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

Real world data

Treatment plan/protocol/regime

Intention-to-treat

Per-protocol

Adherence/ compliance

Terminology check II

Target parameter

Nuisance parameter

Propensity of treatment

Censoring

Competing risk

Efficient 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] Aguayo-Orozco et al. NPJ digital medicine, 4(1):150, 2021.
[3] Alzheimer's and Dementia, 20(12):8661–8672, 2024.

Why not just "adjust for confounders"?

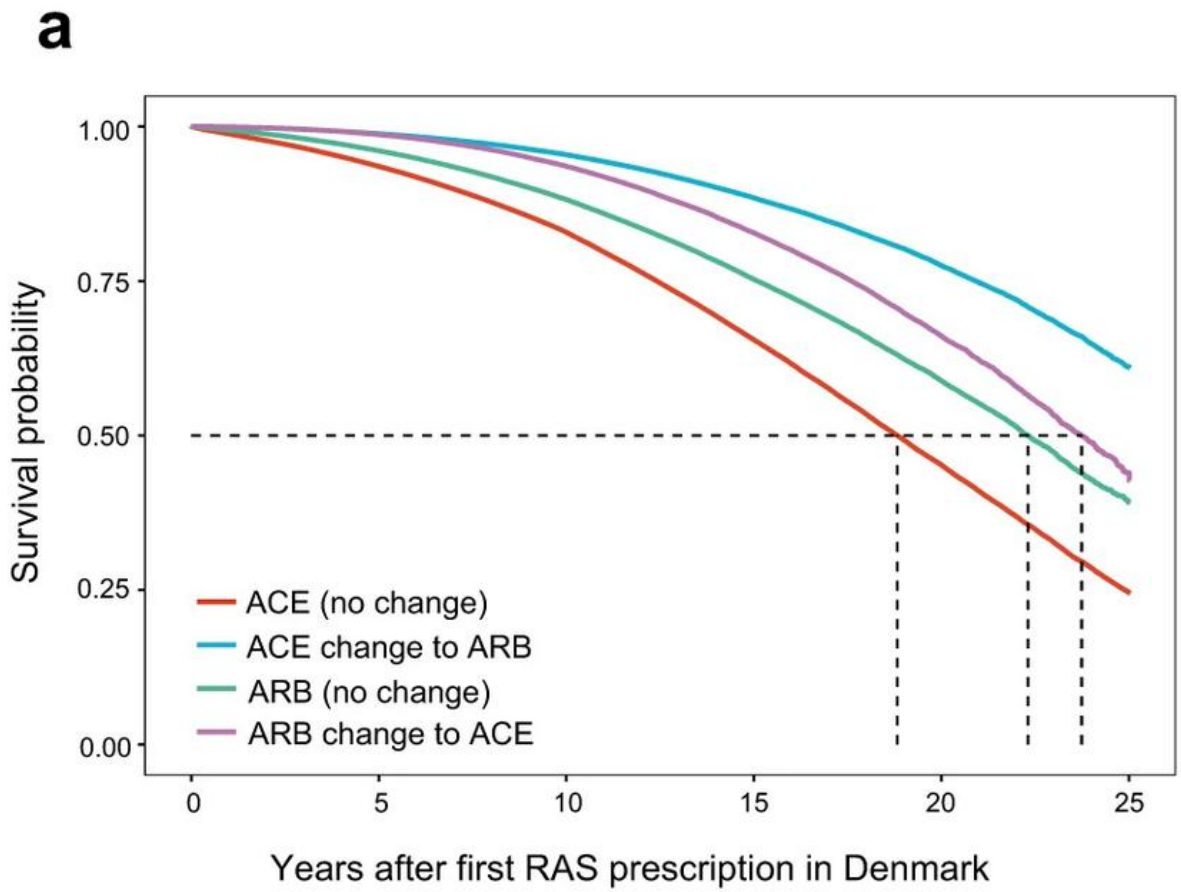
Standard regression analysis should not be applied in longitudinal settings:

- Odds and hazard ratios have unclear clinical interpretation (despite their popularity!)
- 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).

To be avoided: immortal time bias

From: Optimizing drug selection from a prescription trajectory of one patient



Number at risk

ACE (no change)	549,436	410,475	250,056	92,105	31,282	841
ACE change to ARB	229,216	200,730	136,798	54,393	19,223	493
ARB (no change)	287,488	169,789	82,178	43,741	15,545	190
ARB change to ACE	48,054	43,710	34,552	19,865	6,659	46

A wrong conclusion which could have been avoided

"Prescription trajectories can provide novel insights into how individuals' drug use change over time, identify suboptimal or futile prescriptions and suggest initial treatments different from first line therapies. Observations of this kind may also be important when updating treatment guidelines."[2]

A wrong conclusion which could have been avoided

"Prescription trajectories can provide novel insights into how individuals' drug use change over time, identify suboptimal or futile prescriptions and suggest initial treatments different from first line therapies. Observations of this kind may also be important when updating treatment guidelines."[2]

No question was asked!

The result was obtained by data mining!?

The Cox regression model with time-dependent covariates

The outcome-specific hazard rate at time t depends on the treatment and covariate history:

$$\lambda(t|\bar{A}(t), \bar{L}(t)) = \lambda_0(t) \exp\{\beta \bar{A}(t) + \gamma \bar{L}(t)\}$$

The hazard ratio e^{β} can be estimated via maximum partial likelihood, Poisson regression or conditional logistic regression (nested case control design).

The Cox regression model with time-dependent covariates

The outcome-specific hazard rate at time t depends on the treatment and covariate history:

$$\lambda(t|\bar{A}(t), \bar{L}(t)) = \lambda_0(t) \exp\{\beta \bar{A}(t) + \gamma \bar{L}(t)\}$$

The hazard ratio e^{β} can be estimated via maximum partial likelihood, Poisson regression or conditional logistic regression (nested case control design).

Can anyone ask a meaningful question such that e^{β} is the ~~answer~~ 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

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 protocols:
  - Protocol 1: start beta-blockers at discharge and continue for 3 years
  - Protocol 2: no beta-blockers at discharge and do not initiate beta-blockers for 3 years
- 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 people with contraindications (as feasible)
- Treatment protocols:
  - Protocol 1: start beta-blockers at discharge and continue for 3 years
  - Protocol 2: no beta-blockers at discharge and do not initiate beta-blockers for 3 years
- Assignment: not randomized
- Start of follow-up: 30 days after the discharge date
- Outcomes: death and recurrent MI from registries (composite and separate)
- Causal contrast: 5-year risk difference
- Estimation: g-methods to account for confounding

Note that

The target trial should be biologically and physically feasible [4]

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 real world data because:

- there was no protocolled intention to treat
- reasons for deviating from a treatment are very different in RCT vs RWD

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

[4] E.g., cannot randomize people to have a 30% reduced BMI

Active comparator, new user study design

In register studies there is often no natural control group because treatment indication information is poorly registered.

