Robust estimation of the Average Treatment Effect (ATE)
in presence of right-censoring and competing risks

Brice Ozenne[1,2], Thomas Scheike[1], Thomas Gerds[1,3]

1. Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark
2. Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark
3. Danish Heart Foundation, Copenhagen, Denmark

December 20th, 2020 - CMStatistics

A typical request

Population:

▷ patients after atrial fibrillation (n=43299)

Outcome: time to event

▷ time to stroke / thromboembolism

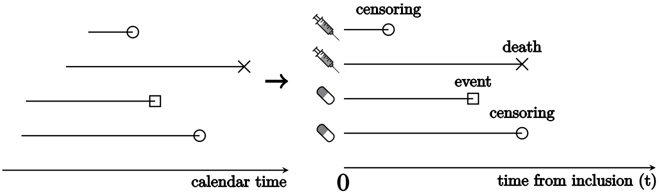
Aim: evaluate differences between alternative drug treatments

▷ non-vitamin K antagonist oral anticoagulants

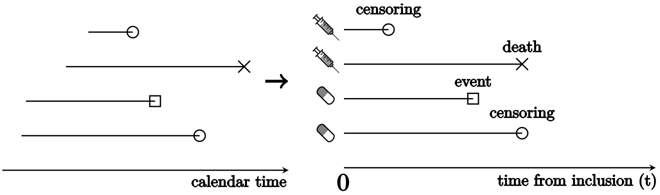
in risk of experiencing the event within τ years.

▷ 1-year risk

Typical data and typical problems



Typical data and typical problems



- right-censoring
  - ▷ assumed at random, e.g. leaving Denmark
- competing risks
  - ▷ mainly death
- non-randomized experiment
  - ▷ accounting for known and observed confounders
- dynamic treatment regimes
  - ▷ intention to treat analysis: differences between alternative drug treatment initiations

Roadmap

- Develop a robust estimator based on:
  - F_1: model the risk of the event(s) using Cox models
  - π: model the treatment allocation using a logistic regression
  - G: model the censoring process using a Cox model

"Clever" combination, unbiased if only "some" mispecifications

- Identify the asymptotic distribution of the estimator
  - via its iid decomposition
  - confidence intervals/p-values

- Implementation
  - `ate` function in the R package `riskRegression`

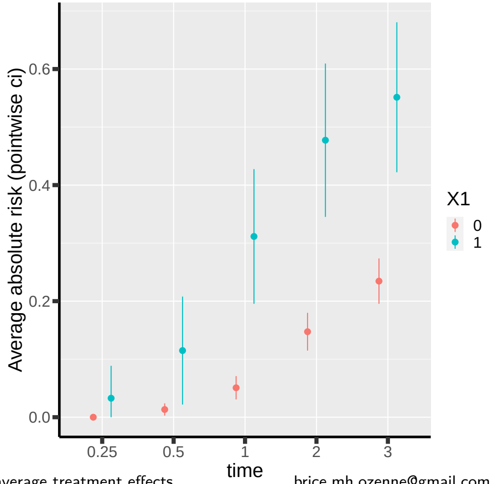
Proposed software

```
e.ate1 <- ate(list(Hist(time,event) ~ strata(X1) + X2,
                                      Hist(time,event) ~ strata(X1) +
     X2 + X6),
    treatment = X1 ~ X2 + X6,
    censor = Surv(time,event==0) ~ X2 + X6,
    data = mydata, times = c(0.25,0.5,1,2,3),
    cause = 1, verbose = FALSE)

summary(e.ate1, type = c("mean","diff"))
```

```
    Average treatment effect for cause 1

 [...]
  time X1=A risk(X1=A) X1=B risk(X1=B) difference            ci  p.value
  0.25    0      0.0000   1     0.0327     0.0327 [-0.02;0.09] 2.51e-01
  0.50    0      0.0133   1     0.1149     0.1016  [0.01;0.20] 3.34e-02
  1.00    0      0.0507   1     0.3115     0.2608  [0.14;0.38] 1.17e-05
  2.00    0      0.1474   1     0.4773     0.3300  [0.20;0.46] 1.39e-06
  3.00    0      0.2345   1     0.5514     0.3168  [0.18;0.45] 2.69e-06
 [...]
```

