Causal Inference for Observational Longitudinal Survival Data

Mark Bech Knudsen

Context

Thesis written in blocks 1 & 2 autumn 2022, supervised by Torben Martinussen & Helene Rytgaard @ Biostatistics KU.

Context

Thesis written in blocks 1 & 2 autumn 2022, supervised by Torben Martinussen & Helene Rytgaard @ Biostatistics KU.

Thesis explores 3 topics relating to causal inference for non-randomized survival data:

Context

Thesis written in blocks 1 & 2 autumn 2022, supervised by Torben Martinussen & Helene Rytgaard @ Biostatistics KU.

Thesis explores 3 topics relating to causal inference for non-randomized survival data:

- Do hazard ratio estimates from the Cox model have a causal interpretation for a time-dependent treatment?

Context

Thesis written in blocks 1 & 2 autumn 2022, supervised by Torben Martinussen & Helene Rytgaard @ Biostatistics KU.

Thesis explores 3 topics relating to causal inference for non-randomized survival data:

- Do hazard ratio estimates from the Cox model have a causal interpretation for a time-dependent treatment?
- A problem with Markov models for survival data with time-dependent covariates.

Context

Thesis written in blocks 1 & 2 autumn 2022, supervised by Torben Martinussen & Helene Rytgaard @ Biostatistics KU.

Thesis explores 3 topics relating to causal inference for non-randomized survival data:

- Do hazard ratio estimates from the Cox model have a causal interpretation for a time-dependent treatment?
- A problem with Markov models for survival data with time-dependent covariates.
- Causal inference in continuous-time using inverse probability of treatment weights.

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

- T^0 is the survival time if the person *does not* receive treatment.

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

- T^0 is the survival time if the person *does not* receive treatment.
- T^1 is the survival time if the person receives treatment.

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

- T^0 is the survival time if the person *does not* receive treatment.
- T^1 is the survival time if the person receives treatment.

The effect of treatment could then be defined as

$$P(T^1 > t) - P(T^0 > t)$$

i.e. how much more likely is it for the person to survive if they receive treatment vs. if they do not?

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

- T^0 is the survival time if the person *does not* receive treatment.
- T^1 is the survival time if the person receives treatment.

The effect of treatment could then be defined as

$$P(T^1 > t) - P(T^0 > t)$$

i.e. how much more likely is it for the person to survive if they receive treatment vs. if they do not?

Or on the population scale: What would the survival rate be if we treated everyone vs. if we did not treat anyone?

Defining causal effects

Assume a binary treatment $A \in \{0, 1\}$ that is given at time $t = 0$. Wish to infer the *effect of treatment on survival*.

For each person, suppose there exists two *potential outcomes* T^0 and T^1:

- T^0 is the survival time if the person *does not* receive treatment.
- T^1 is the survival time if the person receives treatment.