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

Active comparator, new user study design

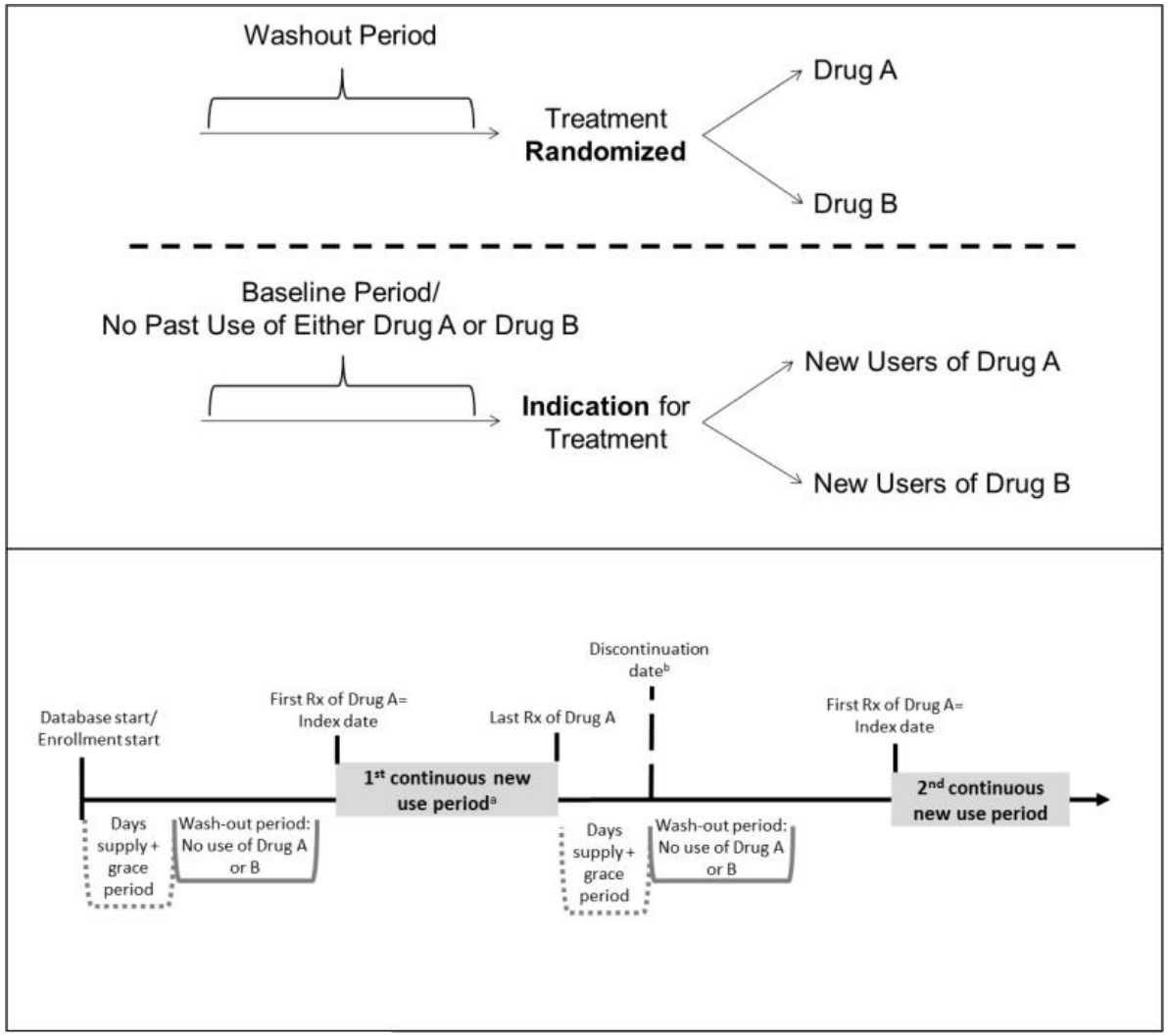
In register studies there is often no natural control group because treatment indication information is poorly registered.

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

Possibilities:

- 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 [5]



[5] 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.

Censoring means *do not observe*.

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!

~~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).

A research report may state:

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

A better way to formulate this is to adapt the description of the intervention/treatment protocol[6]:

Use drug A continously without breaks, do not add drug B (for t-years).

---

[6] We use Robins' G-methods to estimate outcome risks if (hypothetically) all people had followed the protocol.

