Modern Causal Mediation Analysis

A Workshop at SER 2023

Iván Díaz[1], Nima Hejazi[2], Kara Rudolph[3]

updated: June 13, 2023

[1]https://www.idiaz.xyz/
[2]https://nimahejazi.org
[3]https://kararudolph.github.io/

Contents

Welcome!

This open source, reproducible vignette accompanies a half-day workshop on modern methods for *causal mediation analysis*. While we encourage use of this bookdown site, for convenience, we have also made these workshop materials available in PDF[1].

0.1 About this workshop

Causal mediation analysis can provide a mechanistic understanding of how an exposure impacts an outcome, a central goal in epidemiology and health sciences. However, rapid methodologic developments coupled with few formal courses presents challenges to implementation. Beginning with an overview of classical direct and indirect effects, this workshop will present recent advances that overcome limitations of previous methods, allowing for: (i) continuous exposures, (ii) multiple, non-independent mediators, and (iii) effects identifiable in the presence of intermediate confounders affected by exposure. Emphasis will be placed on flexible, stochastic and interventional direct and indirect effects, highlighting how these may be applied to answer substantive epidemiological questions from real-world studies. Multiply robust, nonparametric estimators of these causal effects, and free and open source R packages (medshift[2] and medoutcon[3]) for their application, will be introduced.

To ensure translation to real-world data analysis, this workshop will incorporate hands-on R programming exercises to allow participants practice in implementing the statistical tools presented. It is recommended that participants have working knowledge of the basic notions of causal inference, including counterfactuals and identification (linking the causal effect to a parameter estimable from the observed data distribution). Familiarity with the R programming language is also recommended.

0.2 Tentative schedule

- 08:30A-08:45A: Introductions + mediation set-up

[1] https://code.nimahejazi.org/ser2023_mediation_workshop/causal_mediation.pdf
[2] https://github.com/nhejazi/medshift
[3] https://github.com/nhejazi/medoutcon

- 08:45A-9:15A: Controlled direct effects, natural direct/indirect effects, interventional direct/indirect effects
- 9:15A-9:25A: Choosing an estimand in real-world examples
- 9:25A-10:00A: What is the EIF?!
- 10:00A-10:30A: Break + discussion
- 10:30A-11:05A: Using the EIF to estimate the natural direct effect
- 11:05A-12:00P: Example walkthrough of R packages for effect estimation
- 12:00A-12:30P: Wrap-up

NOTE: All times listed in Pacific Time.

0.3 About the instructors

Iván Díaz[4]

I am an Associate Professor of Biostatistics in the Department of Population Health at the New York University Grossman School of Medicine[5]. My research focuses on the development of non-parametric statistical methods for causal inference from observational and randomized studies with complex datasets, using machine learning. This includes but is not limited to mediation analysis, methods for continuous exposures, longitudinal data including survival analysis, and efficiency guarantees with covariate adjustment in randomized trials. I am also interested in general semi-parametric theory, machine learning, and high-dimensional data.

Nima Hejazi[6]

I am an Assistant Professor of Biostatistics at the Harvard T.H. Chan School of Public Health[7]. My research interests blend causal inference, machine learning, non/semi-parametric inference, and computational statistics, with areas of recent emphasis having included causal mediation analysis, efficient estimation under outcome-dependent sampling designs, estimation of the causal effects of continuous treatments, and machine learning-based sieve estimation in causal inference. My methodological work is motivated principally by scientific collaborations in clinical trials and observational studies of infectious diseases, infectious disease epidemiology, and computational biology. I am also passionate about statistical computing and programming, and the design of open source software tools for reproducible statistical data science.

[4] https://www.idiaz.xyz/
[5] https://med.nyu.edu/faculty/ivan-l-diaz
[6] https://nimahejazi.org
[7] https://connects.catalyst.harvard.edu/Profiles/display/Person/207609

Kara Rudolph[8]

I am an Assistant Professor of Epidemiology in the Mailman School of Public Health at Columbia University[9]. My research interests are in developing and applying causal inference methods to understand social and contextual influences on mental health, substance use, and violence in disadvantaged, urban areas of the United States. My current work focuses on developing methods for transportability and mediation, and subsequently applying those methods to understand how aspects of the school and peer environments mediate relationships between neighborhood factors and adolescent drug use across populations. More generally, my work on generalizing/transporting findings from study samples to target populations and identifying subpopulations most likely to benefit from interventions contributes to efforts to optimally target available policy and program resources.

0.4 Reproducibility

These workshop materials were written using bookdown[10], and the complete source is available on GitHub[11]. This version of the book was built with R version 4.3.0 (2023-04-21), pandoc[12] version 2.19.2, and the following packages:

package	version	source
bookdown	0.26.3	Github (rstudio/bookdown@169c43b6bb95213f2af63a95acd4e977a58a3e1f)
bslib	0.3.1.9000	Github (rstudio/bslib@a4946a49499438e71dce29c810a41e2d05170376)
data.table	1.14.2	CRAN (R 4.3.0)
downlit	0.4.0	CRAN (R 4.3.0)
dplyr	1.0.9	CRAN (R 4.3.0)
ggfortify	0.4.14	CRAN (R 4.3.0)
ggplot2	3.3.6	CRAN (R 4.3.0)
kableExtra	1.3.4	CRAN (R 4.3.0)
knitr	1.39	CRAN (R 4.3.0)
magick	2.7.3	CRAN (R 4.3.0)
medoutcon	0.1.6	Github (nhejazi/medoutcon@49af7b52ad7fc5064e6af984a682118d3463917a)
medshift	0.1.4	Github (nhejazi/medshift@0ae0572fc5e37a8595b798909057afbc48304236)
mvtnorm	1.1-3	CRAN (R 4.3.0)

[8] https://kararudolph.github.io/
[9] https://www.publichealth.columbia.edu/academics/departments/epidemiology
[10] http://bookdown.org/
[11] https://github.com/nhejazi/causal_mediation_workshops
[12] https://pandoc.org/

package	version	source
origami	1.0.5	Github (tlverse/origami@e1b8fe6f5e75fff1d48eed115bb81475c9bd506e)
pdftools	3.2.0	CRAN (R 4.3.0)
readr	2.1.2	CRAN (R 4.3.0)
rmarkdown	2.14	CRAN (R 4.3.0)
skimr	2.1.4	CRAN (R 4.3.0)
sl3	1.4.5	Github (tlverse/sl3@825019f28a650936e24cca421dd155641c860435)
stringr	1.4.0	CRAN (R 4.3.0)
tibble	3.1.7	CRAN (R 4.3.0)
tidyr	1.2.0	CRAN (R 4.3.0)

0.5 Setup instructions

0.5.1 R and RStudio

R and **RStudio** are separate downloads and installations. R is the underlying statistical computing environment. RStudio is a graphical integrated development environment (IDE) that makes using R much easier and more interactive. You need to install R before you install RStudio.

0.5.1.1 Windows

0.5.1.1.1 If you already have R and RStudio installed

- Open RStudio, and click on "Help" > "Check for updates". If a new version is available, quit RStudio, and download the latest version for RStudio.
- To check which version of R you are using, start RStudio and the first thing that appears in the console indicates the version of R you are running. Alternatively, you can type `sessionInfo()`, which will also display which version of R you are running. Go on the CRAN website[13] and check whether a more recent version is available. If so, please download and install it. You can check here[14] for more information on how to remove old versions from your system if you wish to do so.

0.5.1.1.2 If you don't have R and RStudio installed

- Download R from the CRAN website[15].

[13] https://cran.r-project.org/bin/windows/base/
[14] https://cran.r-project.org/bin/windows/base/rw-FAQ.html#How-do-I-UNinstall-R_003f
[15] http://cran.r-project.org/bin/windows/base/release.htm

- Run the .exe file that was just downloaded
- Go to the RStudio download page[16]
- Under *Installers* select **RStudio x.yy.zzz - Windows XP/Vista/7/8** (where x, y, and z represent version numbers)
- Double click the file to install it
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

0.5.1.2 Mac OSX

0.5.1.2.1 If you already have R and RStudio installed

- Open RStudio, and click on "Help" > "Check for updates". If a new version is available, quit RStudio, and download the latest version for RStudio.
- To check the version of R you are using, start RStudio and the first thing that appears on the terminal indicates the version of R you are running. Alternatively, you can type `sessionInfo()`, which will also display which version of R you are running. Go on the CRAN website[17] and check whether a more recent version is available. If so, please download and install it.

