Variable Selection for Emulated Target Trials in Causal Survival Analysis

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

Collaborators: Stefan Michiels & Noemie Simon-Tillaux (Gustave Roussy)

Charlotte Voinot

PhD student at Sanofi R&D and Inria

ISCB 2025

August 28, 2025

Motivation: Tamoxifen Duration in Early-Stage Breast Cancer

Collaboration & Data Source

- Collaboration with **Gustave Roussy** and **Unicancer**.
- CANTO cohort (observational study with 9,000 patients and more than 4,000 variables).
- Study adjuvant **tamoxifen** therapy in breast cancer.

Evidence from Clinical Trials

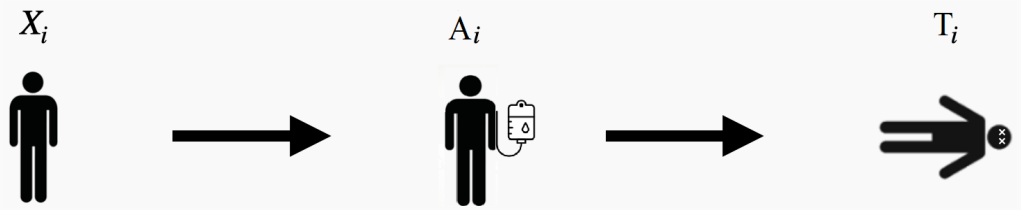
- Delozier et al. (2000): Longer tamoxifen treatment ($>$5 years) improved disease-free survival vs. stopping at 2–3 years.

Real-World Perspective

- Trial inclusion criteria do not fully represent the real-world patient population.
- Clinicians are interested in assessing whether real-world evidence aligns with trial findings.

Aim: **treatment deescalation** without compromising patient outcomes.

Causal Survival Analysis in observational study



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

Let's say that in our example $X_1 =$ sex, $X_2 =$ age and $A = 1$ (resp. $A = 0$) denotes 5-year (resp. 3-year) treatment.

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

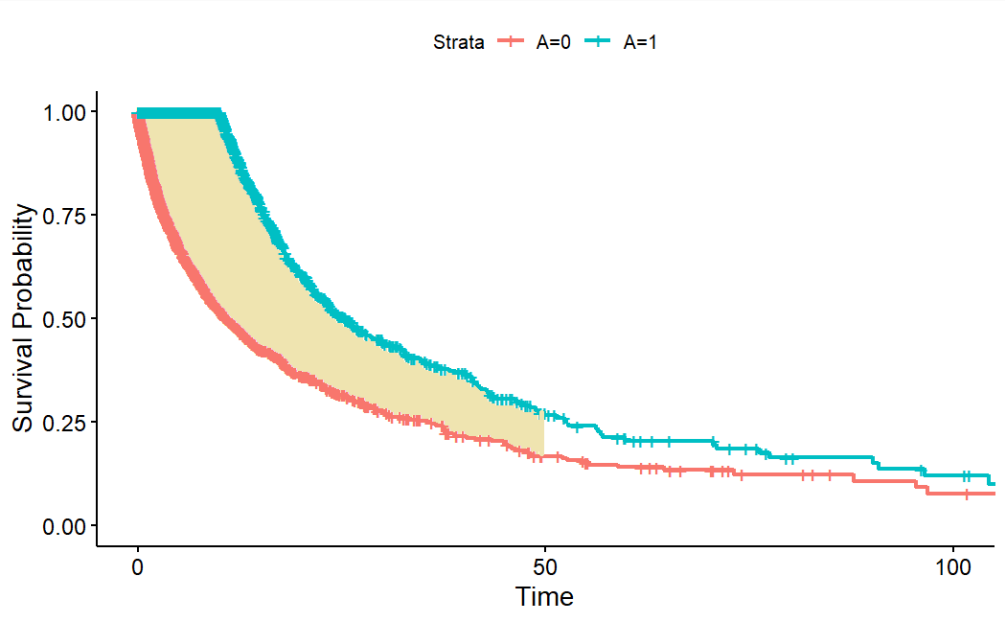
- $\Delta = \mathbb{1}\{C < T\}$ is the censoring indicator.
- $\tilde{T} = min(T, C)$ is the observed outcome.

Estimand : Difference of RMST

Difference in RMST : Average treatment effect in survival analysis

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

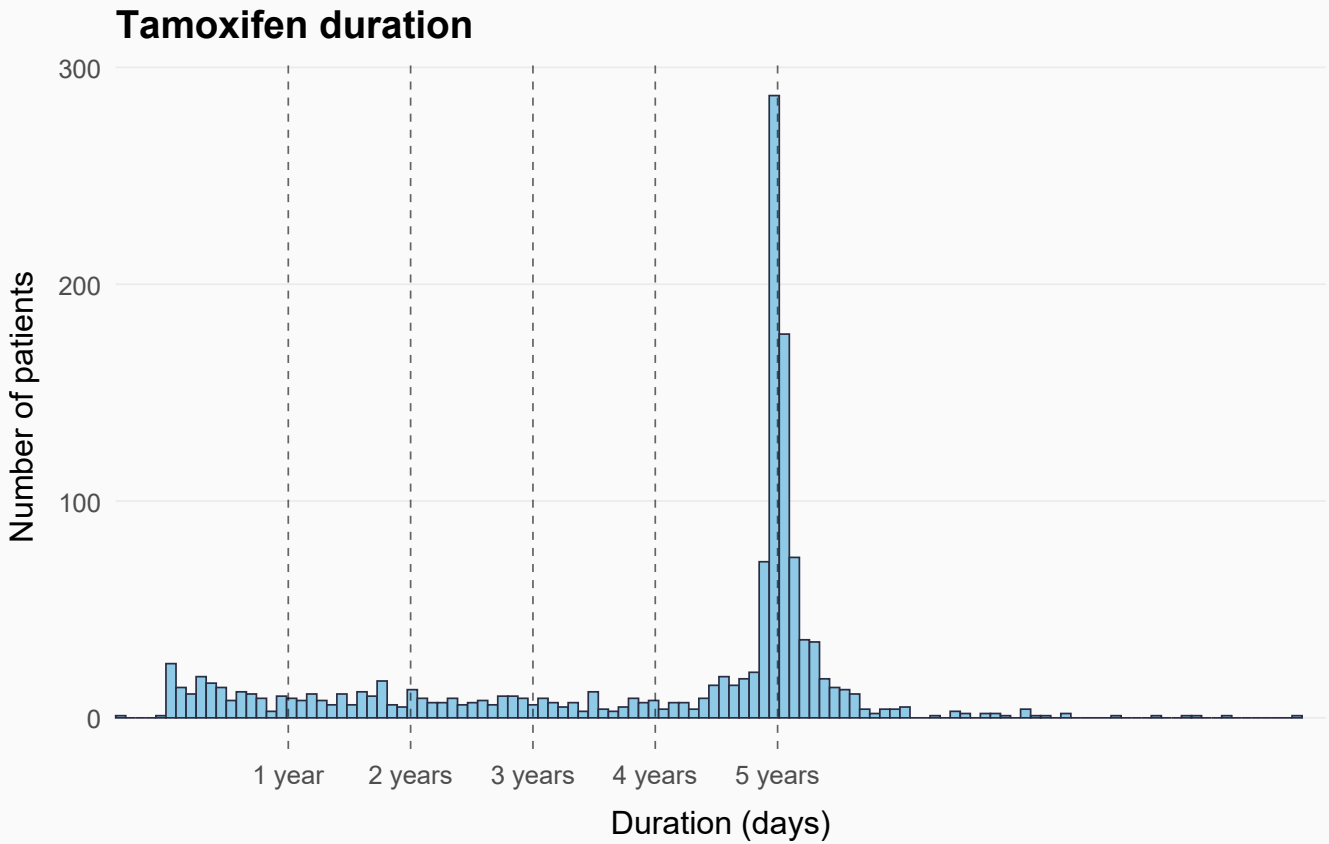
RMST can be defined as a measure of average survival from time 0 to time τ a **fixed time horizon**



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

From Baseline Treatment Assignment to Treatment Duration

- Prior work assumed a baseline "3 vs 5 years" assignment.
- In our data, no such baseline assignment: treatment is **continuous**.
- Observed duration is **post-baseline**.



