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
- Practical I

---

break

- Roadmap of statistical learning
- Practical II

---

lunch

- Longitudinal targeted minimum-loss based estimation
- Practical III

Terminology check I

Randomized clinical trial

Treatment plan/protocol/regime

Intention-to-treat

Per-protocol

Adherence/ compliance

Terminology check II

Target parameter

Nuisance parameter

Censoring

Competing risk

Influence function

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:

> People pause, stop, switch, start, and add to their medical treatments

Observational data are becoming increasingly available through electronic health records.

Question: Can we learn causal effects from them?

Motivation

The *emulated target trial* combined with the *roadmap of statistical learning* helps to:

state clearly what we estimate

avoid avoidable flaws in the epidemiological design

separate what we estimate from how we estimate it

discuss assumptions and limitations of the analysis

avoid spurious goodness-of-fit tests

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 analysis should 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 we 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)

Example: 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 and separate)
- Causal contrast: 5-year risk difference

Example: the emulated trial (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
- Start of follow-up: discharge date (or 30 days later)
- Outcomes: death and recurrent MI from registries
- Estimation: g-methods to account for confounding

Note that

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 because there was no protocol which defined the intention to treat.

Instead one can design a new user analysis and estimate the the effects of starting a treatment vs active comparator.

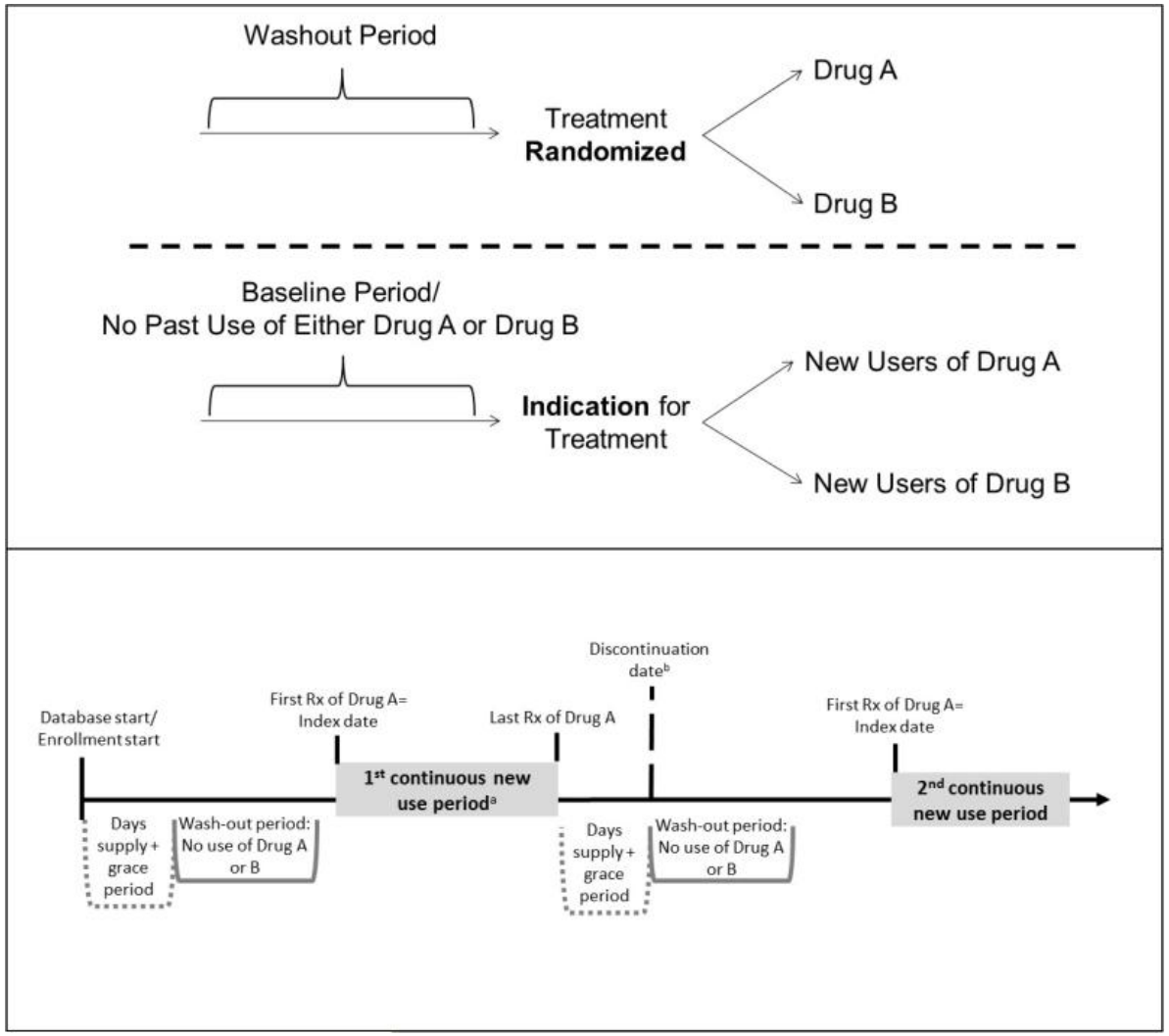
Active comparator, new user study design

