CAUSAL INFERENCE FOR SURVIVAL OUTCOMES: A CENSORED EDITION

Vanessa Didelez
Bianca de Stavola, Els Goetghebeur,
Saskia le Cessie, Ingeborg Waernbaum

STRATOS – TG7 – Causal Inference

SER – virtual – 2021

Formulating Causal Questions

(Goetghebeur et al, 2020, Stat in Med)

To get a causal answer we need to start with a causal question!

1. Define the treatment
2. Define the outcome
3. Specify population(s) of interest
4. Formalise potential outcomes (POs)
5. Specify target causal effect, i.e. the *estimand*, as a (summary) contrast between the PO-distributions
6. Assumptions identifying estimand from available data
7. Statistical inference with suitable methods
8. Evaluate plausibility of assumptions / sensitivity analyses

Target Trial

(Parra et al, 2020, arXiv:2011.11771)

A general principle to elicit & specify a causal question

- The ideal (hypothetical) trial that would answer the research question
  - possibly disregarding practical, ethical, financial constraints
  - … but not disregarding laws of physics (no “turning back time”)

- Especially useful in time-dependent situations
  - fix time ‘zero’
  - prevent immortal time bias etc.

Motivating Example

- End-stage renal disease: which renal replacement therapy (RRT)?
- Pre-emptive transplant (PKT) vs “start with dialysis”?
  - binary point treatment
  - a bit like ITT
- Wanted: “effect” on time to all-cause mortality starting from RRT
  - exact definition of estimand?
  - no competing events here
- *Most studies on the topic suffer from avoidable biases*

(Parra et al, 2020 arXiv:2011.11771)

Motivating Example

Aim:

- Want to mimic RCT ***(target trial)***:
  as if individuals randomised to treatment (PKT) / control (dialysis)
  - need plausible assumptions
  - & suitable methods

Causal Inference - Basics

- A = binary **point** treatment
- Y = outcome (general)

- Y_a = potential outcome if we **set** $A = a$ by (well-defined) intervention

- Common causal contrasts (estimands):

  (total) average effect: $E(Y_1) - E(Y_0)$

  effect on the treated: $E(Y_1 \mid A=1) - E(Y_0 \mid A=1)$

Causal Inference - Basics

- Assumptions:
  - Causal consistency & positivity, no interference
  - **N**o **U**nmeasured **C**onfounding **(NUC)**
  - Some (semi-)parametric model

⇒ Identification

- Many methods for estimation
  - outcome regression, stratification / matching, IPTW, DR
  - with sufficient set of covariates,
    possibly summarised in propensity score
  - check: overlap and balance!

Now: Survival Outcome

- Outcome $Y = T$ = time-to-event
- What's different?
  - Censoring
    - for some units we only know: the event did not occur in some period
  - Dynamics – things happen over time (including treatment)
    - 'mean' not a good summary?

⇒ May want different causal estimands

- assumptions?
- methods?

Survival Outcome - Estimands

Desirable estimand?

- Risk differences at relevant times

$$P(T_1 > t) - P(T_0 > t), \qquad t \text{ in } [0,\mathcal{T}]$$

- i.e. difference in *(marginal)* survival functions of POs
- Interpretation: risk difference for no event by time *t* had random patient been treated versus not
  - ≈ total average causal effects for meaningful time points
  - could easily be by baseline subgroups *(no further details today)*

Survival Outcome - Estimands

Hazard scale? Hazard ratio (HR) / contrast of hazards - popular

- With potential outcomes:

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

- i.e. hazard function in arm '*a*' of our target trial
- Contrasts of $\lambda_1(t)$ vs. $\lambda_0(t)$ are conditional on possibly different 'subgroups' $\{T_1 \geq t\}$ and $\{T_0 \geq t\}$
  - survivors at a given time *t* in the two arms not necessarily comparable anymore even in an RCT

Risks & Hazards – Pros and Cons

- Difficult to interpret causal effects on the hazard scale ***correctly***
  - no such thing as 'the' causal effect
  - 'effect reversals' between hazard and risk scale *possible*

    (Martinussen et al., 2020 LIDA)

  ⇒ must be aware & take into account for correct interpretation of contrasts of hazards

- But $\lambda_a(t)$ as whole function of t : one-to-one relation with $P(T_a > t)$*
  ⇒ hazards still useful modelling tool *(+ model checking etc. well-established)*
  - especially to deal with censoring & include relevant covariates

**in absence of comp.events*

Estimands - Summary

- We like & recommend contrasts on risk scale
  - direct clinical interpretation
  - but may use hazard models as a tool to get there
- There may sometimes be specific reasons to choose hazard contrasts as causal estimands…
  - … but don't let it be just by 'default' or because 'everyone does it'
- Many other estimands – *not enough time today*
  - 'speed' scale (accelerated failure time models) – useful for time-varying treatments
  - restricted mean survival time etc.