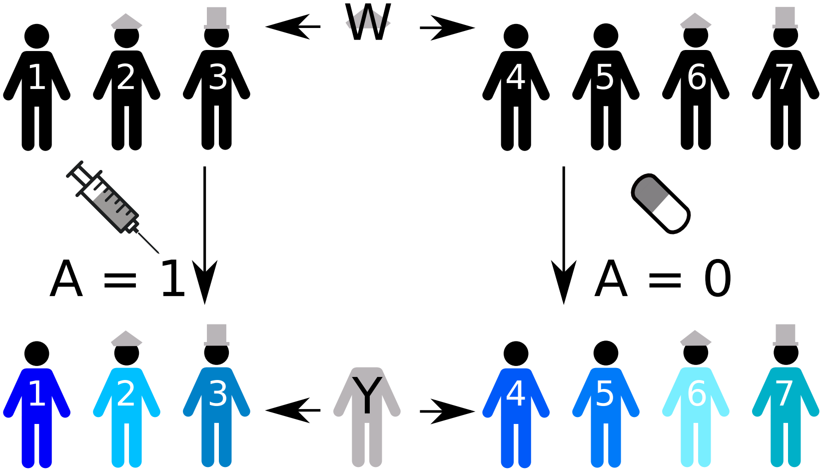
Proposed software

```
ggplot2::autoplot(e.ate1)
```

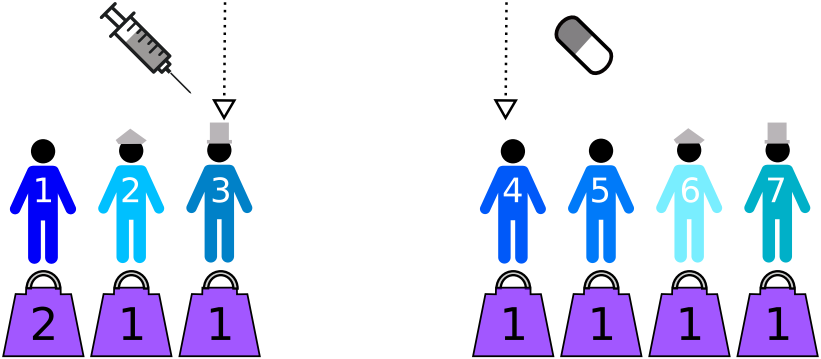


Standard estimators (no censoring)

- intuition
- formally


W
1
2
3
4
5
6
7
A = 1
A = 0
1
2
3
Y
4
5
6
7

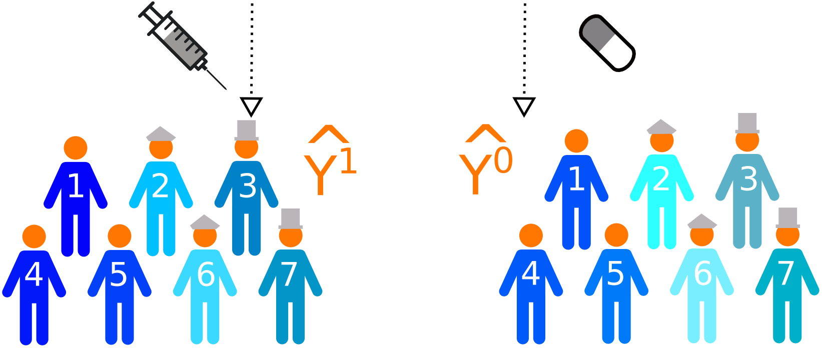
IPTW



Requirement: valid model for the propensity of treatment

$$\pi(W) = \mathbb{P}(A = 1|W)$$

G-formula



Requirement: valid model for the risk

$$F_1(t|A,W) = \mathbb{P}(T \leq t, \Delta = 1|A,W) = \int_0^t S(s-|A,W)\lambda_1(s|A,W)\mathrm{d}s$$

Full data case

Under causal assumptions, the likelihood of the full data O^F is:

$$\mathcal{L}(\Psi|O^F) = \mathbb{P}\left[Y^1(\tau), Y^0(\tau), W|\Psi, \eta_1\right] \mathbb{P}\left[A_i|W, \eta_2\right]$$
$$\propto \mathbb{P}\left[Y^1(\tau), Y^0(\tau), W|\Psi, \eta_1\right]$$

with $\Psi(\tau) = \mathbb{E}\left[Y^1(\tau) - Y^0(\tau)\right]$ ATE
$\eta = (\eta_1, \eta_2)$ nuisance parameters

- estimator: $\widehat{\Psi}^F = \frac{1}{n}\sum_{i=1}^n Y_i^1(\tau) - Y_i^0(\tau)$

