Causal Inference in Epidemiology

Douglas Ezra Morrison

Last modified: 2025-11-13: 1:01:41 (UTC)

Contents

Preface

This will be an online book about causal inference.

Here are some other resources for learning causal inference:

UC Davis courses

- EPI 205[1] "Principles of Epidemiology"
- EPI 206[2] "Epidemiologic Study Design"
- EPI/SPH 207[3] "Advanced Epidemiologic Methodology"
- EPI 225[4] "Advanced Topics in Epidemiology Methods"
- POL 285[5] "Statistics of Causal Inference in Political Science"
- MGB/MGP/MGT 454A[6] "Causal Inference and Statistical Experiments"
  - syllabus: https://webapps.aws.ucdavis.edu/public/documents/4861649/Syllabushttps://schedule.aws.ucdavis.edu/public/documents/5319910/Syllabus
- PSC 204B[7] "Causal Modeling of Correlational Data"
- PSC 205C[8] "Structural Equation Modeling"

Course search options:

- https://schedule.aws.ucdavis.edu/courseScheduling
- https://catalog.ucdavis.edu/course-search/
- https://catalog.ucdavis.edu/courses-subject-code/

Online Videos

- "Introduction to Causal Inference"[9] (slides here[10])
- Online Causal Inference Seminar series[11]

[1] https://catalog.ucdavis.edu/search/?q=EPI+205
[2] https://catalog.ucdavis.edu/search/?q=EPI+206
[3] https://catalog.ucdavis.edu/search/?q=EPI+207
[4] https://catalog.ucdavis.edu/search/?q=EPI+225
[5] https://catalog.ucdavis.edu/search/?q=POL+285
[6] https://catalog.ucdavis.edu/search/?q=MGB+454A
[7] https://catalog.ucdavis.edu/search/?q=PSC+204B
[8] https://catalog.ucdavis.edu/search/?q=PSC+205C
[9] https://ucdhs.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=43e9eb6f-3ed9-41ac-8ad9-ae22016572c8%20
[10] https://health.ucdavis.edu/media-resources/ctsc/documents/pdfs/causal-inference-intro-2022.pdf
[11] https://www.youtube.com/channel/UCiiOj5GSES6uw21kfXnxj3A/videos

UC Davis Datalab learning group

- https://datalab.ucdavis.edu/causal-inference/
  - Reading list[12]

Other links:

- https://cameron.econ.ucdavis.edu/causal/
- https://datalab-icmat.github.io/causal_reading_group.html
- Lab exercises by Ben Noble[13]: https://github.com/bennoble/causal-inference-2022

Books

- Judea Pearl (2016)
- Hernán and Robins (2020)

[12] https://docs.google.com/document/d/1K0QZFSjQIYnOTahpRK7Q83eaiIIFjfa-clSbj_ifgco/edit?tab=t.0#heading=h.farbmh6n76gq
[13] https://benjaminnoble.org/

1 Introduction to causal inference

1.1 Introduction

Hernán and Robins (2020) (page vii):

> Unfortunately, the scientific literature is plagued by studies in which the causal question is not explicitly stated and the investigators' unverifiable assumptions are not declared. This casual attitude towards causal inference has led to a great deal of confusion.

Rigorously defining cause and effect is difficult. Fortunately, many humans have strong intuitions about these concepts. We will make cursory attempts at definitions for the basic terms, and leave the finer points to philosophers.

1.2 Individual causal effects

Definition 1.1 (Action, intervention, exposure, policy, treatment)**.** An **action** (also called an **intervention, exposure, policy,** or **treatment**) is a choice that we consider making.

Definition 1.2 (potential outcome, consequence)**.** A **potential outcome** of action a (also called a **consequence** of a) is the value of a random variable Y that would occur if we were to take action a. The potential outcome of action a on random variable Y is often denoted $Y(a)$, Y^a, or $Y^{A=a}$. We will use notation $Y(a)$.

Definition 1.3 (Factual outcome, observed outcome)**.** A *factual outcome* (or *observed outcome*) is the potential outcome corresponding to an action that was actually taken.