Survival Outcome - Estimands

What about censoring?

- Want estimand ‘outside’ of a study setting, i.e. ‘without’ censoring

- **Target trial:** has no censoring at all

  or at least same complete follow-up $\mathcal{T}$ for everyone

  - aka ‘elimination of censoring’, or *‘complete populations’ (TG8)*
  - careful with special ‘censoring’ events: drop-out, treatment switching, competing events
    - *relevant target trial **without** these types of intercurrent events?*
  - similar ***reasoning & assumptions*** as with counterfactual treatment!
    - *in particular: think about common causes of censoring and outcome event*

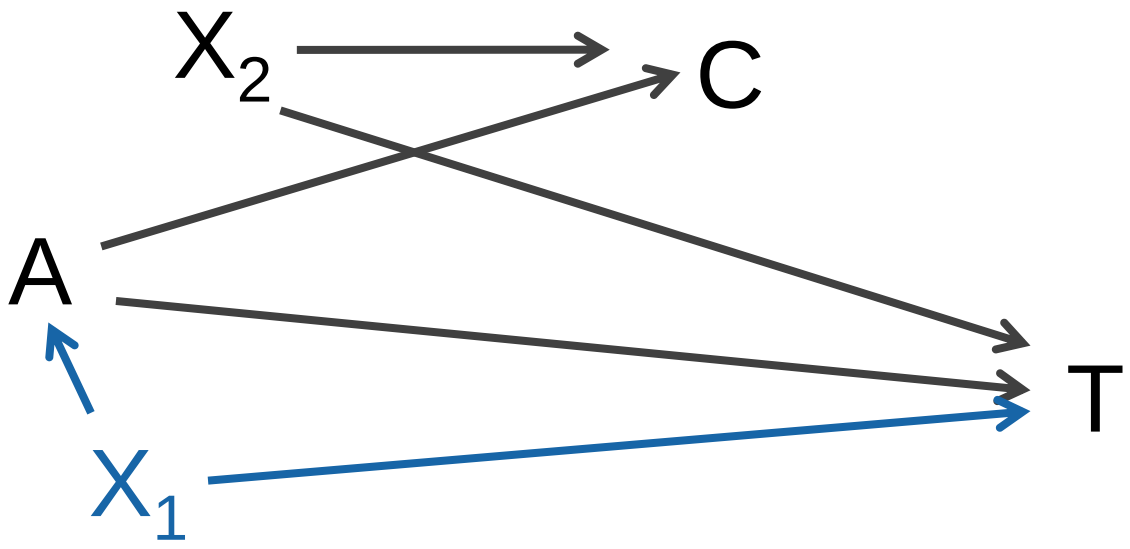
Key Assumptions

- **NUC**: X_1 = sufficient covariate information regarding treatment assignment confounding

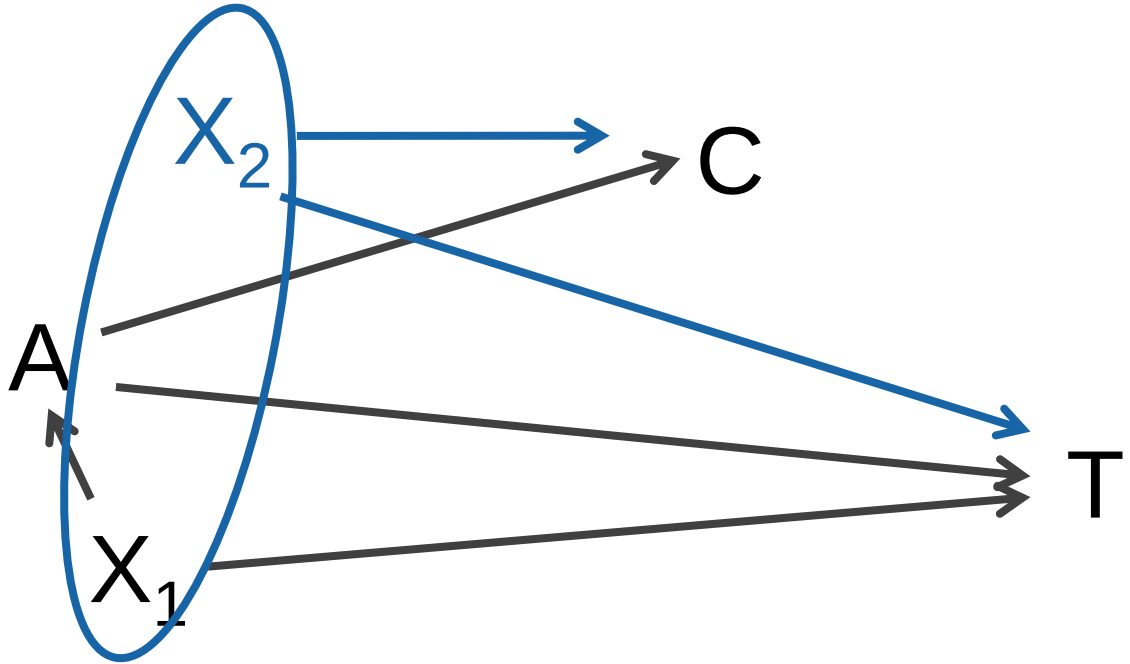
- X_2 = sufficient covariate information regarding possibly (time-varying) 'common causes' of censoring and event
  - ≈ ensure 'conditionally independent' censoring *(TG8)*

