Causal inference for survival outcomes:

a censored edition

Bianca L De Stavola for STRATOS TG 7

Great Ormond Street Institute of Child Health, University College London

RSS Conference 2021, 7^{th} September 2021

STRATOS INITIATIVE

Topic group 7: Causal Inference

Saskia le Cessie	Leiden University Medical Centre & Ghent University
Bianca De Stavola	University College London, UK
Vanessa Didelez	University of Bremen, Germany
Els Goetghebeur	Ghent University, Belgium
Erica Moodie	McGill University, Canada
Ingeborg Waernbaum	Uppsala University, Sweden

STRATOS INITIATIVE

Aims: Overview of the principles that guide current developments in causal inference

- First guidance paper:

Received: 12 May 2019 | Revised: 10 May 2020 | Accepted: 5 August 2020

DOI: 10.1002/sim.8741

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

Formulating causal questions and principled statistical answers

Els Goetghebeur[1,2] | Saskia le Cessie[3] | Bianca De Stavola[4] | Erica EM Moodie[5] | Ingeborg Waernbaum[6] | "on behalf of" the topic group Causal Inference (TG7) of the STRATOS initiative

- Short courses
- Website: ofcaus.org

Basic Principles

- To find a causal answer, start with a causal question. Then:
  1. specify exposure, outcome, population of interest, target causal effects (*e.g.* using potential outcomes)
  2. state assumptions for identification and estimation of effects from the data
  3. interpret results cautiously, aided by sensitivity analyses of assumptions.

Basic Principles

- To find a causal answer, start with a causal question. Then:
  1. specify exposure, outcome, population of interest, target causal effects (*e.g.* using potential outcomes)
  2. state assumptions for identification and estimation of effects from the data
  3. interpret results cautiously, aided by sensitivity analyses of assumptions.
- These are indeed the principles guiding RCTs and, for observational studies, are referred to as "target trial emulation" (TTE).

Basic Principles

- To find a causal answer, start with a causal question. Then:
  1. specify exposure, outcome, population of interest, target causal effects (*e.g.* using potential outcomes)
  2. state assumptions for identification and estimation of effects from the data
  3. interpret results cautiously, aided by sensitivity analyses of assumptions.
- These are indeed the principles guiding RCTs and, for observational studies, are referred to as "target trial emulation" (TTE).
- A major advantage of TTE: avoidance of errors in data manipulation (*e.g.* immortal time bias, treatment assignment).

Potential Outcomes and Estimands

For a binary exposure A and an outcome Y, let:

- Y_a be the potential outcome if we set A to take the value a by a well-defined (hypothetical) intervention.

Potential Outcomes and Estimands

For a binary exposure A and an outcome Y, let:

- Y_a be the potential outcome if we set A to take the value a by a well-defined (hypothetical) intervention.
- Causal effects ("estimands") can then be defined *e.g.* :
  - ACE=$E(Y_1) - E(Y_0)$
  - ATT=$E(Y_1|A=1) - E(Y_0|A=1)$
  - *etc.*

Potential Outcomes and Estimands

For a binary exposure A and an outcome Y, let:

- Y_a be the potential outcome if we set A to take the value a by a well-defined (hypothetical) intervention.
- Causal effects ("estimands") can then be defined *e.g.* :
  - ACE=$E(Y_1) - E(Y_0)$
  - ATT=$E(Y_1|A=1) - E(Y_0|A=1)$
  - *etc.*
- These are comparisons of alternative worlds.

Potential Outcomes and Estimands

For a binary exposure A and an outcome Y, let:

- Y_a be the potential outcome if we set A to take the value a by a well-defined (hypothetical) intervention.
- Causal effects ("estimands") can then be defined *e.g.* :
  - ACE=$E(Y_1) - E(Y_0)$
  - ATT=$E(Y_1|A=1) - E(Y_0|A=1)$
  - *etc.*
- These are comparisons of alternative worlds.
- Identification requires linking observed data to these hypothetical quantities *e.g.* invoking assumptions of no interference, consistency, and positivity.