From Baseline Treatment assignment to Treatment Duration

In CANTO settings, treatment is **not a one-time decision**:

Post-baseline Classification

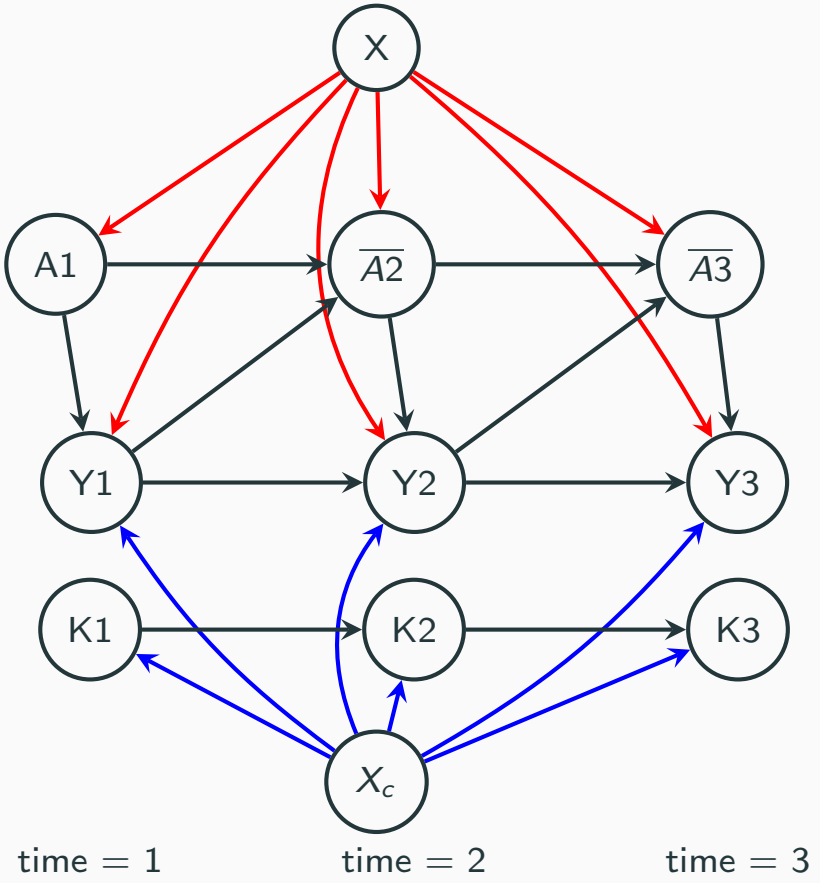
- Classification into "short" vs "long" duration is made **post-baseline**, based on follow-up data.

Time-varying Dependence

Duration is **inherently linked** to survival:

- To have a long duration, the patient must survive long enough (**immortal time**).
- Survival also depends on treatment duration.

Dynamic treatment assignment (Real World assignment)

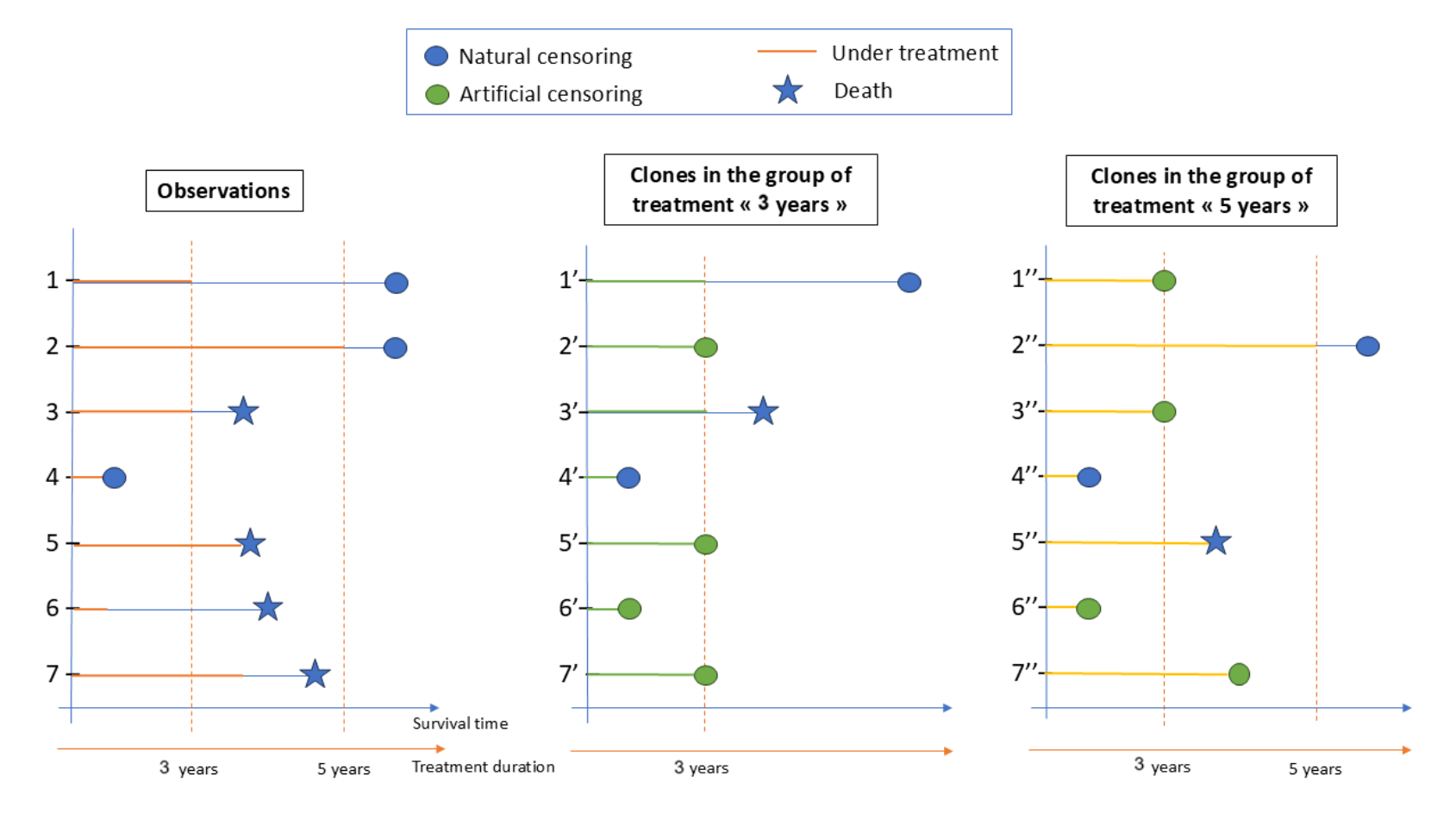


K_t is the indicator of censoring, Y_t is an indicator of the event of interest, A_t the treatment variable, X the baseline confounders and X_c the variables of dependent censoring.