0.5.1.2.2 If you don't have R and RStudio installed

- Download R from the CRAN website[18].
- Select the .pkg file for the latest R version
- Double click on the downloaded file to install R
- It is also a good idea to install XQuartz[19] (needed by some packages)
- Go to the RStudio download page[20]
- Under *Installers* select **RStudio x.yy.zzz - Mac OS X 10.6+ (64-bit)** (where x, y, and z represent version numbers)
- Double click the file to install RStudio
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

0.5.1.3 Linux

- Follow the instructions for your distribution from CRAN[21], they provide information to get the most recent version of R for common distributions. For most distributions, you could use your package manager (e.g., for Debian/Ubuntu run `sudo apt-get install r-base`, and for Fedora `sudo yum install R`), but we don't recommend this approach as the versions provided by this are usually out of date. In any case, make sure you have at least R 3.3.1.

[16] https://www.rstudio.com/products/rstudio/download/#download
[17] https://cran.r-project.org/bin/macosx/
[18] http://cran.r-project.org/bin/macosx
[19] https://www.xquartz.org/
[20] https://www.rstudio.com/products/rstudio/download/#download
[21] https://cloud.r-project.org/bin/linux

- Go to the RStudio download page[22]
- Under *Installers* select the version that matches your distribution, and install it with your preferred method (e.g., with Debian/Ubuntu sudo dpkg –i rstudio –x.yy.zzz–amd64.deb at the terminal).
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

These setup instructions are adapted from those written for Data Carpentry: R for Data Analysis and Visualization of Ecological Data[23].

0.5.2 Virtual Enironment setup with `renv`

These instructions are intended to help with setting up the included renv virtual environment[24], which ensures all participants are using the same exact set of R packages (and package versions). A few important notes to keep in mind:

- When R is started from the top level of this repository, renv is activated automatically. There is no further action required on your part. If renv is not installed, it will be installed automatically, assuming that you have an active internet connection.
- While renv is active, the R session will only have access to the packages (and their dependencies) that are listed in the renv.lock file – that is, you should not expect to have access to any other R packages that may be installed elsewhere on the computing system in use.
- Upon an initial attempt, renv will prompt you to install packages listed in the renv.lock file, by printing a message like the following:

```
# * Project 'PATH/TO/causal_mediation_workshops' loaded. [renv 0.13.2]
# * The project may be out of sync -- use `renv::status()` for more details
> renv::status()
# The following package(s) are recorded in the lockfile, but not installed:
# Use `renv::restore()` to install these packages.
```

In any such case, please call renv::restore() to install any missing packages. Note that you do *not* need to manually install the packages via install.packages(), remotes::install_github(), or similar.

For details on how the renv system works, the following references may be helpful:

1. Collaborating with renv[25]
2. Introduction to renv[26]

[22] https://www.rstudio.com/products/rstudio/download/#download
[23] http://www.datacarpentry.org/R-ecology-lesson/
[24] https://rstudio.github.io/renv/index.html
[25] https://rstudio.github.io/renv/articles/collaborating.html
[26] https://rstudio.github.io/renv/articles/renv.html

In some rare cases, R packages that renv automatically tries to install as part of the renv :: restore () process may fail due to missing systems-level dependencies. In such cases, a reference to the missing dependencies and system-specific instructions their installation involving, e.g., Ubuntu Linux's apt[27] or homebrew for macOS[28], will usually be displayed.

[27] http://manpages.ubuntu.com/manpages/bionic/man8/apt.8.html
[28] https://brew.sh/

Chapter 1

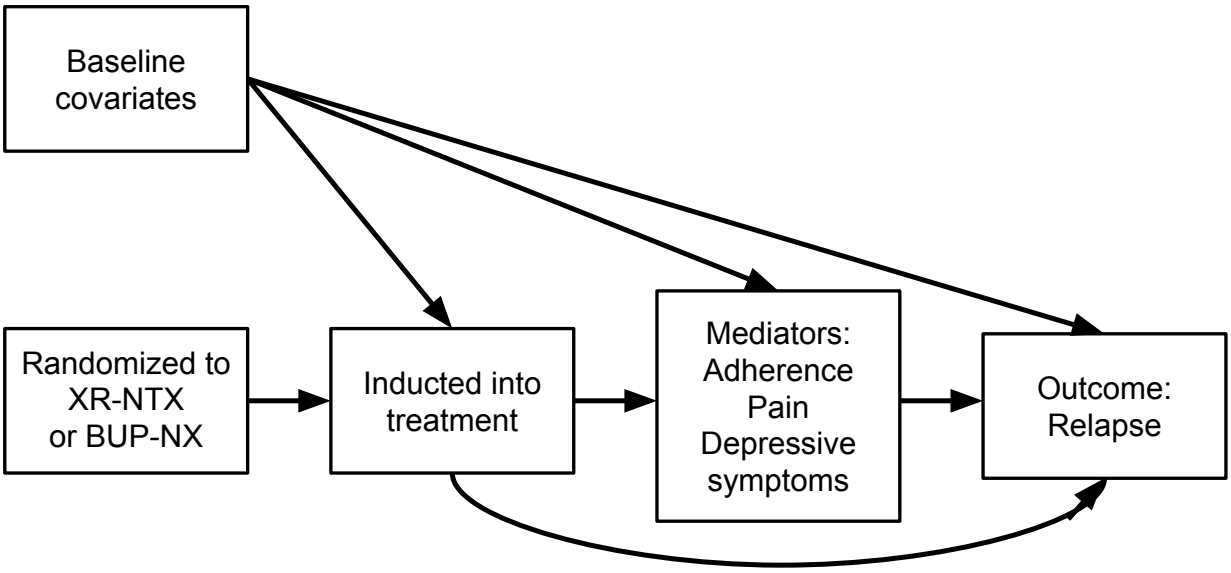
Causal mediation analysis intro

1.1 Motivating study

- A recent, large, multi-site trial (X:BOT) compared the effectiveness of XR-NTX to buprenorphine–naloxone (BUP-NX) in preventing relapse among those with OUD starting medication in inpatient treatment settings.
- An analysis of potential moderators of medication effectiveness found that homeless individuals had a lower risk of relapse on XR-NTX, whereas non-homeless individuals had a lower risk of relapse on BUP-NX.
- The effect sizes were similarly large for these groups but in opposite directions.
- We can to use mediation analysis to explore the mechanisms underlying these differences.
- Key questions:
  - Do differences in the effects of treatment (comparing two medications for opioid use disorder, naltrexone vs buprenorphine) on risk of relapse operate through mediators of adherence, opioid use, pain, and depressive symptoms? (Rudolph et al., 2020)
  - Are those mediated effects different for homeless vs non-homeless individuals?

1.2 What is causal mediation analysis?

- Statistical mediation analyses assess associations between the variables. They can help you establish, for example, if the *association* between treatment and outcome can be mostly explained by an *association* between treatment and mediator
- Causal mediation analyses, on the other hand, seek to asess causal relations. For example, they help you establish whether treatment *causes* the outcome because it *causes* the mediator. To do this, causal mediation seek to understand how the paths behave under circumstances different from the observed circumstances (e.g., interventions)



1.2.1 Why are the causal methods that we will discuss today important?

- Assume you are interested in the effect of treatment assignment A (naltrexone vs. buprenorphine) on an outcome Y (risk of relapse) through mediators M (e.g., opioid use, pain, depressive symptoms)
- We have pre-treatment confounders W
- There is a confounder Z of $M \to Y$ affected by treatment assignment (with adherence)
- We could fit the following models:

$$\mathbb{E}(M \mid A = a, W = w, Z = z) = \gamma_0 + \gamma_1 a + \gamma_2 w + \gamma_3 z \tag{1.1}$$

$$\mathbb{E}(Y \mid M = m, A = a, W = w, Z = z) = \beta_0 + \beta_1 m + \beta_2 a + \beta_3 w + \beta_4 z \tag{1.2}$$

- The product $\gamma_1\beta_1$ has been proposed as a measure of the effect of A on Y through M
- Causal interpretation problems with this method: We will see that this parameter cannot be interpreted as a causal effect

1.2.2 R Example:

- Assume we have a pre-treamtment confounder of Y and M, denote it with W
- For simplicity, assume A is randomized
- We'll generate a really large sample from a data generating mechanism so that we are not concerned with sampling errors

```
n <- 1e6
w <- rnorm(n)
a <- rbinom(n, 1, 0.5)
z <- rbinom(n, 1, 0.2 * a + 0.3)
m <- rnorm(n, w + z)
y <- rnorm(n, m + w - a + z)
```

- Note that the indirect effect (i.e., the effect through M) in this example is nonzero (there is a pathway $A \to Z \to M \to Y$)
- Let's see what the product of coefficients method would say:

```
lm_y <- lm(y ~ m + a + w + z)
lm_m <- lm(m ~ a + w + z)
## product of coefficients
coef(lm_y)[2] * coef(lm_m)[2]
#>           m
#> -0.0014835
```

