Workflow
Targeted learning; an overview
Framework
Simulation and Case study

Causal Inference and the Hypothetical Estimand in Randomised Controlled Trials

November 26, 2025

Background and motivation

- Presenting completed master's thesis work (defended June)
- Thesis done with research partner in collaboration with a pharmaceutical company
  - Also went to University of California, Berkeley
- Motivation: aim to understand direct treatment effects
- Industry interest due to ICH E9(R1) focus on estimands
- Previously approved methods ignored post-ICE data $\rightarrow$ aim to use all collected data

Contents

Workflow
Targeted learning; an overview
Framework
Simulation and Case study

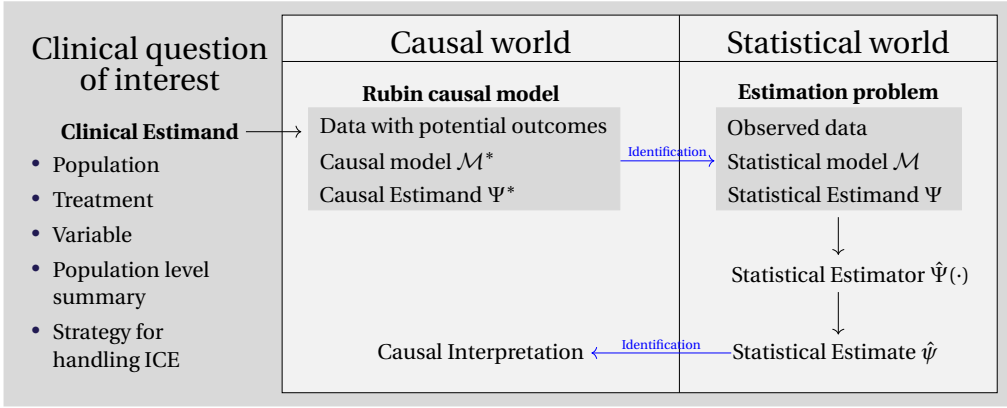
1 Workflow

2 Targeted learning; an overview

3 Framework

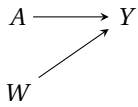
4 Simulation and Case study

Workflow


Clinical question of interest
Clinical Estimand
Population
Treatment
Variable
Population level summary
Strategy for handling ICE
Causal world
Statistical world
Rubin causal model
Data with potential outcomes
Causal model $\mathcal{M}^*$
Causal Estimand Ψ^*
Identification
Estimation problem
Observed data
Statistical model $\mathcal{M}$
Statistical Estimand Ψ
Statistical Estimator $\hat{\Psi}(\cdot)$
Identification
Causal Interpretation
Statistical Estimate $\hat{\psi}$

Notation and setup

- Observed data $O = (W, A, Y)$ from an RCT
- The statistical model $\mathcal{M}$ is all the possible joint probability distributions P of O
- The statistical estimand is a mapping from the model space to $\mathbb{R}$,



$$\Psi : \mathcal{M} \to \mathbb{R}, \quad \text{e.g} \quad \Psi(P) = E_P[E_P[Y \mid W, A = 1]]$$

The joint distribution P has the factorised density

$$p(W, A, Y) = p(W)\, p(A \mid W)\, p(Y \mid A, W)$$

Possibilities now:

- Learn the mechanisms using g-formula, by modelling at all nodes in the DAG
- **Learn, without making parametric assumptions, and update according to our statistical estimand, to obtain an unbiased estimator with the lowest possible variance**

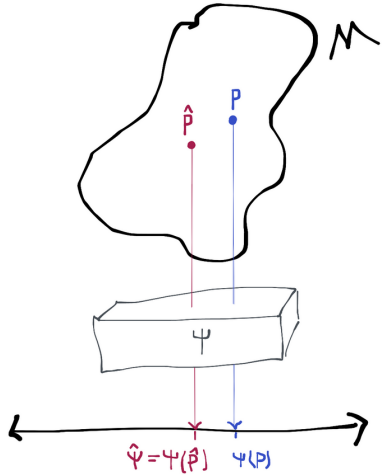
Statistical estimation problem - Initial estimate

Illustrations are borrowed from *Introduction to Modern Causal Inference*, a work-in-progress e-book by Alejandro Schuler and Mark van der Laan

