Faculty of Health Sciences

Assessing treatment effects on registry data in presence of competing risks

Brice Ozenne, Christian Torp-Pedersen, Thomas A. Gerds

May 23, ISCB 2017, Vigo

Danish registry data

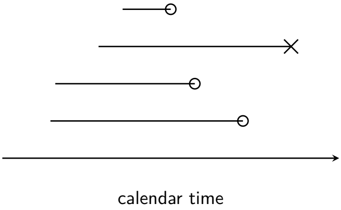
Systematic collection of health data

- observational data
- large sample size, good representativity
  - can be exhaustive!

Danish registry data

Systematic collection of health data

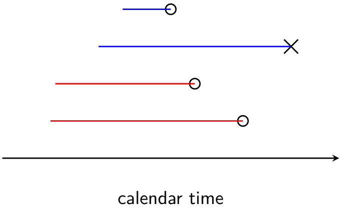
- observational data
- large sample size, good representativity
  - can be exhaustive!



Danish registry data

Systematic collection of health data

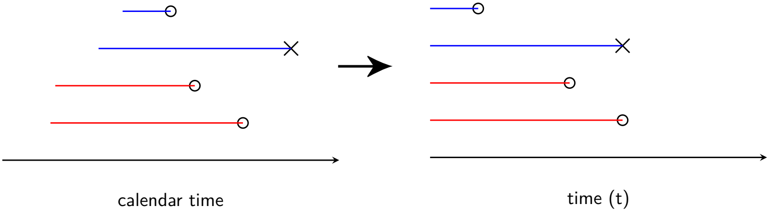
- observational data
- large sample size, good representativity
  - can be exhaustive!



Danish registry data

Systematic collection of health data

- observational data
- large sample size, good representativity
  - can be exhaustive!



Quantity of interest (1)

The absolute risk - or cumulative incidence function:
(Benichou and Gail, 1990)

$$r_1(t|X,Z) = \mathbb{P}\left[T \leq t, D = 1|X,Z\right]$$

T	time to event	
D	event type	$\tilde{D} = 1$ (cause of interest)
t	time horizon	1-year
X, Z	baseline covariates	

Quantity of interest (1)

The absolute risk - or cumulative incidence function:
(Benichou and Gail, 1990)

$$\begin{aligned} r_1(t|X,Z) &= \mathbb{P}\left[T \leq t, D = 1|X,Z\right] \\ &= \int_0^t S(s-|X,Z)\ \lambda_1(s|X,Z) ds \end{aligned}$$

T	time to event	
D	event type	$\tilde{D} = 1$ (cause of interest)
t	time horizon	1-year
X, Z	baseline covariates	
$S(t\|X,Z)$	event-free survival	
$\lambda_1(t\|X,Z)$	cause specific hazard (event of interest)	

Statistical issues

Account for the observational nature of the data

- **Covariate adjustment**: find an appropriate model for $S(t|X,Z)$ and $\lambda_1(t|X,Z)$
- **Causal inference**: standardize the absolute risks to estimate the causal effect of the treatment (under assumptions)

Asymptotics to get p.value, confidence interval/bands

- Bootstrap approaches are very time consuming

(large n, complex model to fit)

Estimation of the absolute risk

Model - hazard

Stratified cause specific Cox model:

$$\lambda_{j,Z}(t|X,Z) = \lambda_{0j,Z}(t) exp(X\beta_j)$$

X	baseline covariates (linear predictor)
Z	baseline covariates (strata variable)
β_j	regression coefficients
$\lambda_{0j,z}$	cause specific baseline hazard for strata z
j	type of event

Model - Survival

Product integral estimator:

$$S(t|X, Z) = \prod_{s \leq t} \left(1 - \sum_{j=1}^{d} d\Lambda_{j,Z}(t|X)\right)$$

$\Lambda_j(t|X)$ cause specific cumulative hazard for the event j

Model - Survival

Product integral estimator:

$$S(t|X,Z) = \prod_{s \le t} \left(1 - \sum_{j=1}^{d} d\Lambda_{j,Z}(t|X)\right)$$

$\Lambda_j(t|X)$ cause specific cumulative hazard for the event j

Exponential approximation:

$$S(t|X,Z) = \exp\left(-\int_0^t \sum_{j=1}^{d} \lambda_{j,Z}(s|X)ds\right)$$

- functional used to derive the asymptotic distribution of $r_1(t|X)$

Estimation of the average treatment effect

Quantity of interest (2)

We are intesting in comparing:

- $r_1(t|X,Z)$ if the patient would receive treatment 1

$$r_1(t|do(T=T_1),X,Z)$$

- $r_1(t|X,Z)$ when the patient would receive treatment 0

$$r_1(t|do(T=T_0),X,Z)$$

Quantity of interest (2)