Among other things, in this workshop:

- We will provide some understanding for why the above method fails in this example
- We will study estimators that are robust to misspecification in the above models

1.3 Causal mediation models

In this workshop we will use directed acyclic graphs. We will focus on the two types of graph:

1.3.1 No intermediate confounders

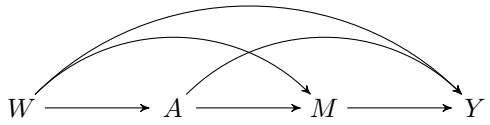


Figure 1.1: Directed acyclic graph under *no intermediate confounders* of the mediator-outcome relation affected by treatment

1.3.2 Intermediate confounders

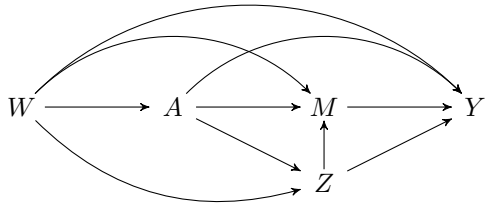


Figure 1.2: Directed acyclic graph under intermediate confounders of the mediator-outcome relation affected by treatment

The above graphs can be interpreted as a *non-parametric structural equation model* (NPSEM), also known as *structural causal model* (SCM):

$$W = f_W(U_W) \tag{1.3}$$
$$A = f_A(W, U_A) \tag{1.4}$$
$$Z = f_Z(W, A, U_Z) \tag{1.5}$$
$$M = f_M(W, A, Z, U_M) \tag{1.6}$$
$$Y = f_Y(W, A, Z, M, U_Y) \tag{1.7}$$

- Here $U = (U_W, U_A, U_Z, U_M, U_Y)$ is a vector of all unmeasured exogenous factors affecting the system
- The functions f are assumed fixed but unknown
- We posit this model as a system of equations that nature uses to generate the data
- Therefore we leave the functions f unspecified (i.e., we do not know the true nature mechanisms)
- Sometimes we know something: e.g., if A is randomized we know $A = f_A(U_A)$ where U_A is the flip of a coin (i.e., independent of everything).

1.4 Counterfactuals

- Recall that we are interested in assessing how the pathways would behave under circumstances different from the observed circumstances
- We operationalize this idea using *counterfactual* random variables
- Counterfactuals are hypothetical random variables that would have been observed in an alternative world where something had happened, possibly contrary to fact

We will use the following counterfactual variables:

- Y_a is a counterfactual variable in a hypothetical world where $\mathbb{P}(A = a) = 1$ with probability one for some value a
- $Y_{a,m}$ is the counterfactual outcome in a world where $\mathbb{P}(A = a, M = m) = 1$
- M_a is the counterfactual variable representing the mediator in a world where $\mathbb{P}(A = a) = 1$.

1.4.1 How are counterfactuals defined?

- In the NPSEM framework, counterfactuals are quantities *derived* from the model.
- Once you define a change to the causal system, that change needs to be progragated downstream.
  - Example: modifying the system to make everyone receive XR-NTX yields counterfactual adherence, mediators, and outcomes.

- Take as example the DAG in Figure 1.2:

$$A = a \tag{1.8}$$
$$Z_a = f_Z(W, a, U_M) \tag{1.9}$$
$$M_a = f_M(W, a, Z_a, U_M) \tag{1.10}$$
$$Y_a = f_Y(W, a, Z_a, M_a, U_Y) \tag{1.11}$$
$$\tag{1.12}$$

- We will also be interested in *joint changes* to the system:

$$A = a \tag{1.13}$$
$$Z_a = f_Z(W, a, U_M) \tag{1.14}$$
$$M = m \tag{1.15}$$
$$Y_{a,m} = f_Y(W, a, Z_a, m, U_Y) \tag{1.16}$$
$$\tag{1.17}$$

- And, perhaps more importantly, we will use *nested counterfactuals*
- For example, if A is binary, you can think of the following counterfactual

$$A = 1 \tag{1.18}$$
$$Z_1 = f_Z(W, 1, U_M) \tag{1.19}$$
$$M = M_0 \tag{1.20}$$
$$Y_{1,M_0} = f_Y(W, 1, Z_1, M_0, U_Y) \tag{1.21}$$
$$\tag{1.22}$$

- Y_{1,M_0} is interpreted as *the outcome for an individual in a hypothetical world where treatment was given but the mediator was held at the value it would have taken under no treatment*
- Causal mediation effects are often defined in terms of the distribution of these nested counterfactuals.
- That is, causal effects give you information about what would have happened *in some hypothetical world* where the mediator and treatment mechanisms changed.

Chapter 2

Types of path-specific causal mediation effects

- Controlled direct effects
- Natural direct and indirect effects
- Interventional direct and indirect effects

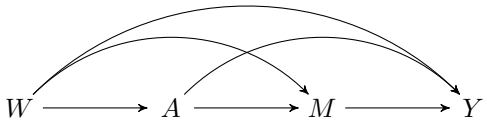
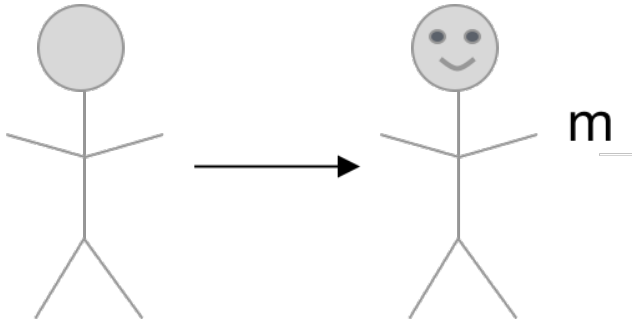


Figure 2.1: Directed acyclic graph under *no intermediate confounders* of the mediator-outcome relation affected by treatment

2.1 Controlled direct effects

$$\psi_{\text{CDE}} = \mathbb{E}(Y_{1,m} - Y_{0,m})$$



- Set the mediator to a reference value $M = m$ uniformly for everyone in the population
- Compare $A = 1$ vs $A = 0$ with $M = m$ fixed

2.1.1 Identification assumptions:

- Confounder assumptions:
  - $A \perp\!\!\!\perp Y_{a,m} \mid W$
  - $M \perp\!\!\!\perp Y_{a,m} \mid W, A$
- Positivity assumptions:
  - $\mathbb{P}(M = m \mid A = a, W) > 0 \; a.e.$
  - $\mathbb{P}(A = a \mid W) > 0 \; a.e.$

Under the above identification assumptions, the controlled direct effect can be identified:

$$\mathbb{E}(Y_{1,m} - Y_{0,m}) = \mathbb{E}\{\mathbb{E}(Y \mid A = 1, M = m, W) - \mathbb{E}(Y \mid A = 0, M = m, W)\}$$

- For intuition about this formula in R, let's continue with a toy example:

```
n <- 1e6
w <- rnorm(n)
a <- rbinom(n, 1, 0.5)
m <- rnorm(n, w + a)
y <- rnorm(n, w + a + m)
```

- First we fit a correct model for the outcome

```
lm_y <- lm(y ~ m + a + w)
```

- Assume we would like the CDE at $m = 0$
- Then we generate predictions $\mathbb{E}(Y \mid A = 1, M = m, W)$ and $\mathbb{E}(Y \mid A = 0, M = m, W)$:

```
pred_y1 <- predict(lm_y, newdata = data.frame(a = 1, m = 0, w = w))
pred_y0 <- predict(lm_y, newdata = data.frame(a = 0, m = 0, w = w))
```

- Then we compute the difference between the predicted values $\mathbb{E}(Y \mid A = 1, M = m, W) - \mathbb{E}(Y \mid A = 0, M$ and average across values of W

```
## CDE at m = 0
mean(pred_y1 - pred_y0)
#> [1] 1.0009
```

2.1.2 Is this the estimand I want?

- Makes the most sense if can intervene directly on M
  - And can think of a policy that would set everyone to a single constant level $m \in \mathcal{M}$.
  - Judea Pearl calls this *prescriptive*.
  - Can you think of an example?
  - Air pollution, rescue inhaler dosage, hospital visits
  - Does not provide a decomposition of the average treatment effect into direct and indirect effects.

What if our research question doesn't involve intervening directly on the mediator?

What if we want to decompose the average treatment effect into its direct and indirect counterparts?