Statistical estimation problem - Initial estimate

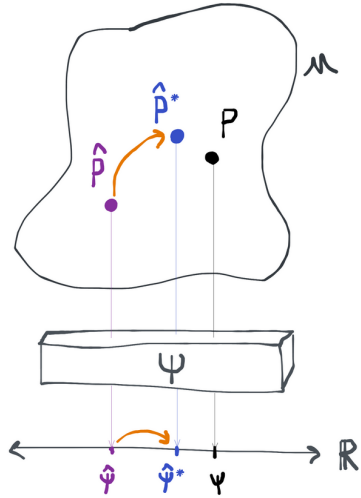
Illustrations are borrowed from *Introduction to Modern Causal Inference*, a work-in-progress e-book by Alejandro Schuler and Mark van der Laan

- $\mathcal{M}$ is the statistical model, containing possible probability distributions
- P is the true underlying distribution
- $\hat{P}$ is the initial estimate of the distribution, e.g. by g-formula or Super Learner
- $\Psi(P) = E_P[E_P[Y \mid W, A = 1]]$ is the statistical estimand
- $\hat{\Psi} = \Psi(\hat{P}) = E_{\hat{P}}[E_{\hat{P}}[Y \mid W, A = 1]]$ is the initial estimate of the statistical estimand



Statistical estimation problem - Updating

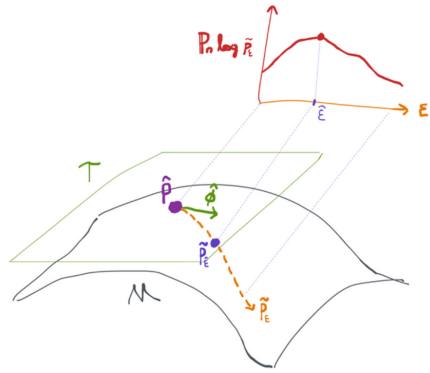
- $\mathcal{M}$ is the statistical model, containing possible probability distributions
- P is the true underlying distribution
- $\hat{P}$ is the initial estimate of the distribution, e.g. by g-formula or Super Learner
- $\Psi(P) = E_P[E_P[Y \mid W, A = 1]]$ is the statistical estimand
- $\hat{P}^*$ is the targeted estimate of the distribution, the way it is updated is tailored to our statistical estimand



Statistical estimation problem - Targeting step

- $\mathcal{M}$ is the statistical model
- $\hat{P}$ is the initial estimate, with density $\hat{p}$
- $\mathcal{T}$ is the tangent space for the statistical model
- $\hat{\phi}$ is the Efficient influence function at $\hat{P}$
- $\{\tilde{P}_\varepsilon \mid \varepsilon \in [0,1]\}$ is a path in the statistical model

First paths, then efficient influence functions

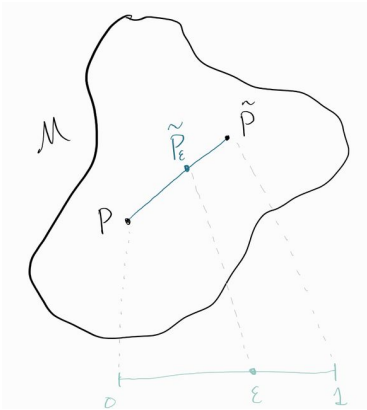


Paths

- A path is a collection of distributions in the statistical model, conveniently parametrised by a one-dimensional parameter.
- If we knew two distributions, P and $\tilde{P}$, in the statistical model, we could use a convex combination as the path; $\tilde{P}_\varepsilon$ is then a distribution with density $\tilde{p}_\varepsilon(o) = \varepsilon\tilde{p}(o) + (1-\varepsilon)p(o)$.

Example

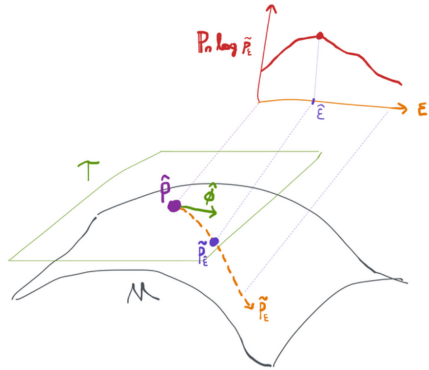
- $P = \mathcal{N}\left(0, \sigma^2\right)$ and $\tilde{P} = \mathcal{N}\left(1, \sigma^2\right)$, where $\mu \in [0, 1]$.