The effect of treatment could then be defined as

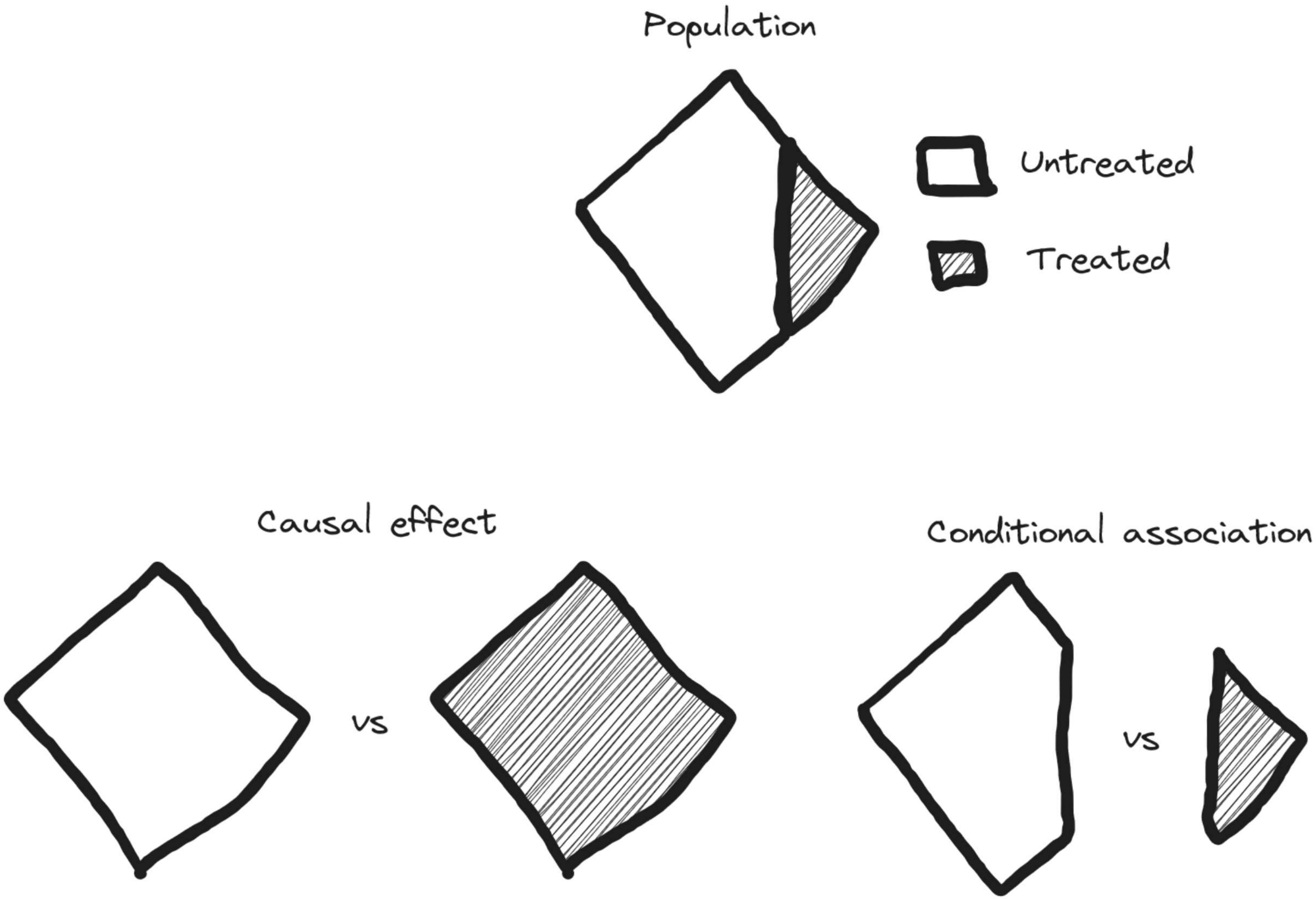
$$P(T^1 > t) - P(T^0 > t)$$

i.e. how much more likely is it for the person to survive if they receive treatment vs. if they do not?

Or on the population scale: What would the survival rate be if we treated everyone vs. if we did not treat anyone?

Different from the conditional association $P(T > t \mid A = 1) - P(T > t \mid A = 0)$.

Causal vs. conditional effects



*Inspired by Miguel A. Hernán: A definition of causal effect for epidemiological research (2004)

Randomized treatment

Assume a randomized treatment A. Then the subpopulation with $A = a$ is *representative* of what would have happened to the rest of the population, *had they been treated*.

Randomized treatment

Assume a randomized treatment A. Then the subpopulation with $A = a$ is *representative* of what would have happened to the rest of the population, *had they been treated*.