Full data case

Under causal assumptions, the likelihood of the full data O^F is:

$$\mathcal{L}(\Psi|O^F) = \mathbb{P}\left[Y^1(\tau), Y^0(\tau), W|\Psi, \eta_1\right]\mathbb{P}\left[A_i|W, \eta_2\right]$$
$$\propto \mathbb{P}\left[Y^1(\tau), Y^0(\tau), W|\Psi, \eta_1\right]$$

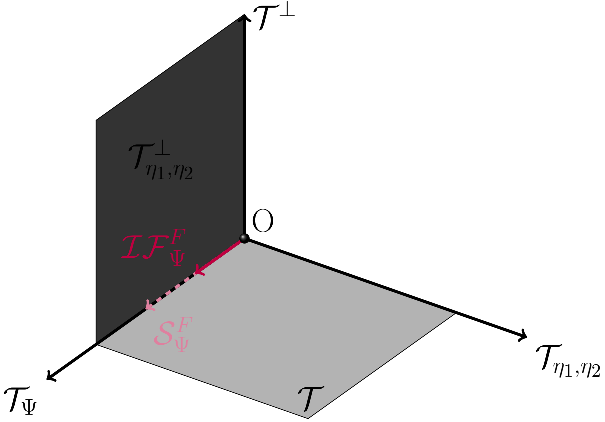
with $\Psi(\tau) = \mathbb{E}\left[Y^1(\tau) - Y^0(\tau)\right]$ ATE
$\eta = (\eta_1, \eta_2)$ nuisance parameters

- estimator: $\widehat{\Psi}^F = \frac{1}{n}\sum_{i=1}^n Y_i^1(\tau) - Y_i^0(\tau)$

- influence function:

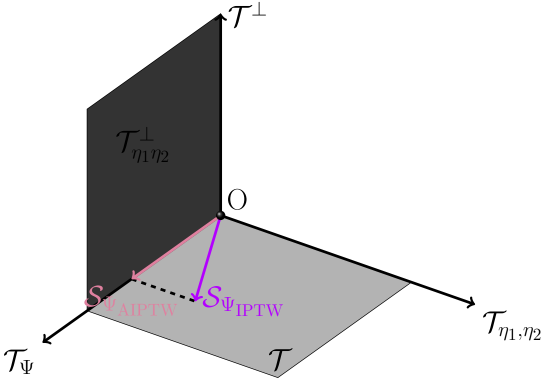
$$\sqrt{n}\left(\widehat{\Psi}^F - \Psi\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{IF}_\Psi^F(O_i^F) + o_p(1)$$
$$\mathcal{IF}_\Psi^F(O_i^F) = Y_i^1(\tau) - Y_i^0(\tau) - \Psi(\tau)$$

Full data case - geometry

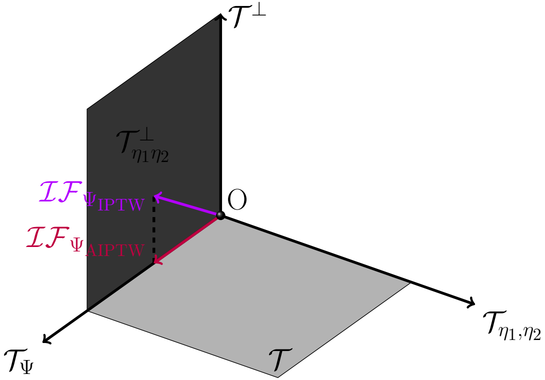


$\mathcal{T}$: tangent space $\approx$ space spanned by the score $\left(\mathcal{S}^F_{\Psi}, \mathcal{S}^F_{\eta_1}, \mathcal{S}^F_{\eta_2}\right)$

Observed data case - geometry



Observed data case - geometry



Notations

$O = \{T_i, \Delta_i, A_i, W_i\}_{i=1}^n$: random iid sample of n individuals:

- T event time
- Δ event type indicator (1 event of interest, 2 competing event)
- A binary treatment variable
- W baseline covariates

- $Y(\tau) = \mathbb{1}_{T \leq \tau, \Delta=1}$ outcome
- $Y^a(\tau)$ potential outcome

- $r^a(\tau) = \mathbb{E}\left[Y^a(\tau)\right]$ average risk under treatment a
- $\Psi(\tau) = r^1(\tau) - r^0(\tau)$ ATE