- Methods must use X_1 & X_2 jointly
  - *(X_1 and X_2 can overlap)*

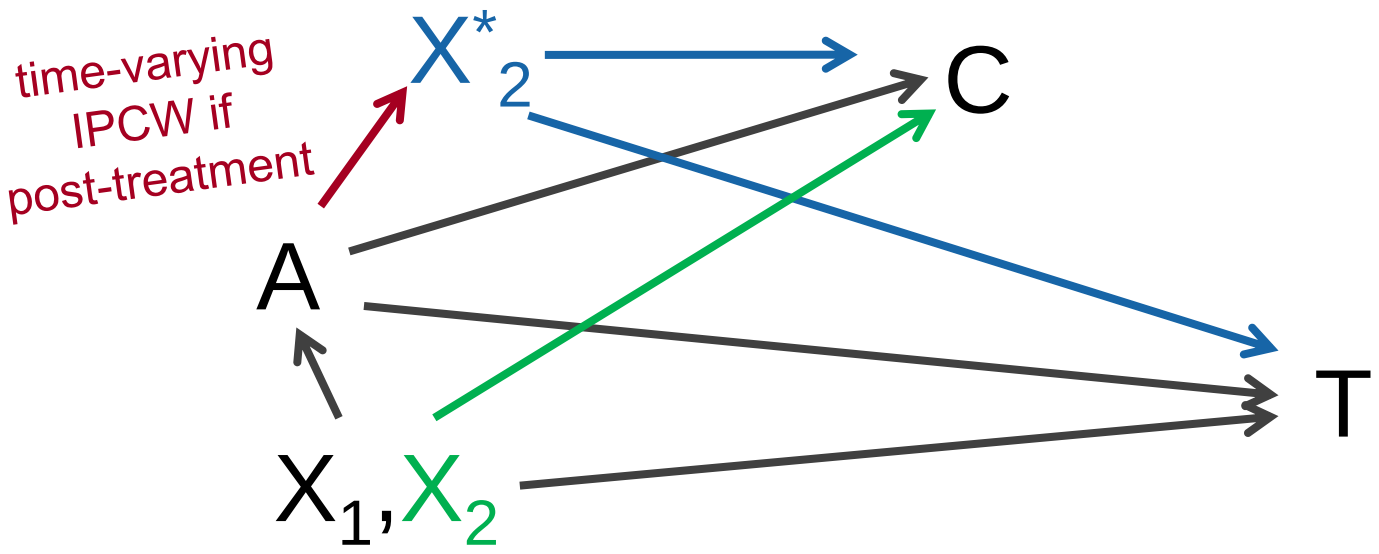
Key Assumptions - DAG



Key Assumptions - DAG

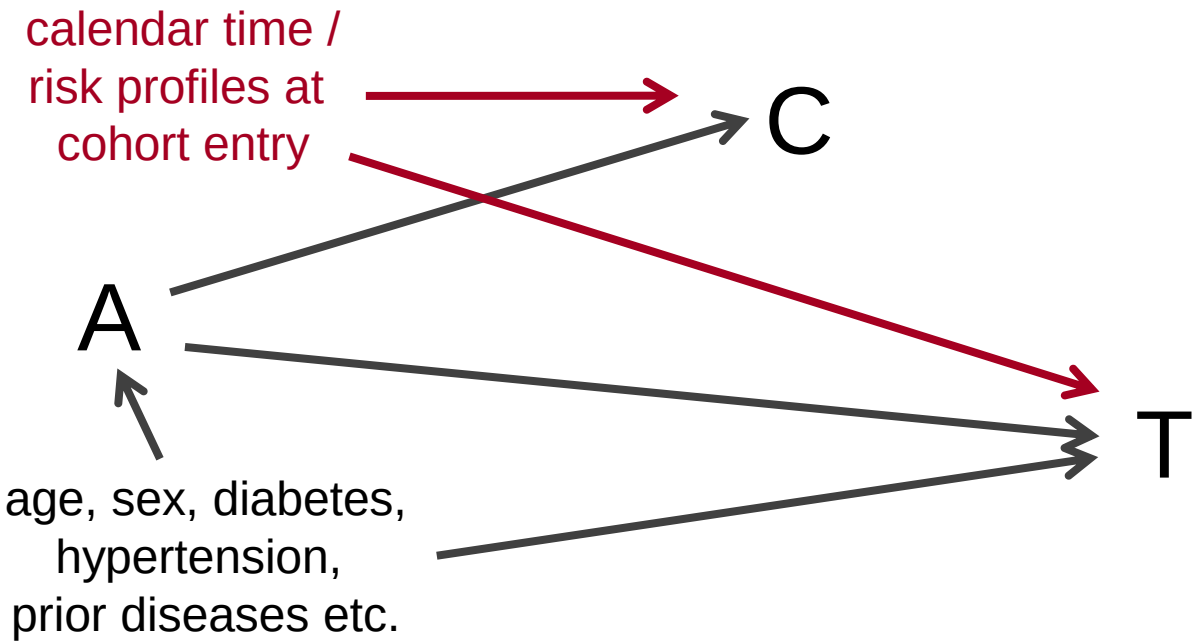


Key Assumptions - DAG



RRT – PKT Example

(Parra et al, 2020, arXiv:2011.11771)



Methods

Model-based marginal counterfactual survival curves:

- (Sufficiently) flexible hazard models
  - possibly separately for treatment groups
  - include both sets of baseline covariates X_1 and X_2

+ transformation to risk scale

+ standardisation

Software:

- R: stdReg *(Sjolander & Dahlqwist);* Stata: stpm2_standsurv *(Lambert)*
- discrete-time-methods: plenty of code (*Hernan & Robins, book)*

Methods

Weighted Kaplan-Meier curves:

- Inverse probability of treatment & censoring weighting
- Note: including covariates in IPTW does not suffice if also needed to adjust for confounding of censoring

⇒ need IPCW too (time-varying) *(Roysland, Didelez et al., 202?- tba)*

Software:

- R: survival library *(Therneau)*, R: ipw *(van der Wal et al, 2011),* R: ahw *(Ryalen, github);* core Stata