2.2 Natural direct and indirect effects

Still using the same DAG as above,

- Recall the definition of the nested counterfactual:

$$Y_{1,M_0} = f_Y(W, 1, M_0, U_Y)$$

- Interpreted as *the outcome for an individual in a hypothetical world where treatment was given but the mediator was held at the value it would have taken under no treatment*
- Recall that, because of the definition of counterfactuals

$$Y_{1,M_1} = Y_1$$

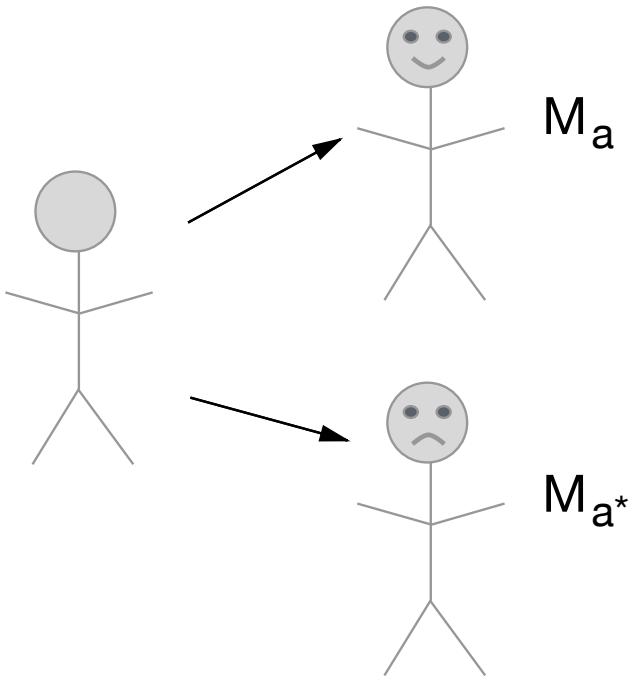
Then we can decompose the *average treatment effect* $E(Y_1 - Y_0)$ as follows

$$\mathbb{E}[Y_{1,M_1} - Y_{0,M_0}] = \underbrace{\mathbb{E}[Y_{1,M_1} - Y_{1,M_0}]}_{\text{natural indirect effect}} + \underbrace{\mathbb{E}[Y_{1,M_0} - Y_{0,M_0}]}_{\text{natural direct effect}}$$

- Natural direct effect (NDE): Varying treatment while keeping the mediator fixed at the value it would have taken under no treatment
- Natural indirect effect (NIE): Varying the mediator from the value it would have taken under treatment to the value it would have taken under control, while keeping treatment fixed

2.2.1 Identification assumptions:

- $A \perp\!\!\!\perp Y_{a,m} \mid W$
- $M \perp\!\!\!\perp Y_{a,m} \mid W, A$
- $A \perp\!\!\!\perp M_a \mid W$
- $M_0 \perp\!\!\!\perp Y_{1,m} \mid W$
- and positivity assumptions



2.2.2 Cross-world independence assumption

What does $M_0 \perp\!\!\!\perp Y_{1,m} \mid W$ mean?

- Conditional on W, knowledge of the mediator value in the absence of treatment, M_0, provides no information about the outcome under treatment, $Y_{1,m}$.
- Can you think of a data-generating mechanism that would violate this assumption?
- Example: in a randomized study, whenever we believe that treatment assignment works through adherence (i.e., almost always), we are violating this assumption (more on this later).
- Cross-world assumptions are problematic for other reasons, including:
  - You can never design a randomized study where the assumption holds by design.

If the cross-world assumption holds, can write the NDE as a weighted average of controlled direct effects at each level of $M = m$.

$$\mathbb{E} \sum_m \{\mathbb{E}(Y_{1,m} \mid W) - \mathbb{E}(Y_{0,m} \mid W)\} \mathbb{P}(M_0 = m \mid W)$$

- If CDE(m) is constant across m, then CDE = NDE.

2.2.3 Identification formula:

- Under the above identification assumptions, the natural direct effect can be identified:

$$\mathbb{E}(Y_{1,M_0} - Y_{0,M_0}) = \mathbb{E}[\mathbb{E}\{\mathbb{E}(Y \mid A = 1, M, W) - \mathbb{E}(Y \mid A = 0, M, W) \mid A = 0, W\}]$$

- The natural indirect effect can be identified similarly.
- Let's dissect this formula in R:

```
n <- 1e6
w <- rnorm(n)
a <- rbinom(n, 1, 0.5)
m <- rnorm(n, w + a)
y <- rnorm(n, w + a + m)
```

- First we fit a correct model for the outcome

```
lm_y <- lm(y ~ m + a + w)
```

- Then we generate predictions
$$\mathbb{E}(Y \mid A = 1, M, W) \text{ and } \mathbb{E}(Y \mid A = 0, M, W)$$
with A fixed but letting M and W take their observed values

```
pred_y1 <- predict(lm_y, newdata = data.frame(a = 1, m = m, w = w))
pred_y0 <- predict(lm_y, newdata = data.frame(a = 0, m = m, w = w))
```

- Then we compute the difference between the predicted values
$$\mathbb{E}(Y \mid A = 1, M, W) - \mathbb{E}(Y \mid A = 0, M, W),$$

- and use this difference as a pseudo-outcome in a regression on A and W:
$$\mathbb{E}\{\mathbb{E}(Y \mid A = 1, M, W) - \mathbb{E}(Y \mid A = 0, M, W) \mid A = 0, W\}$$

```
pseudo <- pred_y1 - pred_y0
lm_pseudo <- lm(pseudo ~ a + w)
```

- Now we predict the value of this pseudo-outcome under $A = 0$, and average the result

```
pred_pseudo <- predict(lm_pseudo, newdata = data.frame(a = 0, w = w))
## NDE:
mean(pred_pseudo)
#> [1] 0.99655
```

2.2.4 Is this the estimand I want?

- Makes sense to intervene on A but not directly on M.
- Want to understand a natural mechanism underlying an association / total effect. J. Pearl calls this *descriptive*.
- NDE + NIE = total effect (ATE).
- Okay with the assumptions.

What if our data structure involves a post-treatment confounder of the mediator-outcome relationship (e.g., adherence)?

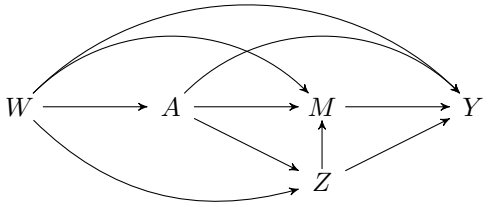
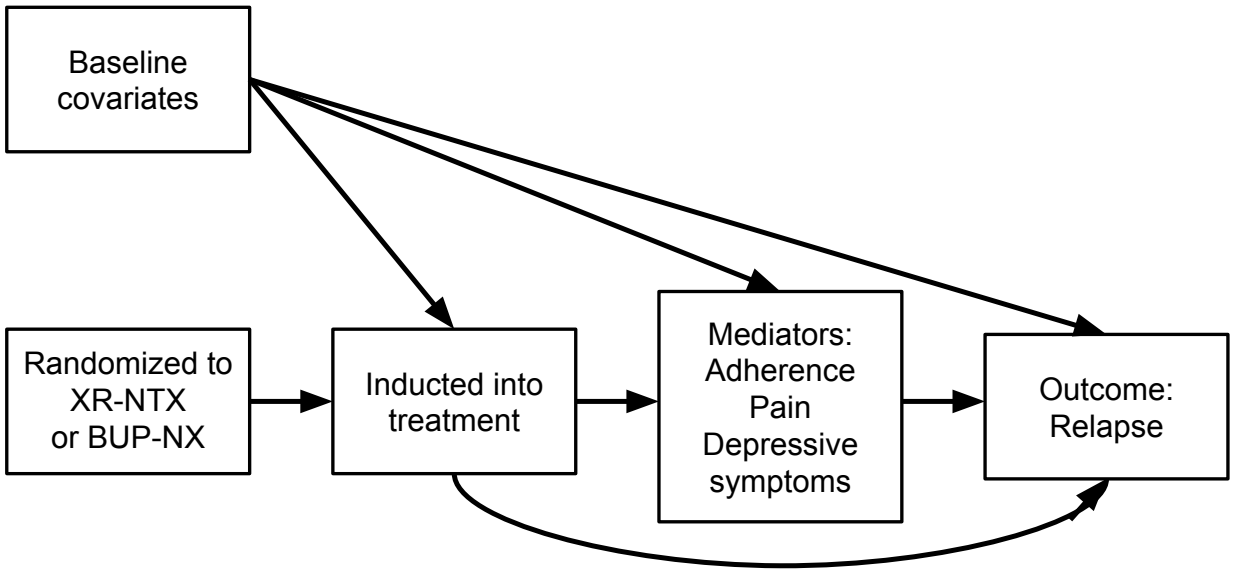


Figure 2.2: Directed acyclic graph under intermediate confounders of the mediator-outcome relation affected by treatment



2.2.5 Unidentifiability of the NDE and NIE in this setting

- In this example, natural direct and indirect effects are not generally point identified from observed data $O = (W, A, Z, M, Y)$.
- The reason for this is that the cross-world counterfactual assumption
$$Y_{1,m} \perp\!\!\!\perp M_0 \mid W$$
does not hold in the above directed acyclic graph.