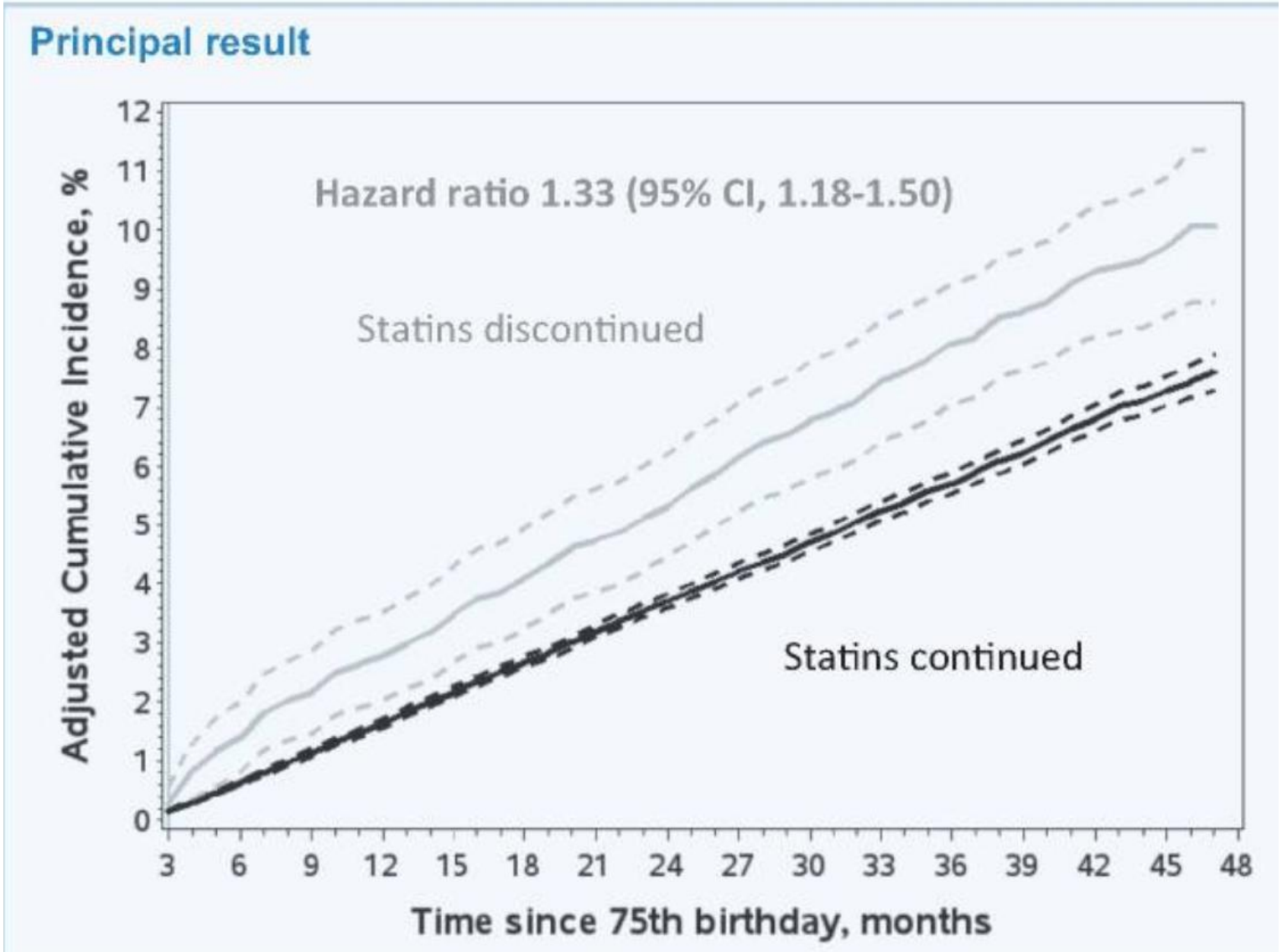
Example: Giral et al (2019)

Cardiovascular effect of discontinuing statins for primary prevention at the age of 75 years: a nationwide population-based cohort study in France [7]

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.

[7] 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

Statistical methods (statins discontinuation paper)

A weighted Cox proportional hazard model was used to estimate the hazard ratio of statin discontinuation vs continuation, controlling for baseline and time-varying confounding. Weighting creates a pseudo-population, in which at each time the probability of being treated is unrelated to [redacted] time-varying confounders. Also, compared conventional multivariate Cox models, weighted Cox proportional hazard models are more flexible in that they do not rely on the assumptions of proportional hazards and no [redacted] In order to [redacted] potentially informative censoring at treatment resumption, the patient's initial weight was multiplied by the inverse probability of remaining uncensored up until month t, based on the patient's covariate history. The resulting weights are called "inverse probability of treatment and censoring weights" (IPTCW). Under the assumptions of no unmeasured confounding, positivity, correct model specification, and consistency, this approach estimates the parameter of a marginal structural model, which can be interpreted as average causal effect of treatment

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

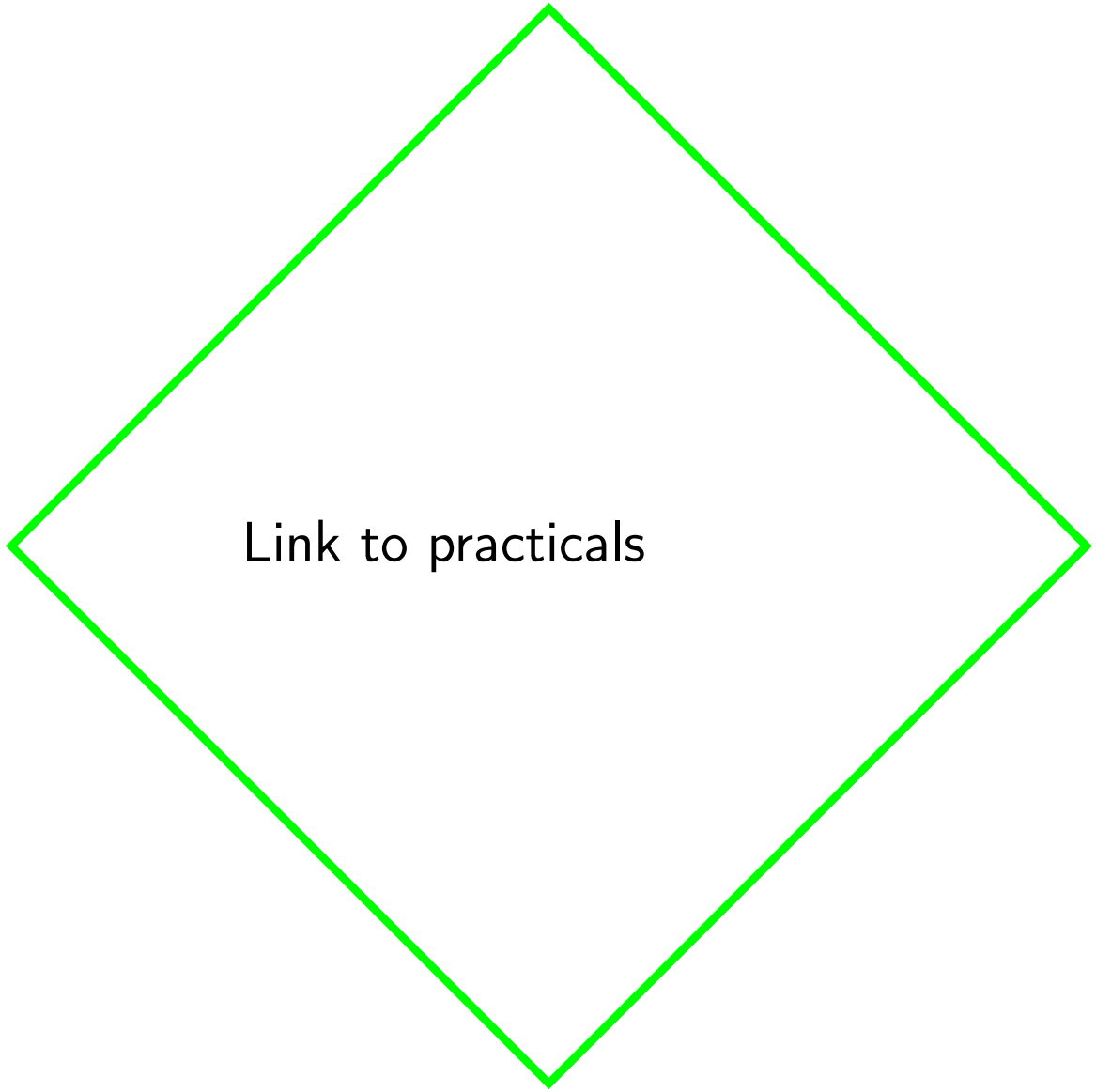
Avoidable flaws in observational analyses[8]

