# Causal Inference

a summary

# Contents

# 1 General

## Causal Roadmap (Petersen and van der Laan, 2014)

systematic approach linking causality to statistical procedures

**1. Specifying Knowledge.** structural causal model (unifying counterfactual language, structural equations, & causal graphs): a set of possible data-generating processes, expresses background knowledge and its limits

**2. Linking Data.** specifying measured variables and sampling specifics (latter can be incorporated into the model)

**3. Specifying Target.** define hypothetical experiment: decide

1. variables to intervene on: one (point treatment), multiple (longitudinal, censoring/missing, (in)direct effects)
2. intervention scheme: static, dynamic, stochastic
3. counterfactual summary of interest: absolute or relative, marginal structural models, interaction, effect modification
4. population of interest: whole, subset, different population

**4. Assessing Identifiability.** are knowledge and data sufficient to derive estimand and if not, what else is needed?

**5. Select Estimand.** current best answer: knowledge-based assumptions + which minimal convenience-based asspumptions (transparency) gets as close as possible

**6. Estimate.** choose estimator by statistical properties, nothing causal here

**7. Interpret.** hierarchy: statistical, counterfactual, feasible intervention, randomized trial

**Notation** chapter 1.1

**average causal effect** chapter 1.2 and 1.3 and 1.4 and 1.5

**randomized experiments (target trial)** 2.1 and 2.2; 3.6

**standardization** 2.3

## IP Weighting 2.4 (adjust for surrogate confounders)

**identifiability conditions** most of 3

**effect modification** chapter 4

**interaction** chapter 5

**causal diagrams** chapter 6, include swigs from 7.5 and that one technical point

**confounding** chapter 7

**selection bias** chapter 8

**measurement bias** chapter 9

**random variabilty** chapter 10

# 2 Models

## Modeling

data are a sample from the target population

| | | |
|---|---|---|
| *estimand:* | quantity of interest, | e. g. $\mathrm{E}[Y \mid A = a]$ |
| *estimator:* | function to use, | e. g. $\widehat{\mathrm{E}}[Y \mid A = a]$ |
| *estimate:* | apply function to data, | e. g. 4.1 |

**model**: a priori restriction of joint distribution/dose-response curve; *assumption:* no model misspecification (usually wrong)
**non-parametric estimator:** no restriction (saturated model) = *Fisher consistent estimator* (entire population data → true value)
**parsimonious model:** few parameters estimate many quantities
**bias-variance trade-off:**
wiggliness ↑ → misspecification bias ↓, CI width ↑

## IP Weighting

/ marginal structural models

(comparison at 13.4) (maybe censoring as own paragraph)

**standardization / parametric g-formula** chapter 13

**g-estimation** chapter 14

**Outcome regression** chapter 15

**instrumental variable estimation** chapter 16

**causal survival analysis** chapter 17

**variable selection** beginning of chapter 18

**machine learning in CI** end of chapter 18

# 3 Longitudinal Data

**time-varying treatments** beginning chapter 19

**identifiability** middle chapter 19

**treatment-confounder feedback** end chapter 19 and chapter 20

**g-formula** chapter 21.1

**IP weighting** chapter 21.2

**doubly robust estimators** chapter 21.3

**g-estimation** chapter 21.4

**censoring** chapter 21.5

**target trial** chapter 22 (does that even really fit in here, maybe push to 3rd paragraph in without models)

# References

*If no citation is given, the source is (Hernán and Robins, 2023)*

Hernán, M. A. and Robins, J. M. (2023). *Causal inference: what if.* Boca Raton: Chapman & Hall/CRC.

Petersen, M. L. and van der Laan, M. J. (2014). Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass.)*, 25(3):418–426.