- To give intuition, we focus on the counterfactual outcome $Y_{A=1,M_{A=0}}$.
  - This counterfactual outcome involves two counterfactual worlds simultaneously: one in which $A = 1$ for the first portion of the counterfactual outcome, and one in which $A = 0$ for the nested portion of the counterfactual outcome.
  - Setting $A = 1$ induces a counterfactual treatment-induced confounder, denoted $Z_{A=1}$. Setting $A = 0$ induces another counterfactual treatment-induced confounder, denoted $Z_{A=0}$.
  - The two treatment-induced counterfactual confounders, $Z_{A=1}$ and $Z_{A=0}$ share unmeasured common causes, U_Z, which creates a spurious association.
  - Because $Z_{A=1}$ is causally related to $Y_{A=1,M=m}$, and $Z_{A=0}$ is also casually related to $M_{A=0}$, the path through U_Z means that the backdoor criterion is not met for identification of $Y_{A=1,M_{A=0}}$, i.e., $M_0 \not\perp\!\!\!\perp Y_{A=1,m} \mid W$, where W denotes baseline covariates.

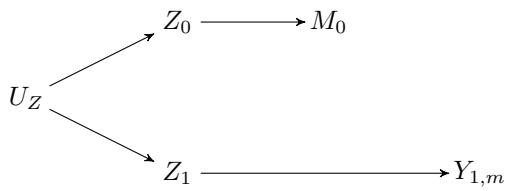


Figure 2.3: Parallel worlds model of the scenario considered.

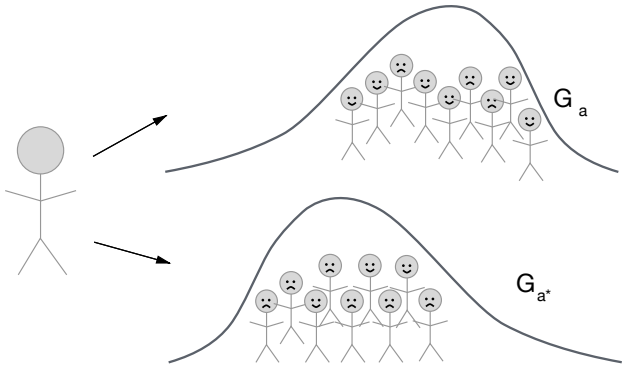
However:

- We can actually actually identify the NIE/NDE in the above setting if we are willing to invoke monotonicity between a treatment and one or more binary treatment-induced confounders (Tchetgen Tchetgen and VanderWeele, 2014).
- Assuming monotonicity is also sometimes referred to as assuming "no defiers"—in other words, assuming that there are no individuals who would do the opposite of the encouragement.
- Monotonicity may seem like a restrictive assumption, but may be reasonable in some common scenarios (e.g., in trials where the intervention is randomized treatment assignment and the treatment-induced confounder is whether or not treatment was actually taken–in this setting, we may feel comfortable assuming that there are no "defiers", frequently assumed when using IVs to identify causal effects)

Note: CDEs are still identified in this setting. They can be identified and estimated similarly to a longitudinal data sructure with a two-time-point intervention.

2.3 Interventional (in)direct effects

- Let G_a denote a random draw from the distribution of $M_a \mid W$



- Define the counterfactual Y_{1,G_0} as the counterfactual variable in a hypothetical world where A is set $A = 1$ and M is set to $M = G_0$ with probability one.

- Define Y_{0,G_0} and Y_{1,G_1} similarly
- Then we can define:

$$\mathbb{E}[Y_{1,G_1} - Y_{0,G_0}] = \underbrace{\mathbb{E}[Y_{1,G_1} - Y_{1,G_0}]}_{\text{interventional indirect effect}} + \underbrace{\mathbb{E}[Y_{1,G_0} - Y_{0,G_0}]}_{\text{interventional direct effect}}$$

- Note that $\mathbb{E}[Y_{1,G_1} - Y_{0,G_0}]$ is still a *total effect* of treatment, even if it is different from the ATE $\mathbb{E}[Y_1 - Y_0]$
- We gain in the ability to solve a problem, but lose in terms of interpretation of the causal effect (cannot decompose the ATE)

2.3.1 An alternative definition of the effects:

- Above we defined G_a as a random draw from the distribution of $M_a \mid W$
- What if instead we define G_a as a random draw from the distribution of $M_a \mid (Z_a, W)$
- It turns out the indirect effect defined in this way only measures the path $A \to M \to Y$, and not the path $A \to Z \to M \to Y$
- There may be important reasons to choose one over another (e.g., survival analyses where we want the distribution conditional on Z, instrumental variable designs where it doesn't make sense to condition on Z)

2.3.2 Identification assumptions:

- $A \perp\!\!\!\perp Y_{a,m} \mid W$
- $M \perp\!\!\!\perp Y_{a,m} \mid W, A, Z$
- $A \perp\!\!\!\perp M_a \mid W$
- and positivity assumptions.

Under these assumptions, the population interventional direct and indirect effect is identified:

$$\mathbb{E}(Y_{a,G_{a'}}) = \mathbb{E}\left[\mathbb{E}\left\{\sum_z \mathbb{E}(Y \mid A=a, Z=z, M, W)\mathbb{P}(Z=z \mid A=a, W) \mid A=a', W\right\}\right]$$

- Let's dissect this formula in R:

```
n <- 1e6
w <- rnorm(n)
a <- rbinom(n, 1, 0.5)
z <- rbinom(n, 1, 0.5 + 0.2 * a)
m <- rnorm(n, w + a - z)
y <- rnorm(n, w + a + z + m)
```

- Let us compute $\mathbb{E}(Y_{1,G_0})$ (so that $a=1$, and $a'=0$).
- First, fit a regression model for the outcome, and compute

$$\mathbb{E}(Y \mid A=a, Z=z, M, W)$$

for all values of z

```
lm_y <- lm(y ~ m + a + z + w)
pred_a1z0 <- predict(lm_y, newdata = data.frame(m = m, a = 1, z = 0, w = w))
pred_a1z1 <- predict(lm_y, newdata = data.frame(m = m, a = 1, z = 1, w = w))
```

- Now we fit the true model for $Z \mid A, W$ and get the conditional probability that $Z=1$ fixing $A=1$

```
prob_z <- lm(z ~ a)
pred_z <- predict(prob_z, newdata = data.frame(a = 1))
```

- Now we compute the following pseudo-outcome:

$$\sum_z \mathbb{E}(Y \mid A=a, Z=z, M, W)\mathbb{P}(Z=z \mid A=a, w)$$

```
pseudo_out <- pred_a1z0 * (1 - pred_z) + pred_a1z1 * pred_z
```

- Now we regress this pseudo-outcome on A, W, and compute the predictions setting $A=0$, that is,

$$\mathbb{E}\left\{\sum_z \mathbb{E}(Y \mid A=a, Z=z, M, W)\mathbb{P}(Z=z \mid A=a, w) \mid A=a', W\right\}$$

```
fit_pseudo <- lm(pseudo_out ~ a + w)
pred_pseudo <- predict(fit_pseudo, data.frame(a = 0, w = w))
```

- And finally, just average those predictions!

```
## Mean(Y(1, G(0)))
mean(pred_pseudo)
#> [1] 1.1979
```

- This was for $(a, a') = (1, 0)$. Can do the same with $(a, a') = (1, 1)$, and $(a, a') = (0, 0)$ to obtain an effect decomposition

$$\mathbb{E}[Y_{1,G_1} - Y_{0,G_0}] = \underbrace{\mathbb{E}[Y_{1,G_1} - Y_{1,G_0}]}_{\text{interventional indirect effect}} + \underbrace{\mathbb{E}[Y_{1,G_0} - Y_{0,G_0}]}_{\text{interventional direct effect}}$$

2.3.3 Is this the estimand I want?

- Makes sense to intervene on A but not directly on M.
- Goal is to understand a descriptive type of mediation.
- Okay with the assumptions!

2.3.4 But, there is an important limitation of interventional effects

Miles (2022) recently uncovered an important limitation of these effects, which can be described as follows. The sharp mediational hull hypothesis can be defined as

$$H_0 : Y(a, M(a')) = Y(a, M(a^\star)); \text{ for all } a, a', a^\star.$$

The problem is that interventional effects are not guaranteed to be null when the sharp mediational hypothesis is true.