Strategy for 3 years vs 5 years of treatment

Cloning–Censoring–Weighting (Cain et al., 2010)

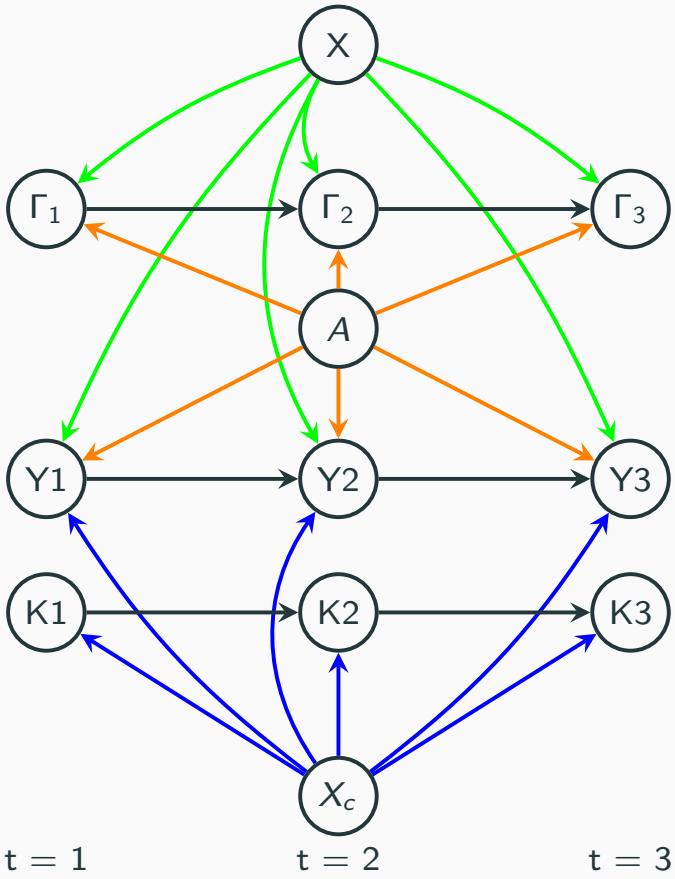
Clone each subject for each strategy; *artificially censor* at deviation:



Strategy for 3 years vs 5 years of treatment

Cloning–Censoring–Weighting (Cain et al., 2010)

G is artificial informative censoring ($G = min(t \in [\tau], \Gamma_t = 1)$), C is natural informative censoring ($C = min(t \in [\tau], K_t = 1)$), T is the time to event outcome ($T = min(t \in [\tau], Y_t = 1)$).

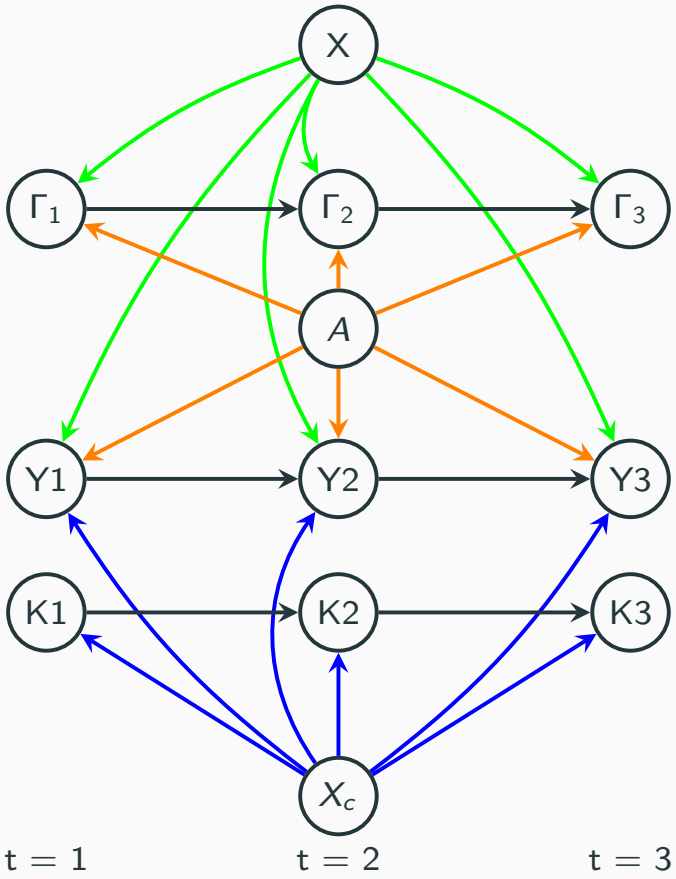


DAG after cloning (A=0 if under first strategy and A=1 if under second strategy).
Γ_t is the indicator of artificial censoring.

Strategy for 3 years vs 5 years of treatment

Cloning–Censoring–Weighting (Cain et al., 2010)

G is artificial informative censoring ($G = min(t \in [\tau], \Gamma_t = 1)$), C is natural informative censoring ($C = min(t \in [\tau], K_t = 1)$), T is the time to event outcome ($T = min(t \in [\tau], Y_t = 1)$).



DAG after cloning (A=0 if under first strategy and A=1 if under second strategy).
Γ_t is the indicator of artificial censoring.

Weighted Kaplan–Meier [1]:

Two informative censorings ⇒ compute weights

- Artificial censoring:
$$W^A_{i,a}(t) = \frac{\mathbf{1}\{ G_i \geq t \}}{\prod_{j=0}^{t} Pr[\Gamma_{i,j}=0 | \Gamma_{i,j-1}=0, Y_{i,j}=0, X, A_i=a]}$$
- Natural censoring:
$$W^N_{i,a}(t) = \frac{\mathbf{1}\{ C_i \geq t \}}{\prod_{j=0}^{t} Pr[K_{i,j}=0 | K_{i,j-1}=0, Y_{i,j}=0, X_c, A_i=a]}$$

Then weighted KM by $w_{i,a}(t) = W^A_{i,a}(t).W^N_{i,a}(t)$ and compute θ_{RMST}, the difference of RMST.

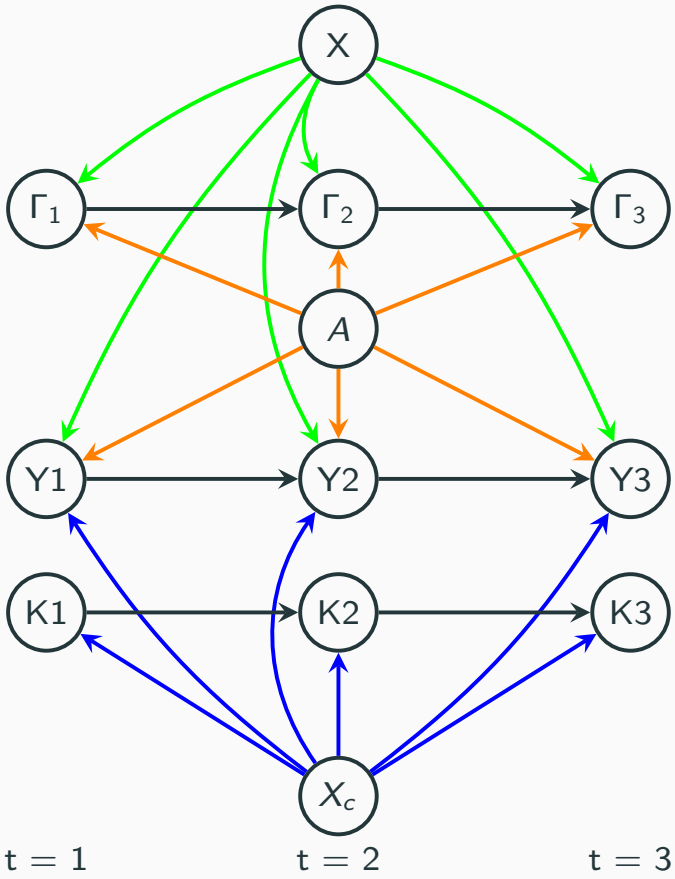
[1] Identifiability: exchangeability, positivity, consistency.