In register studies there is often no natural control group of patients who are not exposed to one of the drugs, because treatment indication information is poorly registered.

That is, we do not know if a patient did consider to take a drug and then decided not to.

- A medical diagnosis or a biomarker above threshold can approximate treatment indication.
- Matching can be used to obtain non-users (e.g., exposure density sampling).
- Comparative effectiveness study: active comparator arms.

Active comparator, new user study design [4]



[4] Curr Epidemiol Rep. 2015 Sep 30;2(4):221–228.

Censoring clash

Censored data are the existence of survival analysis!

We define the survival probabilities in the uncensored world and then use the censored data to estimate them.

The survival probabilities do not depend on the censoring mechanism.

When investigators decide to "censor" patients at intermediate events (such as side effects of a drug or change of treatment) they implicitly change the protocol of the target trial and the interpretation of the target parameter!

Hazard rates

$$P(T \in \mathrm{d}t | T \geq t, A(0) = 1) \qquad \text{hazard rate of new users}$$

"Censor" when $A(t) = 0$ happens

$$P(T \in \mathrm{d}t | T \geq t, \bar{A}(t) = 1) \qquad \text{hazard rate "while treated"}$$

The treatment adherence/compliance is an internal covariate which affects the disease progress and also depends on the disease progress.

The inherent selection bias is called

- survival of the fittest
- depletion of the susceptible

~~Censoring~~ Intervening on intermediate events

At any time during the person-specific followup period of an emulated trial, a person can be non-adherent or non-compliant and violate the treatment protocol (pause, stop, start, switch, add).

In the analysis of drug A, we censored people when they stopped taking added drug B

A better way to formulate this is to adapt the description of the intervention/treatment protocol:

Use drug A continously without breaks, do not add drug B

The use Robins' G-methods to estimate risk if hypothetically all people had followed the protocol.