This could present a problem in practice if some subgroup of the population has a relationship between A and M, but not between M and Y. Then, another distinct subgroup of the population has a relationship between M and Y but not between A and M. In such a scenario, the interventional indirect effect would be nonzero, but there would be no one person in the population whose effect of A on Y would be mediated by M.

More details in the original paper.

2.4 Estimand Summary

Table 1. Mediation Estimand Definitions, Descriptions, and Assumptions

Estimand	Description	Identifying Assumptions in Addition to Posit Consistency
Controlled direct effect $E(Y_{a,m}) - E(Y_{a^*,m})$	Difference in the expected value of Y setting A to a versus a^* and in both cases setting M to m	1. No unmeasured confounding between A ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M ($M \perp Y_{a,m} \mid W, A$).
Natural direct effect $E(Y_{a,M_{a^*}}) - E(Y_{a^*,M_{a^*}})$	Difference in the expected value of Y setting A to a versus a^* and in both cases letting M be the value that it would naturally be under a^*	1. No unmeasured confounding between A ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M ($M \perp Y_{a,m} \mid W, A$). 3. No unmeasured confounding of $A - M$ ($A \perp M_a \mid W$). 4. No measured or unmeasured posttreatm confounding of the $M - Y$ relationship ($M_{a^*} \perp Y_{a,m} \mid W$). 5. Y_a is equivalent to Y_{a,M_a}.
Natural indirect effect $E(Y_{a,M_a}) - E(Y_{a,M_{a^*}})$	Difference in the expected value of Y in both cases setting A to a and contrasting M under a versus a^*	
Interventional direct effect $E(Y_{a,g_{M\mid a^*,W}}) - E(Y_{a^*,g_{M\mid a^*,W}})$	Difference in the population average of Y setting A to a versus a^* and in both cases drawing the value of M from a distribution of M conditional on $A = a^*$ and the individual's set of covariate values, W	1. No unmeasured confounding between A ($A \perp Y_{a,m} \mid W$). 2. No unmeasured confounding between M ($M \perp Y_{a,m} \mid W, A$). 3. No unmeasured confounding of $A - M$ ($A \perp M_a \mid W$).
Interventional indirect effect $E(Y_{a,g_{M\mid a,W}}) - E(Y_{a,g_{M\mid a^*,W}})$	Difference in the population average of Y in both cases setting A to a and contrasting drawing the value of M from a distribution of M conditional on $A = a$ versus $A = a^*$ and the individual's set of covariate values, W	

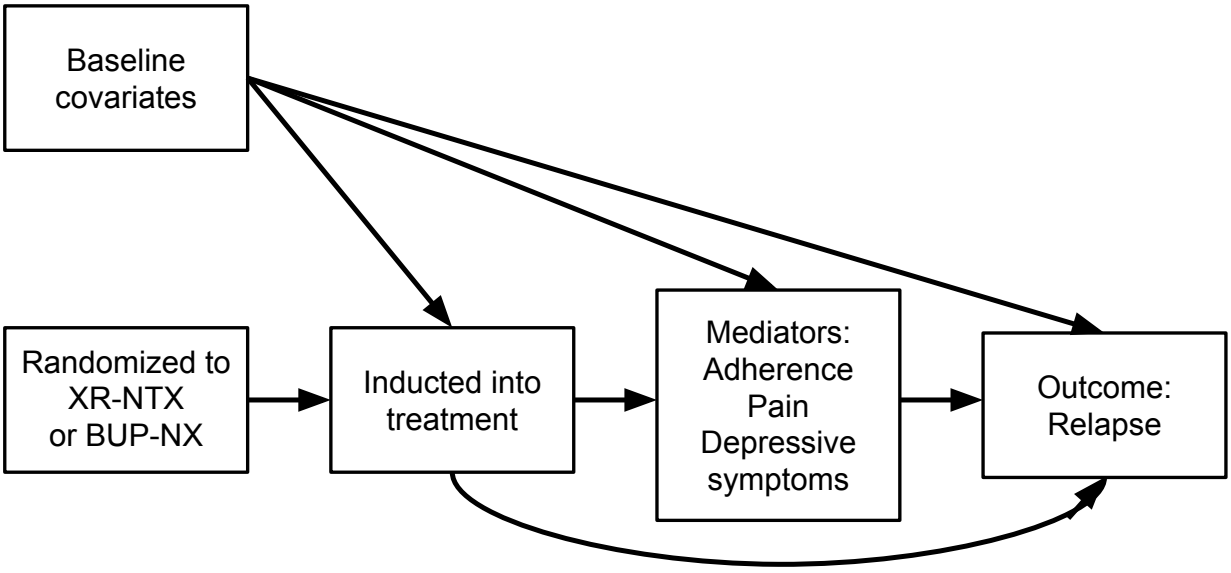
Abbreviations: A, treatment; M, mediator; W, covariates; Y, outcome.

Figure 2.4: Excerpted from @rudolph2019causal

Chapter 3

How to choose an estimand: Real-world example

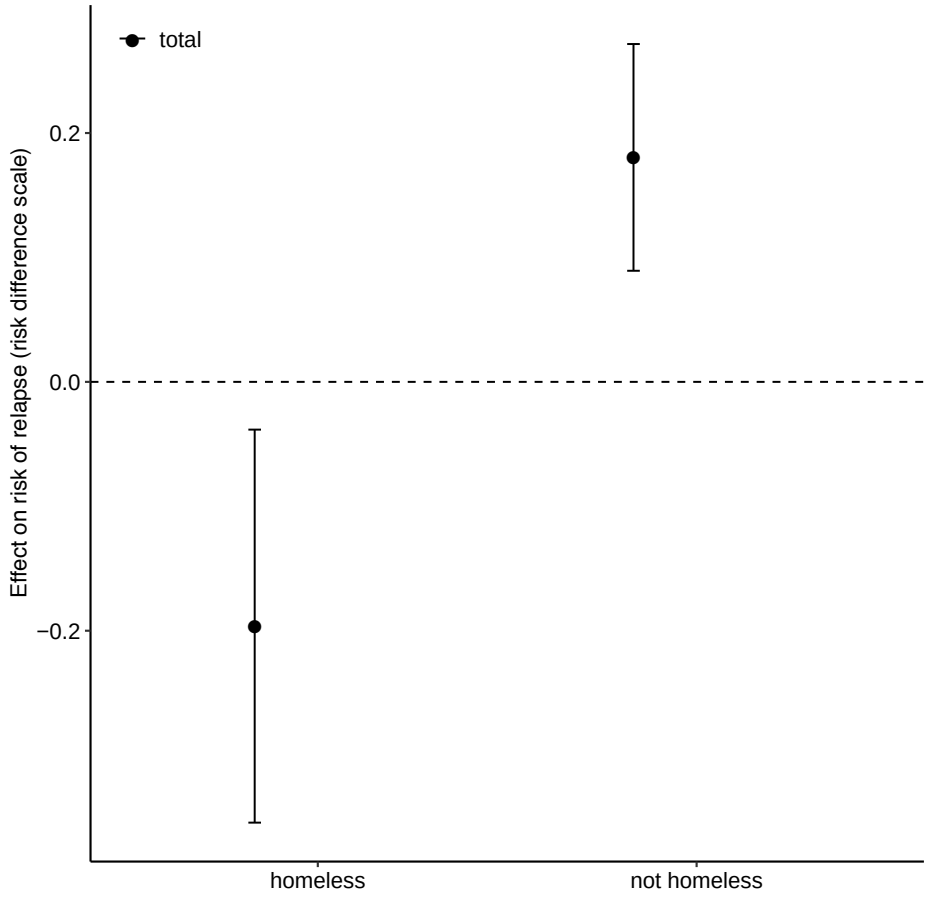
3.1 Comparative effectiveness of two medications for opioid use disorder (OUD)



Motivation: Opposite overall treatment effects for homeless versus nonhomeless participants. This application was explored in detail by Rudolph et al. (2020).

3.1.1 Getting specific about the question

To what extent does the indirect effect through mediators of adherence, pain, and depressive symptoms explain the differences in treatment effects on OUD relapse for homeless and nonhomeless individuals?



What estimand do we want?

- Can we set $M = m$ (i.e., same value) for everyone?
- Are we interested in estimating indirect effects?

→ So, *not* controlled direct effect.

- Do we have an intermediate confounder?
  - Yes, and it's important.
- Do we have a binary treatment assignment variable and a binary intermediate confounder?
  - Yes.
- Can we assume monotonicity?
  - Yes.