[2] Adjust on X_c and X

Strategy for 3 years vs 5 years of treatment

Cloning–Censoring–Weighting (Cain et al., 2010)

G is artificial informative censoring ($G = min(t \in [\tau], \Gamma_t = 1)$), C is natural informative censoring ($C = min(t \in [\tau], K_t = 1)$), T is the time to event outcome ($T = min(t \in [\tau], Y_t = 1)$).



DAG after cloning (A=0 if under first strategy and A=1 if under second strategy).
Γ_t is the indicator of artificial censoring.

Weighted Kaplan–Meier [1]:

Two informative censorings ⇒ compute weights

- Artificial censoring:
$$W^A_{i,a}(t) = \frac{\mathbf{1}\{ G_i \geq t \}}{\prod_{j=0}^{t} Pr[\Gamma_{i,j}=0 | \Gamma_{i,j-1}=0, Y_{i,j}=0, X, A_i=a]}$$
- Natural censoring:
$$W^N_{i,a}(t) = \frac{\mathbf{1}\{ C_i \geq t \}}{\prod_{j=0}^{t} Pr[K_{i,j}=0 | K_{i,j-1}=0, Y_{i,j}=0, X_c, A_i=a]}$$

Then weighted KM by $w_{i,a}(t) = W^A_{i,a}(t).W^N_{i,a}(t)$ and compute θ_{RMST}, the difference of RMST.

G-formula estimator[1] [2]:

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

with $\widehat{\mu}(x_i, a) = \int_0^\tau S(t|X = x_i, A = a)dt$ the integral of the **conditional survival function** truncated at τ.

[1] Identifiability: exchangeability, positivity, consistency.

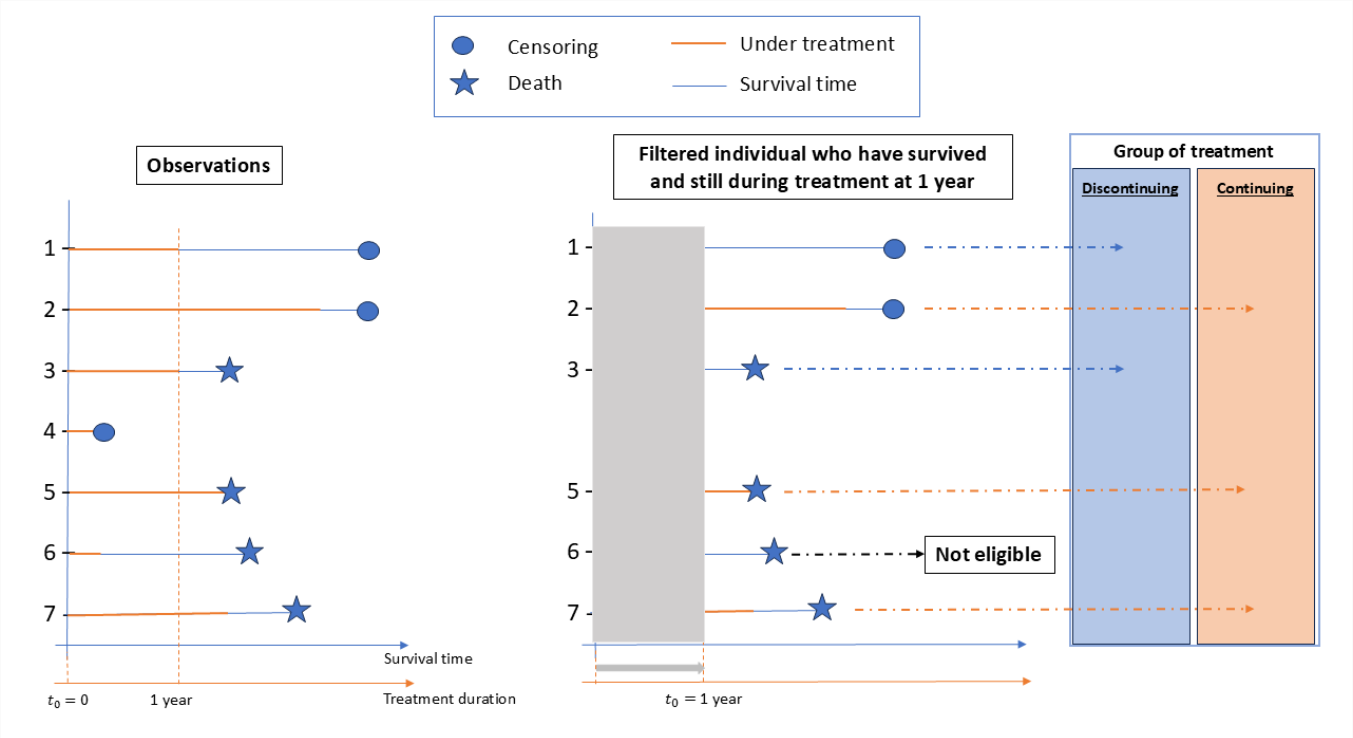
[2] Adjust on X_c and X

Strategy for finding optimal treatment duration

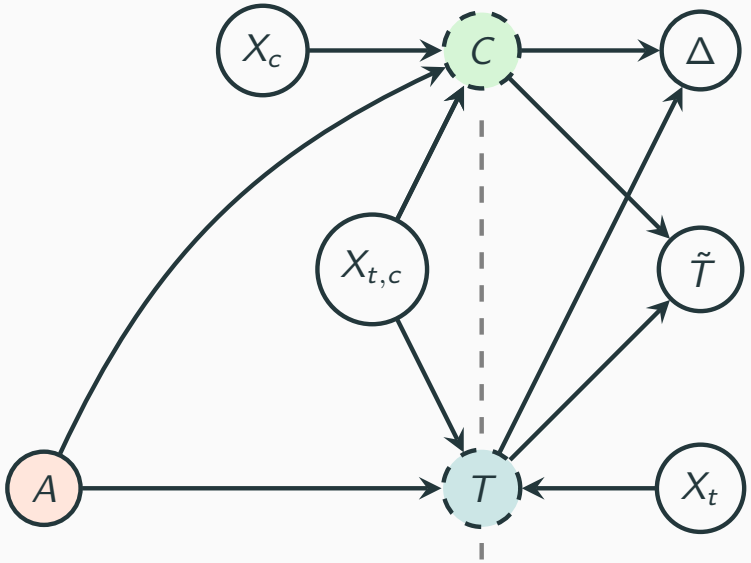
Evaluate at different sequential time t_1, **among patients under treatment**, compare stop vs continue treatment to learn the **optimal duration**.