Efficient influence function

- $\hat{\phi}$ is the Efficient influence function at $\hat{P}$
  - Determines the direction of the path in the model (illustrated by orange line)
  - Informs how much each observation deviates from the distribution
  - Determines the variance of the estimator
- Maximising with respect to empirical likelihood along the path with densities

$$\tilde{p}_\varepsilon = (1 + \varepsilon\hat{\phi})\hat{p}$$



Estimation part

Recall the statistical estimand

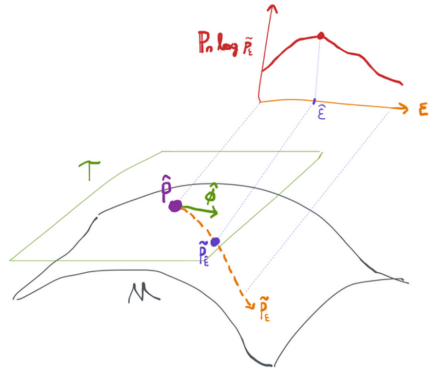
$$\Psi(P) = E[E[Y \mid W, A = 1]]$$

where the corresponding efficient influence function is defined as

$$\phi_P(W, A, Y) = \frac{\mathbb{1}[A = 1]}{\pi(W)}(Y - \mu(W)) + \mu(W) - \Psi(P)$$

with

$$\pi(W) = P(A \mid W)$$
$$\mu(W) = E[Y \mid W, A = 1]$$



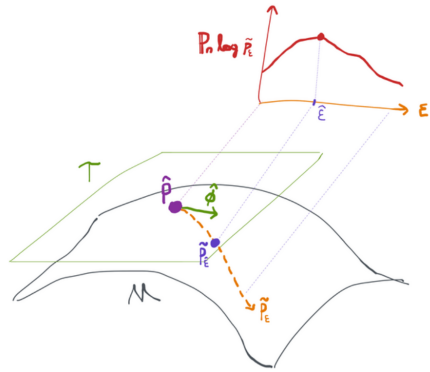
Estimation part

Update along the path $\tilde{p}_\varepsilon = (1 + \varepsilon\hat{\phi})\hat{p}$ determined by the efficient influence $\hat{\phi} = \phi_{\hat{P}}$ to obtain

- A new estimator $\hat{\mu}^*(W) = E_{\hat{P}^*}[Y \mid W, A = 1]$
- Solves the estimating equation $\frac{1}{n}\sum_{i=1}^{n} \hat{\phi}_{P^*}(o_i) = 0$

The estimator

$$\hat{\Psi}_{\text{TMLE}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}^*(w_i)$$



A summary

Targeted learning and TMLE provides

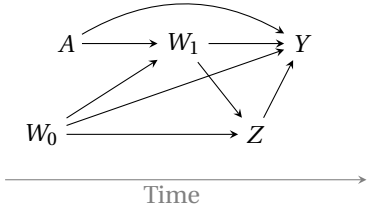
- Unbiased estimators
- Well-performing estimators in small samples
- Possibility to incorporate non-parametric models and machine learning models
- But it requires a lot of prerequisites

Now; The application!

Notation and framework

Randomised trial designed to study the efficacy of semaglutide compared to placebo in patients with type 2 diabetes.

- **Covariates** (W_0, W_1) including HbA1c measurements
- **Treatment** A is binary; $A = 1$ is assignment to treatment, $A = 0$ is placebo
- **Indicator** Z is binary; indicator of initialisation of rescue medication
- **Outcome** Y is measurement of HbA1c at the end of trial
  - $Y(a, z)$ potential outcome when $A = a$ and $Z = z$.



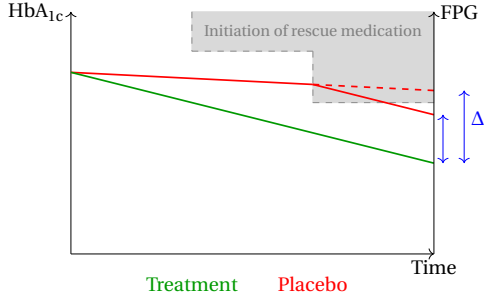
Checklist at visit 1:

1. What are the blood glucose levels?
2. Is the participant in need of rescue medication?

Assumption: Can only change "state" during visits.