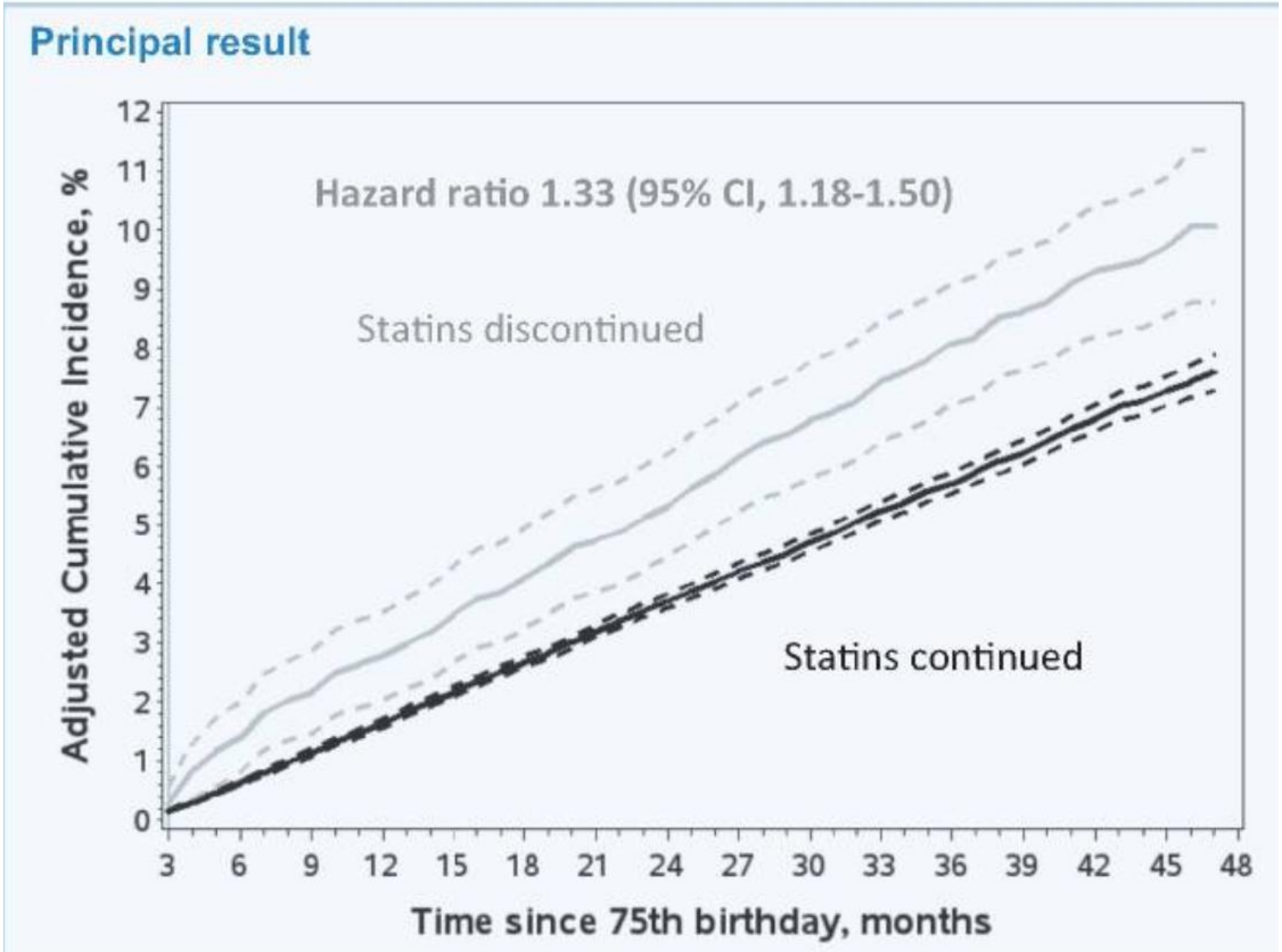
Example

Cardiovascular effect of discontinuing statins for primary prevention at the age of 75 years: a nationwide population-based cohort study in France [5]

Aims	The effect of statin discontinuation on cardiovascular outcomes
Enrollment	All statin users who turned 75 in 2012-2014
Methods?	The hazard ratio comparing statin discontinuation with continuation was estimated using a marginal structural model adjusting for both baseline and time-varying covariates
Results??	Statin discontinuation was associated with a 33% (HR: 1.33) increased risk of admission for cardiovascular event in 75-year-old primary prevention patients.

[5] Giral et al. European heart journal, 40(43):3516–3525, 2019.

Giral et al.


Principal result
Hazard ratio 1.33 (95% CI, 1.18-1.50)
Statins discontinued
Statins continued
Adjusted Cumulative Incidence, %
0
1
2
3
4
5
6
7
8
9
10
11
12
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
Time since 75th birthday, months

Certain potential biases (statins discontinuation paper)

With the study design adopted, time-related bias, in particular immortal time bias, was avoided by . . .

. . . Potential bias due to treating competing events as censoring events as in our analyses is also small: in the main analysis, only 2.7% of patients were censored for death and, . . .

. . . The percentage of competing events was therefore much lower than 10%, the critical value reported in the literature (Austin et al., 2016).

Certain potential biases (statins discontinuation paper)

With the study design adopted, time-related bias, in particular immortal time bias, was avoided by . . .