We are intesting in comparing:

- $r_1(t|X,Z)$ if the patient would receive treatment 1

$$r_1(t|do(T=T_1),X,Z)$$

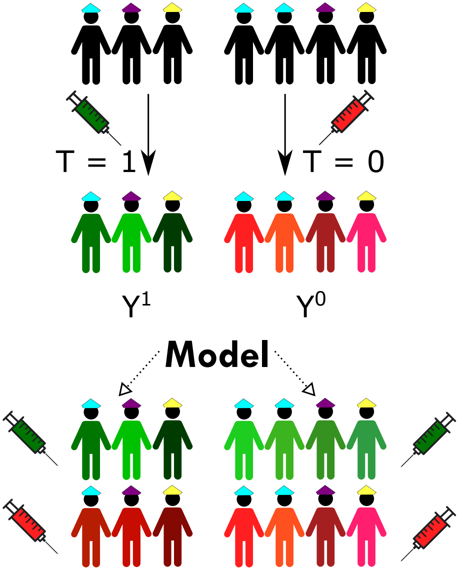
- $r_1(t|X,Z)$ when the patient would receive treatment 0

$$r_1(t|do(T=T_0),X,Z)$$

(Individual) causal treatment effect:

$$CTE(t,T_1,T_0|X,Z) = r_1(t|do(T=T_1),X,Z) - r_1(t|do(T=T_0),X,Z)$$

G formula



Quantity of interest (2) - feasible

Average treatment effect

$$ATE(t, T_1, T_0) = \mathbb{E}_{X,Z}[r_1(t|T = T_1, X, Z) \\ - r_1(t|T = T_0, X, Z)]$$

Assumptions

- no unmeasured confounders
- positivity
- well-defined intervention
- correctly specified model for r_1

Application - context

Objective:

- to compare 3 antiplatelet regimens using the danish registry
- n = 19223 patients
- period 2007-2010
- time horizon: 1 year

Outcome

- date of first stroke event after atrial fibrilation

(n=1610, 8.4%)

Competing event:

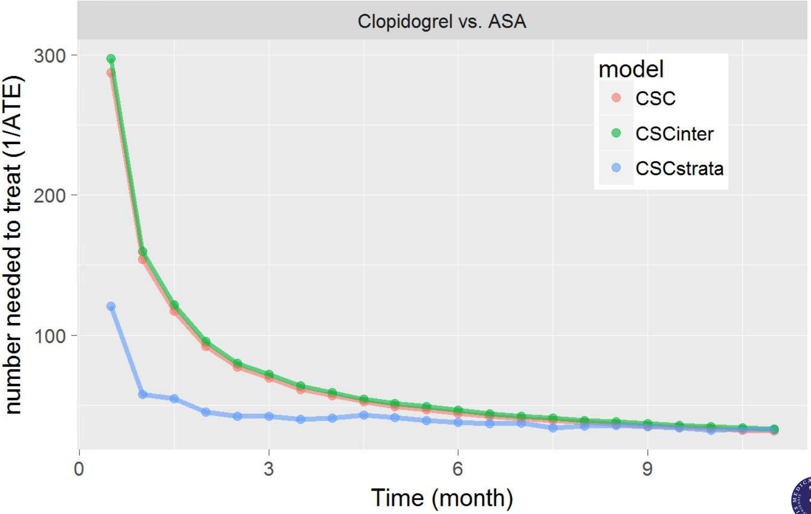
- death (n=677, 3.5%)

Covariates: age, period, gender + other risk factors

Application - model

CSC	Two cause specific Cox model
CSC inter	CSC
	+ interactions between treatment and gender, age, year
	+ cubic spline on age
CSC strata	stratified CSC on treatment, gender, year

Application - results



Asymptotics

Asymptotics (for r_1)

We observe a sample $(\mathcal{X}_i)_{i \in \{1,\ldots,n\}}$ of n replication of $(\tilde{T}, \tilde{D}, X, Z)$
$\tilde{T}$ observed time to event
$\tilde{D}$ observed event type

Asymptotics (for r_1)

We observe a sample $(\mathcal{X}_i)_{i\in\{1,\dots,n\}}$ of n replication of $(\tilde{T},\tilde{D},X,Z)$
$\tilde{T}$ observed time to event
$\tilde{D}$ observed event type

Assumption: independent and identically distributed replications
no tied event

Asymptotics (for r_1)

We observe a sample $(\mathcal{X}_i)_{i\in\{1,\dots,n\}}$ of n replication of $(\tilde{T}, \tilde{D}, X, Z)$
$\tilde{T}$ observed time to event
$\tilde{D}$ observed event type