Potential Outcomes and Estimands

For a binary exposure A and an outcome Y, let:

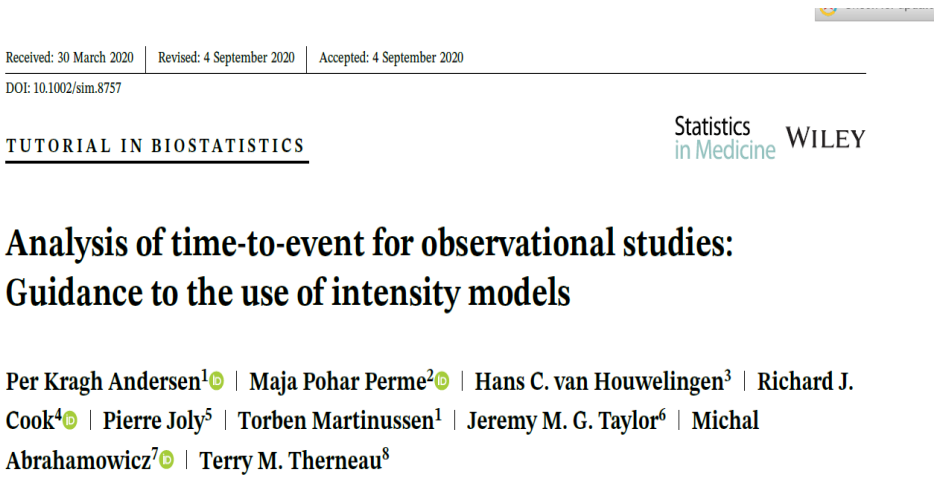
- Y_a be the potential outcome if we set A to take the value a by a well-defined (hypothetical) intervention.
- Causal effects ("estimands") can then be defined *e.g.* :
  - ACE=$E(Y_1) - E(Y_0)$
  - ATT=$E(Y_1|A=1) - E(Y_0|A=1)$
  - *etc.*
- These are comparisons of alternative worlds.
- Identification requires linking observed data to these hypothetical quantities *e.g.* invoking assumptions of no interference, consistency, and positivity.
- Choice of estimation methods: each requiring additional assumptions (*e.g.* no unmeasured confounding, correct (semi-)parametric models).

Time-to-event outcomes

- Well known challenges:
  - Time origin and scale: from birth/entry/surgery?
  - Censoring: information on whether event is observed at the end.

Time-to-event outcomes

- Well known challenges:
  - Time origin and scale: from birth/entry/surgery?
  - Censoring: information on whether event is observed at the end.
- Discussed by TG8-Survival Analysis:



Received: 30 March 2020 | Revised: 4 September 2020 | Accepted: 4 September 2020
DOI: 10.1002/sim.8757

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

Analysis of time-to-event for observational studies: Guidance to the use of intensity models

Per Kragh Andersen[1] | Maja Pohar Perme[2] | Hans C. van Houwelingen[3] | Richard J. Cook[4] | Pierre Joly[5] | Torben Martinussen[1] | Jeremy M. G. Taylor[6] | Michal Abrahamowicz[7] | Terry M. Therneau[8]

Which aspects of time should we focus on when comparing alternative worlds?

Estimands

Let T_a be the potential survival time if we set A to take the value a by a well-defined (hypothetical) intervention.

(A) Risk scale:
Differences in survival probabilities at relevant times

$$\text{ACE}(t) \;=\; P(T_1 > t) - P(T_0 > t), \quad t \;\text{in}[0, \tau]$$

This is the difference in (marginal) survival functions of POs,

- Interpretation:
risk difference for no event by time t had random patient been treated versus not.

Estimands

(B) Hazard scale:

Contrast of hazards, for example $\frac{\lambda_1(t)}{\lambda_0(t)}$, where

$$\lambda_a(t) = lim_{h \to 0} \frac{1}{h} P(t \leq T_a < t + h | T_a \geq t)$$

- Complication:

  Interpretational difficulties because of the built-in selection due to the conditioning on different subgroups ($T_0 \geq t$ and $T_1 \geq t$)).

Estimands

(B) Hazard scale:
Contrast of hazards, for example $\frac{\lambda_1(t)}{\lambda_0(t)}$, where

$$\lambda_a(t) = lim_{h \to 0} \frac{1}{h} P(t \leq T_a < t + h | T_a \geq t)$$

- Complication:
  Interpretational difficulties because of the built-in selection due to the conditioning on different subgroups ($T_0 \geq t$ and $T_1 \geq t$)).