. . . Potential bias due to treating competing events as censoring events as in our analyses is also small: in the main analysis, only 2.7% of patients were censored for death and, . . .

. . . The percentage of competing events was therefore much lower than 10%, the critical value reported in the literature (Austin et al., 2016).

Censor the nonsense!

The study by Giral et al is half-baked

The authors define a proper framework for survival analysis:

- population
- time zero
- time to cardiovascular outcome

The authors use a marginal structural model and inverse probability weighting to deal with time-varying confounding.

But:

- What is the interpretation of the adjusted cumulative incidence curve and the adjusted hazard ratio?
- When is the time of statins discontinuation?
- They ignored the competing risk of death due to other causes

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

Practicals (2)

Statins discontinuation or 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?

Roadmap of targeted learning

Road map for targeted learning [6]

Once we have designed an emulated target trial it remains to analyse the data:

- Define causible target parameters (estimands) that possess a desired clinical interpretation in the hypothetical target trial.
- State and discuss identifiability assumptions needed for causal interpretation of the estimates
- Specify models for all the nuisance parameters: regression models for outcome, propensity of treatment, probability of censoring
- Estimate the target parameters with LTMLE
- Estimate standard errors based on the efficient influence function or using a cheap subsampling bootstrap algorithm

[6] van der Laan & Gruber (2012) The international journal of biostatistics