→ So, could estimate natural (in)direct effects under monotonicity.

What if we don't want to assume monotonicity, or if we do not have a binary treatment assignment variable and binary intermediate confounder?

→ Interventional direct and indirect effects.

- Do we want to estimate the path through treatment initiation (Z)?
  - Yes, so, *not* the conditional versions of these effects.
  - Estimands:
    - Direct effect: $\mathbb{E}(Y_{1,G_0} - Y_{0,G_0})$
    - Indirect effect: $\mathbb{E}(Y_{1,G_1} - Y_{1,G_0})$
  - Here G_a is a draw from the distribution of $M_a \mid W$.
- Need to incorporate multiple and continuous mediators

What if the positivity assumption $\mathbb{P}(A = a \mid W) > 0$ violated?

→ Can't identify or estimate any of the above effects

- But we can estimate the effect of some stochastic interventions, e.g., IPSIs
- Tradeoff between feasibility and interpretation

What if the exposure variable is continuous?

→ All the above effects are defined for binary exposures

- But we can estimate the effect of some stochastic interventions
- Work in progress (including upcoming R software)

What if the exposure is actually time-varying? What if the mediators and/or intermediate confounders are actually time-varying?

- Longitudinal causal mediation
- Recent paper extending interventional (in)direct effects to longitudinal setting with R software https://arxiv.org/abs/2203.15085

Chapter 4

Estimation preliminaries: review of doubly robust estimators for the average treatment effect

Recall our motivation for doing mediation analysis. We would like to decompose the total effect of a treatment A on an outcome Y into effects that operate through a mediator M vs effects that operate independently of M.

Recall that we define the *average treatment effect* as $E(Y_1 - Y_0)$, and decompose it as follows

$$\mathbb{E}[Y_{1,M_1} - Y_{0,M_0}] = \underbrace{\mathbb{E}[Y_{1,M_1} - Y_{1,M_0}]}_{\text{natural indirect effect}} + \underbrace{\mathbb{E}[Y_{1,M_0} - Y_{0,M_0}]}_{\text{natural direct effect}}$$

To introduce some of the ideas that we will use for estimation of the NDE, let us first briefly discuss estimation of $\mathbb{E}(Y_1)$ (estimation of $\mathbb{E}(Y_0)$ can be performed analogously).

First, notice that under the assumption of no unmeasured confounders ($Y_1 \perp\!\!\!\perp A \mid W$), we have

$$\mathbb{E}(Y_1) = \mathbb{E}[\mathbb{E}(Y \mid A = 1, W)]$$

4.1 Option 1: G-computation estimator

The first estimator of $\mathbb{E}[\mathbb{E}(Y \mid A = 1, W)]$ can be obtained in a three step procedure:

- Fit a regression for Y on A and W
- Use the above regression to predict the outcome mean if everyone's A is set to $A = 1$
- Average these predictions

In formulas, this estimator can be written as

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}(Y \mid A_i = 1, W_i)$$

- Note that this is just a plug-in estimator for the above formula (called the g-formula): $\mathbb{E}[\mathbb{E}(Y \mid A = 1, W)]$
- This estimator requires that the regression model for $\hat{\mathbb{E}}(Y \mid A_i = 1, W_i)$ is correctly specified.
- Downside: If we use machine learning for this model, we do not have general theory for computing standard errors and confidence intervals

4.2 Option 2: Inverse probability weighted estimator

An alternative method of estimation can be constructed after noticing that

$$\mathbb{E}[\mathbb{E}(Y \mid A = 1, W)] = \mathbb{E}\left[\frac{A}{\mathbb{P}(A = 1 \mid W)}Y\right],$$

using the following procedure:

- Fit a regression for A and W
- Use the above regression to predict the probability of treatment $A = 1$
- Compute the inverse probability weights $A_i/\hat{\mathbb{P}}(A_i = 1 \mid W_i)$.
- This weight will be zero for untreated units, and the inverse of the probability of treatment for treated units.
- Compute the weighted average of the outcome:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\hat{\mathbb{P}}(A_i = 1 \mid W_i)}Y_i$$

- This estimator requires that the regression model for $\hat{\mathbb{P}}(A = 1 \mid W_i)$ is correctly specified.
- Downside: If we use machine learning for this model, we do not have general theory for computing standard errors and confidence intervals

4.3 Option 3: Augmented inverse probability weighted estimator

Fortunately, we can combine these two estimators to get an estimator with improved properties.

The improved estimator can be seen both as a *corrected* (or augmented) IPW estimator:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\hat{\mathbb{P}}(A_i = 1 \mid W_i)}Y_i}_{\text{IPW estimator}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mathbb{E}}(Y \mid A_i = 1, W_i)}{\hat{\mathbb{P}}(A_i = 1 \mid W_i)}[A_i - \hat{\mathbb{P}}(A_i = 1 \mid W_i)]}_{\text{Correction term}}$$

or

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{\mathbb{E}}(Y \mid A_i = 1, W_i)}_{\text{G-comp estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\hat{\mathbb{P}}(A_i = 1 \mid W_i)}[Y_i - \hat{\mathbb{E}}(Y \mid A_i = 1, W_i)]}_{\text{Correction term}}$$

This estimator has some desirable properties:

- It is robust to misspecification of at most one of the models (outcome or treatment) (Q: can you see why?)
- It is distributed as a normal random variable as sample size grows. This allows us to easily compute confidence intervals and do hypothesis tests
- It allows us to use machine learning to estimate the treatment and outcome regressions to alleviate model misspecification bias

Next, we will work towards constructing estimators with these same properties for the mediation parameters that we have introduced.

Chapter 5

Construction of G-computation and weighted estimators for the NDE: The case of the natural direct effect

5.1 Recap of definition and identification of the natural direct effect

Recall:

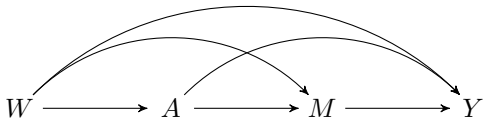


Figure 5.1: Directed acyclic graph under *no intermediate confounders* of the mediator-outcome relation affected by treatment

- Assuming a binary A, we define the natural direct effect as: NDE $= \mathbb{E}(Y_{1,M_0} - Y_{0,M_0})$.
- and the natural indirect effect as: NIE $= \mathbb{E}(Y_{1,M_1} - Y_{1,M_0})$.
- The observed data is $O = (W, A, M, Y)$

This SCM is represented in the above DAG and the following causal models:

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ M &= f_M(W, A, U_M) \\ Y &= f_Y(W, A, M, U_Y), \end{aligned}$$

where (U_W, U_A, U_M, U_Y) are exogenous random errors.

Recall that we need to assume the following to identify the above caual effects from our observed data:

- $A \perp\!\!\!\perp Y_{a,m} \mid W$
- $M \perp\!\!\!\perp Y_{a,m} \mid W, A$
- $A \perp\!\!\!\perp M_a \mid W$
- $M_0 \perp\!\!\!\perp Y_{1,m} \mid W$
- and positivity assumptions

Then, the NDE is identified as

$$\psi(\mathbb{P}) = \mathbb{E}[\mathbb{E}\{\mathbb{E}(Y \mid A = 1, M, W) - \mathbb{E}(Y \mid A = 0, M, W) \mid A = 0, W\}]$$

5.2 From causal to statistical quantities

- We have arrived at identification formulas that express quantities that we care about in terms of observable quantities
- That is, these formulas express what would have happened in hypothetical worlds in terms of quantities observable in this world.
- This required **causal assumptions**
  - Many of these assumptions are empirically unverifiable
  - We saw an example where we could relax the cross-world assumption, at the cost of changing the parameter interpretation (when we introduced randomized interventional direct and indirect effects).
  - We also include an extra section at the end about **stochastic** randomized interventional direct and indirect effects, which allow us to relax the positivity assumption, also at the cost of changing the parameter interpretation.
- We are now ready to tackle the estimation problem, i.e., how do we best learn the value of quantities that are observable?
- The resulting estimation problem can be tackled using **statistical assumptions** of various degrees of strength
  - Most of these assumptions are verifiable (e.g., a linear model)
  - Thus, most are unnecessary (except for convenience)
  - We have worked hard to try to satisfy the required causal assumptions
  - This is not the time to introduce unnecessary statistical assumptions
  - The estimation approach we will minimizes reliance on these statistical assumptions.