We can write:

$$r_1(t|X,Z) = \int_0^t S(s-|X)\, d\Lambda_{1,Z}(s|X)$$
$$= \phi(\Lambda_{01,Z}, \Lambda_{02,Z}, \beta_1, \beta_2, X)$$

Asymptotics (for r_1)

We observe a sample $(\mathcal{X}_i)_{i\in\{1,\dots,n\}}$ of n replication of $(\tilde{T},\tilde{D},X,Z)$
$\tilde{T}$ observed time to event
$\tilde{D}$ observed event type

We can write:

$$
\begin{aligned}
r_1(t|X,Z) &= \int_0^t S(s-|X)\, d\Lambda_{1,Z}(s|X) \\
&= \phi(\Lambda_{01,Z},\Lambda_{02,Z},\beta_1,\beta_2,X)
\end{aligned}
$$

- We know the asymptotics of $\Lambda_{01,Z},\Lambda_{02,Z},\beta_1,\beta_2$
- Can we infer the asymptotics of r_1?

Functional Delta method (Van der Vaart, 2000)

Given the following iid decomposition for the statistic T:

$$\sqrt{n}(\hat{T} - T) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{IF}_T(\mathcal{X}_i) + o_p(1)$$

where $\mathcal{IF}_T(\mathcal{X}_i)$ is the influence of a sample $\mathcal{X}_i$ on the statistic T.

Functional Delta method (Van der Vaart, 2000)

Given the following iid decomposition for the statistic T:

$$\sqrt{n}(\hat{T} - T) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{IF}_T(\mathcal{X}_i) + o_p(1)$$

where $\mathcal{IF}_T(\mathcal{X}_i)$ is the influence of a sample $\mathcal{X}_i$ on the statistic T.

Then:

$$\sqrt{n}(\phi(\hat{T}) - \phi(T)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi'(T).\mathcal{IF}_T(\mathcal{X}_i) + o_p(1)$$

where ϕ is Hadamard differentiable

Implementation

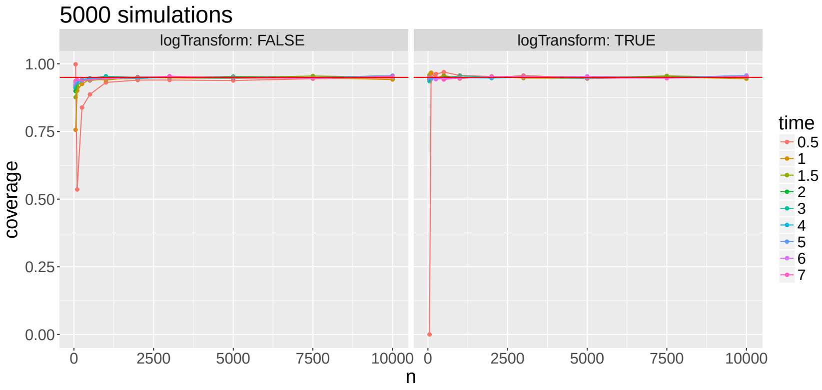
Given a Cox model:

- we can compute the baseline hazard $\Lambda_{0j,z}$ and its influence function $\mathcal{IF}_{\Lambda_{0j,z}}$
- we can compute the $\mathcal{IF}_{\beta_j} = f(\mathcal{IF}_{\Lambda_{0j,z}}, \beta_j, \mathcal{I}_j)$
- we can compute the $\mathcal{IF}_{r_1} = f(\mathcal{IF}_{\beta_1}, \mathcal{IF}_{\beta_2}, \mathcal{IF}_{\Lambda_{01,z}}, \mathcal{IF}_{\Lambda_{02,z}}, \Lambda_{01,z}, \Lambda_{02,z}, \beta_1, \beta_2, X)$ and obtain σ_{r_1}

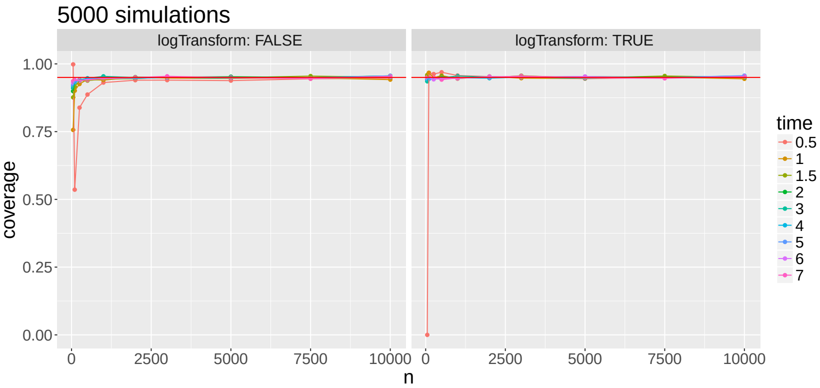
Confidence intervals:

- original scale: $[r_1 \pm 1.96\sigma_{r_1}] \cap [0;1]$
- log-log scale: $\left[\exp\left(-\exp\left(\log(-\log(r_1)) \pm 1.96\sigma_{\log(-\log(r_1))}\right)\right)\right]$

Simulation study



Simulation study



- the log-log transformation improves the coverage in small samples

Influence test

Influence test

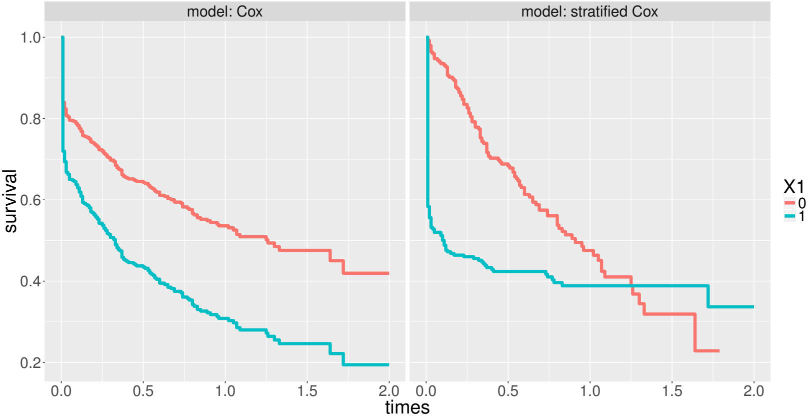
What if we would like to compare the ATE obtained by two models?

Under a model $\mathcal{M}$:

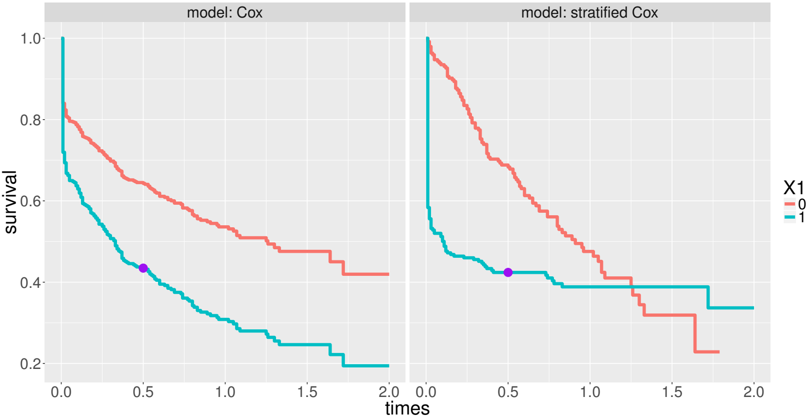
$$\sqrt{n}\left(\widehat{ATE}^{\mathcal{M}}(t,T_1,T_0) - ATE^{\mathcal{M}}(t,T_1,T_0)\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathcal{IF}_{ATE}^{\mathcal{M}}(\mathcal{X}_i|t,T_1,T_0) + o_p(1)$$

So under $\mathcal{H}_0$: $ATE^{\mathcal{M}_0}(t,T_1,T_0) - ATE^{\mathcal{M}_1}(t,T_1,T_0) = 0$
we can compute the asymptotic variance of the test statistic

Influence test - Example



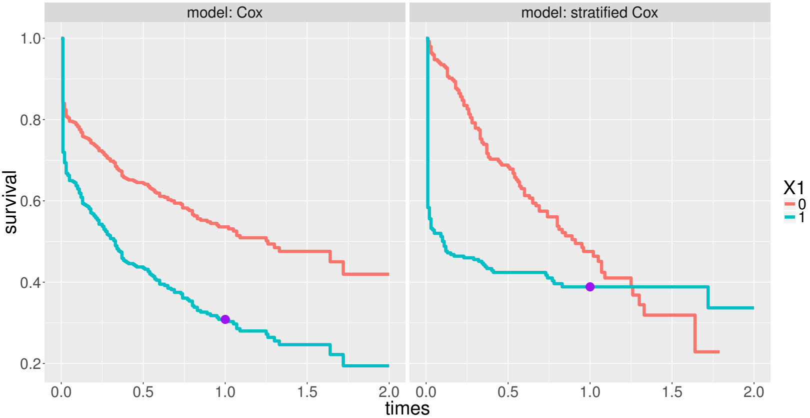
Influence test - Example



- influence test at time 0.5:

```
          delta   se.delta   t.delta   p.value
1 -0.01063889 0.03655373 -0.291048 0.7710146
```

Influence test - Example



- influence test at time 1:

```
         delta   se.delta  t.delta    p.value
1 0.08007145 0.03849627 2.07998 0.03752741
```

Summary

Absolute risks can be used to:

- assess patient-specific risks
- compare treatment effects

The influence function is a versatile tool:

- estimate the asymptotic variance
- compare an estimate across models
- and identify influential observations!