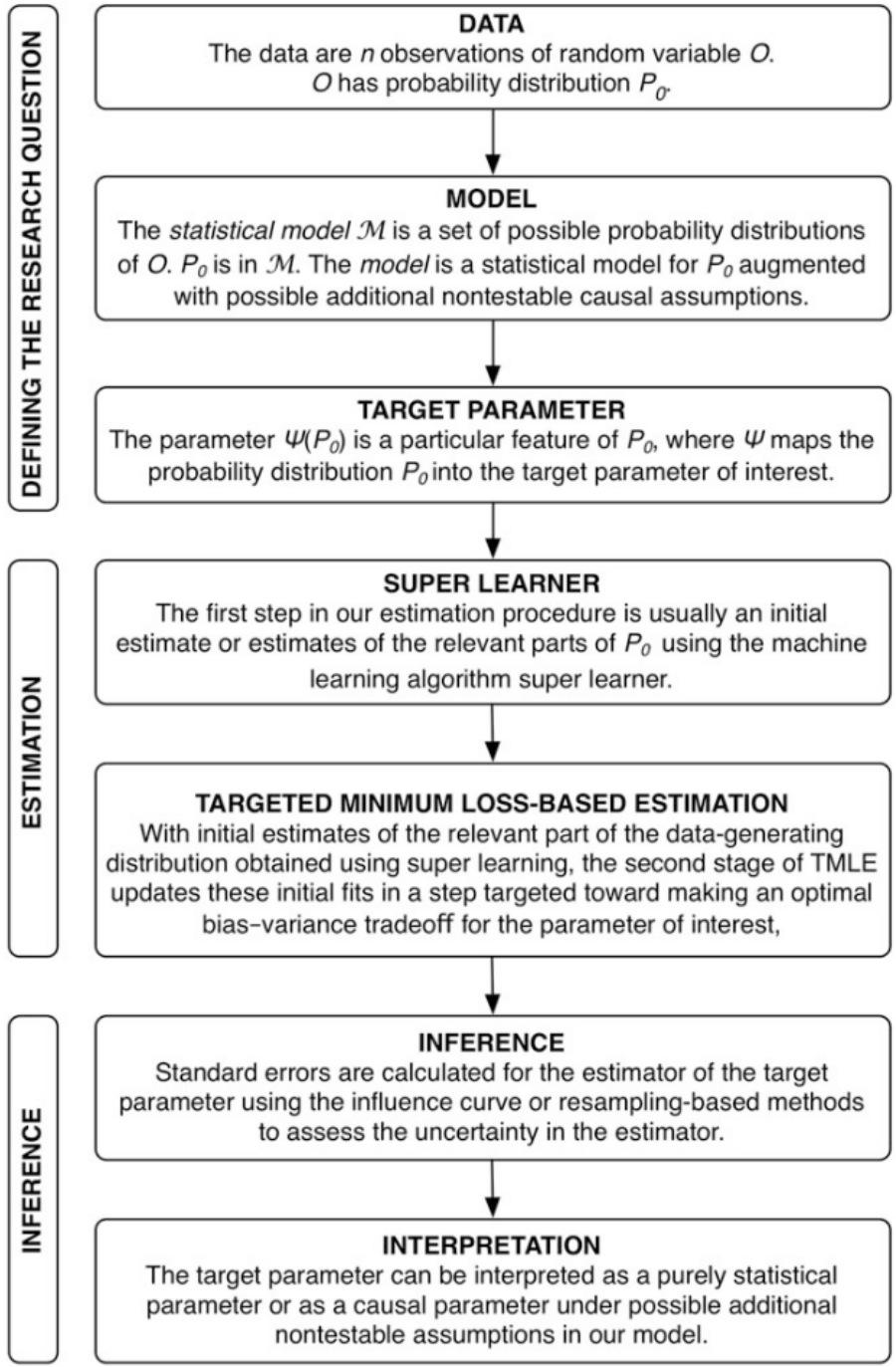


Fig. 1.1 Road map for targeted learning

Introduction to the notation

Uncensored data, one interval $[0, t_1]$ [7]

$$X = (L_0, A_0, Y_1)$$

Observed likelihood

$$\mathrm{P}X = \underbrace{\mathrm{P}_{Y_1|A_0,L_0}}_{F_1} \underbrace{\mathrm{P}_{A_0|L_0}}_{\pi_0} \underbrace{\mathrm{P}_{L_0}}_{H_0}$$

Likelihood in the target trial

$$P^* = F_1 \, \pi_0^* \, H_0$$

[7] $Y_1 = 1\{T \leq t_1\})$

Longitudinal setting for the observational data [8]

Discretized time scale:

$$[0 \cdots\cdots\cdots t_1 \cdots\cdots\cdots t_2]$$

Data for two time intervals:

$$X = (L_0, A_0, Y_1, L_1, A_1, Y_2).$$

The joint probability distribution:

$$\mathrm{P}_X = \mathrm{P}_{Y_2|A_1,L_1,Y_1,A_0,L_0}\mathrm{P}_{A_1|L_1,Y_1,A_0,L_0}\mathrm{P}_{L_1|Y_1,A_0,L_0}\mathrm{P}_{Y_1|A_0,L_0}\mathrm{P}_{A_0|L_0}\mathrm{P}_{L_0}$$

[8] no censoring, no competing risks, $Y_1 = 1\{T \leq t_1\}$, $Y_2 = 1\{T \leq t_2\}$

Longitudinal setting for the observational data [8]

Discretized time scale:

$$[0 \cdots\cdots\cdots t_1 \cdots\cdots\cdots t_2]$$

Data for two time intervals:

$$X = (L_0, A_0, Y_1, L_1, A_1, Y_2).$$

The joint probability distribution:

$$\mathrm{P}X = \underbrace{\mathrm{P}_{Y_2|A_1,L_1,Y_1,A_0,L_0}}_{F_2}\underbrace{\mathrm{P}_{A_1|L_1,Y_1,A_0,L_0}}_{\pi_1}\underbrace{\mathrm{P}_{L_1|Y_1,A_0,L_0}}_{H_1}\underbrace{\mathrm{P}_{Y_1|A_0,L_0}}_{F_1}\underbrace{\mathrm{P}_{A_0|L_0}}_{\pi_0}\underbrace{\mathrm{P}_{L_0}}_{H_0}$$

[8] no censoring, no competing risks, $Y_1 = 1\{T \le t_1\}$, $Y_2 = 1\{T \le t_2\}$

Intervention: Treatment assignment in the target trial

The protocols of the target trial dictate the treatment(s) given at any time (doctor visit) during the target trial period.