(C) Other scales, *e.g.* speed from Accelerated Failure Time Models.

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.
- Whichever one is chosen, definitions above have no consideration of the impact of censoring.

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.
- Whichever one is chosen, definitions above have no consideration of the impact of censoring.
  - Do we want to quantify causal effects in the absence in censoring?

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.
- Whichever one is chosen, definitions above have no consideration of the impact of censoring.

- Do we want to quantify causal effects in the absence in censoring?

$$T_{A=a,C=0}$$

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.
- Whichever one is chosen, definitions above have no consideration of the impact of censoring.

- Do we want to quantify causal effects in the absence in censoring?

$$\boxed{T_{A=a,C=0}}$$

- It depends on the source of censoring: for some it does not always make clinical sense to remove them,
  - Administrative reasons
  - Loss to follow-up
  - Treatment switching
  - Competing event

Comments

- The choice between these estimands should be guided by their clinical relevance.
- In most settings these are contrasts on risk scale.
- Note however that hazard models are useful to derive such contrasts.
- Whichever one is chosen, definitions above have no consideration of the impact of censoring.

- Do we want to quantify causal effects in the absence in censoring?

$$\boxed{T_{A=a,C=0}}$$

- It depends on the source of censoring: for some it does not always make clinical sense to remove them,
  - Administrative reasons ✓
  - Loss to follow-up ✓
  - Treatment switching ??
  - Competing event ??

Assumptions

For identification of causal effects for a time-to-event outcome:

- No interference, consistency and positivity of the exposure
- No unmeasured confounding (NUC), *i.e.* : sufficient covariate information regarding treatment assignment confounding
- In the presence of censoring we also require: sufficient covariate information regarding (possibly time-varying) 'common causes' of censoring and event.

The simulation learner

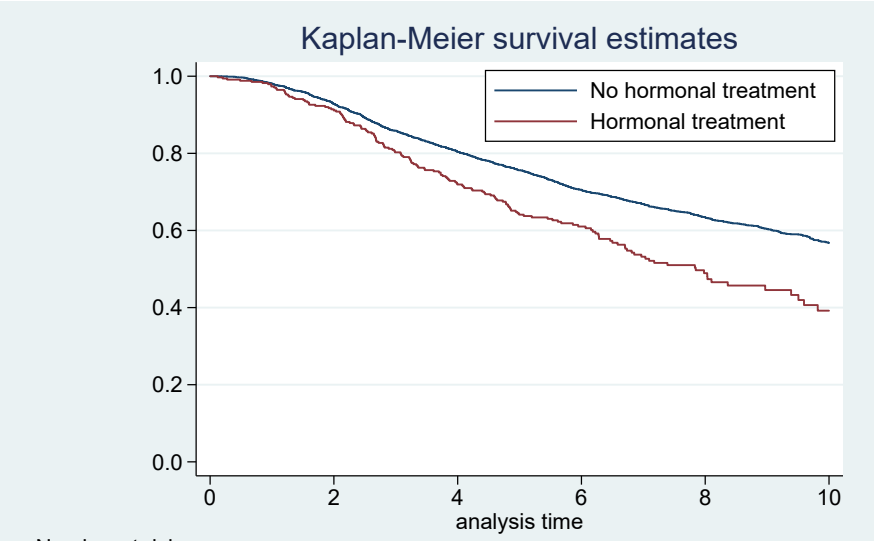
Inspiration: the Rotterdam study [Royston & Lambert, 2009; Sjolander, 2016]

- About 3000 women who had undergone surgery for breast cancer and, for some, hormonal therapy was offered in 1978 to 1993
- Outcome of interest: overall mortality
- Strong negative confounding of the association between therapy and mortality
- Informative loss to follow-up driven by age and year of surgery
- Lack of positivity for younger women and earlier patients

