Bridging the gap between causal inference and survival analysis: a censored edition

STRATOS TG 7

Saskia le Cessie	Leiden University Medical Centre and 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

- STRengthening Analytical Thinking for Observational Studies
- Tremendous amount of development of new statistical methods
- In practice only few of these methods are used
- Big gap between methods development and methods applications
- Aim: provide guidence for statisticians and other data analysists

STRATOS: 9 topic groups

1. Missing data
2. Selection of variables and functional forms in multivariable analysis
3. Initial data analysis
4. Measurement error and misclassification
5. Study design
6. Evaluating diagnostic tests and prediction models
7. Causal inference
8. Survival analysis
9. High-dimensional data

And several panels (in total 11 e.g. simulation panel)
More in Biometrics Bulletins and https://stratos-initiative.org/

Some recent (2020) highlights of STRATOS

- TG 2: Overview of variable and Function Selection (Diagnostic and Prognostic Research 2020)
- TG 3: Review on reporting initial data analysis (BMC Medical Research Methodology 2020)
- TG 4: Two guidance papers on measurement error (Stat in Med 2020)
- Involved in SISAQOL, large European project on setting standards for analysis of Patients Reported Outcomes in Cancer Trials

MORE IN STRATOS MINI-SYMPOSIUM ON THURSDAY

Topic group 7: causal inference

- courses
- didactic material (case studies)
- a simulation learner
- website. (ofcaus.org)
- talks at meetings (+ organizing meetings)
- papers

Tutorial on causal questions and principled answers will appear shortly in Statistics in Medicine

Key steps of causal inference

1. Define treatment with relevant levels/values corresponding to scientific question of study.
2. Define outcome.
3. Define the population(s) of interest.
4. Formalise potential outcomes
5. Specify the target causal effect ('Estimand')

Key steps of causal inference

1. Define treatment with relevant levels/values corresponding to scientific question of study.
2. Define outcome.
3. Define the population(s) of interest.
4. Formalise potential outcomes
5. Specify the target causal effect ('Estimand')
6. State assumptions validating the causal effect estimation from the available data.
7. Estimate target causal effect.
8. Evaluate the validity of the assumptions and perform sensitivity analyses as needed.

The Simulation Learner

- Simulate dataset based on an existing study
- Augment data with potential outcomes and potential exposures
- Aim: Illustrates concepts and methods on data

Estimands for time to event outcome

Needed:

- Relevant time-to-event $\tilde{T}$ as outcome variable;
- A clear starting point (time zero);
- Population of interest (whole/ marginal, subpopulation/ conditional, the treated);
- (point) exposure/treatment of interest with interventions to compare: e.g. treat versus not-treat
- Potential outcomes: $\tilde{T}^a$ is potential time-to-event under treatment setting $A = a$.

Estimands for time to event outcome

Needed:

- Relevant time-to-event $\tilde{T}$ as outcome variable;
- A clear starting point (time zero);
- Population of interest (whole/ marginal, subpopulation/ conditional, the treated);
- (point) exposure/treatment of interest with interventions to compare: e.g. treat versus not-treat
- Potential outcomes: $\tilde{T}^a$ is potential time-to-event under treatment setting $A = a$.

Concept of emulating Target Trials can be very helpful (Hernán(2016))

Estimands: censoring

- Censoring makes the event of interest 'invisible';
- Interest in inference for an uncensored population;
- Estimand should not depend on censoring aspects;