Notation:

$$\pi^*(t \mid L(t-), A(t-))$$

Examples:

Protocol	Type of intervention	$\pi^*(t \mid L(t-), A(t-))$
Never treat	Static	1
Always treat	Static	0
Treat for 2 years	Static	$1\{t \leq 2\}$
Treat if $L(t-) > \xi$	Dynamic	$1\{L(t-) > \xi\}$
Treat with probability 0.8	Stochastic	0.8
If $L(t-) > \xi$ treat with probability 0.8	Stochastic & dynamic	$1\{L(t-) > \xi\}0.8$

Treatment assignment in the target trial

Example of treatment regimens:

- **Protocol A** Patients should use GLP1-RA continuously for 3 years and not intensify with SGLT2i
- **Protocol B** Patients should use SGLT2i continuously for 3 years and not intensify with GLP1-RA

Protocol A assigns 100% probability for GLP1-RA and 0% probability for SGLT2i:

$$\pi^{A*}(t) = (1, 0).$$

Protocol B is defined similarly:

$$\pi^{B*}(t) = (0, 1)$$

The target parameter (aka the estimand)

The analysis of the emulated target trial estimates the absolute risks of the outcome(s) if hypothetically all patients had followed the treatment protocols (per-protocol effects).

Example: 3-year risk of cardiovascular disease under π^{j*} and differences thereof:

$$\mathrm{P}_{\pi^{A*}}(Y_2 = 1) - \mathrm{P}_{\pi^{B*}}(Y_2 = 1)$$

The analyst uses the information of the time-varying covariates (comorbidity, co-medicine) to achieve a good compromise between:

- the available data
- the desire of the investigators
- the causal assumptions: *positivity*, *sequential coarsening at random (NUC)*, *consistency*.

Causible parameters

A target parameter is called *causible* if it permits a causal interpretation in the target trial.

Hazard ratios are not causible. [9]

Some of the causal assumptions are not testable and some will be violated to some extent, so that an estimate of a causible parameter based on an emulated trial does not necessarily allow a causal interpretation.

[9] Hernan (2010) The hazards of hazard ratios. Epidemiology.

LTMLE

LTMLE for breakfast

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 motivation and rationale behind Robin's iterative regression formula [10] and the LTMLE estimator of van der Laan [11]

[10] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973, 2005.

[11] Targeted minimum loss based estimation of causal effects of multiple time point interventions. van der Laan and Gruber (2012). The International Journal of Biostatistics, 8(1), 2012.

Robins g-methods[12]

Data	$\{(L_{0i}, A_{0i}, Y_{1i})\}_{i=1}^n = \{X_i\}_{i=1}^n$
Outcome regression model	$\hat{F}_1$
Empirical distribution	$\hat{H}_0 = \frac{1}{n}\sum_{i=1}^n L_{0,i}$
Empirical distribution	$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n X_i$
Propensity of treatment	$\hat{\pi}_0$

Estimator 1: g-formula

$$\hat{\Psi}_g^* = \hat{F}_1 \pi_0^* \hat{H}_0$$

Estimator 2: Inverse probability weighting

$$\hat{\Psi}_{\text{IPTW}}^* = \frac{\mathbb{P}_n \pi_0^*}{\hat{\pi}_0}$$

[12] Robins (1986) Mathematical modelling, 7(9-12):1393–1512

Longitudinal setting [13]

Discretized time scale:

$$[0 \cdots\cdots\cdots t_1 \cdots\cdots\cdots t_2]$$

Data for two time intervals:

$$X = (L_0, A_0, Y_1, L_1, A_1, Y_2).$$

The joint probability distribution:

$$\mathrm{P}_X = \mathrm{P}_{Y_2|A_1,L_1,Y_1,A_0,L_0}\mathrm{P}_{A_1|L_1,Y_1,A_0,L_0}\mathrm{P}_{L_1|Y_1,A_0,L_0}\mathrm{P}_{Y_1|A_0,L_0}\mathrm{P}_{A_0|L_0}\mathrm{P}_{L_0}$$

[13] no censoring, no competing risks, $Y_1 = 1\{T \leq t_1\}$, $Y_2 = 1\{T \leq t_2\}$

Longitudinal setting [13]