If we consider taking action a or an alternative a', and we actually take action $A = a$, then $Y(a)$ is the factual outcome.

Definition 1.4 (Counterfactual outcome)**.** A counterfactual outcome is a potential outcome corresponding to an action that was not actually taken.

If we consider taking action a or an alternative a', and we actually take action $A = a$, then $Y(a')$ is a counterfactual outcome.

There might be more than one counterfactual outcome, depending on how many action options were considered, but there can only ever be one factual outcome per random variable.

Definition 1.5 (Random variable)**.** A **random variable** is a variable that may have different values for different individuals and/or for different actions/exposures that an individual might experience. [1]

Definition 1.6 (Cause)**.** Action a **causes** outcome y (or is a cause of y) if:

- outcome y would occur if we were to take action a
  and
- outcome y would not occur if we did not take action a.

In other words, if:

- $Y(a) = y$
  and
- $\exists a' \neq a : Y(a') \neq y$

Definition 1.7 (effect)**.** The effect of action a on outcome Y, relative to a given alternative action a', is the contrast in potential outcomes, $Y(a)$ versus $Y(a')$.

Definition 1.8 (Consistency)**.** Consistency is the assumption that if we observe an action a, then the observed outcome Y is equal to the "factual potential outcome" corresponding to action a; in other words, if $A = a$, then $Y(a) = Y$.

---

[1] adapted from Hernán and Robins (2020) p3

Definition 1.9 (Exchangeability)**.** Subpopulations defined by exposure X are exchangeable with respect to a potential outcome $Y(x)$ if the distribution of $Y(x)$ does not depend on the observed exposure X:

$$Y(x) \perp\!\!\!\perp X$$

Theorem 1.1. *If subpopulations defined by values of exposure X are exchangeable with respect to potential outcome $Y(x)$, then the expected value of $Y(x)$ does not depend on the observed value of X:*

$$E[Y(x)|X = x'] = E[Y(x)|X = x]$$

Definition 1.10 (Conditional exchangeability)**.** Subpopulations defined by exposure X are exchangeable with respect to a potential outcome $Y(x)$ if the distribution of $Y(x)$ does not depend on the observed exposure X, conditional on covariate(s) Z:

$$Y(x) \perp\!\!\!\perp X|\tilde{Z}$$

Theorem 1.2. *If subpopulations defined by values of exposure X are conditionally exchangeable with respect to potential outcome $Y(x)$ given covariate $\tilde{Z}$, then the expected value of $Y(x)$ does not depend on the observed value of X:*

$$E\Big[Y(x)|X = x', \tilde{Z} = \tilde{z}\Big] = E\Big[Y(x)|X = x, \tilde{Z} = \tilde{z}\Big]$$

Definition 1.11 (Fundamental problem of causal inference)**.** The *fundamental problem of causal inference* is that only one potential outcome (the factual outcome) can be observed per person (or per sampling unit, more generally) (Holland 1986). The other, counterfactual outcomes, are all missing data, and thus, the causal effects are all missing data as well.

2 Difference in differences analyses

Many approaches to causal inference assume exchangeability (Definition 1.10) and exploit its consequence (Theorem 1.1):

$$\mathrm{E}[Y(x)|X = x'] = \mathrm{E}[Y(x)|X = x]$$

Difference-in-differences makes a weaker exchangeability assumption:

$$\mathrm{E}[Y_t(0) - Y_{t'}(0)|X = 1] = \mathrm{E}[Y_t(0) - Y_{t'}(0)|X = 0]$$

3 Summary

In summary, this book has no content whatsoever.

```
1 + 1
#> [1] 2
```

References

Hernán, MA, and J Robins. 2020. "Causal Inference: What If. Boca Raton: Chapman & Hill/Crc.(2020)." *Publisher's Note Springer Nature Remains Neutral with Regard to Jurisdictional Claims in Published Maps and Institutional Affiliations*. https://miguelhernan.org/whatifbook.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.1080/01621459.1986.10478354.

Judea Pearl, Nicholas P. Jewell, Madelyn Glymour. 2016. *Causal Inference in Statistics: A Primer*. 1st ed. Chicester: Wiley.