Scale up subpopulation to full population size with weights

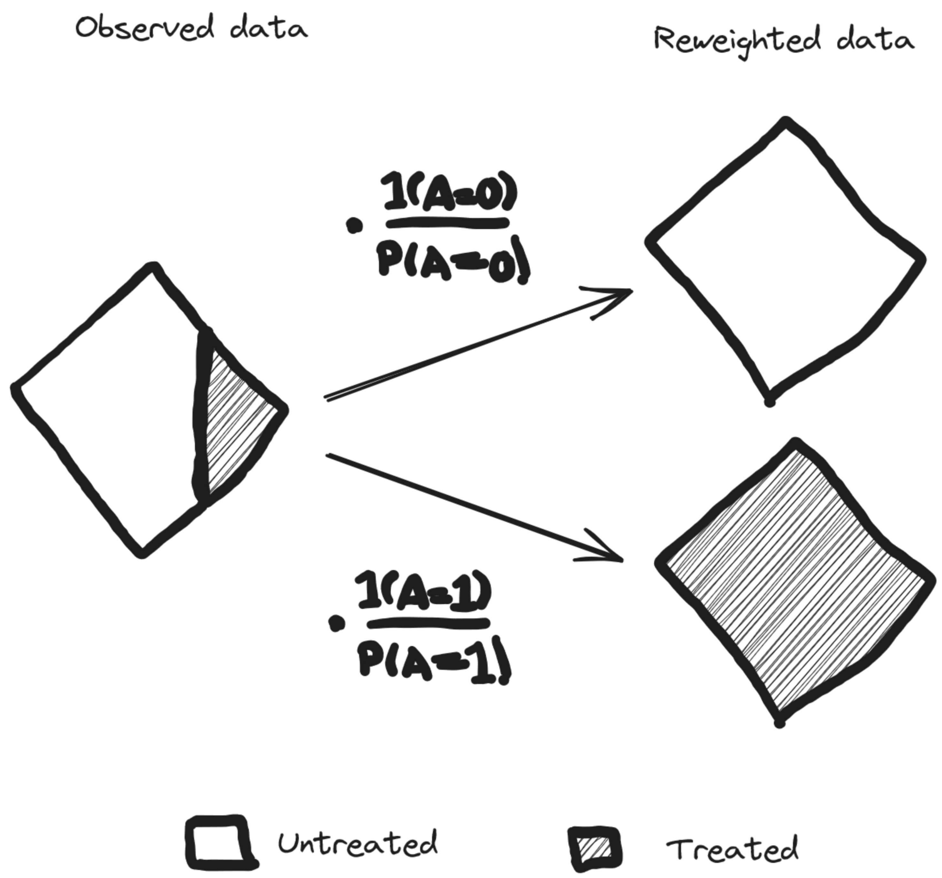
$$W_i^a = \frac{1(A_i = a)}{P(A = a)}$$

Randomized treatment

Assume a randomized treatment A. Then the subpopulation with $A = a$ is *representative* of what would have happened to the rest of the population, *had they been treated*.

Scale up subpopulation to full population size with weights

$$W_i^a = \frac{1(A_i = a)}{P(A = a)}$$



Randomized treatment

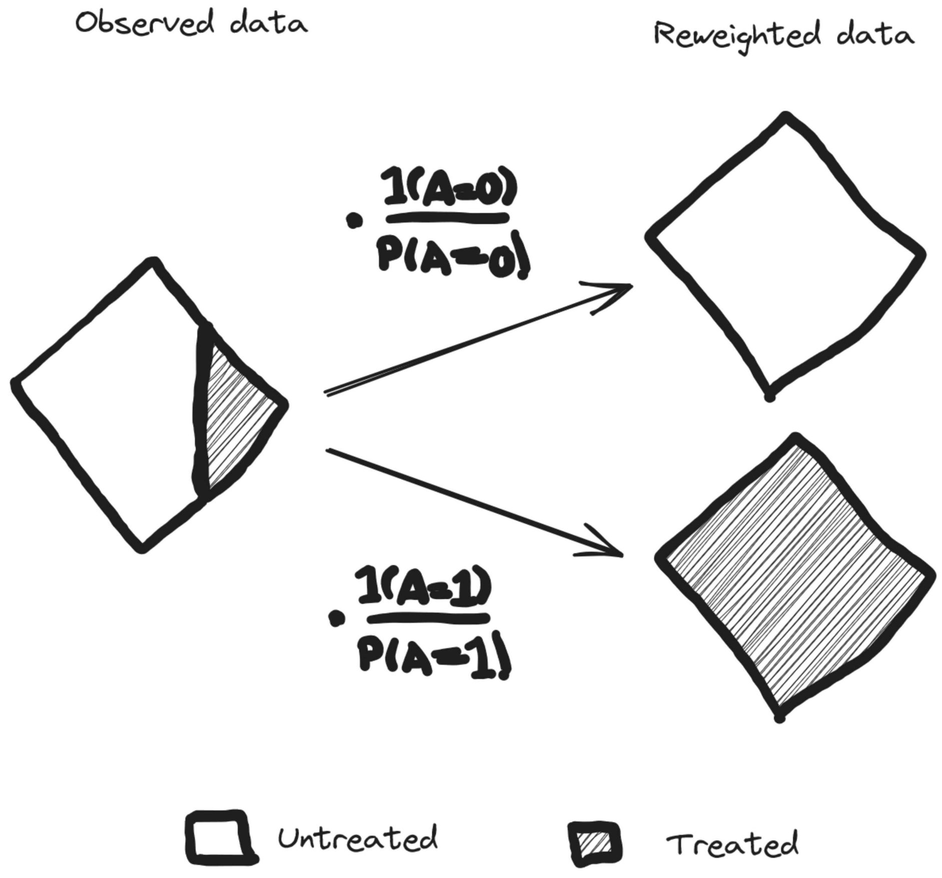
Assume a randomized treatment A. Then the subpopulation with $A = a$ is *representative* of what would have happened to the rest of the population, *had they been treated*.

Scale up subpopulation to full population size with weights

$$W_i^a = \frac{1(A_i = a)}{P(A = a)}$$

An estimate of the probability of survival, had everyone received treatment a:

$$P(T^a > t) = \frac{1}{n}\sum_{i=1}^{n} W_i^a 1(T_i > t)$$



Randomized treatment

Assume a randomized treatment A. Then the subpopulation with $A = a$ is *representative* of what would have happened to the rest of the population, *had they been treated*.

