Right censored survival outcomes: on the use of logistic regression and longitudinal causal inference

Part 2: longitudinal causal inference

Thomas Alexander Gerds

Section of Biostatistics, University of Copenhagen

Overview

- Target trial emulation
- Longitudinal register data
- Roadmap of statistical learning
- Iterative sequential regression formula
- Longitudinal targeted minimum-loss based estimator
- Summary

RCT: Randomized clinical trial
Treatment plan/protocol/regime
Intention-to-treat
Per-protocol
Adherence (in patient care): "The patient showed good adherence to the treatment plan, taking all medication as prescribed". (Voluntary behavior change)

Compliance (in patient care): "The doctor was concerned about the patient's low compliance with the recommended lifestyle changes". (Mandated or expected behavior that may or may not be

fully adopted)

Motivation

Randomized trials = gold standard for causal inference

But: trials may be infeasible, unethical, or too expensive, and people cannot be forced to comply with a randomized treatment strategy: *they can stop, switch, start and pause treatment*

Observational data are becoming increasingly available through electronic health records.

Question: Can we learn causal effects from them?

Examples of research questions for emulated trials

Does menopausal hormone therapy increase the risk of cardiovascular disease? [1]

Comparative effectiveness of omalizumab, mepolizumab, and dupilumab in asthma [2]

Can diabetes medicine prevent Alzheimer's disease? [3]

---

[1] British Medical Journal 2024, 387, e078784.

[2] Journal of Allergy and Clinical Immunology, 151(5):1269–1276, 2023.

[3] Alzheimer's and Dementia, 20(12):8661–8672, 2024.

Why not just "adjust for confounders"?

- Standard regression may not be applied in longitudinal settings
- Common epidemiological designs lead to odds and hazard ratios with unclear clinical interpretation
- Confounders are often time-varying, affected by past treatments, and affecting future treatments
- Avoidable biases arise when naïve adjustment is used

The emulated trial framework helps us to explicitly state the assumptions needed to estimate the target parameters and to avoid some of the more severe pitfalls (such as immortal time bias).

Target trial emulation (Step I)

Define the target trial, that is the ideal trial which would like to do but cannot for whatever reasons (ethical, financial, lack of time):

- Eligibility criteria
- Treatment regimens: protocols dictate treatment during followup
- Assignment procedure: enrollment period, hypothetical randomization scheme
- Follow-up period: when does it start and how long does it last
- Outcomes: primary and secondary, competing risks
- Causal contrasts (AKA estimands, target parameters)

Target trial emulation (Step II)

Define the emulated trial by aligning the observational data with the target trial:

- Choose study period in calendar time, note administrative end of followup
- Time zero and how eligibility criteria are approximated
- Handling of run-in/grace periods (which the target trial does not need)
- What information about treatment is available and how is "being on treatment" defined
- How are outcomes and competing risks are assessed (in the absence of clinical data)

Important notes

The target trial should be biologically and physically feasible.

The treatment regimens/protocols of the emulated trial may deviate from the protocols of the target trial.

An intention-to-treat-analysis-analog does usually not exist in the emulated trial analysis of observational data.

Censoring means "do not observe" due to administrative end-of-followup or emigration but is sometimes (mis)-used to also describe an intervention on treatment, i.e., followup is censored when a person does not comply (adhere) with the treatment protocol

Practicals (1)

1. Read the abstract of the paper by Finkelstein & Robins
2. Read from the middle of the first column of page 780 starting with "More precisely, . . . " until
3. The authors describe 4 analyses which differ regarding how and when the followup data are censored. Describe corresponding treatment regimens/protocols.
4. Start treatment then do not control treatment (use standard care) $\neq$ intention to treat
5. Ask chatGPT

There is no gentle introduction to longitudinal targeted minimum loss based estimation. Here is one for the rough and tough.

In the following we discuss the rationale and details behind Robin's iterative regression formula in discrete time, and then discusses the

transition to continuous time.

Notation

The target parameter

Identification of the target parameters

Potential outcomes

- Y^a: outcome if assigned treatment strategy a
- Causal effect = contrast of potential outcomes
- Target trial defines these outcomes explicitly

Identification assumptions

- Exchangeability (no unmeasured confounding)
- Positivity
- Consistency
- These assumptions guide valid trial emulation

The clinical question

- Does prescribing beta-blockers after myocardial infarction (MI) reduce mortality and recurrence risk?

The target trial (ideal RCT)

- Eligibility: patients hospitalized with acute MI, no contraindications
- Treatment strategies:
  - Strategy 1: start beta-blocker at discharge
  - Strategy 2: no beta-blocker at discharge
- Assignment: randomization at discharge
- Start of follow-up: hospital discharge
- Outcomes: death, recurrent MI (composite or separate)
- Causal contrast: 5-year risk difference or hazard ratio

Emulating with observational data

- Data source: hospital registers, prescription databases, national death registry
- Eligibility: patients with first MI hospitalization, exclude contraindications (as feasible)
- Treatment strategies: classify patients by filled prescription at discharge
- Assignment: not randomized $\rightarrow$ need to adjust for confounding
- Start of follow-up: discharge date
- Outcomes: death and recurrent MI from registries
- Estimation: g-methods to account for confounding

Student exercise (beta-blockers after MI)

- Fill in the standard table:
  - Left column: target trial specification
  - Right column: emulated trial specification
- Discuss: what is easy to emulate? what is hard?

Template table

Component	Target Trial
Eligibility criteria	MI patients, no contraindications
Treatment strategies	Randomize to beta-blocker vs no beta-blocker
Assignment procedure	Randomization
Start of follow-up	Discharge date
Outcomes	Death, recurrent MI
Causal contrast	5-year risk difference / hazard ratio

Discussion prompts

- What assumptions are needed for the emulated trial to recover the target trial effect?
- What aspects of the target trial cannot be perfectly emulated?
- Where might bias enter?

Three estimation strategies

- Inverse Probability Weighting (IPW)
- G-computation formula
- Targeted Maximum Likelihood Estimation (TMLE)
- All aim to reproduce what randomization would have given us

Single time point

- $E[Y^a] = \sum_l E[Y \mid A = a, L = l]P(L = l)$
- Average predicted outcomes across covariate distribution
- Think of it as "standardization"

Motivation

- Treatments and confounders vary over time
- We need methods aligned with the design of longitudinal target trials

Iterative sequential regression (longitudinal g-formula)

- Works backwards in time
- Recursively predict outcomes under interventions
- Natural extension of single-time g-formula
- Provides explicit link to target trial emulation