A special type of intermediate event

Estimands: defining time zero

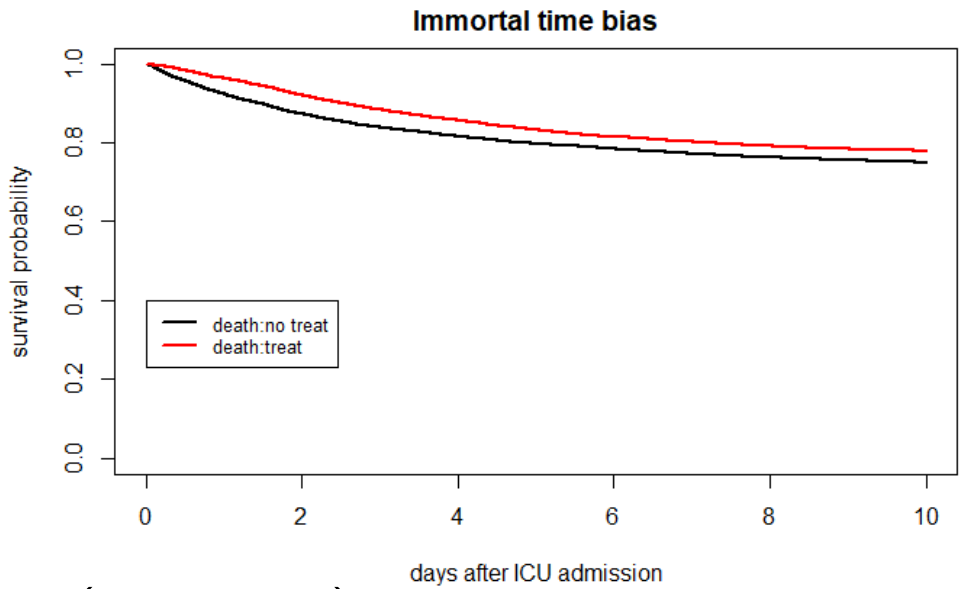
- In RCTs: time zero is clear (start of randomisation)
- In observational studies often less so
- Different time scales: calendar time versus start treatment versus age
- Prevalent instead of incident users $\Rightarrow$ selection bias
- Users defined later in follow up $\Rightarrow$ immortal time bias

Immortal time bias: a common made mistake

Simulation example: based on recent COVID research

- Population: patients admitted to ICU
- A treatment without an effect
- Treatment started between day 0 and 5

Ignoring delayed start of treatment yields:



Observed effect (HR=0.88), while no effect was present

The COVID example: issues when defining estimand

- What would be the population of interest? All patients at the ICU? The treated subpopulation? Or those "admissible for treatment"?
- What are the levels of treatment? "Start immediately" versus "no treatment"? Or some dynamic treatment strategy(e.g start treatment if oxygen levels are very low)

Estimands: the scale of the contrast

1. Survival/risk scale
   - Differences/ratios in survival probabilities at specific time point(s) $P(\tilde{T}^1 > t)$ versus $P(\tilde{T}^0 > t)$;
   - Difference in survival curves;
   - Difference in median survival time;
   - Difference in (restricted) mean survival time $E(min(\tilde{T}^1, t^*))$ versus $E(min(\tilde{T}^0, t^*))$, with t^* a predefined time horizon

Estimands: hazard scale

Hazard ratios are extremely popular in survival analysis

- They can be estimated (assuming non informative censoring) without specifying baseline hazard
- Theory and software of Cox proportional hazard model are well developed
- Often low dimensional models fit well

But

- Hazard ratios are non collapsible (conditional hazards differ from marginal hazards)
- Hazards and hazard ratios have difficult causal interpretation (hazard of hazard ratios (Hernán (2010); Aalen et al(2015); Stensrud et al(2019); Martinussen et al (2020), Young et al (2020))

A causal view on hazards

Causal hazard function:

$$\lambda^a(t) = \lim_{h \to 0} \frac{1}{h} P(t \leq \tilde{T}^a < t + h \mid \tilde{T}^a \geq t).$$

- Interest in contrast between $\lambda^1(t)$ and $\lambda^0(t)$.
- When patients have different frailties: populations with $\tilde{T}^1 \geq t$ and $\tilde{T}^0 \geq t$ after some time no longer exchangeable
- Relatively more frail people at risk in treated group.
- Conditioning $\Rightarrow$ Selection bias
- Counter-intuitive things can happen:
  - $\tilde{T}_i^1 > \tilde{T}_i^0$ for all individuals i
  - while $\lambda^1(t) \leq \lambda^0(t)$ for some t .
- Even if A is randomised.

