Design of an emulated trial using modern causal inference

Thomas Alexander Gerds

Section of Biostatistics, University of Copenhagen

May 22, 2023

The overall idea

In causal inference, an emulated target trial is a method for estimating the effects of a treatment or intervention.[1]

To conduct an emulated target trial, researchers use observational data and statistical methods to construct a hypothetical trial that emulates the conditions of a randomized controlled trial.

This involves selecting a study population that is similar to a population that would have been eligible for the hypothetical randomized trial and modeling the treatment assignment process.

The researchers then simulate the outcomes that would have been observed if the hypothetical trial had been conducted.

[1] Hernan and Robins. Am J Epi, 183(8):758–764, 2016.

(Limitations of) Register data

The most important challenges that have to be resolved are:

1. how to identify when a person is eliglible?
2. identify variables that were used (in the real world) to decide whether the person initiated treatment
3. collect data to determine adherence to treatment during followup
4. collect variables that were used to decide treatment changes
5. collect predictors for outcome(s), censoring, and competing risks
6. force all this information into a discrete time grid

The statistical analysis plan

Design

1. Inclusion/Exclusion $\rightarrow$ time zero
2. Width of time grid intervals
3. Treatment regimens (static, stochastic, dynamic)
4. Target parameters (estimands, contrasts)

Analysis

- Specification of nuisance parameter estimation, history adjusted:
  - propensity score models
  - censoring models
  - outcome models
- List of learners for super learning (pre-specified ands much more flexible than a single regression model)

When is time zero?

When is time zero

In the trials actually designed and carried out in the real world, a prespecified trial protocol defines the target population and how a study population would be recruited from eligible patients, and therefore the definition of time zero of any participant already in the study is obvious.

However, during the recruitment process of a hypothetical trial, there always exist multiple possibilities in defining a study population and the associated time zeros.

When is time zero: register data

Strategy 1 include everyone in the register who passes some given eligibility criteria:

a. set time zero at the first pass and look back in time to define treatment status
b. grace period: set time zero x days after date of first pass. exlude previously treated. treatment initiation happens in grace period

When is time zero: register data

Strategy 1 include everyone in the register who passes some given eligibility criteria:

- a set time zero at the first pass and look back in time to define treatment status
- b grace period: set time zero x days after date of first pass. exlude previously treated. treatment initiation happens in grace period

Strategy 2 include everyone in the register who starts the treatment:

- a compare with patients who start a different treatment
- a match with other patients who could have started the treatment but did not (not yet)

When is time zero: illustration

Example inclusion

Include all outcome-free patients above 50 who have a 5 year or longer history of diabetes treatment when they have a recent HbA1c measurement above 58 mmol/mol.

[2] A real randomized trial would not include the same patient multiple times

[3] Subset analyses (e.g., age $>$ 60 or HbA1c $>$ 65) can potentially include a patient who is not in this subgroup for the main analysis at a different date.

When is time zero: illustration

Example inclusion

Include all outcome-free patients above 50 who have a 5 year or longer history of diabetes treatment when they have a recent HbA1c measurement above 58 mmol/mol.

Challenges

- not all patients get their HbA1c measured regularly
- the register data do not contain all HbA1c measurements
- there can be HbA1c above 58 during the first 5 diabetes years
- a single patient can fulfill the eligibility criteria multiple times according to the registered data[2, 3]

[2] A real randomized trial would not include the same patient multiple times

[3] Subset analyses (e.g., age $>$ 60 or HbA1c $>$ 65) can potentially include a patient who is not in this subgroup for the main analysis at a different date.

Treatment regimens of the hypothetical target trial

Treatment regimens

Static treatment regimen

A fixed treatment strategy that does not change based on individual patient responses/characteristics.

Dynamic treatment regimen

A personalized treatment strategy that adapts treatment decisions based on individual patient responses over time. Used in situations where patients require sequential or multiple treatments, and the optimal treatment may depend on how the patient responded to previous treatments.

Stochastic treatment regimen

Assigns treatments to patients with varying probabilities based on a probabilistic algorithm. Stochastic treatment regimens are useful when the optimal treatment is uncertain, and there are multiple treatments with similar expected outcomes.

Examples

Static treatment regimens

- Patients continue treatment A for x years
- Patients continue to not use treatment A for x years (placebo is not an option!)
- Patients continue treatment A for x years but are not allowed to use treatment B

Dynamic treatment regimens

- If measured `HbA1c > 60` then intensify dose of treatment A or add treatment B
- If patient has a diagnosis with myocardiac infarction then add treatment B

Stochastic treatment regimens

- Patients add treatment B with a given probability conditional on patient characteristics, treatment history and comorbidity

Target parameters

Target parameters

For one treatment regimen, the expected outcome t-years after time zero.

For two or multiple treatment regimens, target causal effects (estimands) include

- Average Treatment Effect (ATE): The average difference in outcomes[4] t-years after time zero between two treatment regimens in the eligible patients
- Average Treatment Effect on the Treated (ATT): same as ATE but in individuals who received the treatment

[4] Competing risks affect the interpretation!