The increasing availability of large healthcare databases is fueling an intense debate on whether real-world data should play a role in the assessment of the benefit-risk of medical treatments.

In many observational studies, for example, statin users were found to have a substantially lower risk of cancer than in meta-analyses of randomized trials. While such discrepancies are often attributed to a lack of randomization in the observational studies, they may be explained by flaws that can be avoided by

explicitly emulating a target trial.

[8] Dickerman, . . . , Hernan. Nature medicine, 25(10):1601–1606, 2019.


Link to practicals

Roadmap for targeted learning (on a discretized time-scale)

Roadmap for targeted learning [9]

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

[9] van der Laan and Rose. Targeted learning in data science. Springer, 2018.

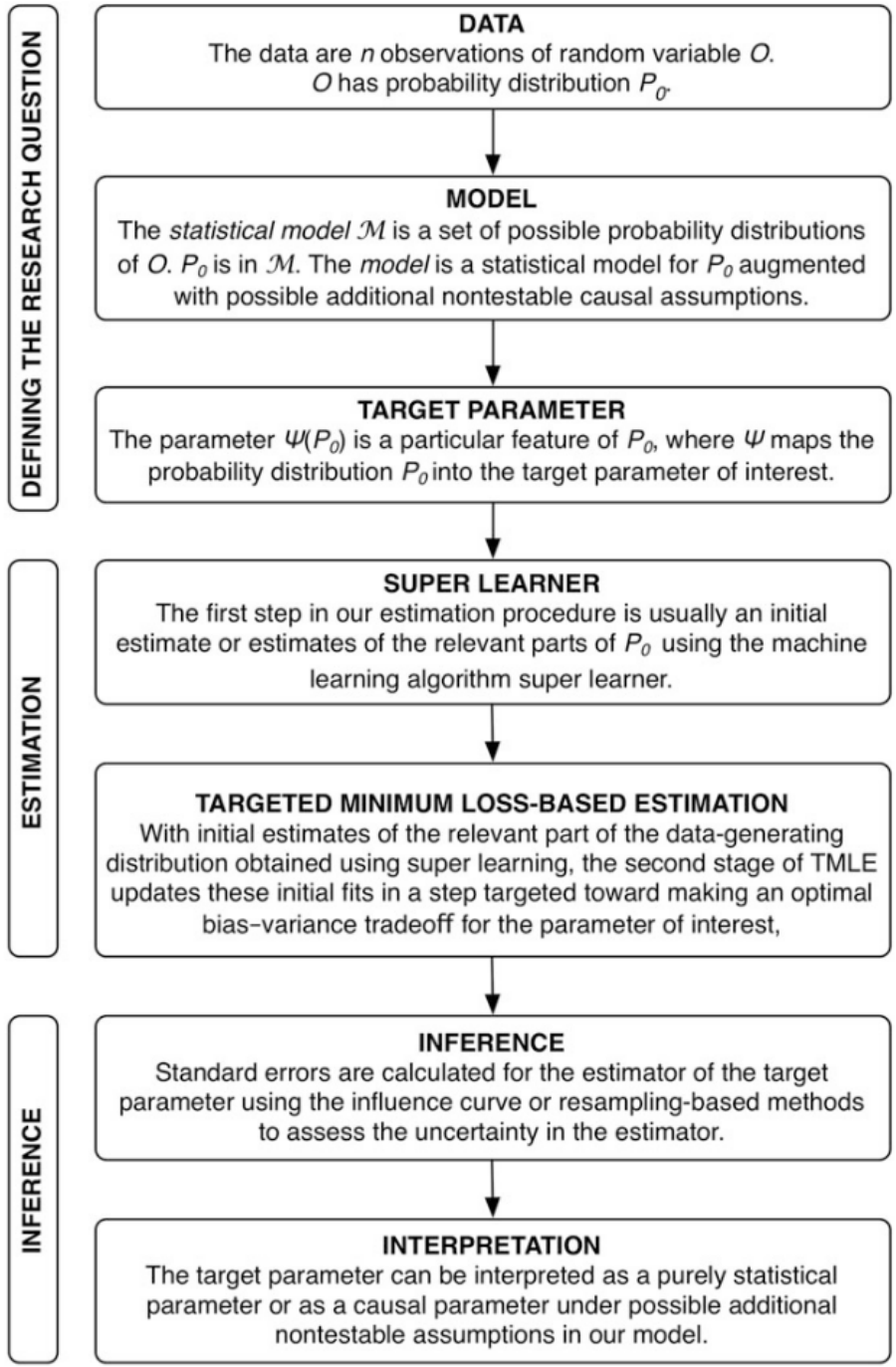


Fig. 1.1 Road map for targeted learning

Register data for an emulated target trial may look like this:

id	sex	age at enrollment	date of enrollment
1	woman	74	2019-07-02
2	man	69	2016-01-13
3	woman	88	2021-11-27

Treatment:

id	date	drug
1	2019-07-02	GLP1-RA
1	2019-10-08	GLP1-RA
1	2021-01-07	GLP1-RA
2	2016-01-13	SGLT2i
2	2016-08-27	SGLT2i
3	2021-11-28	GLP1-RA
3	2021-12-03	SGLT2i

Hospital diagnoses:

id	hospital admission date	diagnosis
2	2018-02-15	hypertension
2	2018-09-03	hypertension
3	2022-01-02	arterial fibrillation