Estimands on 'speed' scale

Accelerated failure times

$$T_i^*(1) = T_i^*(0)\exp(\psi_i).$$

or

$$\log(T_i^*(1)) = \log(T_i^*(0)) + \psi_i.$$

An estimand in the population could be
$\psi = E(\log T^*(1)) - E(\log T^*(0))$

General assumptions needed for estimation

- Positivity; each individual should be able to receive all treatment levels
- Causal consistency: observed outcome is equal to an potential outcome when A is set to observed treatment level
- No interference: impact of treatment on outcome is not affected by other individuals being exposed or not.

Other assumptions needed for estimation

- No unmeasured confounding (NUC); received treatment is independent of the potential outcomes, given baseline covariates.
- Or alternative sets of assumptions (e.g. instrumental variable methods, discontinuity designs)

Methods based on NUC: 1. Outcome regression with standardisation

- Predict survival curves for each individual, given all (measured) confounders, separately for treated and untreated level
- Then average over empirical covariate distribution, separate for treated and untreated
- Direct standardisation
- Yields average survival curves in total population
- Easy R package stdReg (Sjolander (2016)) or stpm2 in Stata

Methods based on NUC: 2.Propensity score methods

- Fit a model for the propensity of receiving treatment conditional on the covariates, $e(L) = P(A = 1|L)$
- Check if fitted PS is adequate (well balanced)
- Then use PS methods (e.g matching, stratification, inverse probability weighting)

Different underlying assumptions

- Outcome regression:
  - Correct formulation of survival model
  - Non informative censoring, conditional on covariates in model

  Propensity score methods:
  - Correct propensity model
  - Non informative censoring

Example: pre-emptive kidney transplant and survival

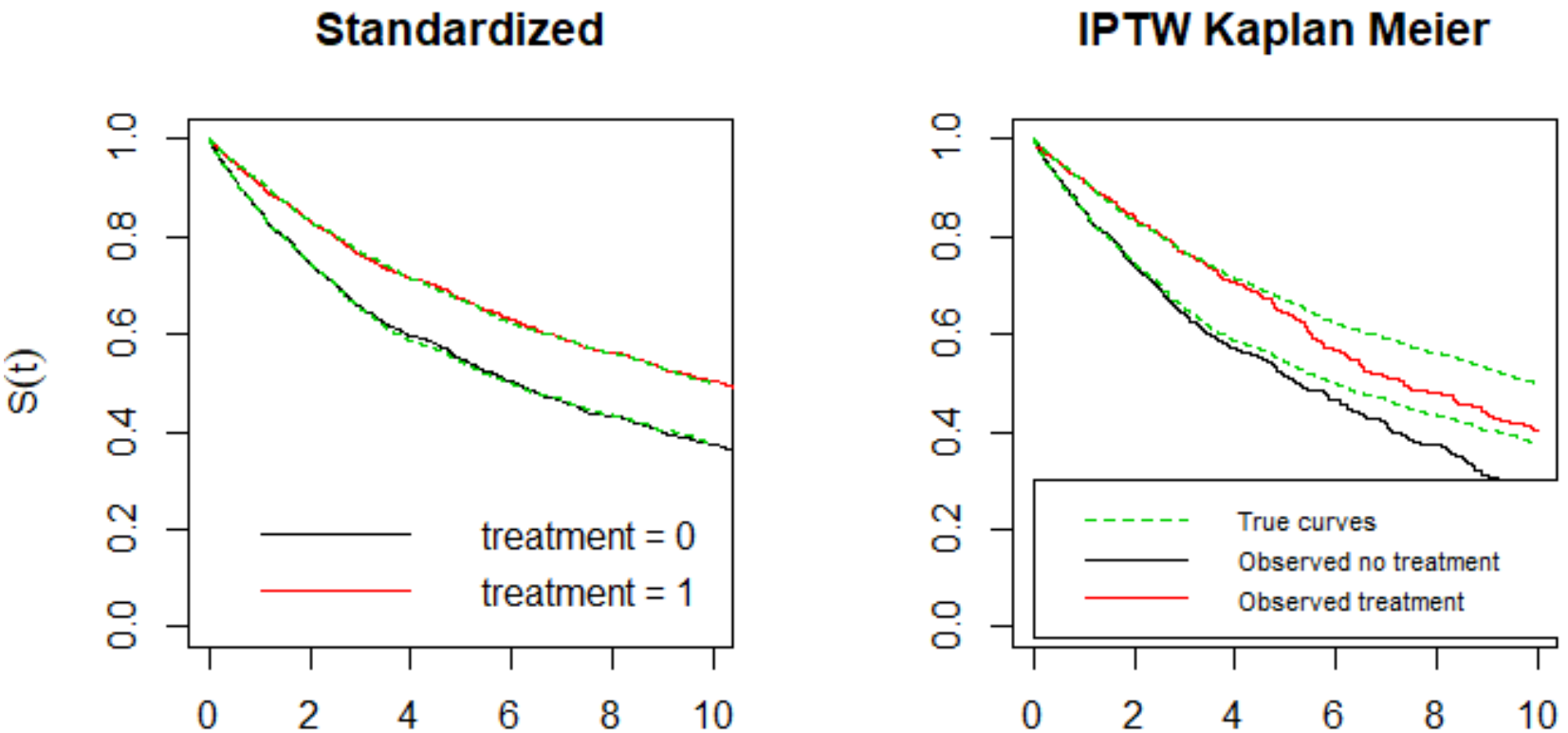
Generate a simulation setting, motivated by real data example