R package `riskRegression` (version $\geq$ 1.4.3)

- available on CRAN
- software paper submitted

Reference I

Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, pages 813–826.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509.

Gerds, T. and Schumacher, M. (2001). On functional misspecification of covariates in the cox regression model. *Biometrika*, pages 572–580.

Levine, M. N. and Julian, J. A. (2008). Registries that show efficacy: good, but not good enough. *Journal of Clinical Oncology*, 26(33):5316–5319.

Reid, N. and Crépeau, H. (1985). Influence functions for proportional hazards regression. *Biometrika*, 72(1):1–9.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Other issues (won't be discussed)

- **quality of the data**

e.g. negative age `n = 3069/318,458`
event posterior to death `n = 6852/318,458`

- **definition changing in time**

e.g. outcome /exposure $t \in [11/1982; 12/2015]$

- **unmeasured confounders**

e.g. confounding by indication

- **missing data**

I have no missing data!

Product integral vs. exponential approximation

Product integral estimator:

- $S(t|X,Z) + r_1(t|X,Z) + r_2(t|X,Z) = 1$
- Numerically instable at the end of the follow-up

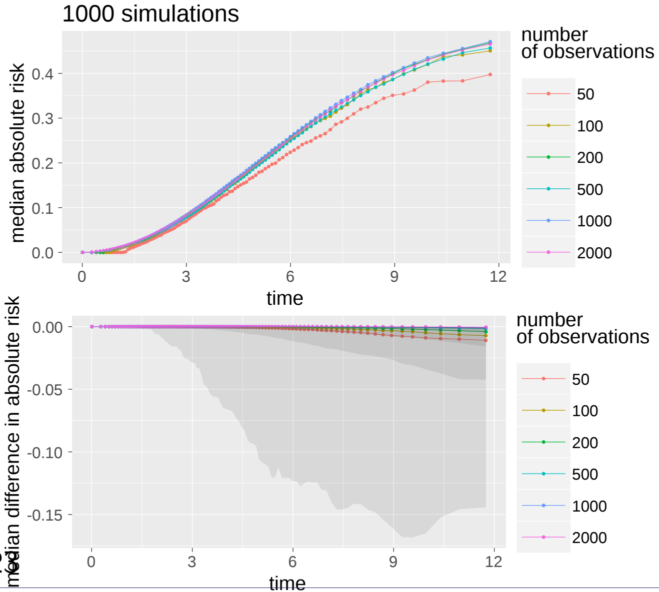
Exponential approximation:

- $S(t|X,Z) + r_1(t|X,Z) + r_2(t|X,Z) \approx 1$
- Easier to work with to derive asymptotic properties for $r_1(t|X,Z)$
- Efron method to handle tied events

Simulation study:

```
e.CSC <- CSC(Hist(time,event) ~ X1 + X6, data = dt,
    iid = FALSE)
```

Product integral vs. exponential approximation



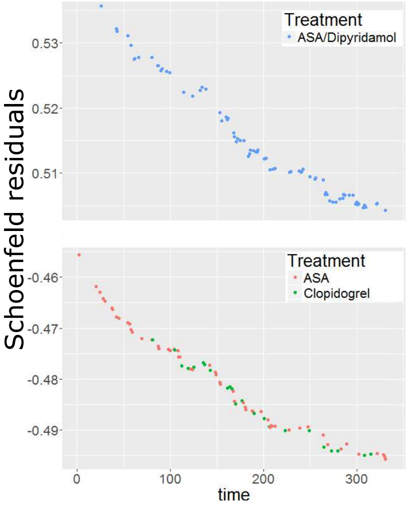
Application - checking assumptions

Check proportionality assumption:

$$\mathbb{E}\left[r_{ij}\right] \approx \beta_j(t_i) - \widehat{\beta_j}$$

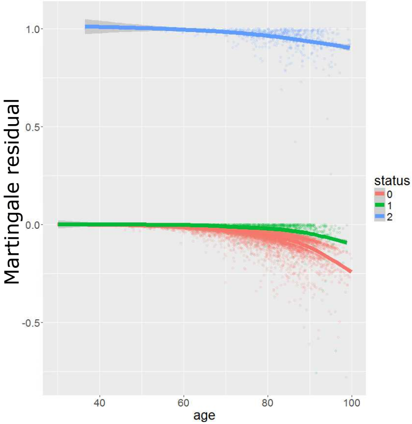
- only rejected for one treatment modality

We can stratify on treatment!