Target Trial Emulation

1. Specify the *target trial (e.g. time-zero $t_0 \in \{1, 2, 3, 4, 5\}$ years; eligible patients are those alive and on treatment at t_0, target 10-year survival)*
2. Decide whether a **grace period** (may lead to immortal time bias) is used to classify individuals into "stop" vs "continue".



RCT with dependent censoring: Variable Selection



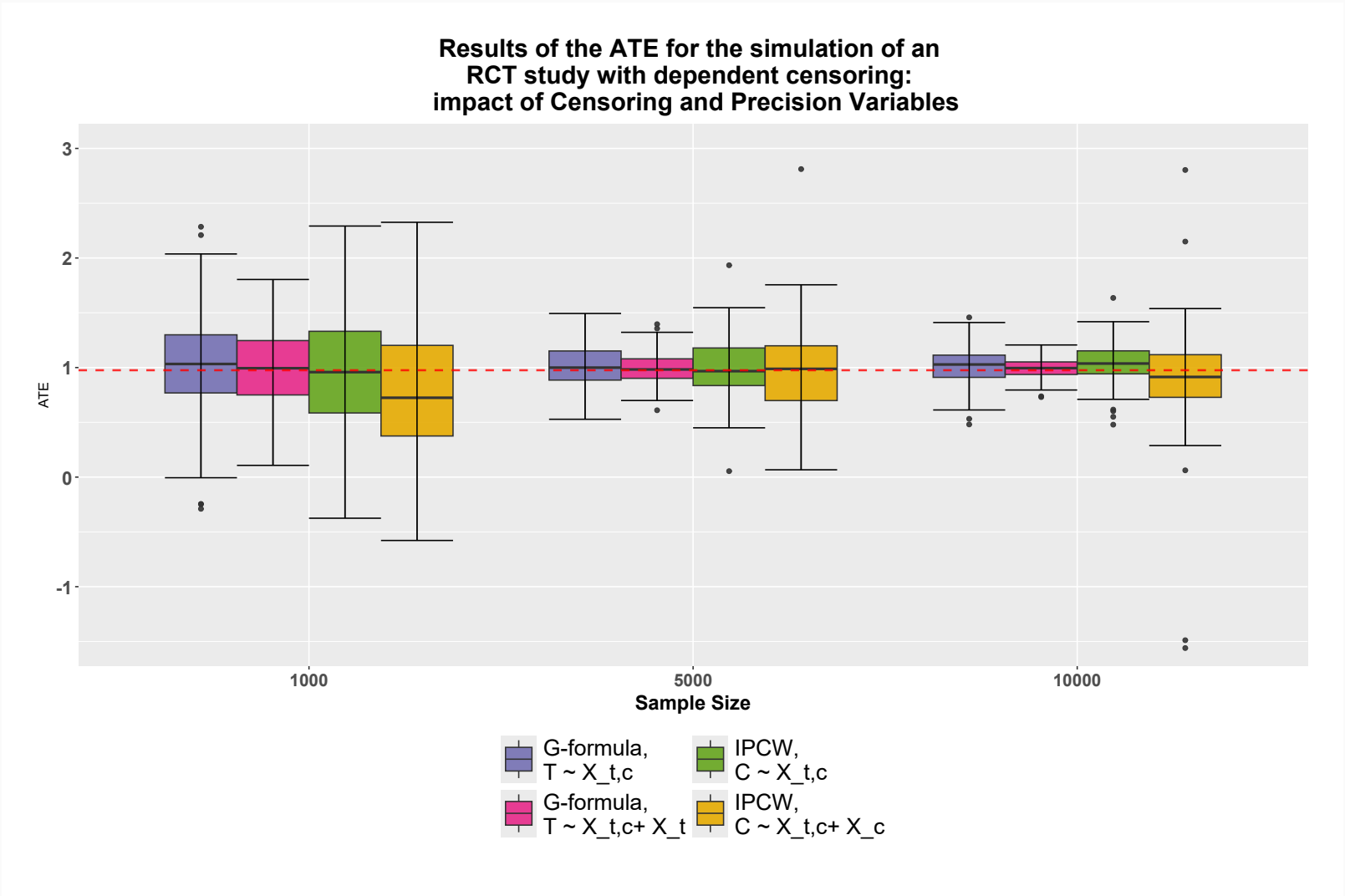
DAG of a RCT with dependent censoring (A is the treatment, T is the time to event outcome and C is the censoring time).

To ensure causal identifiability in RCT and dependent censoring, adjust each model for:

- **Censoring model and Outcome model:** $X_{t,c}$ (remove selection bias from informative censoring)

However: Adding extra variables (precision, or censoring-related) can affect estimator variance.

Baseline Treatment Assignment: Variable Selection



The simulations show that:

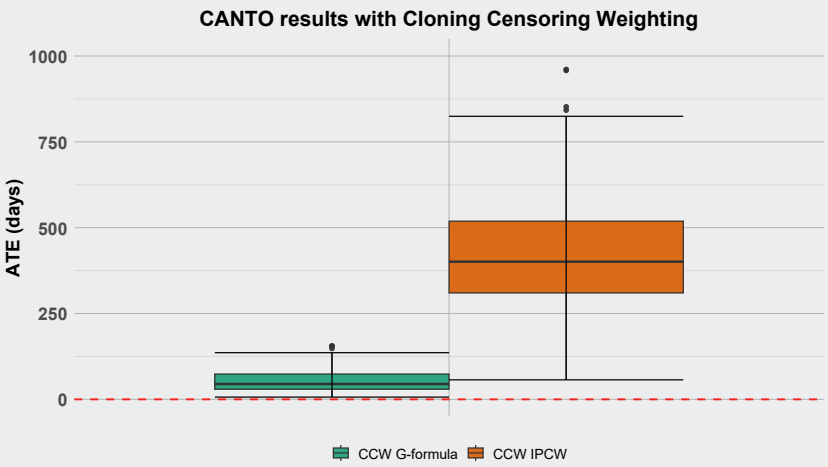
- Precision variables (X_t) (Purple vs Pink): decrease the variance.
- Censoring-related variables (X_c) (Green vs Yellow): increase the variance.

RMST Difference Between Tamoxifen Durations

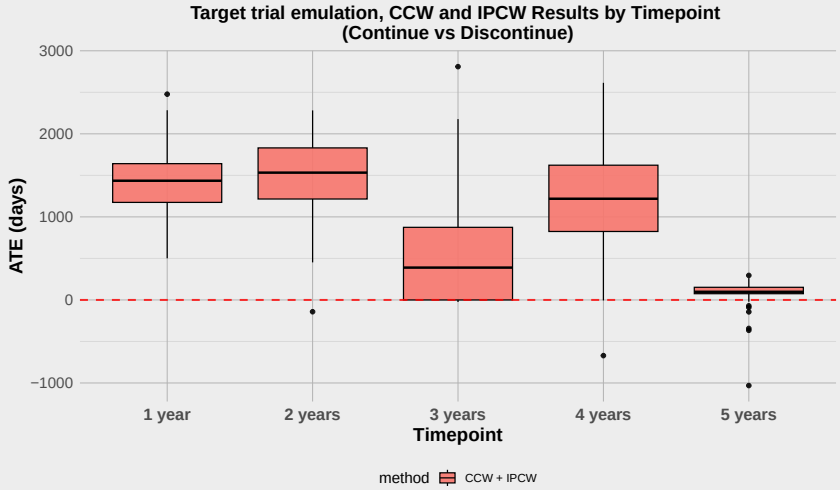
Setup

- **Objective:** Estimate the ATE on 10-year OS of **3-year** vs **5-year** treatment duration.
- **Grace period:** 6 months after (and/or before) the timepoint.
- **Adjustment:** baseline confounders only (no time-varying covariates).
- **Uncertainty:** boxplots over **150** bootstrap samples.