- Patients become eligible for kidney transplant
- Treatment: transplanted immediately versus later
- Generated 2000 patients, included between 2001 and 2017
- Treatment depends on age and calender year
- Survival (exponential) depends on treatment, age, interaction between age and treatment and year
- Over time mean age in population increases
- Administrative censoring in 2020

Estimation

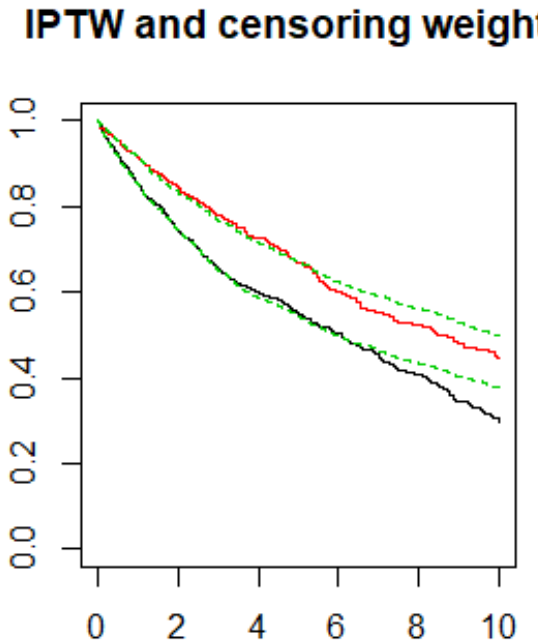
- Outcome regression, with standardisation
- Propensity with inverse probability weighting
- In both models age and year as covariates

Dealing with informative censoring



Propensity methods need to account for informative censoring

Using IPTW + censoring weights



- Censoring model I used, is probably not perfect
- If censoring depends on time varying covariates, both methods need censoring weights

Competing events

- Competing event Z, with time to occurrence V
- Here Z can affect $\tilde{T}^a$, as the event of interest will never happen after Z occurs.
- Different estimands may be of interest
  1. contrasts between $\tilde{T}^1$ versus $\tilde{T}^0$
     - the 'total' effect of A on $\tilde{T}$.
     - corresponds to "competing risk cumulative incidences".
  2. contrasts between $\tilde{T}^{1,V=\infty}$ versus $\tilde{T}^{0,V=\infty}$
     - 'Direct effect' effect of A on $\tilde{T}^a$.
     - What would happen if we eliminate Z?

  Estimand 2 is what typically be used by many as default, without being aware of it.

Conclusion

Think careful about estimand of interest, and about corresponding analysis with underlying assumptions

www.ofcaus.org

References

- Goetghebeur et al. Stat in Med. 2020 To appear
- Aalen et al. Lifetime Data Anal 2015
- Hernán. Epidemiology 2010
- Hernán, Robins. Am J Epidemiol 2016
- Martinussen Lifetime Data Anal 2020
- Sjolander Eur J Epidemiol 2016
- Stensrud et al. Eur Heart J.2019
- Young et al. Stat in Med 2020