Standard estimators

$$\widehat{\Psi}_{\text{IPTW}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i(\tau)\left(\frac{A_i}{\widehat{\pi}_n(W_i)} - \frac{1-A_i}{1-\widehat{\pi}_n(W_i)}\right)\right\}.$$

$$\widehat{\Psi}_{\text{G}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{F}_{1n}(\tau|A=1,W_i) - \hat{F}_{1n}(\tau|A=0,W_i)\right\}.$$

$$\begin{aligned}\widehat{\Psi}_{\text{AIPTW}}(\tau) =& \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Y_i(\tau)A_i}{\widehat{\pi}_n(W_i)} + \hat{F}_{1n}(\tau|A=1,W_i)\left(1-\frac{A_i}{\widehat{\pi}_n(W_i)}\right)\right.\\ &\left. -\frac{Y_i(\tau)(1-A_i)}{1-\widehat{\pi}_n(W_i)} - \hat{F}_{1n}(\tau|A=0,W_i)\left(1-\frac{1-A_i}{1-\widehat{\pi}_n(W_i)}\right)\right\}\\ =& \widehat{r}^{1}_{\text{AIPTW}}(\tau) - \widehat{r}^{0}_{\text{AIPTW}}(\tau)\end{aligned}$$

Proposed estimator (censoring)

- First attempt
- Augmented estimator

Notations

$O = \{(\tilde{T}_i, \tilde{\Delta}_i, A_i, W_i)\}_{i=1}^n$: random iid sample of

- C censoring time
- $\tilde{T} = T \wedge C$ right-censored event time
- $\tilde{\Delta} = \Delta \mathbb{1}_{T \leq C}$ observed event type indicator
  (0 censoring, 1 event of interest, 2 competing event)

- $G(t|A, W) = \mathbb{P}(C > t|A, W)$ model for the censoring process
  (η_3 nuisance parameter for the censoring process)

Note: $Y(\tau)$ cannot be computed for the censored observations.

IPW estimator

We cannot compute:

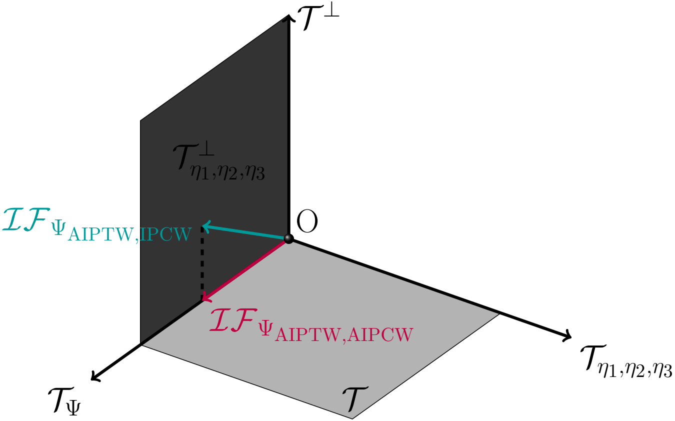
$$\widehat{r}_{1,\text{AIPTW}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Y_i(\tau)A_i}{\widehat{\pi}_n(W_i)} + \hat{F}_{1n}(\tau|A=1,W_i)\left(1-\frac{A_i}{\widehat{\pi}_n(W_i)}\right)\right\}$$

but we can compute:

$$\widehat{r}^{1}_{\text{AIPTW,IPCW}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Y_i(\tau)A_i\mathbb{1}_{C_i>T_i\wedge\tau}}{\widehat{\pi}_n(W_i)\widehat{G}_n(\tilde{T}_i|A_i,W_i)} + \hat{F}_{1n}(\tau|A=1,W_i)\left(1-\frac{A_i}{\widehat{\pi}_n(W_i)}\right)\right\}$$

(using that $\mathbb{1}_{C>T\wedge\tau}Y(\tau) = \underbrace{\mathbb{1}_{\tilde{T}>\tau}Y(\tau)}_{=0} + \mathbb{1}_{\tilde{T}\leq\tau,\tilde{\Delta}\neq 0}Y(\tau)$)

Geometric point of vue



Augmentation term

Using chapter 10 of Tsiatis (2007), the influence function for $\widehat{r}^1$ is:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{1}_{C_i>T_i\wedge\tau}\mathcal{IF}_{\widehat{r}^1_{\text{AIPTW}}(\tau)}(O_i)}{\widehat{G}_n(\tilde{T}_i|A_i,W_i)} + \int_0^{\tau\wedge\tilde{T}_i}\frac{\mathbb{E}\left[\mathcal{IF}_{\widehat{r}^1_{\text{AIPTW}}(t)}(O_i)\middle| T>t,A_i,W_i\right]}{G(t|A_i,W_i)}dM_i^C(t)\right\}$$

