_t$	$C \sim X_{t,c}$	$A \sim X_{a,t} + X_t$

Table 1: The different adjustment set tested for AIPTW-AIPCW

Focus on AIPTW-AIPCW (Doubly Robust Estimator)



Conclusion

It appears that insights from variable selection in standard causal inference can be extended to the causal survival setting.

- **Instrumental variables** should not be included in the propensity score model, since they can inflate variance.
- **Precision variables** may be included in both outcome and propensity score models to enhance efficiency.
- **Variables that influence only censoring** are generally unsuitable for the censoring model, as they tend to increase estimator variance.
- **Variables that affect both censoring and the outcome** behave like precision variables and often reduce variance in the outcome model.

Thank you for your attention !

Here's a summary of our previous review on practical recommendations for causal survival estimators:

Arxiv article (submitted)

Github repository

Feel free to give me some feedback or contact me for any question:

charlotte.voinot@sanofi.com

Disclosure

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

References i

Bhattacharya, Jay and William B Vogt (2007). **Do instrumental variables belong in propensity scores?**

Buckley and James (Dec. 1979). **"Linear regression with censored data".** In: *Biometrika* 66.3, pp. 429–436.

Fan, Jianqing and Irene Gijbels (1994). **Local polynomial modelling and its applications.**

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.

Lunceford, Jared K. and Marie Davidian (2004). **"Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study".** In: *Statistics in Medicine* 23.19, pp. 2937–2960.

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.

Rubin, Donald B. (Oct. 1974). **"Estimating causal effects of treatments in randomized and nonrandomized studies.".** en. In: *Journal of Educational Psychology* 66.5, pp. 688–701.

References ii

Zhou, Kangjie and Jinzhu Jia (2021). **Variance Reduction for Causal Inference.**

Appendix

Simulation of observational study ((semi)-parametric DGP)

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

- $X_{a,t} \sim \mathcal{N}\left(\mu = [1,1]^\top, \Sigma = I_2\right)$
- $X_{t,c} \sim \mathcal{N}\left(\mu = [1,1]^\top, \Sigma = I_2\right)$
- $X_t \sim \mathcal{N}\left(\mu = [1.5,1.5]^\top, \Sigma = I_2\right)$
- $X_a \sim \mathcal{N}\left(\mu = [1,1]^\top, \Sigma = I_2\right)$
- $X_c \sim \mathcal{N}\left(\mu = [1,1]^\top, \Sigma = I_2\right)$
- $logit(e(X)) = \beta_A^T X_{a,t} + \gamma_A^T X_a$ where $\beta_A = (1, 0.5), \gamma_A = (1, -1)$ and $logit(p) = log(p/(1-p))$ the logistic function.
- Then $A_i \sim$ Bernoulli(e_i), $\quad \forall i \in \{1, \ldots, n\}$
- $\lambda^{(0)}(t|X) = 0.01 \cdot \exp\left\{\beta_T^T X_t + \gamma_T^T X_{a,t} + \delta_T^T X_{t,c}\right\}$ hazard for the event time $T(0)$ with $\beta_T = (0.5, -0.05), \gamma_T = (0.5, 0.05), \delta_T = (1, 0.01)$
- $T(1) = T(0) + 10$
- The hazard for the censoring time C: $\lambda_c(X) = 0.01 \cdot \exp\left\{\beta_c^T X_c + \gamma_C^T X_{t,c} + 0.5A\right\}$ with $\beta_c = (1, -0.5), \gamma_c = (0.5, -0.5)$.
- The threshold time τ is set to 15.

Identifiability assumptions in each context

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** $0 < P(C > t \mid X = x, A = a) < 1$

Estimators based on Censoring unbiased transformation

The basic requirement is to create an fully observable variable T^* such that $E(T^*|X,A) = E(T \wedge \tau|X,A)$. [Fan and Gijbels 1994]

IPC transformation [Koul, Susarla, and Ryzin 1981]:

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

with $G(\widetilde{T} \wedge \tau|X,A) := \mathbb{P}(C \geqslant t|X,A)$ being the left limit of the **conditional survival function of the censoring**. $\Delta^\tau := I\{T \wedge \tau \leqslant C\}$ is the censoring indicator of the restricted time.