Concomitant medical treatment:

id	redemption date	drug
2	2018-12-24	ACE/ARP
3	2022-02-11	NOAC
3	2022-08-10	NOAC

Outcome

id	cardiovascular disease	date of death	end of followup
1	NA	NA	2024-09-17
2	NA	2019-03-03	2024-09-17
3	2023-02-07	2023-02-07	2024-09-17

new R-package 'rtmle' (work in progress)

```
library(rtmle)
library(data.table)
x <- rtmle_init(intervals=6,
                name_id="id",
                name_outcome="cvd_death",
                name_competing="Dead",
                name_censoring="Censored",
                censored_label="censored")
D <- function(...){as.Date(c(...))}
x <- add_baseline_data(x,data.table(id=1:3,
                                    sex=c("woman","man","woman"),
                                    age=c(74,69,88),
                                    start_followup_date=D("2019-07-02",
                                                          "2016-01-13",
                                                          "2021-11-27")))
treatment_data <- list("GLP1RA"=data.table(id=c(1,1,1,2),
                                           date=D("2019-07-02",
                                                  "2019-10-08",
                                                  "2021-01-07",
                                                  "2021-11-28")),
                       "SGLT2i"=data.table(id=c(2,3,3),
                                           date=D("2016-01-13",
                                                  "2016-08-27",
                                                  "2021-12-03")))
timevar_data <- list("hypertension"=data.table(id=c(2,2),
                                               date=D("2018-02-15",
                                                      "2018-09-03")),
                     "af"=data.table(id=3,date=D("2022-01-02")))
```

new R-package 'rtmle' (work in progress)

```
outcome_data <- data.table(id = 1,
                           date = D("2023-02-07"))
censored_data <- data.table(id = c(1,2,3),
                            date = rep(D("2024-09-17"),3))
competing_data <- data.table(id = c(2,3),
                             date = D("2019-03-03",
                                      "2023-02-07"))
x <- add_long_data(x,
                   outcome_data,
                   censored_data,
                   competing_data,
                   c(timevar_data,treatment_data))
# 3 years followup discretized with six 6 months long intervals
x <- long_to_wide(x,intervals = seq(0,30.45*36,30.45*6),"start_followup_
    date")
# protocol for emulated trial
x <- protocol(x,name = "always GLP1RA never SGLT2i",
              treatment_variables = c("GLP1RA","SGLT2i"),
              intervention = c(1,0))
x <- target(x,name = "Outcome_risk",
            estimator = "tmle",
            protocols = "always GLP1RA never SGLT2i")
x <- prepare_data(x, intervals = seq(0,30.45*36,30.45*6))
# data in wide format
x$prepared_data
```

The register data forced onto a discretized time scale

```
       id    sex    age hypertension_0  af_0 GLP1RA_0 SGLT2i_0 Censored_1 cvd_death_1 Dead_
   <char> <char> <num>          <num> <num>    <num>    <num>     <fctr>       <num>  <num
1:      1  woman    74              0     0        1        0 uncensored           0
2:      2    man    69              0     0        0        1 uncensored           0
3:      3  woman    88              0     0        0        0 uncensored           0
   hypertension_1  af_1 GLP1RA_1 SGLT2i_1 Censored_2 cvd_death_2 Dead_2 hypertension_2  af
            <num> <num>    <num>    <num>     <fctr>       <num>  <num>          <num> <nu
1:              0     0        1        0 uncensored           0      0              0
2:              0     0        0        0 uncensored           0      0              0
3:              0     1        0        1 uncensored           0      0              0
   GLP1RA_2 SGLT2i_2 Censored_3 cvd_death_3 Dead_3 hypertension_3  af_3 GLP1RA_3 SGLT2i_3
      <num>    <num>     <fctr>       <num>  <num>          <num> <num>    <num>    <num>
1:        0        0 uncensored           0      0              0     0        0        0
2:        0        0 uncensored           0      0              0     0        0        0
3:        0        0 uncensored           0      1             NA    NA       NA       NA
   Censored_4 cvd_death_4 Dead_4 hypertension_4  af_4 GLP1RA_4 SGLT2i_4 Censored_5 cvd_dea
       <fctr>       <num>  <num>          <num> <num>    <num>    <num>     <fctr>       <
1: uncensored           0      0              0     0        1        0 uncensored
2: uncensored           0      0              0     0        0        0 uncensored
3:       <NA>           0     NA             NA    NA       NA       NA       <NA>
   Dead_5 hypertension_5  af_5 GLP1RA_5 SGLT2i_5 Censored_6 cvd_death_6
    <num>          <num> <num>    <num>    <num>     <fctr>       <num>
1:     0              0     0        1        0 uncensored           0
2:     0              1     0        0        0 uncensored           0
3:    NA             NA    NA       NA       NA       <NA>           0
```

Introduction to the notation

Uncensored data, one interval $[0, t_1]$ [10]

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

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

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

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

Longitudinal setting for the observational data [11]

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:

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

[11] 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 [11]

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:

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

[11] 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 [12]

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

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

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

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

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	0
Always treat	Static	1
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:

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

Causible parameters

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

Hazard ratios are not causible. [13]

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.

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

The target parameter (aka the estimand)

The analysis of the emulated target trial aims to estimate the absolute risks of the outcome(s) if hypothetically all patients had followed the treatment protocols (per-protocol effects [14]).

Example: 2-year risk of cardiovascular disease under π^{j*} and differences thereof:

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

where $\mathrm{P}_{\pi^{A*}} = F_2 \pi_1^{A*} H_1 F_1 \pi_0^{A*} H_0$

[14] Not to be confused with adhoc "per-protocol analyses" of randomized trials which include only patients that complied their treatment protocol

The target parameter (aka the estimand)

The analysis of the emulated target trial aims to estimate the absolute risks of the outcome(s) if hypothetically all patients had followed the treatment protocols (per-protocol effects [14]).

Example: 2-year risk of cardiovascular disease under π^{j*} and differences thereof:

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

where $\mathrm{P}_{\pi^{A*}} = F_2 \pi_1^{A*} H_1 F_1 \pi_0^{A*} H_0$

This parameter is causible! But, is it identifiable from the observational data?