Discretized time scale:

$$[0 \cdots\cdots\cdots t_1 \cdots\cdots\cdots t_2]$$

Data for two time intervals:

$$X = (L_0, A_0, Y_1, L_1, A_1, Y_2).$$

The joint probability distribution:

$$\mathrm{P}X = \underbrace{\mathrm{P}_{Y_2|A_1,L_1,Y_1,A_0,L_0}}_{F_2}\underbrace{\mathrm{P}_{A_1|L_1,Y_1,A_0,L_0}}_{\pi_1}\underbrace{\mathrm{P}_{L_1|Y_1,A_0,L_0}}_{H_1}\underbrace{\mathrm{P}_{Y_1|A_0,L_0}}_{F_1}\underbrace{\mathrm{P}_{A_0|L_0}}_{\pi_0}\underbrace{\mathrm{P}_{L_0}}_{H_0}$$

[13] no censoring, no competing risks, $Y_1 = 1\{T \le t_1\}$, $Y_2 = 1\{T \le t_2\}$

Longitudinal causal inference on discretized time scale

Uncensored data, two intervals [14]

$$X = (L_0, A_0, Y_1, L_1, A_1, Y_2)$$

Observed likelihood

$$\mathrm{P}X = \underbrace{\mathrm{P}_{Y_2|A_1,L_1,Y_1,A_0,L_0}}_{F_2}\underbrace{\mathrm{P}_{A_1|L_1,Y_1,A_0,L_0}}_{\pi_1}\underbrace{\mathrm{P}_{L_1|Y_1,A_0,L_0}}_{H_1}\underbrace{\mathrm{P}_{Y_1|A_0,L_0}}_{F_1}\underbrace{\mathrm{P}_{A_0|L_0}}_{\pi_0}\underbrace{\mathrm{P}_{L_0}}_{H_0}$$

Likelihood in the target trial

$$\mathrm{P}^* = F_2\,\pi_1^*\,H_1\,F_1\,\pi_0^*\,H_0$$

[14] Changes of A(t) and L(t) in last interval $(t_{1,t2}]$ are ignored.

Robins g-methods (continued)

Estimator 1: g-formula

$$\hat{\mathrm{P}}^*_g = \hat{F}_2 \pi^*_1 \hat{H}_1 \hat{F}_1 \pi^*_0 \hat{H}_0$$

Estimator 2: Inverse probability weighting

$$\hat{\mathrm{P}}^*_{\mathsf{IPTW}} = \frac{\mathbb{P}_n \, \pi^*_0 \pi^*_1}{\hat{\pi}_0 \hat{\pi}_1}$$

Robins g-methods (continued)

Estimator 1: g-formula

$$\hat{\mathrm{P}}^*_g = \hat{F}_2 \pi^*_1 \hat{H}_1 \hat{F}_1 \pi^*_0 \hat{H}_0$$

Estimator 2: Inverse probability weighting

$$\hat{\mathrm{P}}^*_{\mathrm{IPTW}} = \frac{\mathbb{P}_n \pi^*_0 \pi^*_1}{\hat{\pi}_0 \hat{\pi}_1}$$

Estimator 3: Iterative conditional expectations AKA Sequential regression

$$\begin{aligned}
\mathrm{E}[Y_2] &= \mathrm{E}[\mathrm{E}[Y_2|L_0]] \\
&= \mathrm{E}[\mathrm{E}[\mathrm{E}[Y_2|L_0, A_0]|L_0]] \\
&= \mathrm{E}[\mathrm{E}[\mathrm{E}[\mathrm{E}[Y_2|L_0, A_0, L_1]|L_0, A_0]|L_0]] \\
&= \mathrm{E}[\mathrm{E}[\mathrm{E}[\mathrm{E}[\mathrm{E}[Y_2|L_0, A_0, L_1, A_1]|L_0, A_0, L_1]|L_0, A_0]|L_0]]
\end{aligned}$$

Iterative conditional expectations: discretized time

Estimator 3: Robins (1999), Bang & Robins (2005)

$$\mathrm{E}[Y_2] = \mathrm{E}[\underbrace{\mathrm{E}[\mathrm{E}[\mathrm{E}[\underbrace{\mathrm{E}[Y_2|L_0, A_0, L_1, A_1]}_{Q_2(L_0,A_0,L_1,A_1)}|L_0, A_0, L_1]|L_0, A_0]|L_0]}_{Q_1(L_0,A_0)}]$$