Motivation

- Protocols: Participants initiate rescue based on exceeding some threshold of FPG (Fasting Plasma Glucose) or HbA1c (Long term blood glucose levels) or in case of a specific safety concern.



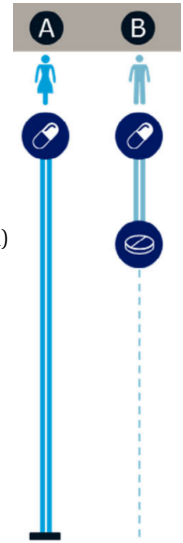
- Estimand of interest: Hypothetical strategy
  Treatment effect Δ *in the hypothetical scenario where participants do not initiate rescue medication.*

Mixed Models for Repeated Measures

- Data from participants **without rescue medication**.

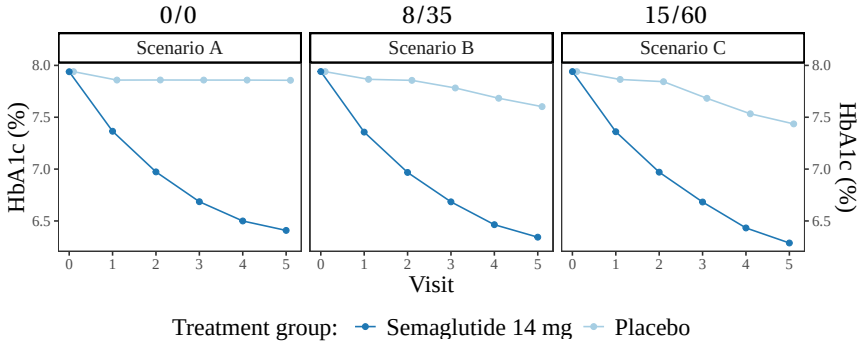
$$\text{HbA1c}_{\text{Visit}} \sim (\text{HbA1c}_{BL} + \text{Treatment} + \text{Region}) \times \text{Visit} + us(\text{Visit} \mid \text{id})$$

- *What would the treatment effect be had patients not needed rescue medication and* ***behaved like other patients who did not take rescue medication?***
- Assumption: Missing at Random



Simulation study

- 1000 two-armed RCTs with 400 participants
- 5 visits post-baseline
- both measured and unmeasured baseline covariates
- treatment effect of -1.5
- varying amount of rescue medication introduced



How is TMLE applied to this case?

Under extended identifying assumptions we are able to write the estimand as iterative conditional expectations, one for every visit:

$$\Psi(P) = E[\cdots E[Y \mid W_{0:4}, A = a, Z_{0:4} = 0]\cdots]$$

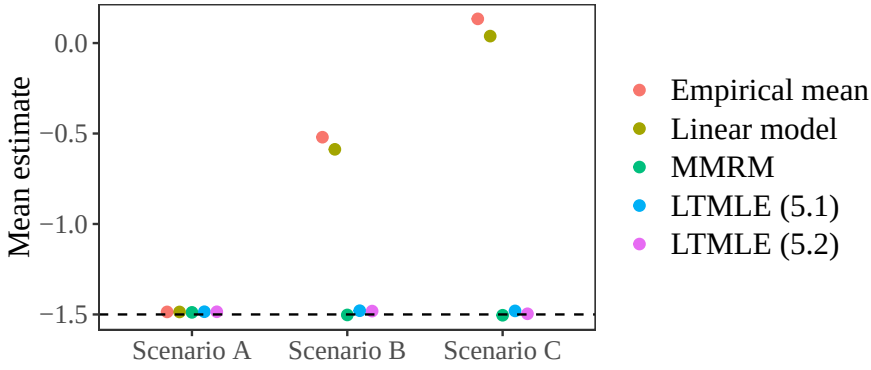
In addition, we investigated two approaches

- Without any assumptions, neither on dependency structure nor on distributions; LTMLE (Figure 5.1)
- With some assumptions, distribution of the treatment mechanism and Markov-like property; LTMLE (Figure 5.2)