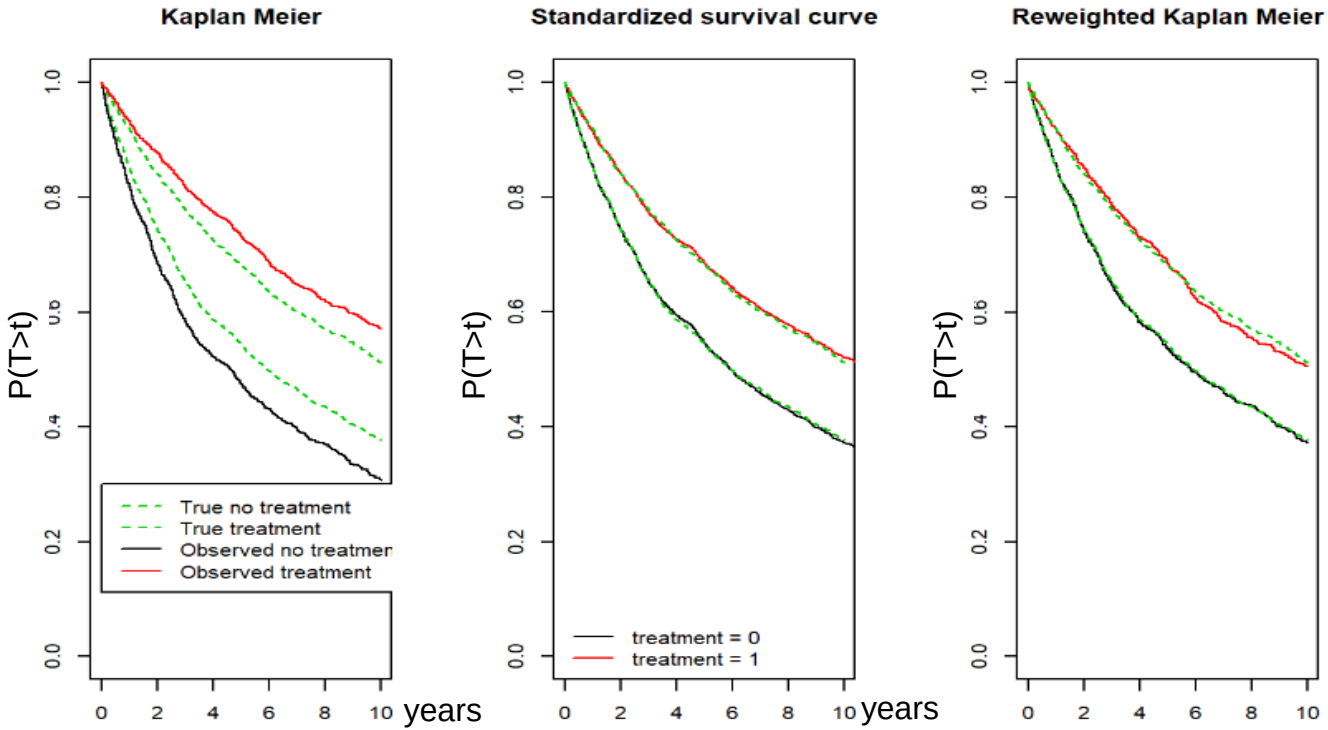
Illustration

Simulated data inspired by RRT data (but somewhat simplified)

N=2000

Confounding by observed covariates & no censoring

(very basic programming)



Illustration

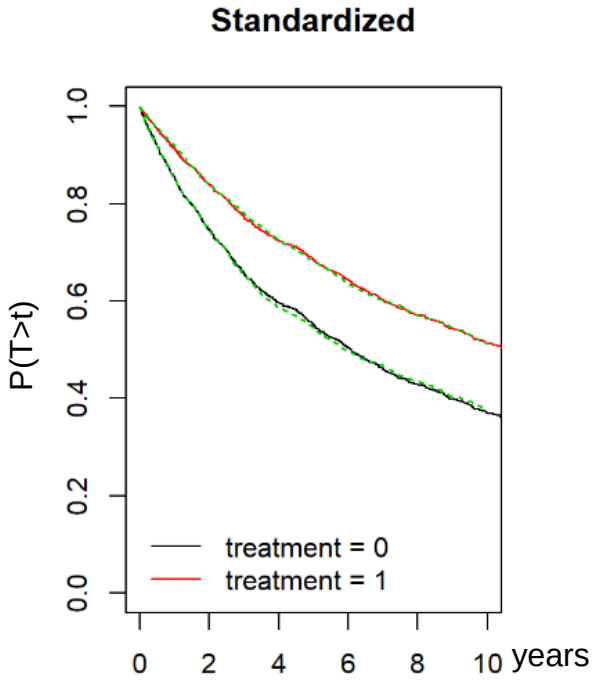
Simulated data inspired by RRT data (but somewhat simplified)

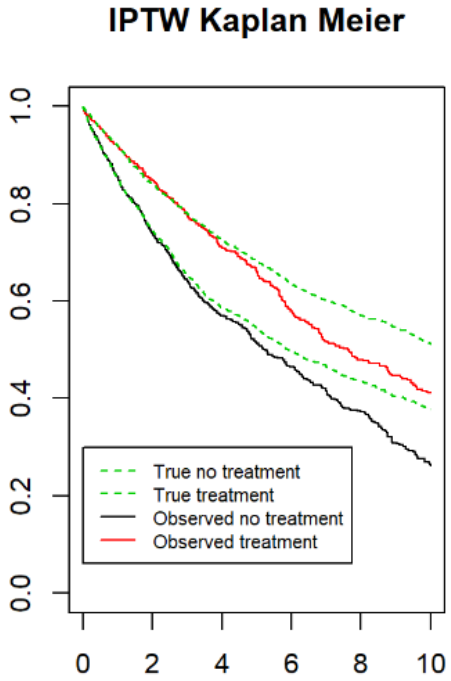
N=2000

Confounding
by observed
covariates
& with censoring

Only using IPTW
not good enough

Some improvement
with IPCW (not shown)





Summary

Key messages

- Can & should choose meaningful, clinically relevant causal estimands for survival outcomes
  - target trial should also address censoring
- Hazard models well-established – only need to be suitably transformed
- Think ‘causally’ about censoring to justify key assumptions
  - in addition to ‘no unmeasured confounding’
- More details on (simple) implementation / software in paper – forthcoming!

STRATOS INITIATIVE

Contact
Vanessa Didelez
Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen
didelez@leibniz-bips.de

BIPS Leibniz