[14] Not to be confused with adhoc "per-protocol analyses" of randomized trials which include only patients that complied their treatment protocol

Causal assumptions

At any time during the followup period:

positivity there is a positive probability to comply with the treatment protocol (uncensored) for all people in the population

sequential coarsening at random (AKA no unmeasured confounders) the probability of deviation from treatment and the probability of being uncensored depend only on the observed history

consistency treatment definition is not ambiguous (e.g., hormon therapy vs standard of care), no interference (what happens to one person does affect the outcome of another person), treatment versions have the same effect (e.g., infusion vs pill)

Identifying equation: g-formula

The target parameter is the expectation of the outcome at the last time point under the discretized-time probability distribution of the target trial:

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

To identify the target parameter under the causal assumptions we need consistent estimators of the nuisance parameter models:

$$\begin{aligned} F_2 &= \mathrm{P}(Y_2 = 1 | \text{history}) \\ H_1 &= \mathrm{P}(L_1 \in \mathrm{d}\ell | \text{history}) \\ F_1 &= \mathrm{P}(Y_1 = 1 | \text{history}) \\ H_0 &= \mathrm{P}(L_0 \in \mathrm{d}\ell | \text{history}) \end{aligned}$$

Note that π^* is known and does not have to be estimated.

Identifying equation: IPW

The observed data likelihood can be written as:

$$\begin{aligned} \mathrm{P}_X &= F_2 \pi_1 H_1 F_1 \pi_0 H_0 \\ &= \frac{\pi_0}{\pi_0^*} \frac{\pi_1}{\pi_1^*} F_2 \pi_1^* H_1 F_1 \pi_0^* H_0 \end{aligned}$$

and hence

$$\begin{aligned} \mathrm{P}_{\pi^*} &= F_2 \pi_1^* H_1 F_1 \pi_0^* H_0 \\ &= \frac{\pi_0^*}{\pi_0} \frac{\pi_1^*}{\pi_1} \mathrm{P}_X \end{aligned}$$

Identifying equation: IPW

The observed data likelihood can be written as:

$$\begin{aligned} \mathrm{P}_X &= F_2 \pi_1 H_1 F_1 \pi_0 H_0 \\ &= \frac{\pi_0}{\pi_0^*} \frac{\pi_1}{\pi_1^*} F_2 \pi_1^* H_1 F_1 \pi_0^* H_0 \end{aligned}$$

and hence

$$\begin{aligned} \mathrm{P}_{\pi^*} &= F_2 \pi_1^* H_1 F_1 \pi_0^* H_0 \\ &= \frac{\pi_0^*}{\pi_0} \frac{\pi_1^*}{\pi_1} \mathrm{P}_X \end{aligned}$$

We need consistent estimators of the nuisance parameter models:

$$\begin{aligned} \pi_1 &= \mathrm{P}(A_1 = a | \text{history}) \\ \pi_0 &= \mathrm{P}(A_0 = a | \text{history}) \end{aligned}$$

Roadmap philosophy

Inference is based on the efficient influence function for which we need estimators of all nuisance parameter models (doubly robust style).

In order to obtain (asymptotically) valid inference for the target parameter we need to pre-specify estimators for all nuisance parameter models.

Method 1 use subject matter knowledge to specify a single (semi-)parametric model for each nuisance parameter

Method 2 specify multiple models [15] and estimate via cross-validation (cross-fitting)

[15] which differ in number of predictors, non-linear effects, interactions and perhaps also some machine learning algorithms

Roadmap summary

Step 1: We say what we estimate: target parameter / estimand

Step 2: We explain how we would interprete an estimate of the target parameter if the causal assumptions were satified

Step 3: We derive an estimator which depends on nuisance parameter models

Step 4: We pre-specify estimators for the nuisance parameter models

Step 5: We estimate the target parameter

Step 6: We estimate standard errors based on an estimate of the efficient influence function or with the bootstrap

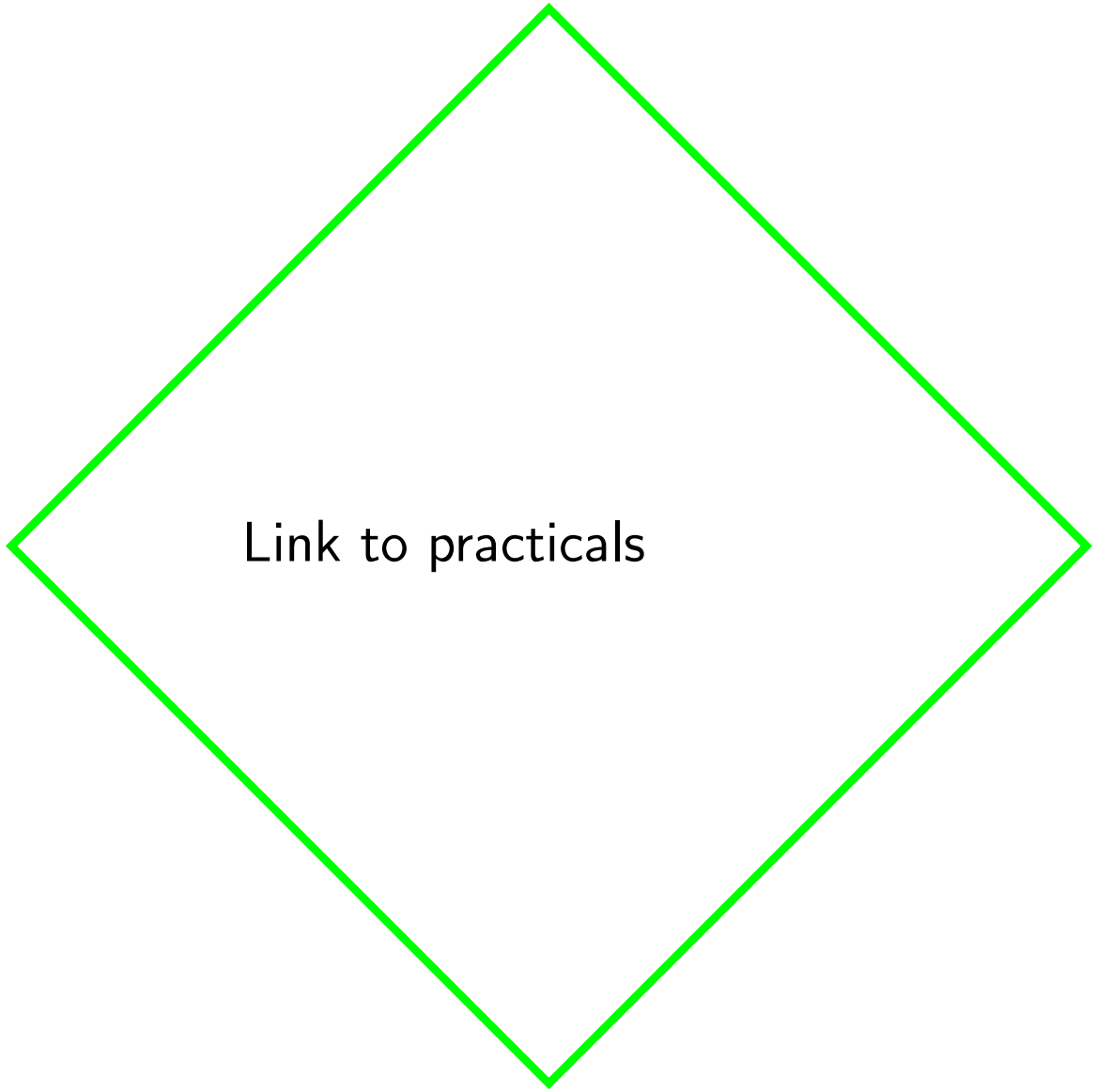
Roadmap left-overs (incomplete)