The simulation learner

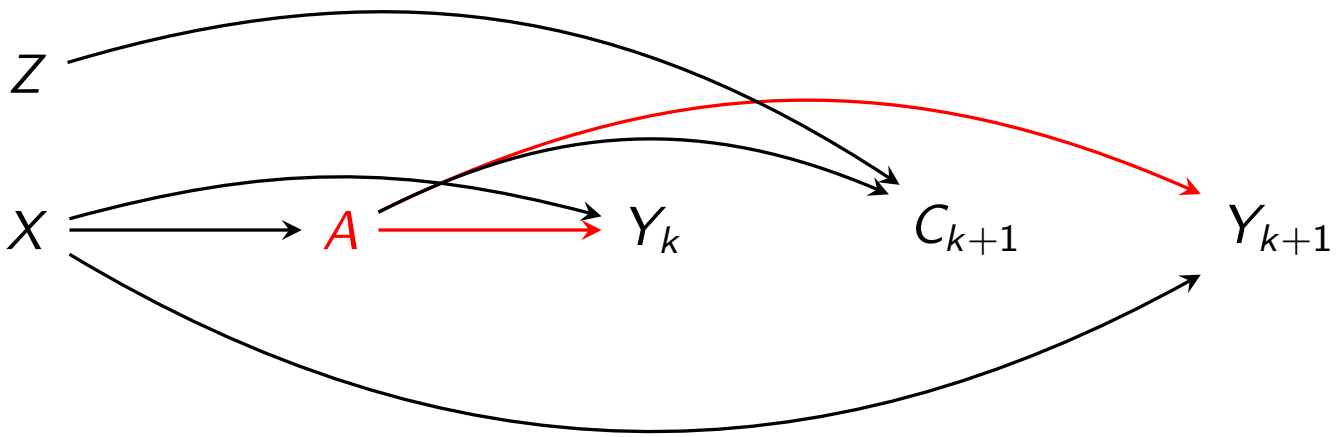
Inspiration: the Rotterdam study [Royston & Lambert, 2009; Sjolander, 2016]

- About 3000 women who had undergone surgery for breast cancer and, for some, hormonal therapy was offered in 1978 to 1993
- Outcome of interest: overall mortality
- Strong negative confounding of the association between therapy and mortality
- Informative loss to follow-up driven by age and year of surgery
- Lack of positivity for younger women and earlier patients



The simulation learner

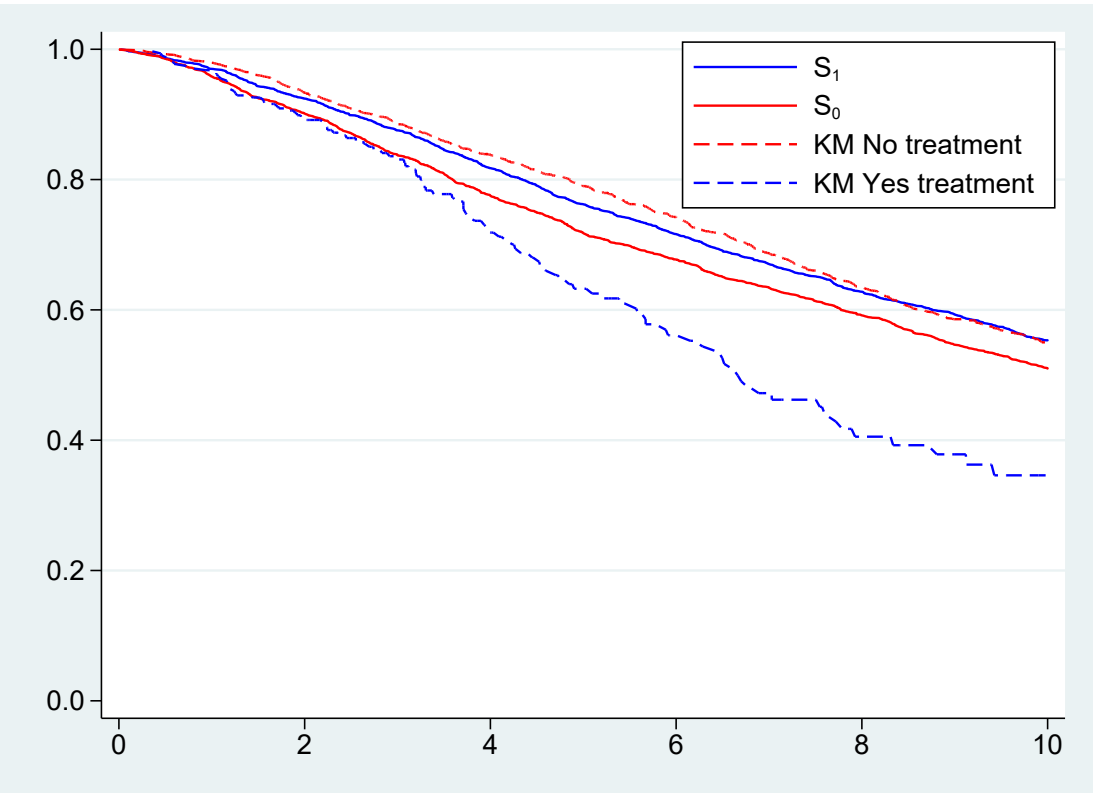
- Excluded women younger than 40 and with surgery before 1982
- Retained all the original confounders data