Buckley-James transformation [Buckley and James 1979]:

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

where, for $t \leqslant \tau$, $Q_S(t|X,A) := \mathbb{E}[T \wedge \tau|X,A, T \wedge \tau > t]$ estimated with the **conditional survival function**.

Estimators based on Censoring unbiased transformation

The basic requirement is to create an fully observable variable T^* such that $E(T^*|X, A) = E(T \wedge \tau|X, A)$. [Fan and Gijbels 1994]

IPC transformation [Koul, Susarla, and Ryzin 1981]:

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

with $G(\widetilde{T} \wedge \tau|X, A) := \mathbb{P}(C \geqslant t|X, A)$ being the left limit of the **conditional survival function of the censoring**. $\Delta^\tau := I\{T \wedge \tau \leqslant C\}$ is the censoring indicator of the restricted time.

Buckley-James transformation [Buckley and James 1979]:

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

where, for $t \leqslant \tau$, $Q_S(t|X, A) := \mathbb{E}[T \wedge \tau|X, A, T \wedge \tau > t]$ estimated with the **conditional survival function**.

Estimators based on Censoring unbiased transformation

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

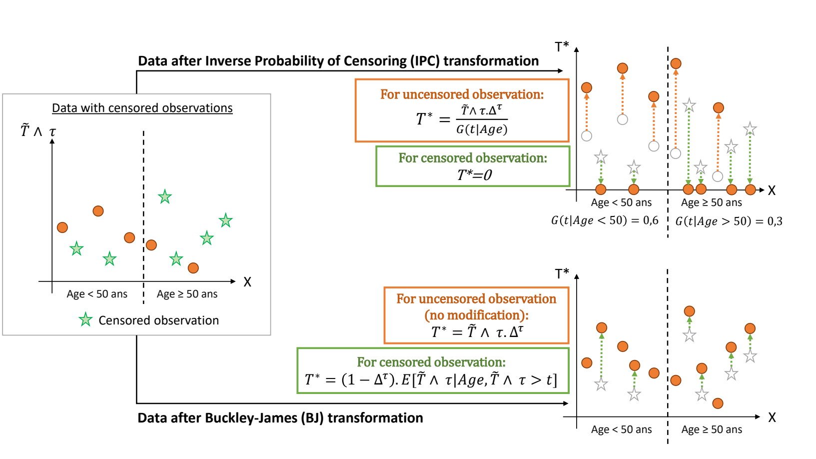


Figure 2: Illustration of Inverse-Probability-of-Censoring and Buckley-James transformations

Estimator based on survival functions

G-formula estimator

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

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

- Various potential methods to estimate the **conditional survival function**:
  - using a parametric model (Weibull, Exponential)
  - using a semi-parametric model (Cox model)
  - using a "model free" (Survival Forest)
- Estimates are consistent under the some standard causal inference, survival and estimation assumptions (Unconfoundedness, Consistency, Overlap, Conditionally Independent Censoring and that the quantities are consistently estimated).

Adjustment set to test

Estimator	Outcome model	Censoring model	Treatment model
Minimal sufficient set	$T \sim X_{a,t}$	$C \sim X_{t,c}$	$A \sim X_{a,t}$
Extended set with X_t	$T \sim X_{a,t} + X_t$	$C \sim X_{t,c} + X_t$	$A \sim X_{a,t} + X_t$
Extended set with X_a	$T \sim X_{a,t} + X_a$	$C \sim X_{t,c} + X_a$	$A \sim X_{a,t} + X_a$
Extended set with X_c	$T \sim X_{a,t} + X_c$	$C \sim X_{t,c} + X_c$	$A \sim X_{a,t} + X_c$
Other unexplored Extended set	$T \sim X_{a,t} + X_{t,c}$ $T \sim X_{a,t} + X_{t,c} + X_t$		

Table 2: Different adjustment set for causal survival analysis.