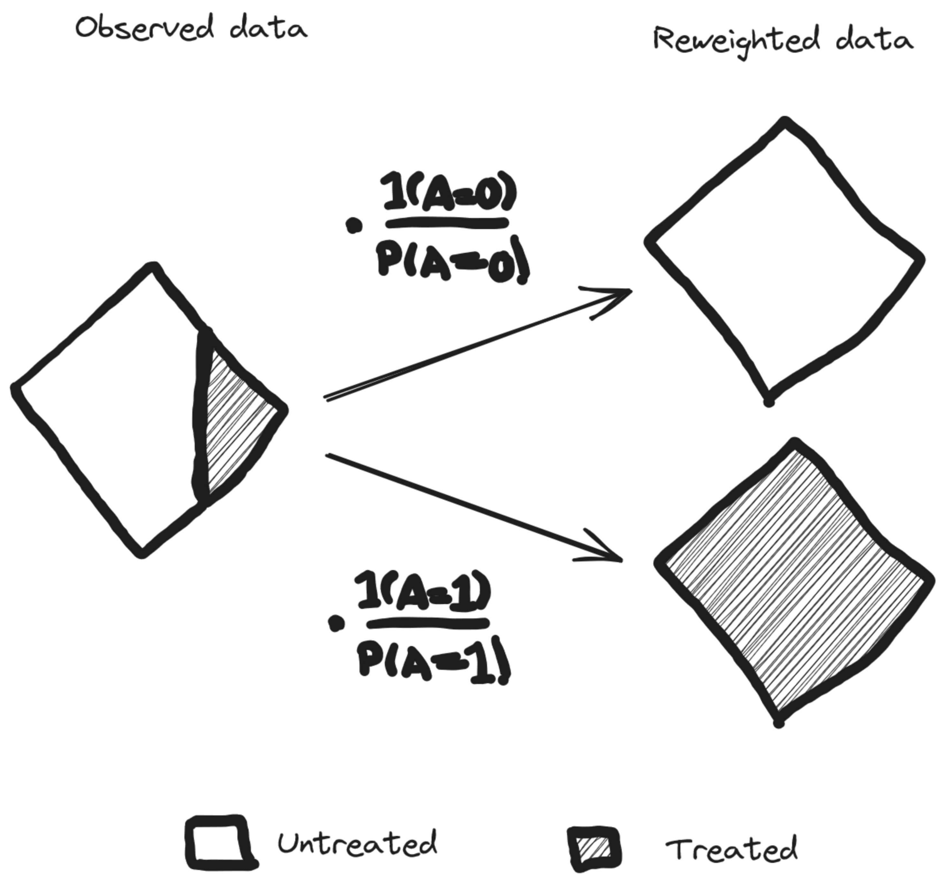
Scale up subpopulation to full population size with weights

$$W_i^a = \frac{1(A_i = a)}{P(A = a)}$$

An estimate of the probability of survival, had everyone received treatment a:

$$P(T^a > t) = \frac{1}{n}\sum_{i=1}^{n} W_i^a 1(T_i > t)$$

The W_i^a are referred to as *inverse probability of treatment weights (IPTW)*.



Confounding

Suppose now that a *confounder* $X \in \{0, 1\}$ (low/high risk) affects both treatment and outcome (non-randomized data).

Confounding

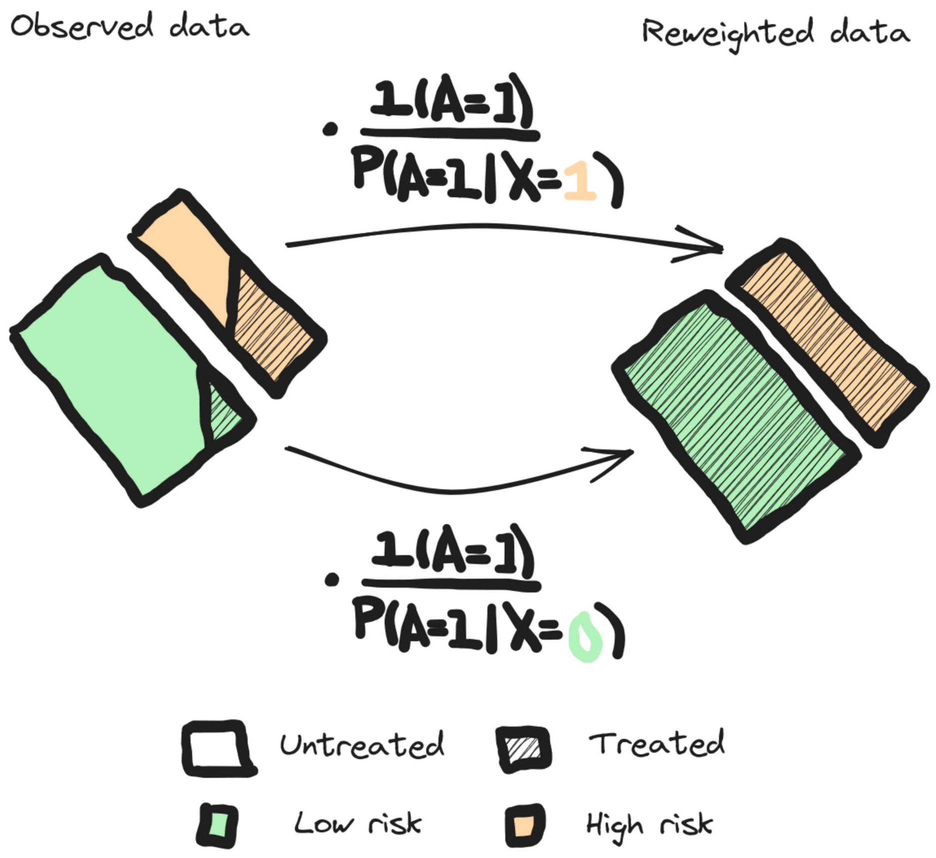
Suppose now that a *confounder* $X \in \{0, 1\}$ (low/high risk) affects both treatment and outcome (non-randomized data).