We need to map the observed data (drug exposure, covariates, outcome, competing risks) onto a discretized time scale. For this we need to choose a discrete time scale.

We need to achieve a feel-good balance for the causal assumptions by adjusting the

- the inclusion criteria of the emulated target trial
- the calendar time period
- the ambitions of the subject matter researchers

We need to test if the estimators are sensitive to variations of the random seed


Link to practicals

LTMLE

LTMLE just after lunch

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 [16] and the LTMLE estimator of van der Laan [17]

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

[17] 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[18]

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

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

Estimator 2: Inverse probability weighting

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

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

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

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

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

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

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.[19]

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

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:

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

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

Statins discontinuation paper

Table S6. Sensitivity analysis: Impact of weight truncation on the estimated effect

Truncation percentiles 1-p, p[a]	HR of statin discontinuation vs continuation (95% CI)	Estimated weights	
		Mean (SD)	Minimum/ maximum
0, 100[b]	1.43 (1.23-1.66)	1.01 (0.30)	0.01/77.63
0.1, 99.9	1.37 (1.21-1.56)	1.00 (0.19)	0.07/3.65
0.5, 99.5	1.33 (1.18-1.50)	1.00 (0.15)	0.23/2.14
1, 99	1.31 (1.17-1.46)	1.00 (0.12)	0.43/1.66
5, 95	1.28 (1.15-1.42)	1.00 (0.04)	0.91/1.11
10, 90	1.27 (1.15-1.41)	0.99 (0.03)	0.94/1.05
25, 75	1.27 (1.14-1.41)	0.99 (0.02)	0.97/1.01
50, 50[c]	1.26 (1.14-1.40)	1.00 (0.00)	1.00/1.00

- [a] Weights were truncated by resetting the value of the weights greater (less) than the p (1-p) percentile to the value of the p (1-p) percentile.
- [b] No weight truncation.
- [c] Constant weights, i.e. no adjustment.

From the statins discontinuation paper . . .

This table illustrates the tradeoff between bias and variance: with increasing weight truncation, confidence intervals were shorter, but bias increased

Assuming estimation without weight truncation is unbiased.

In this case, it can be reasonably argued to report the result with the weights truncated at the 0.5% and 99.5% percentiles, on the basis of centering of the weights at a value of 1 and the reduction in the 1/minimum and maximum weights.

The gbounds issue

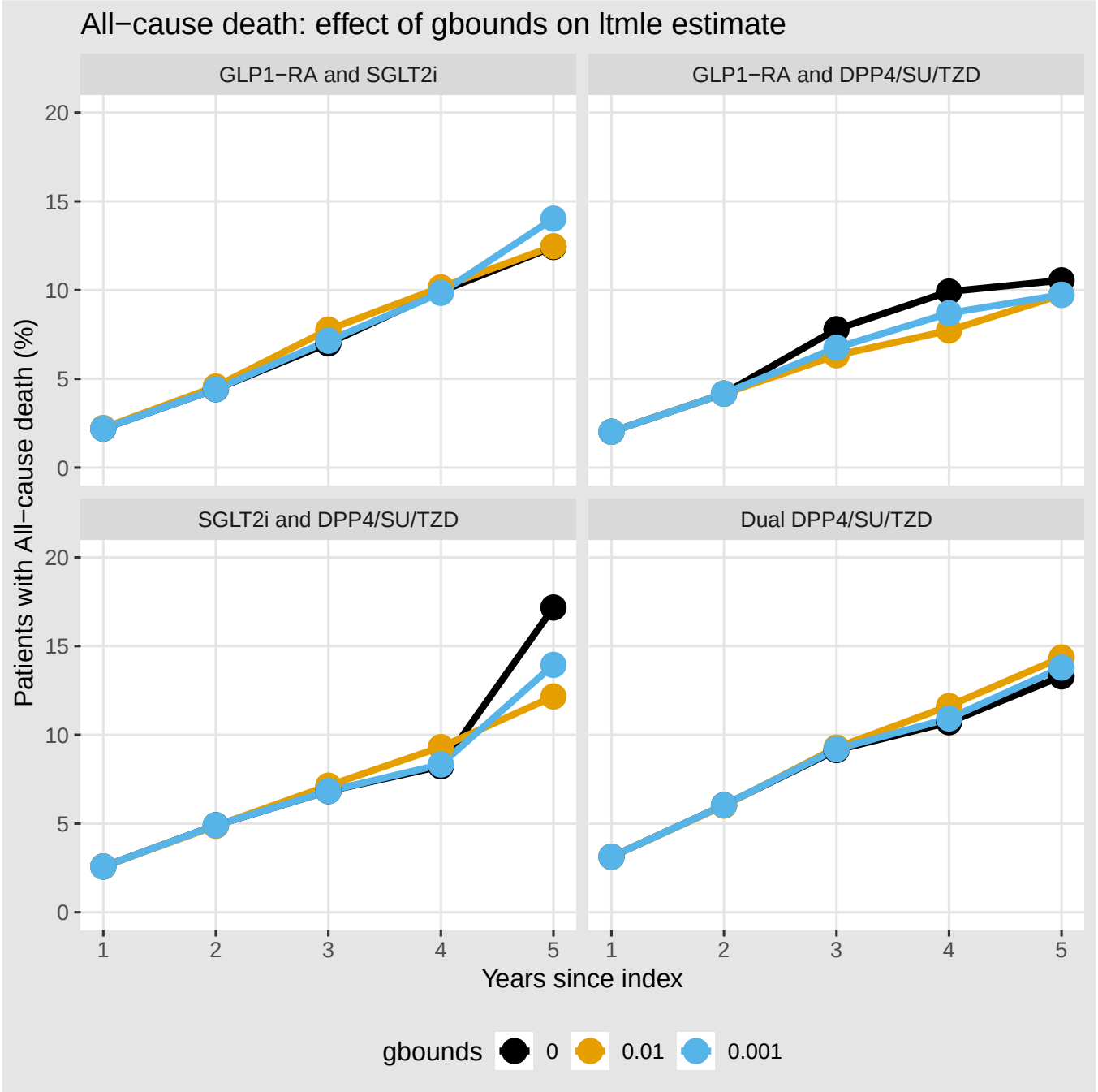
The default behavior of the R-package `ltmle` makes sure that (the product of) the estimated propensity scores and censoring probabilities cannot become smaller than (by default)