Application - checking assumptions

Check linearity assumption:



Variable age

- additional risk after 75 years
- approx. linear

IF - Notations

$\mathbb{F}$	cumulative distribution function (CDF)
$T = \phi(\mathbb{F})$	statistic
$(\mathcal{X}_i)_{i \in \{1,\ldots,n\}}$	an iid random sample of $\mathbb{F}$
$\mathbb{F}_n$	empirical CDF of $(\mathcal{X}_i)_{i \in \{1,\ldots,n\}}$ $\hat{\mathbb{F}}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$ (univariate) $\hat{\mathbb{F}}_n(A) = \frac{1}{n} \sum_{i=1}^n \delta_{\mathcal{X}_i}(A)$
$\hat{T} = \phi(\hat{\mathbb{F}}_n)$	empirical statistic

$\delta_X(A) = \mathbb{1}_A(X)$ is the Dirac measure

IF - Example of statistical functional

Breslow estimator (no strata, no covariates):

$$\begin{aligned}\Lambda_{01}(t) &= \int_0^t \frac{\text{death at time s}}{\text{at risk at time s}} ds \\ &= \int_0^t \frac{d\mathbb{F}(\tilde{T} = s, \tilde{D} = 1)}{\int \mathbb{1}_{u \geq s} d\mathbb{F}(\tilde{T} = u)} \\ &= \phi(\mathbb{F})(t) = T(t)\end{aligned}$$

$$\begin{aligned}\hat{\Lambda}_{01}(t) &= \sum_{s \in \{(\tilde{T}_j)_{j \in \{1,\dots,n\}}; \tilde{T}_j < t\}} \frac{\sum_{i=1}^n \mathbb{1}_{\tilde{T}_i \leq t, \tilde{D}_i = 1}}{\sum_{i=1}^n \mathbb{1}_{\tilde{T}_i \geq s}} \\ &= \phi((\tilde{T}_i, \tilde{D}_i)_{i \in \{1,\dots,n\}})(t) \\ &= \phi(\hat{\mathbb{F}}_n)(t) = \hat{T}(t)\end{aligned}$$

$T : (\mathbb{F}, \beta) \mapsto \Lambda_{01,z}$ is a statistical functional
i.e. the mapping of a function ($\mathbb{F}$) to a statistic ($\Lambda_{01,z}$).

IF - Von Mises expansion (Van der Vaart, 2000)

First order Taylor expansion:
$f(x+th) = f(a) + tf'(a).h + o(t||h||)$

Under regularity condition for ϕ (Hadamard differentiability):

$$\phi(\hat{\mathbb{F}}_n) = \phi(\mathbb{F}) + \frac{1}{\sqrt{n}}\phi'(\mathbb{F}).\sqrt{n}\left(\hat{\mathbb{F}}_n - \mathbb{F}\right) + o_p(1)$$

$$\sqrt{n}\left(\phi(\hat{\mathbb{F}}_n) - \phi(\mathbb{F})\right) = \sqrt{n}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}\phi'(\mathbb{F}).(\delta_{\mathcal{X}_i} - \mathbb{F})}_{\text{average of iid random variables}}\right) + o_p(1)$$

$\phi'(\mathbb{F}).(\delta_{\mathcal{X}_i} - \mathbb{F})$ is called influence function

IF - Influence function

The influence of a sample $\mathcal{X}_i$ on the functional ϕ is given by:

$$\begin{aligned}\mathcal{IF}_\phi(\mathcal{X}_i) &= \phi'(\mathbb{F}).(\delta_{\mathcal{X}_i} - \mathbb{F}) \\ &= \frac{d}{dt}\bigg|_{t=0} \phi((1-t)\mathbb{F} + t\delta_{\mathcal{X}})\end{aligned}$$

The empirical values of the influence function approximate (Reid and Crépeau, 1985):

$$\widehat{\mathcal{IF}}_\phi(\mathcal{X}_i) \approx (n-1)\hat{\phi} - \hat{\phi}_{-i}$$

where $\hat{\phi}_{-i}$ is the estimate of ϕ when the ith observation is deleted

Also:

$$\sqrt{n}\,(\phi(\mathbb{P}_n) - \phi(P)) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \frac{1}{n}\sum_{i=1}^{n} \mathcal{IF}_\phi(\mathcal{X}_i)^2\right)$$