If X is the *only* confounder, then we still have representativity *within levels of* X. So we can pretend randomized treatment within levels of X.

Confounding

Suppose now that a *confounder* $X \in \{0, 1\}$ (low/high risk) affects both treatment and outcome (non-randomized data).

If X is the *only* confounder, then we still have representativity *within levels of* X. So we can pretend randomized treatment within levels of X.



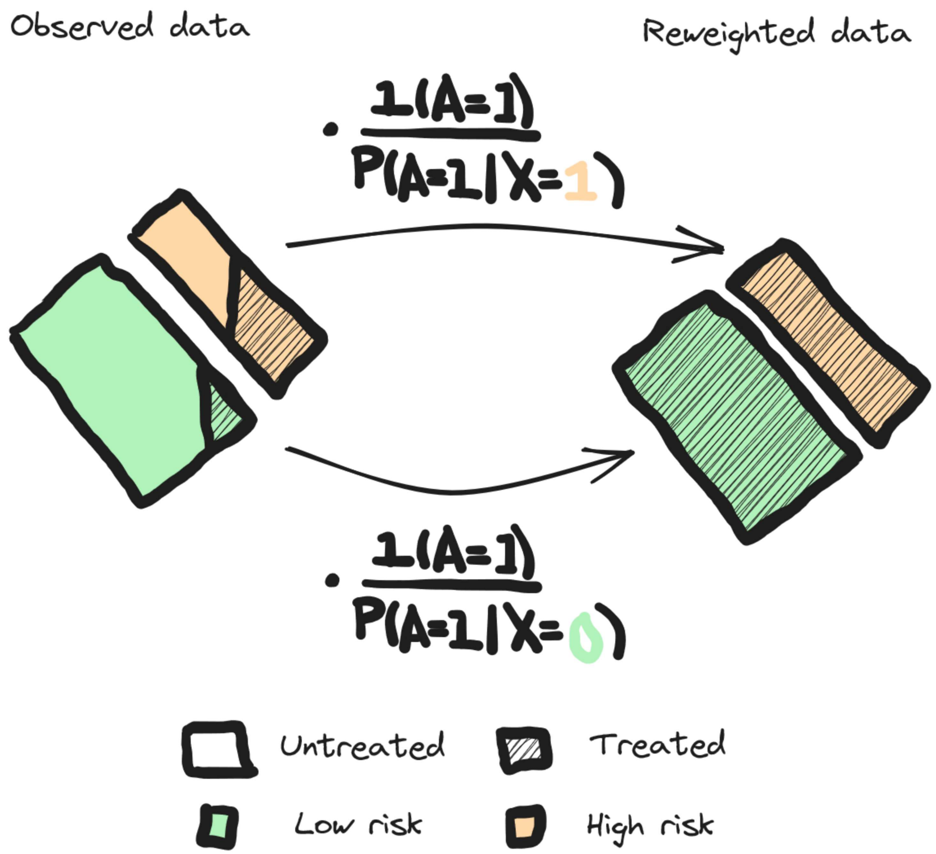
Confounding

Suppose now that a *confounder* $X \in \{0, 1\}$ (low/high risk) affects both treatment and outcome (non-randomized data).

If X is the *only* confounder, then we still have representativity *within levels of* X. So we can pretend randomized treatment within levels of X.

So we assign different weights depending on X:

$$W_i^a = \frac{1(A_i = a)}{P(A = a \mid X_i)}$$



Confounding

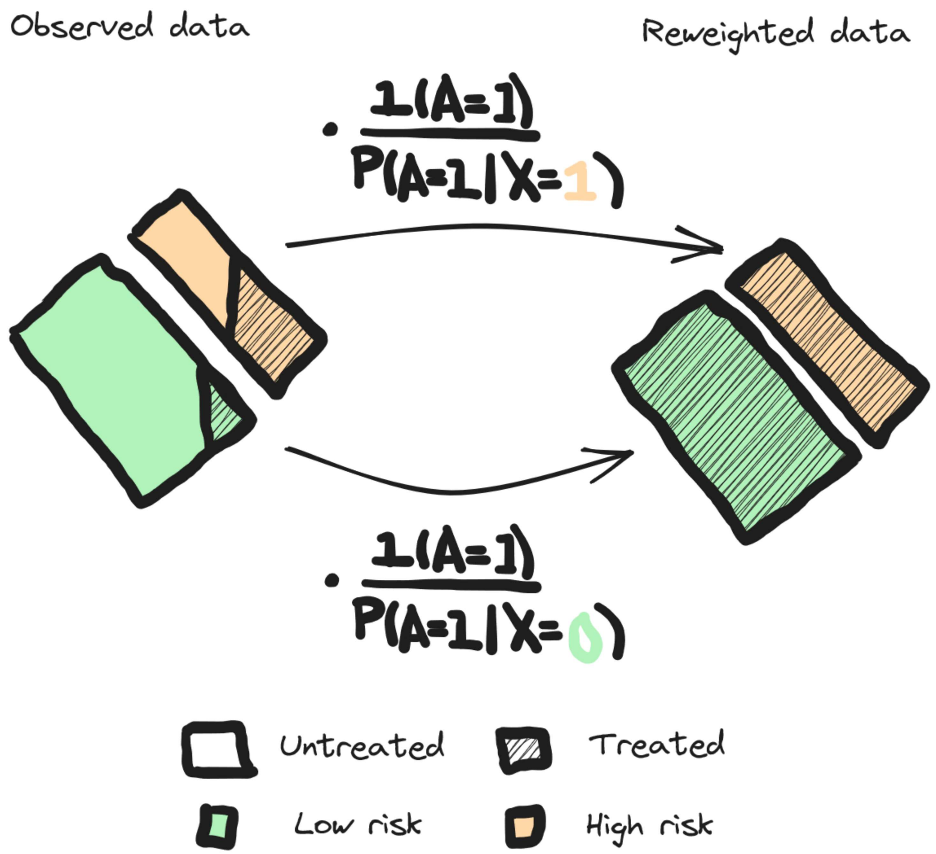
Suppose now that a *confounder* $X \in \{0, 1\}$ (low/high risk) affects both treatment and outcome (non-randomized data).

If X is the *only* confounder, then we still have representativity *within levels of* X. So we can pretend randomized treatment within levels of X.

So we assign different weights depending on X:

$$W_i^a = \frac{1(A_i = a)}{P(A = a \mid X_i)}$$

This works regardless of the dimension of X, and also for continuous covariates.



Time-varying treatment

Consider a case where treatment is a binary counting process $A(t)$ that starts at 0 and jumps to 1 at time τ when a treatment is undertaken (e.g. surgery).

Time-varying treatment

Consider a case where treatment is a binary counting process $A(t)$ that starts at 0 and jumps to 1 at time τ when a treatment is undertaken (e.g. surgery).

Covariates $X(t)$ are also time-dependent.

Time-varying treatment

Consider a case where treatment is a binary counting process $A(t)$ that starts at 0 and jumps to 1 at time τ when a treatment is undertaken (e.g. surgery).

Covariates $X(t)$ are also time-dependent.

Can generalize to a weight *process*

$$W_i(t) = \frac{1}{\lambda^A(\tau \mid \mathcal{F}_{\tau-})^{A(t)} e^{-\Lambda^A(t \wedge \tau)}}$$

where λ^A is the hazard/rate of treatment initiation and

$$\Lambda^A(t) = \int_0^t \lambda^A(s \mid \mathcal{F}_{s-}) \mathrm{d}s$$

Time-varying treatment

Consider a case where treatment is a binary counting process $A(t)$ that starts at 0 and jumps to 1 at time τ when a treatment is undertaken (e.g. surgery).

Covariates $X(t)$ are also time-dependent.

Can generalize to a weight *process*

$$W_i(t) = \frac{1}{\lambda^A(\tau \mid \mathcal{F}_{\tau-})^{A(t)} \mathrm{e}^{-\Lambda^A(t \wedge \tau)}}$$

where λ^A is the hazard/rate of treatment initiation and

$$\Lambda^A(t) = \int_0^t \lambda^A(s \mid \mathcal{F}_{s-}) \mathrm{d}s$$

The processes $W_i(t)$ are quite nice (exponential martingales), and can be estimated from data.

Time-varying treatment

Consider a case where treatment is a binary counting process $A(t)$ that starts at 0 and jumps to 1 at time τ when a treatment is undertaken (e.g. surgery).

Covariates $X(t)$ are also time-dependent.