3 years vs 5 years



Continue vs Discontinue



Results

- **ATE** > 0 **across analyses**, better 10-year OS under the 5-year strategy.

Conclusion & Perspectives

Theoretical

- *Prove* the variance increase from including censoring-related variables in the censoring model.
- *Formalize* the Cloning–Censoring–Weighting approach.

Data Analyses

- Incorporate *time-varying covariates* (longitudinal history).
- Identify potential *precision covariates*.

Challenges

- Treatment dates often uncertain (missing day and/or month).
- Time-varying covariates are difficult to construct (continuous variables, choice of evaluation timepoints).

Thank you for your attention !
Here's a summary of our previous contribution 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

Lauren E Cain, James M Robins, Emilie Lanoy, Roger Logan, Dominique Costagliola, and Miguel A Hernán. When to start treatment? a systematic approach to the comparison of dynamic regimes using observational data. *The international journal of biostatistics*, 6(2), 2010.

T Delozier, M Spielmann, J Macé-Lesec'h, M Janvier, C Hill, B Asselain, J P Julien, B Weber, L Mauriac, J C Petit, P Kerbrat, J P Malhaire, P Vennin, B Leduc, and M Namer. Tamoxifen adjuvant treatment duration in early breast cancer: initial results of a randomized study comparing short-term treatment with long-term treatment. fédération nationale des centres de lutte contre le cancer breast group. *J. Clin. Oncol.*, 18(20):3507–3512, October 2000.

S. Greenland, J. M. Robbins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, 14:29–46, January 1999.

Miguel A Hernán. The hazards of hazard ratios. *Epidemiology*, 21(1):13–15, January 2010.

A. Huitfeldt, M. Stensrud, and E. Suzuki. On the collapsibility of measures of effect in the counterfactual causal framework. *Emerging Themes in Epidemiology*, 16:1–12, January 2019.

Torben Martinussen and Stijn Vansteelandt. On collapsibility and confounding bias in cox and aalen regression models. *Lifetime Data Anal.*, 19(3):279–296, July 2013.

References ii

Torben Martinussen, Stijn Vansteelandt, and Per Andersen. Subtleties in the interpretation of hazard contrasts. *Lifetime Data Analysis*, 26, 10 2020. doi: 10.1007/s10985-020-09501-5.

Patrick Royston and Mahesh K B Parmar. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med. Res. Methodol.*, 13(1):152, December 2013.

Mats J Stensrud, John M Aalen, Odd O Aalen, and Morten Valberg. Limitations of hazard ratios in clinical trials. *European heart journal*, 40(17):1378–1383, 2019.

Appendix

IPCW Kaplan-Meier

Dynamic G-formula (no time-dependent confounding)

Context. We compare two *static treatment duration strategies* (e.g., 3y vs 5y). Since there is no time-dependent confounding, only **baseline covariates** X need to be adjusted for.

Principle.

- Fit survival models $P(T \mid A, X)$ within each interval of follow-up.
- Create counterfactual datasets where treatment trajectories are **forced**:

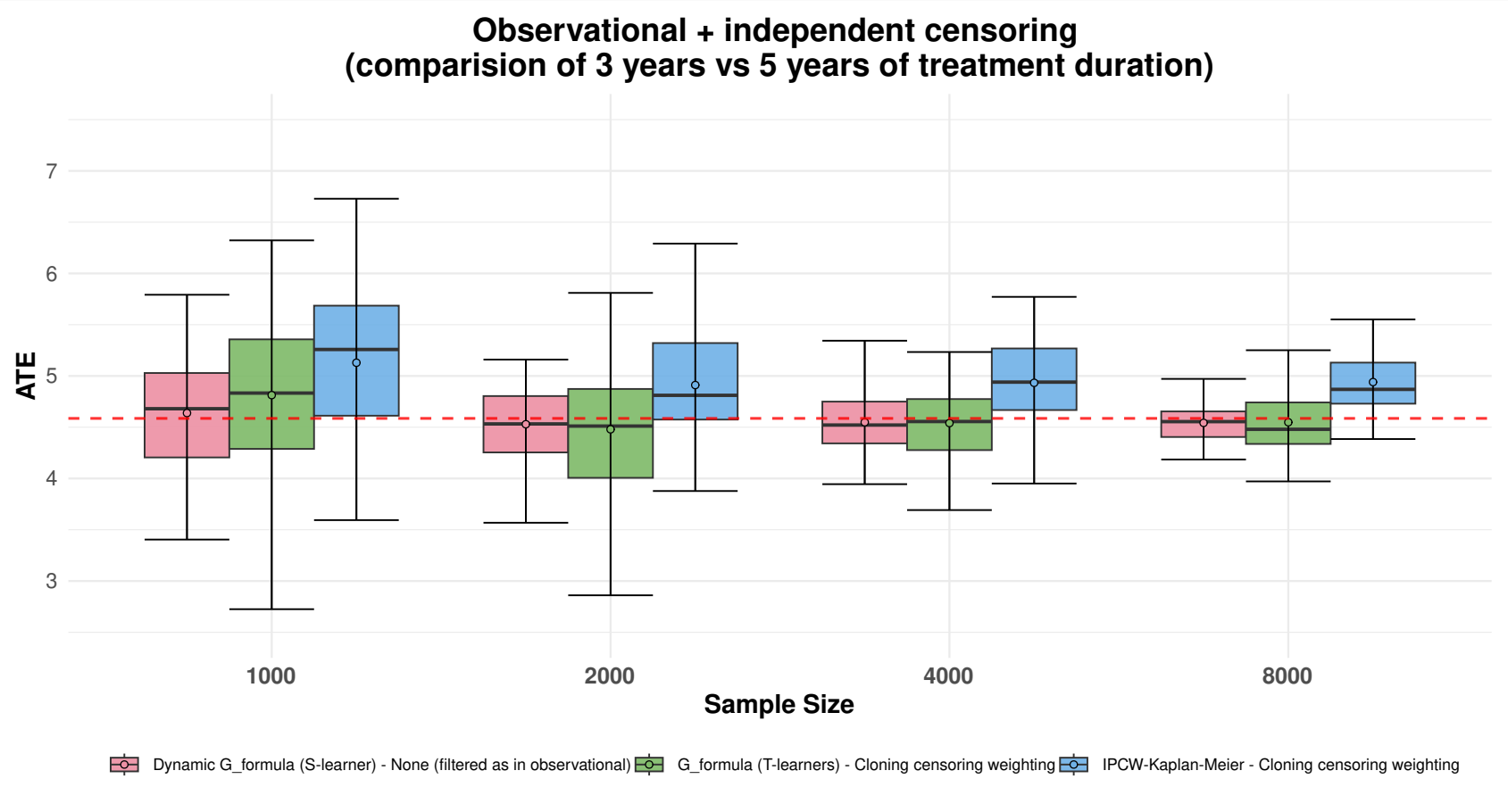
$$\bar{A}^{(0)} = (1, 1, 1, 0, 0), \quad \bar{A}^{(1)} = (1, 1, 1, 1, 1).$$

- Use fitted models to predict survival for all individuals under each strategy.
- Aggregate to obtain survival curves $S^{(0)}(t), S^{(1)}(t)$ and functionals (e.g., RMST difference).