Functional delta method - Application

We know the influence function of the coefficients (Reid and Crépeau, 1985):

$$\sqrt{n}(\hat{\beta}_j - \beta_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{IF}_{\beta_j}(\mathcal{X}_i) + o_p(1)$$

and of the baseline cumulative hazard (Gerds and Schumacher, 2001):

$$\sqrt{n}(\hat{\Lambda}_{0j,z}(t) - \Lambda_{0j,z})(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{IF}_{\Lambda_{0j,z}}(\mathcal{X}_i; t) + o_p(1)$$

Since:

$$\Lambda_{j,z}(t|X = x) = \Lambda_{0j,z}(t) \exp(x\beta_j)$$

we get:

$$\mathcal{IF}_{\Lambda_{j,z}}(\mathcal{X}; t, x) = \exp(x\beta_j)\left(\mathcal{IF}_{\Lambda_{0j,z}}(\mathcal{X}; t) + \Lambda_{0j,z}(t)\; x\, \mathcal{IF}_{\beta_j}(\mathcal{X})^{\intercal}\right).$$

Functional delta method - Application

Finally considering only 2 competing events:

$$\mathcal{IF}_{r_1}(\mathcal{X};t,x,z) = \int_0^t S(s-|x,z) d\mathcal{IF}_{\Lambda_{1,z}}(\mathcal{X};s,x)$$
$$-\int_0^t S(s-|x,z)\left(\lambda_{1,z}(s|x)\mathcal{IF}_{\Lambda_{1,z}}(\mathcal{X};s,x)ds + \lambda_{1,z}(s|x)\mathcal{IF}_{\Lambda_{2,z}}(\mathcal{X};s,x)\right)ds$$

where $S(t|X,Z) = \exp(-\Lambda_{1,z}(t|x) - \Lambda_{2,z}(t|x))$

Simulation study

Given a timepoint τ and some covariates X_0, repeat for each sample size:

- Simulate data
- Fit cause specific Cox model

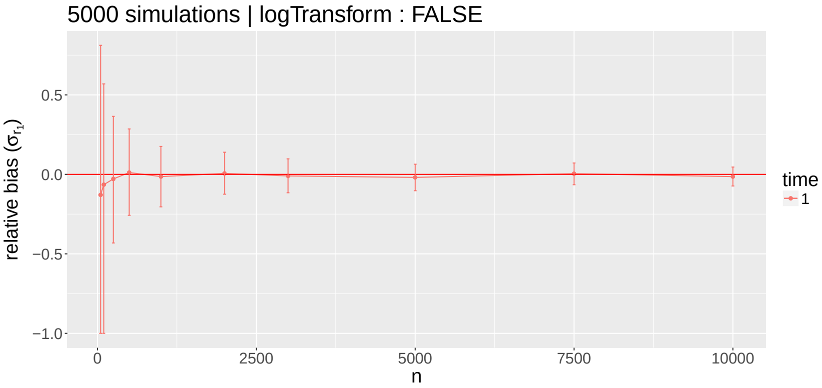
```
CSC.fit <- CSC(Hist(time,event)~ X1+X2,data=d,
               method = "breslow", iid = TRUE)
```

- Estimate the absolute risk $\hat{r}_1(\tau, X_0)$ for X_0 at τ

with its standard error $\hat{\sigma}_{r_1}(\tau, X_0)$ and its confidence interval.

True risk: $r_1(\tau, X_0) = \mathbb{E}\left[\hat{r}_1(\tau, X_0)\right]$
True standard error: $\sigma_{r_1}(\tau, X_0) = \sqrt{\mathbb{V}ar\left[\hat{r}_1(\tau, X_0)\right]}$

Simulation study



- very large CI for $n < 2500$

Asymptotics

ATE is a functionnal of r_1 that is Hadamard differentiable:

$$ATE(t, T_1, T_0) = \int r_1(t|T = T_1, X, Z) - r_1(t|T = T_0, X, Z)]d\mathbb{F}_{X,Z}$$
$$= \phi(r_1, \mathbb{F})$$

$$\sqrt{n}\left(\widehat{ATE}(t, T_1, T_0) - ATE(t, T_1, T_0)\right)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \mathcal{IF}_{r_1}(\mathcal{X}_i|t, T_1, X_j, Z_j) - \mathcal{IF}_{r_1}(\mathcal{X}_i|t, T_0, X_j, Z_j)$$
$$+ r_1(t, T_1, X_i, Z_i) - r_1(t, T_0, X_i, Z_i) - ATE(t, T_0, T_1)\Big) + o_p(1)$$