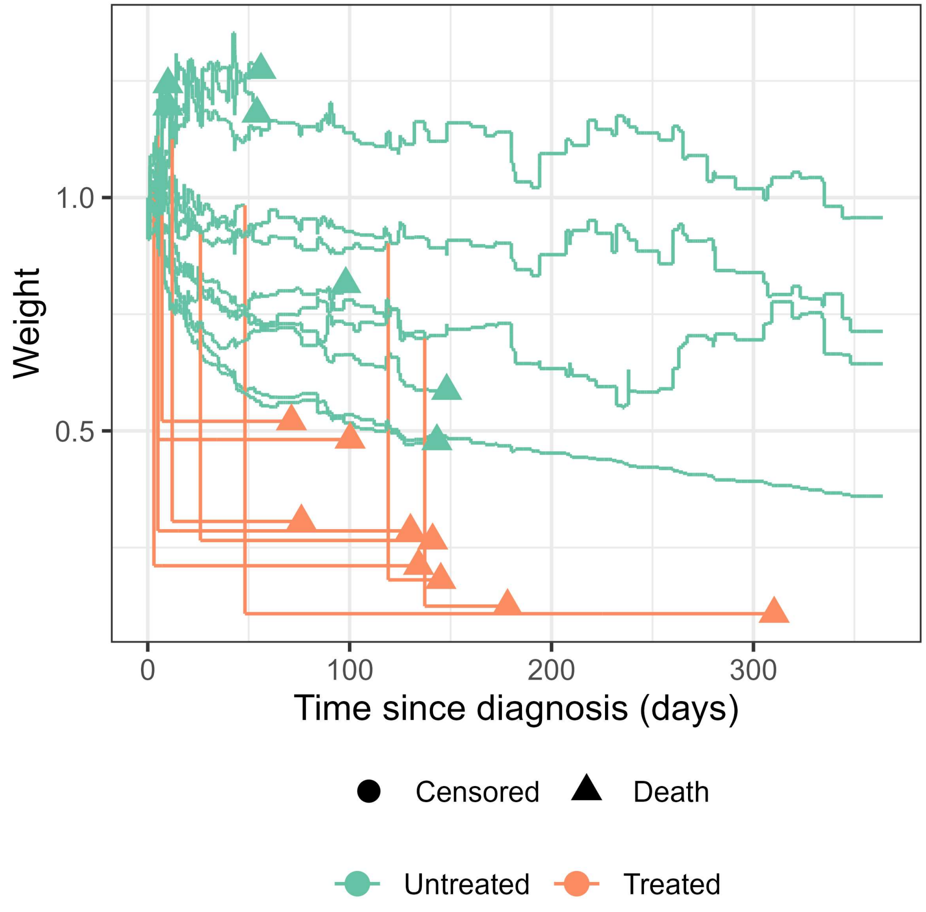
Can generalize to a weight *process*

$$W_i(t) = \frac{1}{\lambda^A(\tau \mid \mathcal{F}_{\tau-})^{A(t)} e^{-\Lambda^A(t \wedge \tau)}}$$

where λ^A is the hazard/rate of treatment initiation and

$$\Lambda^A(t) = \int_0^t \lambda^A(s \mid \mathcal{F}_{s-}) \mathrm{d}s$$

The processes $W_i(t)$ are quite nice (exponential martingales), and can be estimated from data.



A data example

Observational data ($n = 601$) from oncologists at Rigshospitalet.

A data example

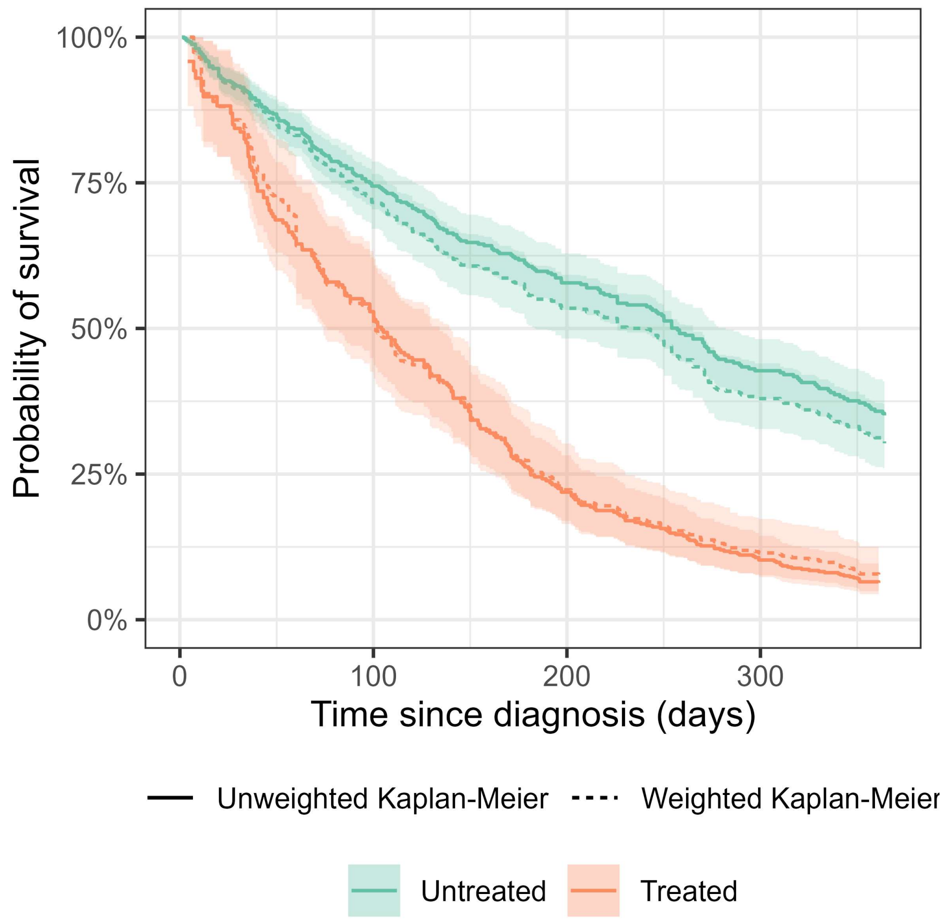
Observational data ($n = 601$) from oncologists at Rigshospitalet.