For a subset of the follow-up:

X: baseline confounders; Z: baseline predictors of censoring and death; A: treatment; C_k: censoring indicator at time t_k; Y_k: outcome at time t_k.

The simulation learner



Estimation of ACE(t)

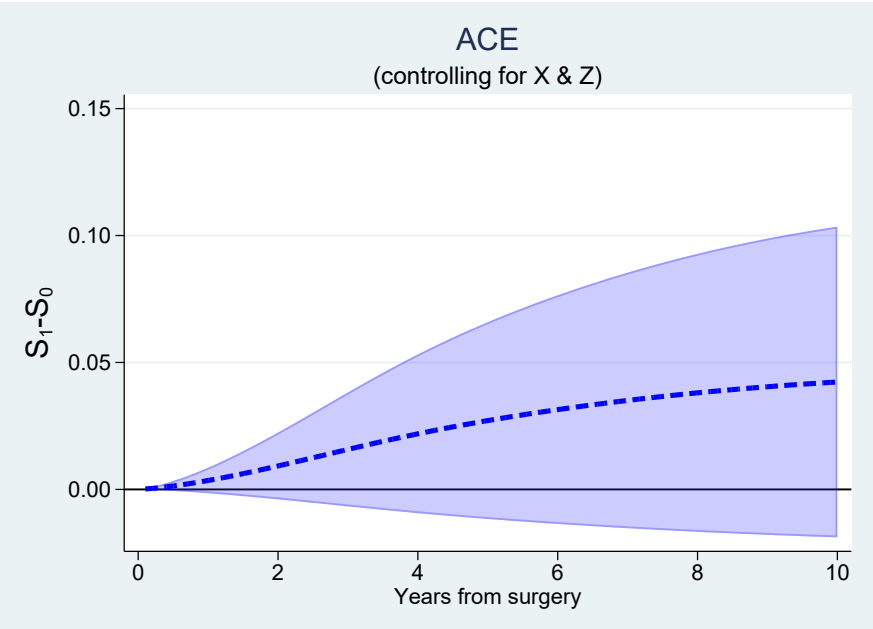
1. Model-based marginal counterfactual survival curves:
   - (Sufficiently) flexible hazard models
   - Derive individual-level predicted potential survival curves
   - Standardisation to the distribution of the observed confounders
   - Compute difference at selected values of t

Estimation of ACE(t)

2. Weighted Kaplan-Meier curves:
   - Fit propensity score model for A and save predicted scores
   - Fit pooled logistic regression model for C and save predicted (time-varying) probabilities
   - Combine the weights
   - Estimate K-M curves using the inverse of these combined values as weights