Augmentation term

Using chapter 10 of Tsiatis (2007), the influence function for $\widehat{r}^1$ is:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{1}_{C_i > T_i \wedge \tau}\mathcal{IF}_{\widehat{r}^1_{\text{AIPTW}}(\tau)}(O_i)}{\widehat{G}_n(\tilde{T}_i|A_i,W_i)} + \int_0^{\tau\wedge\tilde{T}_i}\frac{\mathbb{E}\left[\mathcal{IF}_{\widehat{r}^1_{\text{AIPTW}}(t)}(O_i)\middle| T > t, A_i, W_i\right]}{G(t|A_i,W_i)}dM_i^C(t)\right\}$$

Loosely speaking:

$$\mathcal{IF}_{\widehat{r}_{\text{AIPTW}}(\tau)}(O_i) \propto \frac{Y_i(\tau)A_i}{\widehat{\pi}_n(W_i)} + \hat{F}_{1n}(\tau|A=1,W_i)\left(1-\frac{A_i}{\widehat{\pi}_n(W_i)}\right)$$

AIPTW,AIPCW estimator

We get, after some calculations:

$$\widehat{\Psi}_{\text{AIPTW,AIPCW}}(\tau) =$$

$$\frac{1}{n}\sum_{i=1}^{n}\left\{ \frac{\mathbb{1}_{\tilde{T}_i\leq\tau,\tilde{\Delta}_i\neq 0}}{\widehat{G}_n(\tilde{T}_i|A_i,W_i)} Y_i(\tau)\left(\frac{A_i}{\widehat{\pi}_n(W_i)} - \frac{1-A_i}{1-\widehat{\pi}_n(W_i)}\right)\right.$$

$$+\hat{F}_{1n}(\tau|A=1,W_i)\left(1-\frac{A_i}{\widehat{\pi}_n(W_i)}\right) - \hat{F}_{1n}(\tau|A=0,W_i)\left(1-\frac{1-A_i}{1-\widehat{\pi}_n(W_i)}\right)$$

$$\left. +\hat{I}(\tilde{T}_i,\tau|A_i,W_i)\left(\frac{A_i}{\widehat{\pi}_n(W_i)} - \frac{1-A_i}{1-\widehat{\pi}_n(W_i)}\right)\right\}$$

where:

$$\hat{I}(\tilde{T}_i,\tau|A_i,W_i) = \int_0^{\tilde{T}_i\wedge\tau} \frac{\hat{F}_{1n}(\tau|A_i,W_i) - \hat{F}_{1n}(t|A_i,W_i)}{\widehat{S}_n(t|A_i,W_i)} \frac{1}{\widehat{G}_n(t|A_i,W_i)} d\hat{M}_i^C(t)$$

Asymptotic properties

- Bias
- Uncertainty

Properties of the AIPTW,AIPCW estimator

Denoting by * the large sample limit of the estimators

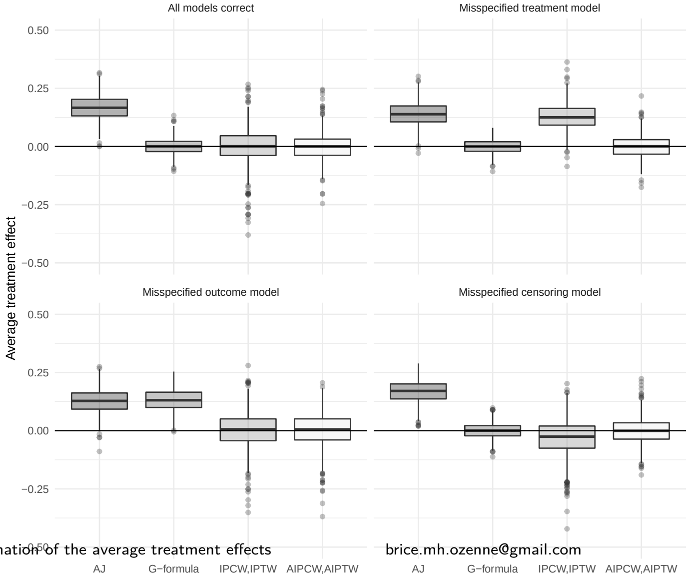
▷ $F_1^*(\tau|a,w) = F_1(\tau,a,w)$ means the outcome model is correctly specified

Theorem

The estimator $\widehat{\Psi}_{\text{AIPTW,AIPCW}}(\tau)$ *is consistent whenever one of the following conditions is satisfied for all* $s \in [0,\tau]$, $a \in \{0,1\}$ *and almost all* w*:*