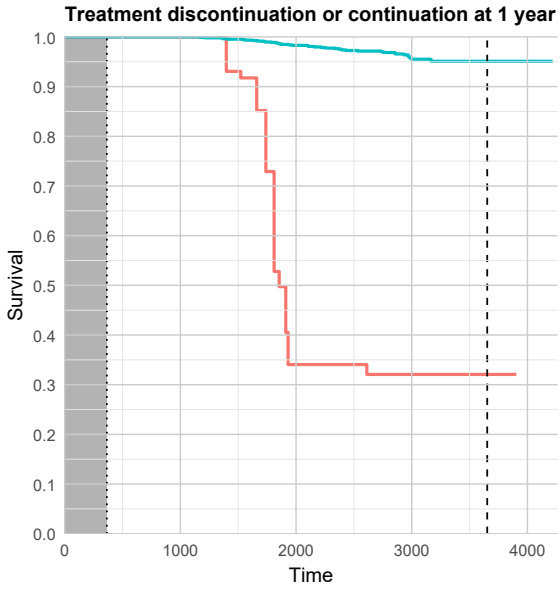
Why dynamic g-formula here?

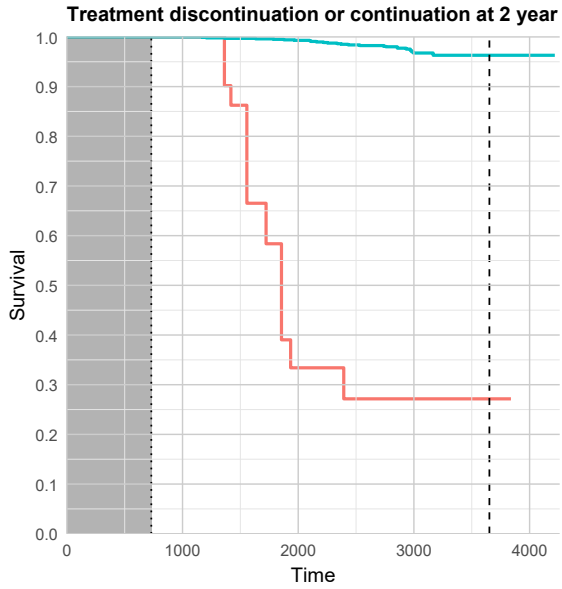
- Still framed longitudinally (assign $\bar{A}_t$ at each visit),
- but no need to model covariate evolution since L_t does not depend on A.

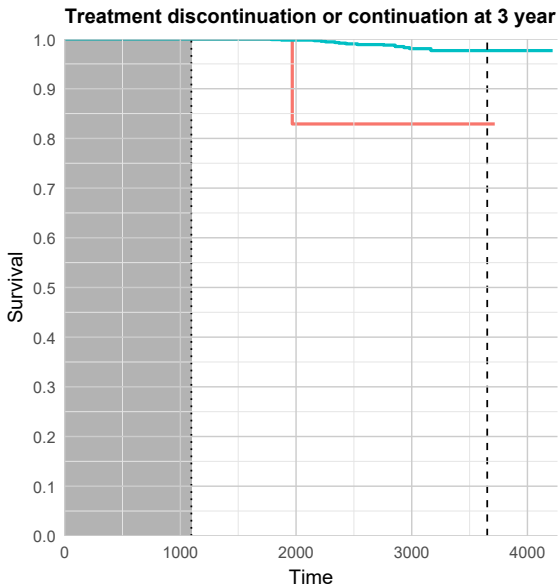
Comparision of strategies to counter immortal time bias