- T time from diagnosis of incurable throat cancer to death.
- $A(t)$ insertion of metalic tube (stent) into throat.
- A bunch of covariates $X(t)$, mostly measured at baseline.

A data example

Observational data ($n = 601$) from oncologists at Rigshospitalet.

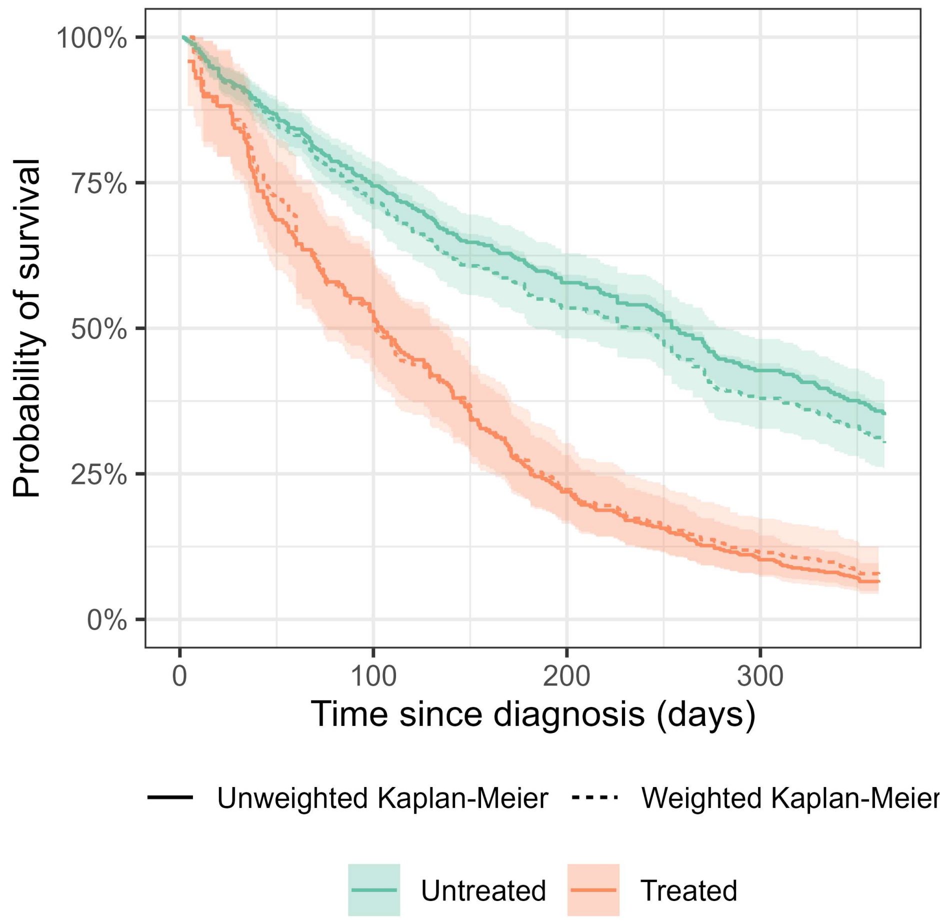
- T time from diagnosis of incurable throat cancer to death.
- $A(t)$ insertion of metalic tube (stent) into throat.
- A bunch of covariates $X(t)$, mostly measured at baseline.



A data example

Observational data ($n = 601$) from oncologists at Rigshospitalet.

- T time from diagnosis of incurable throat cancer to death.
- $A(t)$ insertion of metalic tube (stent) into throat.
- A bunch of covariates $X(t)$, mostly measured at baseline.



Weighting does not change the conclusion much. But we did not have enough information about *time-dependent* confounders.

Thank you!

And good luck when you write your thesis.