- $G^*(s|a,w) = G(s|a,w)$ *and* $F_1^*(\tau|a,w) = F_1(\tau,a,w)$
- $G^*(s|a,w) = G(s|a,w)$ *and* $\pi^*(w) = \pi(w)$
- $F_1^*(s|a,w) = F_1(s|a,w)$ *and* $S^*(s|a,w) = S(s|a,w)$

Does it really works? Simulation study with $n = 500$



Back to the G-formula estimator

$$\widehat{\Psi}_{\mathrm{G}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{F}_{1n}(\tau|A=1,W_i) - \hat{F}_{1n}(\tau|A=0,W_i)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\{F_1(\tau|A=1,W_i,\hat{\eta}_1) - F_1(\tau|A=0,W_i,\hat{\eta}_1)\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} f(W_i,\hat{\eta}_1)$$

Two (correlated) sources of uncertainty:

- Averaging over the empirical distribution of W, H_n (instead of the expectation over H, the true distribution of W)
- Plug-in the estimated parameters for computing F_1 (instead of η_1, the true value of the parameters)

Functional delta method

$$\widehat{\Psi}_G(\tau) - \Psi^*_G(\tau) = \int \underbrace{\left(f(w,\hat{\eta}_1) - f(w,\eta_1)\right)}_{\text{uncertainty "nuisance parameters"}} dH_n(w)$$

$$+ \int f(w,\eta_1) d \underbrace{\left(H_n(w) - H(w)\right)}_{\text{uncertainty "averaging"}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} \nabla_{\eta_1} f(W_j,\eta_1) \right) IF_{\hat{\eta}_1}(O_i)$$

$$+ \frac{1}{n}\sum_{i=1}^{n} f(W_i,\eta_1) - \Psi^*_G(\tau) + o_p\left(n^{-\frac{1}{2}}\right)$$

Average of independent terms:

- asymptotically normally distributed
- variance estimator: average of the squared terms

Take-home message

Several estimator of the average treatment effect:

- no censoring: G-formula; ~~IPTW~~; AIPTW
- censoring: G-formula; ~~IPTW,IPCW~~; AIPTW,AIPCW

These estimators are asymptotically normally distributed:

- "closed form formula" for the asymptotic variance

G-formula vs. AIPTW/AIPTW,AIPCW:

- bias-variance tradeoff between
- prior knowledge

Implemented in the function `ate` in `riskRegression`.

- more details/reference (Ozenne et al., 2020)

Reference I

Ozenne, B. M. H., Scheike, T. H., Stærk, L., and Gerds, T. A. (2020). On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biometrical Journal*, 62(3):751–763.

Tsiatis, A. (2007). *Semiparametric theory and missing data*. Springer Science & Business Media.

Properties of the AIPTW,AIPCW estimator

Proof:

- correctly specified censoring model:
  $\widehat{\Psi}_{\text{AIPTW,AIPCW}}(\tau)$ and $\widehat{\Psi}_{\text{AIPTW}}(\tau)$ have the same limit.

- correctly specified "outcome" models:
  $\widehat{\Psi}_{\text{AIPTW,AIPCW}}(\tau)$ and $\widehat{\Psi}_{G-formula}(\tau)$ have the same limit.

$$
\begin{aligned}
&\frac{Y_i(\tau)\mathbb{1}_{\tilde{T}_i\leq\tau,\tilde{\Delta}\neq 0}}{G^*(\tilde{T}_i|A_i,W_i)} + \int_0^{\tilde{T}_i\wedge\tau} \frac{F_1^*(\tau|A_i,W_i)-F_1^*(t|A_i,W_i)}{S^*(t|A_i,W_i)G^*(t|A_i,W_i)} dM_i^{C,*}(t) \\
=&Y_i(\tau) + \int_0^{\tilde{T}_i\wedge\tau} \frac{\frac{F_1(\tau|A_i,W_i)-F_1(t|A_i,W_i)}{S(t|A_i,W_i)} - Y_i(\tau)}{G^*(t|A_i,W_i)} dM_i^{C,*}(t) \\
=&Y_i(\tau) + \int_0^{\tilde{T}_i\wedge\tau} \frac{\mathbb{E}\left[Y_i(\tau)|T_i>t,A_i,W_i\right]-Y_i(\tau)}{G^*(t|A_i,W_i)} dM_i^{C,*}(t).
\end{aligned}
$$