$$\texttt{gbounds} = 0.01.$$

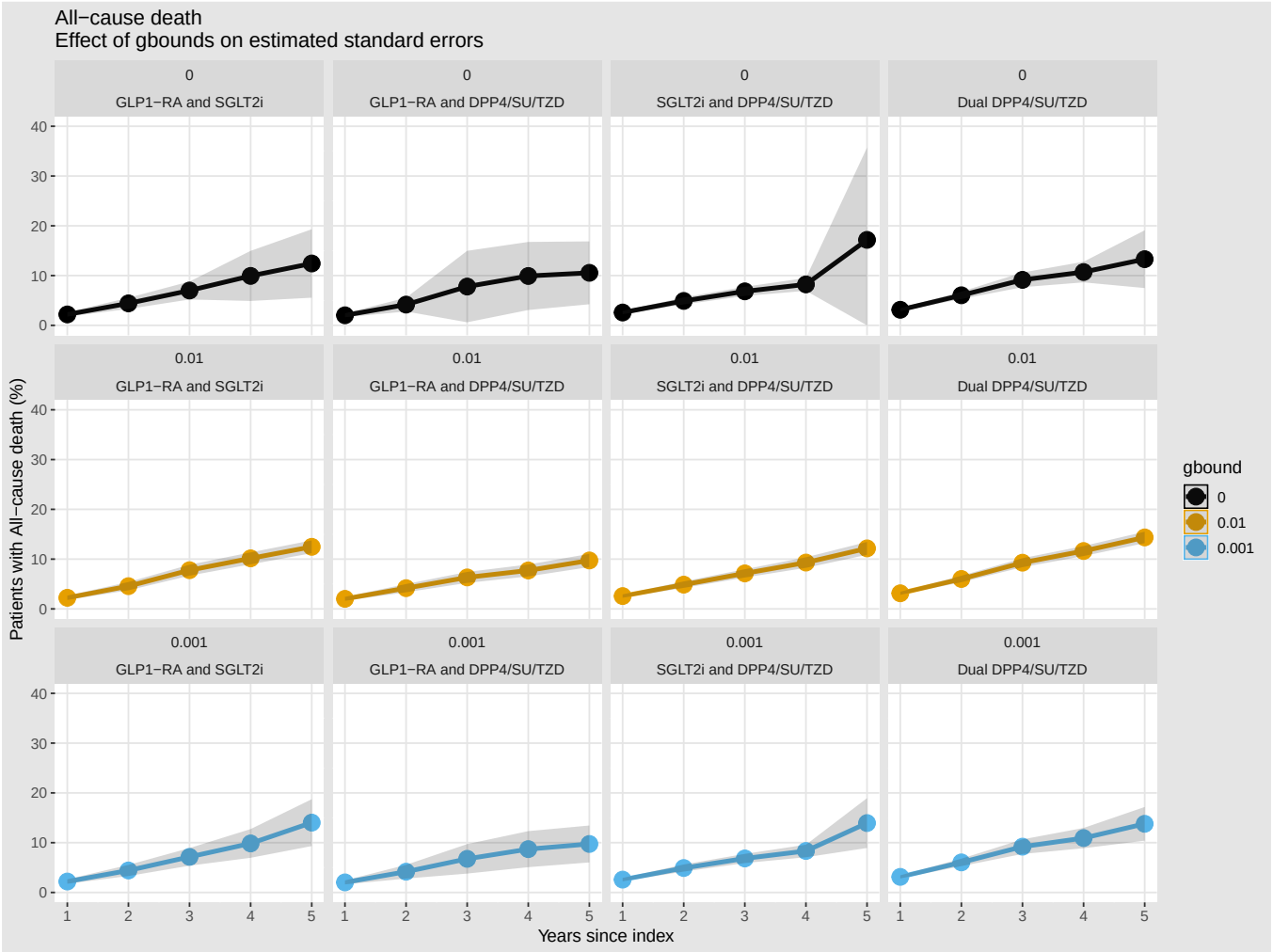
There is no good reason for us to use that particular value, or any other value, hence we subsequently set it to `0`.

The `gbounds` argument affects both the point estimate via the clever weights of the tmle-update step BUT even more the influence function based plug-in estimate of the standard errors.

Effect of gbounds on LTMLE point estimate



Effect of gbounds on standard errors (via influence function)



Cheap subsampling bootstrap confidence interval[20]

1. Draw without replacement $m < n$ times from the observations to obtain a subsample $\mathcal{D}_m^*$. Estimate $\Psi(P)$ on this subsample to obtain $\hat{\Psi}_m^*$ of $\Psi(P)$. Repeat this $B \geq 1$ times to obtain the resamples $\hat{\Psi}_{(m,1)}^*, \ldots, \hat{\Psi}_{(m,B)}^*$. Calculate also $\hat{\Psi}_n$ based on the full sample $\mathcal{D}_n$.
2. Calculate $S = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\Psi}_{(m,b)}^* - \hat{\Psi}_n\right)^2}$ and obtain the confidence interval

$$\mathcal{I}_{(m,n)} = \left(\hat{\Psi}_n - t_{B,1-\alpha/2}\sqrt{\frac{m}{n-m}}S, \hat{\Psi}_n + t_{B,1-\alpha/2}\sqrt{\frac{m}{n-m}}S\right), \tag{1}$$

where $t_{B,1-\alpha/2}$ is the $1-\alpha/2$ quantile of a t - distribution with B degrees of freedom.

[20] Work by Johan Ohlendorff (2024) based on the general approach of Henry Lam (2022, A Cheap Bootstrap Method for Fast Inference)