Asymptotics (for ATE)

ATE is a functionnal of r_1 that is Hadamard differentiable:

$$
\begin{aligned}
ATE(t, T_1, T_0) &= \int r_1(t|T = T_1, X, Z) - r_1(t|T = T_0, X, Z)] d\mathbb{F}_{X,Z} \\
&= \phi(r_1, \mathbb{F})
\end{aligned}
$$

$$
\sqrt{n}\left(\widehat{ATE}(t, T_1, T_0) - ATE(t, T_1, T_0)\right)
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \mathcal{IF}_{r_1}(\mathcal{X}_i | t, T_1, X_j, Z_j) - \mathcal{IF}_{r_1}(\mathcal{X}_i | t, T_0, X_j, Z_j) \right.
$$

$$
\left. + r_1(t, T_1, X_i, Z_i) - r_1(t, T_0, X_i, Z_i) - ATE(t, T_0, T_1) \right) + o_p(1)
$$

Confidence bands

What if we would like to compare two treatments over a period of time?

Considering a time interval $\mathcal{T} = [\tau_1; \tau_2]$, the normalized process:

$$\psi_{ATE}(\mathcal{X}_i; t, T_0, T_1) = \mathcal{IF}_{ATE}(\mathcal{X}_i; t, T_0, T_1) / \sigma_{ATE}(t, T_0, T_1)$$

converges weakly to a gaussian process (mean 0, variance 1).

Then the $1 - \alpha/2$ quantile of:

$$\sup_{t \in \mathcal{T}} |\psi_{ATE}(\mathcal{X}_i; t, T_0, T_1)|$$

can be used to construct the confidence bands.

- estimated by simulation

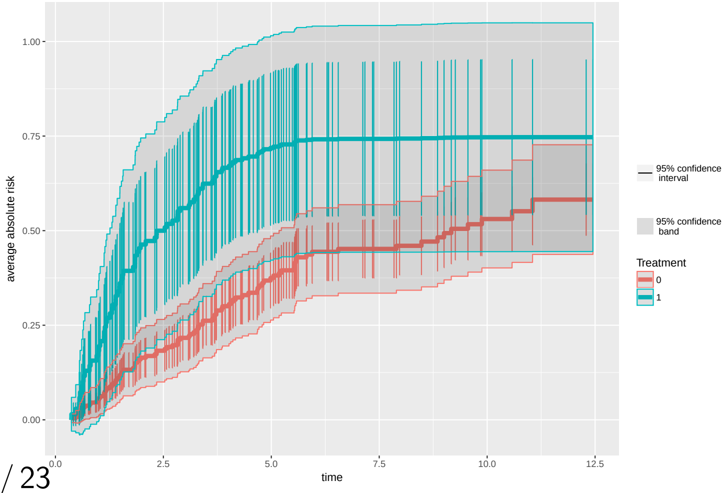
Confidence bands - Example

```
library(riskRegression) ; library(survival)
set.seed(10)

dt.data <- sampleData(2e2,outcome="competing.risks")
fit <- CSC(formula = Hist(time,event)~ X1 + X2,
       data=dt.data)

seqTimes <- sort(unique(fit$eventTimes))

system.time(
ateFit <- ate(fit, dt.data, treatment = "X1",
  cause = 1, times = seqTimes, band = TRUE)
)
```

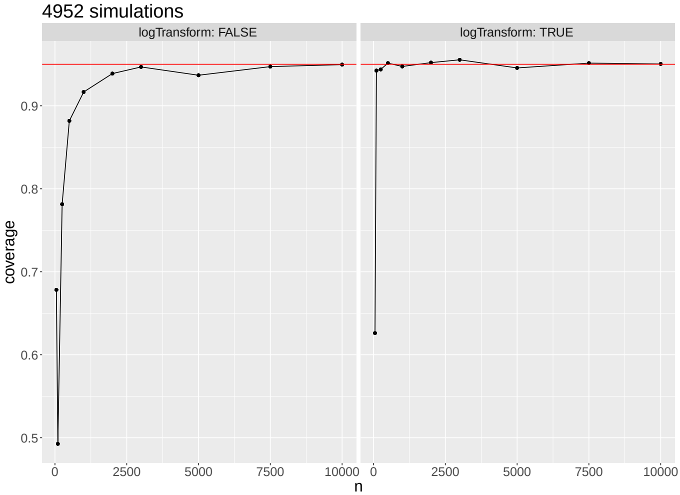
```
user   system elapsed
4.22    0.08    4.30
```

Confidence bands - Example

```
gg <- plot(ateFit, band = TRUE, ci = TRUE)
```



Confidence bands - Simulation



- log transformation seems beneficial