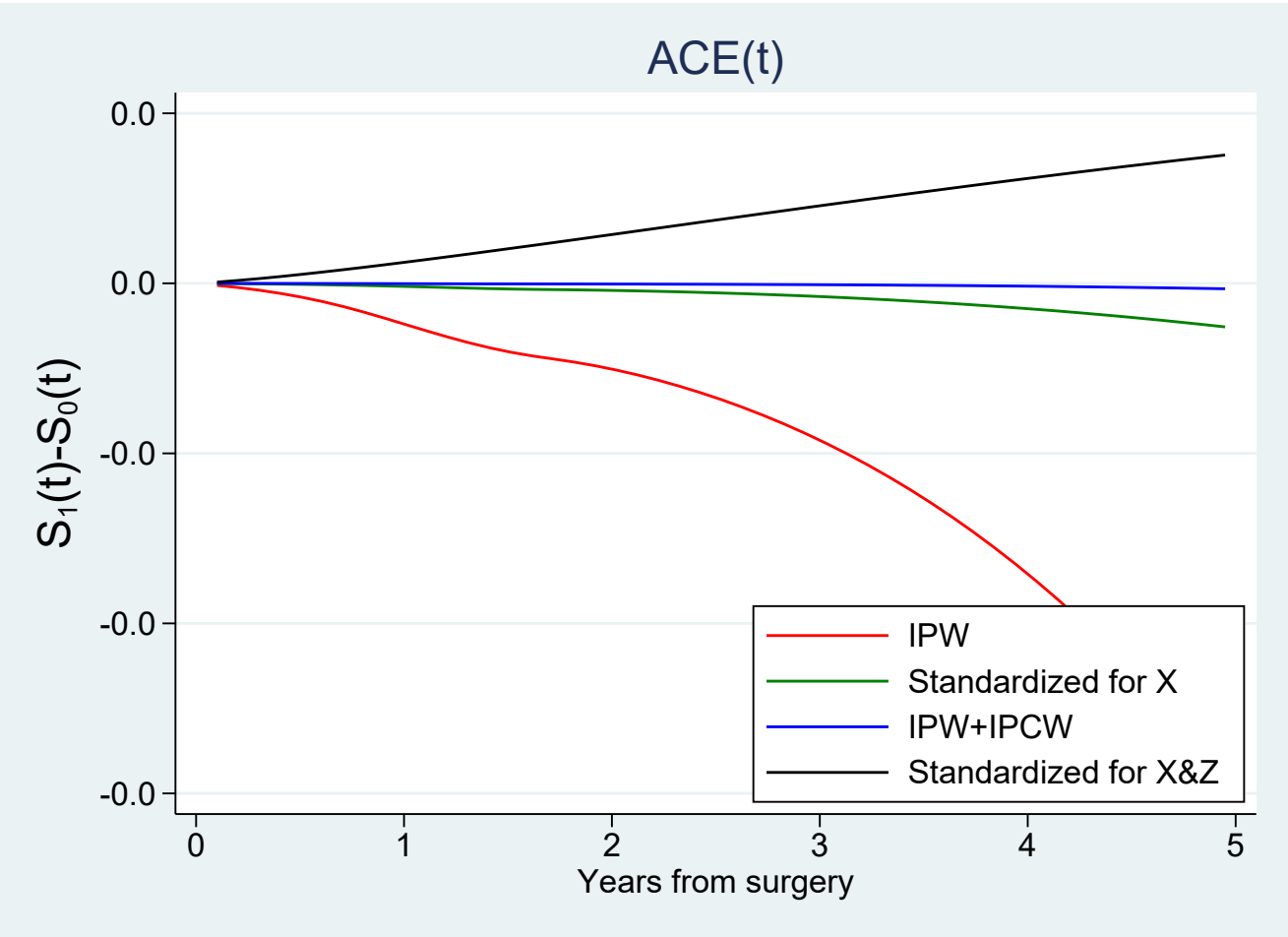
Results

Model-based marginal counterfactual survival curves



Results

All methods together



Summary

- Counterfactual-based causal inference has shifted the focus from model-based parameters to estimands defined irrespectively of any model: we should choose meaningful, clinically relevant quantities

Summary

- Counterfactual-based causal inference has shifted the focus from model-based parameters to estimands defined irrespectively of any model: we should choose meaningful, clinically relevant quantities
- This should free us from necessarily wanting to express causal effects on the hazard scale:
  - challenges to be interpreted causally
  - useful for deriving the causal estimates of interest.

Summary

- Counterfactual-based causal inference has shifted the focus from model-based parameters to estimands defined irrespectively of any model: we should choose meaningful, clinically relevant quantities
- This should free us from necessarily wanting to express causal effects on the hazard scale:
  - challenges to be interpreted causally
  - useful for deriving the causal estimates of interest.
- Dealing with censoring calls upon a careful choice of potential outcomes.