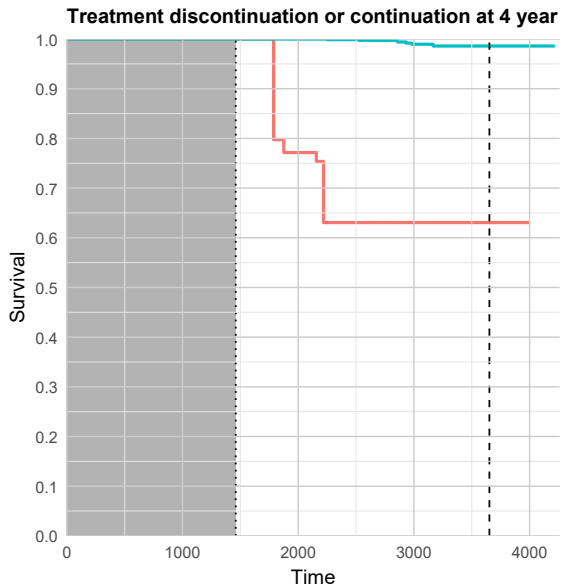


Survival curves of Tamoxifen duration: continue vs discontinue

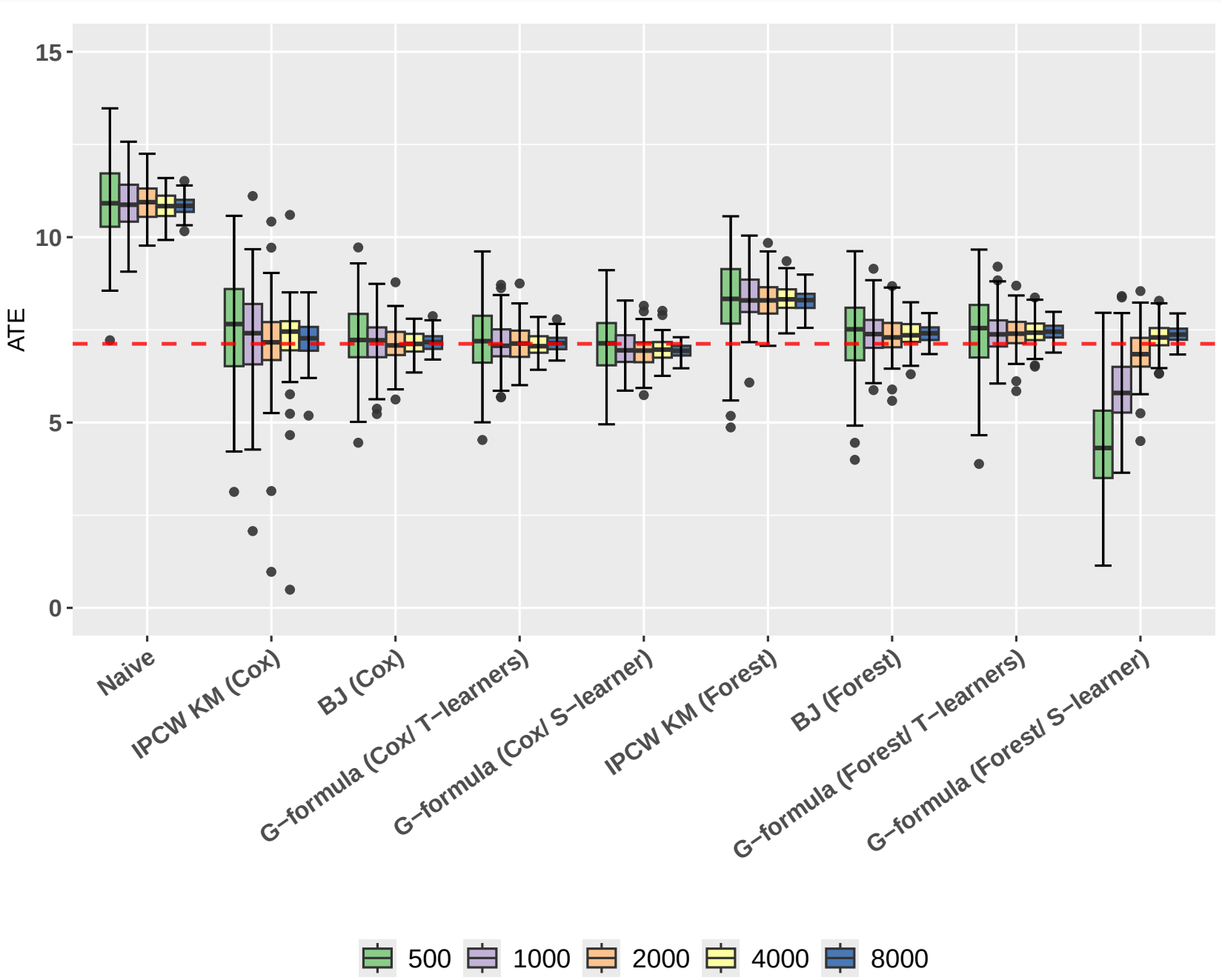








Estimators consistent in RCT with dependent censoring



Why using RMST instead of Hazard Ratio ?

Limitations of the Hazard Ratio (HR):

- Assumes **proportional hazards**.
- Hazard ratios **are not causally interpretable** (even in RCT unless there is no treatment effect) due to a built-in selection bias. [Martinussen and Vansteelandt (2013); Martinussen et al. (2020)]
- **Not collapsible**.[Huitfeldt et al. (2019); Greenland et al. (1999)]
- Difficult to interpret clinically. [Hernán (2010)]

Advantages of Restricted Mean Survival Time (RMST):

- **Causal measure** as it is the extension of Average Treatment Effect. [Royston and Parmar (2013)]
- Does not necessarily require the assumption of proportional risks.
- Provides an **absolute measure**.
- **Clinically interpretable**: "How much longer does a patient survive on average within a given time frame?"

Built-in selection bias HR

Here is the definition of Hazard ratio:

$$\exp(\beta) = \frac{\log P(T^1 > t)}{\log P(T^0 > t)}$$

Under proportional hazard assumption, it goes:

$$\begin{aligned}\exp(\beta) &= \frac{\lim\limits_{h\to 0} P(t \leqslant T < t+h | T \geqslant t, A = 1)}{\lim\limits_{h\to 0} P(t \leqslant T < t+h | T \geqslant t, A = 0)} \\ &= \frac{\lim\limits_{h\to 0} P(t \leqslant T^1 < t+h | T^1 \geqslant t)}{\lim\limits_{h\to 0} P(t \leqslant T^0 < t+h | T^0 \geqslant t)}\end{aligned}$$

First, the right-hand expression shows that $\exp(\beta)$ contrasts the hazard functions **with** and **without** intervention for two separate groups of individuals who survive time $t > 0$.
If it exists a treatment effect then, population **with** and **without** intervention cannot be compared anymore.

Built-in selection bias HR

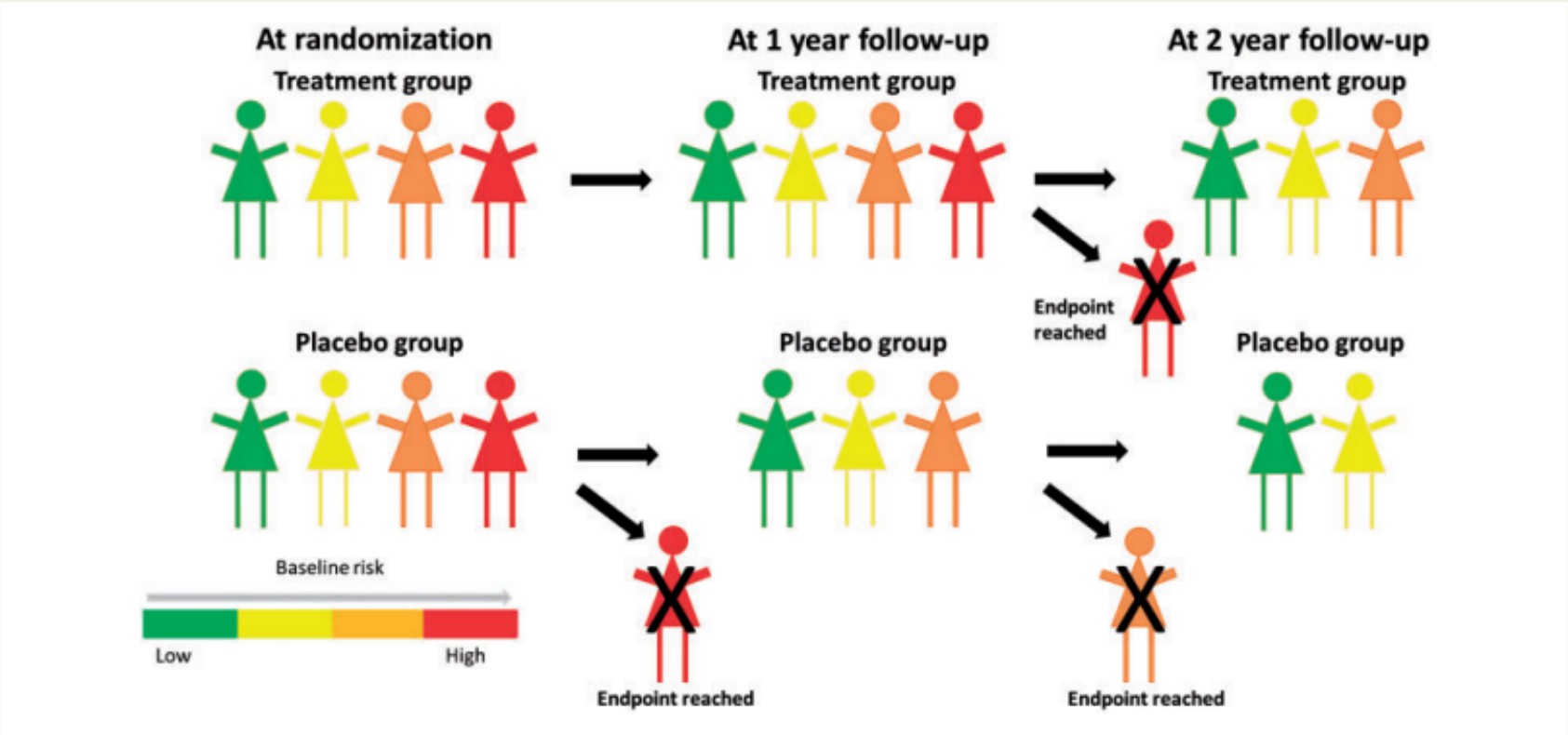


Figure 2: This schematic drawing illustrates the built-in selection bias in population-level hazard ratios: by definition, the population-level hazard ratio at a given time point is based on individuals who survived up to that time point, thereby it is a comparison between the unbalanced groups (from Stensrud et al. (2019))