Step 1 $\hat{Q}_2$: Regress Y_2 on L_0, A_0, L_1, A_1

Step 2 Integrate $\hat{Q}_2$ with respect to π_1^*

Step 3 $\hat{Q}_1$: Regress result of Step 2 on L_0, A_0, L_1, A_1

Step 4 Integrate $\hat{Q}_1$ with respect to π_0^*

Step 5 Average with respect to $\hat{H}_0$

Longitudinal targeted minimum loss based estimator

A motivation for the roadmap of targeted learning is the problem that the various nuisance parameter regression models could be misspecified.

The targeted minimum loss based estimator can be consistent even if some of the nuisance parameter models are misspecified.

For longitudinal data analysis, the sequential regression estimator is moved closer to the unknown truth by sequential updating with

"clever covariates"

which depend on the inverse propensity of treatment weights.[6]

Longitudinal targeted minimum loss based estimator

Under the usual causal assumptions and if the convergence rate of the estimators of the nuisance parameters is sufficiently fast we can estimate the target parameter:

$$\psi : \mathcal{M} \mapsto \mathbb{R},$$

where ψ is a suitably smooth functional defined on a set of probability measures, at the $\sqrt{n}$-rate

$$\sqrt{n}(\psi(\hat{\mathrm{P}}^*_{\mathsf{LTMLE}}) - \psi(\mathrm{P})) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{IF}_{\mathrm{P}}(X_i) + o_{\mathrm{P}}(1)$$

Longitudinal targeted minimum loss based estimator

The efficient influence function for the target parameter (two intervals):

$$\mathrm{IF}_{\mathrm{P}}(X) = (Y_2 - Q_2)\frac{\pi_1^*\pi_0^*}{\pi_1\pi_0} + (Q_2 - Q_1)\frac{\pi_0^*}{\pi_0} + Q_1 - \psi(\mathrm{P})$$

Targeting algorithm:

- Step 1 Initial estimators: $\hat{\pi}_1, \hat{\pi}_0$
- Step 2 $\hat{Q}_2$: Regress Y_2 on L_0, A_0, L_1, A_1
- Step 3 $\hat{Q}_2^*$: TMLE update: Loss function and parametric fluctuation model to solve the current part of the efficient influence function
- Step 4 Integrate $\hat{Q}_2^*$ with respect to π_1^*
- Step 5 Regress result of Step 4 on A_0, L_0
- Step 6 ...

Censoring and competing risks in LTMLE

The order of the data in wide format with censoring and competing risks:

$$X = (L_0, A_0, C_1, Y_1, D_1, L_1, A_1, C_2, Y_2, D_2, \ldots, Y_K)$$

- probabilities are modelled conditional on the history (variables to the left). We denote by G_k the conditional distribution of being uncensored.
- competing risk probabilities are not modelled but D_k is used to remove patients from being at risk. (However, the presence of competing risks is important for the clinical interpretation!)
- when outcome event Y_k and censoring C_k occur in same interval, then evidently outcome is not censored!

LTMLE algorithm

1. Prepare algorithm by estimating all propensity score and censoring models conditional on current history.
2. Fit model for outcome at last time point conditional on past
3. Get targeted predictions for previous time point via TMLE update step with fluctuation model and clever covariates:

$$\prod_{k=0}^{K} \frac{\pi_k^* 1\{C_k = 0\}}{\hat{\pi}_k \hat{G}_k}$$

4. Fit model with targeted predictions as outcome conditional on past ...

LTMLE algorithm

1. Prepare algorithm by estimating all propensity score and censoring models conditional on past.
2. Fit model for outcome at last time point conditional on past
3. Get targeted predictions for previous time point via TMLE update step with fluctuation model and clever covariates:

$$\texttt{gbounds:} \prod_{k=0}^{K} \frac{\pi_k^* 1\{C_k = 0\}}{\max(0.01, \hat{\pi}_k \hat{G}_k)}$$

1. Fit model with targeted predictions as outcome conditional on past ...