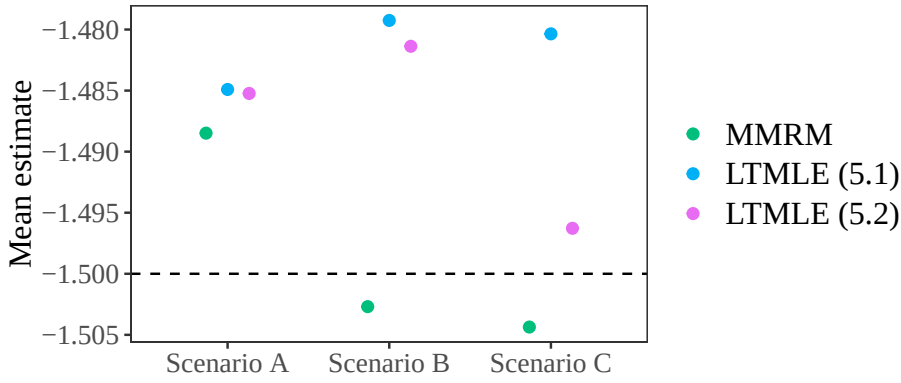
Results

Scenario	Method	Mean estimate	Mean bias	RMSE	Coverage
A	**Empirical mean**	−1.486	0.01410	0.3467	95.4%
	ANCOVA	−1.486	0.01385	0.3392	95%
	MMRM	−1.488	0.01151	0.3314	95.2%
	LTMLE (Figure 5.1)	−1.485	0.01509	0.01042	95.1%
	LTMLE (Figure 5.2)	−1.485	0.01477	0.01041	95.1%
B	**Empirical mean**	−0.5204	0.9796	1.038	19.5%
	ANCOVA	−0.5874	0.9126	0.9728	24.9%
	MMRM	−1.503	−0.002691	0.3545	94%
	LTMLE (Figure 5.1)	−1.479	0.02074	0.01236	94.3%
	LTMLE (Figure 5.2)	−1.481	0.01863	0.01291	94.8%
C	**Empirical mean**	0.1337	1.634	1.676	1.4%
	ANCOVA	0.03843	1.538	1.585	3.1%
	MMRM	−1.504	−0.004362	0.3949	93%
	LTMLE (Figure 5.1)	−1.480	0.01964	0.01307	90.1%
	LTMLE (Figure 5.2)	−1.496	0.003724	0.01428	93.9%

Results visually



Results visually

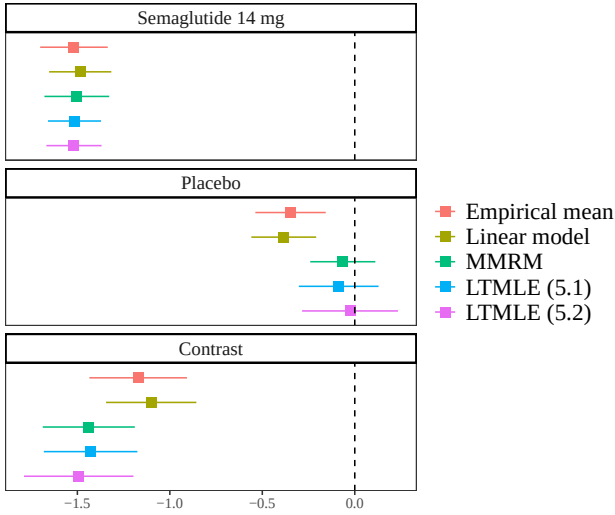


Results from case study

We were provided data from the PIONEER 1 trial, by Novo Nordisk A/S.

Data used	Method	Semaglutide 14 mg	Placebo	Contrast
Endpoint only	**Empirical mean**	-1.51 $[-1.68, -1.32]$	-0.35 $[-0.55, -0.14]$	-1.17 $[-1.44, -0.91]$
	Linear model	-1.49 $[-1.65, -1.32]$	-0.38 $[-0.56, -0.21]$	-1.1 $[-1.34, -0.86]$
Repeated measures	**MMRM**	-1.50 $[-1.68, -1.33]$	-0.06 $[-0.24, 0.11]$	-1.44 $[-1.69, -1.19]$
	LTMLE (**Figure 5.1**)	-1.52 $[-1.66, -1.37]$	-0.09 $[-0.30, 0.13]$	-1.43 $[-1.68, -1.18]$
	LTMLE (**Figure 5.2**)	-1.52 $[-1.67, -1.37]$	-0.03 $[-0.28, 0.23]$	-1.49 $[-1.79, -1.20]$

Results from case study



References

Schuler, A. & van der Laan, M. *Introduction to Modern Causal Inference*, work-in-progress digital book. `https://alejandroschuler.github.io/mci/introduction-to-modern-causal-inference.html`