5.3 Computing identification formulas if you know the true distribution

- The mediation parameters that we consider can be seen as a function of the joint probability distribution of observed data $O = (W, A, Z, M, Y)$

- For example, under identifiability assumptions the natural direct effect is equal to
$$\psi(\mathbb{P}) = \mathbb{E}[\mathbb{E}\{\mathbb{E}(Y \mid A=1, M, W) - \mathbb{E}(Y \mid A=0, M, W) \mid A=0, W\}]$$

- The notation $\psi(\mathbb{P})$ means that the parameter is a function of $\mathbb{P}$ – in other words, that it is a function of this joint probability distribution
- This means that we can compute it for any distribution $\mathbb{P}$
- For example, if we know the true $\mathbb{P}(W, A, M, Y)$, we can comnpute the true value of the parameter by:
  - Computing the conditional expectation $\mathbb{E}(Y \mid A=1, M=m, W=w)$ for all values (m, w)
  - Computing the conditional expectation $\mathbb{E}(Y \mid A=0, M=m, W=w)$ for all values (m, w)
  - Computing the probability $\mathbb{P}(M=m \mid A=0, W=w)$ for all values (m, w)
  - Compute
$$\mathbb{E}\{\mathbb{E}(Y \mid A=1, M, W) - \mathbb{E}(Y \mid A=0, M, W)| A=0, W\} = \sum_m \{\mathbb{E}(Y \mid A=1, m, w) - \mathbb{E}(Y \mid A=0, m, w)\}\mathbb{P}(M=m \mid A=0, W=w)$$
  - Computing the probability $\mathbb{P}(W=w)$ for all values w
  - Computing the mean over all values w

5.4 Plug-in (a.k.a g-computation) estimator

The above is how you would compute the *true value* **if you know** the true distribution $\mathbb{P}$

- This is exactly what we did in our R examples before
- But we can use the same logic for estimation:
  - Fit a regression to estimate, say $\hat{\mathbb{E}}(Y \mid A=1, M=m, W=w)$
  - Fit a regression to estimate, say $\hat{\mathbb{E}}(Y \mid A=0, M=m, W=w)$
  - Fit a regression to estimate, say $\hat{\mathbb{P}}(M=m \mid A=0, W=w)$
  - Estimate $\mathbb{P}(W=w)$ with the empirical distribution
  - Evaluate
$$\psi(\hat{\mathbb{P}}) = \hat{\mathbb{E}}[\hat{\mathbb{E}}\{\hat{\mathbb{E}}(Y \mid A=1, M, W) - \hat{\mathbb{E}}(Y \mid A=0, M, W) \mid A=0, W\}]$$

- This is known as the G-computation estimator.

5.5 First weighted estimator (akin to inverse probability weighted)

- An alternative expression of the parameter functional (for the NDE) is given by
$$\mathbb{E}\left[\left\{\frac{\mathbb{I}(A=1)}{\mathbb{P}(A=1 \mid W)}\frac{\mathbb{P}(M \mid A=0, W)}{\mathbb{P}(M \mid A=1, W)} - \frac{\mathbb{I}(A=0)}{\mathbb{P}(A=0 \mid W)}\right\} \times Y\right]$$

- Thus, you can also construct a weighted estimator as
$$\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\frac{\mathbb{I}(A_i=1)}{\hat{\mathbb{P}}(A_i=1\mid W_i)}\frac{\hat{\mathbb{P}}(M_i\mid A_i=0,W_i)}{\hat{\mathbb{P}}(M_i\mid A_i=1,W_i)}-\frac{\mathbb{I}(A_i=0)}{\hat{\mathbb{P}}(A_i=0\mid W_i)}\right\}\times Y_i\right]$$

5.6 Second weighted estimator

- The parameter functional for the NDE can also be expressed as a combination of regression and weighting:
$$\mathbb{E}\left[\left\{\frac{\mathbb{I}(A=0)}{\mathbb{P}(A=0\mid W)}\right\}\times\mathbb{E}(Y\mid A=1,M,W)-\mathbb{E}(Y\mid A=0,M,W)\right]$$

- Thus, you can also construct a weighted estimator as
$$\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\frac{\mathbb{I}(A_i=0)}{\hat{\mathbb{P}}(A_i=0\mid W_i)}\right\}\times\hat{\mathbb{E}}(Y\mid A=1,M_i,W_i)-\hat{\mathbb{E}}(Y\mid A=0,M_i,W_i)\right]$$

5.7 How can g-estimation and weighted estimation be implemented in practice?

- There are two possible ways to do G-computation or weighted estimation:
  - Using parametric models for the above regressions
  - Using flexible data-adaptive regression (aka machine learning)

5.8 Pros and cons of G-computation and weighting parametric models

- Pros:
  - Easy to understand
  - Ease of implementation (standard regression software)
  - Can use the Delta method or the bootstrap for computation of standard errors
- Cons:
  - Unless W and M contain very few categorical variables, it is very easy to misspecify the models
  - This can introduce sizable bias in the estimators
  - This modelling assumptions have become less necessary in the presence of data-adaptive regression tools (a.k.a., machine learning)

5.9 An example of the bias of a g-computation estimator of the natural direct effect

- The following R chunk provides simulation code to exemplify the bias of a G-computation parametric estimator in a simple situation

```
mean_y <- function(m, a, w) abs(w) + a * m
mean_m <- function(a, w) plogis(w^2 - a)
pscore <- function(w) plogis(1 - abs(w))
```

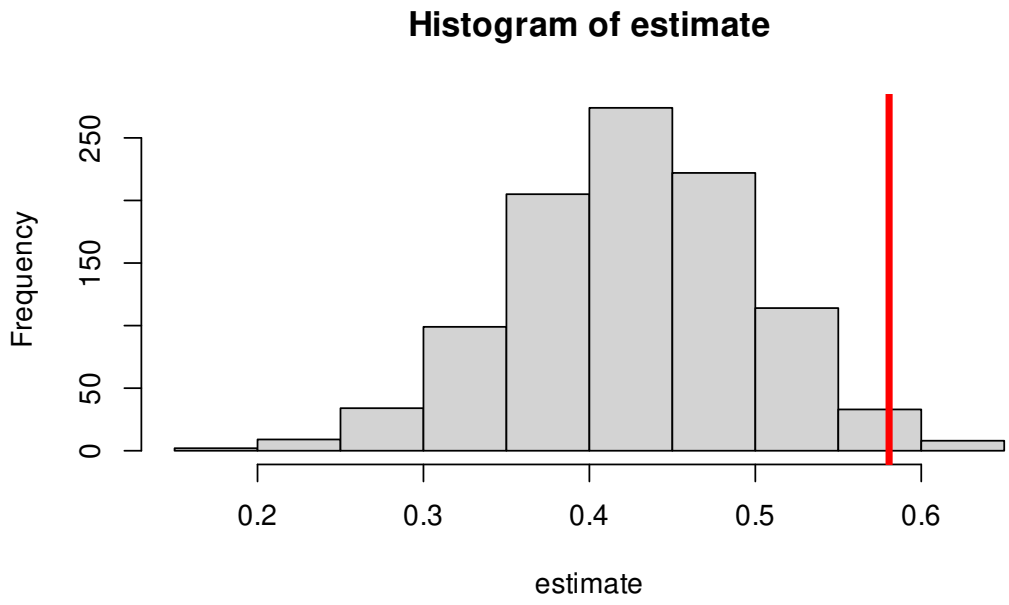
- This yields a true NDE value of 0.58048

```
w_big <- runif(1e6, -1, 1)
trueval <- mean((mean_y(1, 1, w_big) - mean_y(1, 0, w_big)) *
  mean_m(0, w_big) + (mean_y(0, 1, w_big) - mean_y(0, 0, w_big)) *
    (1 - mean_m(0, w_big)))
print(trueval)
#> [1] 0.58062
```

- Let's perform a simulation where we draw 1000 datasets from the above distribution, and compute a g-computation estimator based on

```
gcomp <- function(y, m, a, w) {
  lm_y <- lm(y ~ m + a + w)
  pred_y1 <- predict(lm_y, newdata = data.frame(a = 1, m = m, w = w))
  pred_y0 <- predict(lm_y, newdata = data.frame(a = 0, m = m, w = w))
  pseudo <- pred_y1 - pred_y0
  lm_pseudo <- lm(pseudo ~ a + w)
  pred_pseudo <- predict(lm_pseudo, newdata = data.frame(a = 0, w = w))
  estimate <- mean(pred_pseudo)
  return(estimate)
}

estimate <- lapply(seq_len(1000), function(iter) {
  n <- 1000
  w <- runif(n, -1, 1)
  a <- rbinom(n, 1, pscore(w))
  m <- rbinom(n, 1, mean_m(a, w))
  y <- rnorm(n, mean_y(m, a, w))
  est <- gcomp(y, m, a, w)
  return(est)
})
```

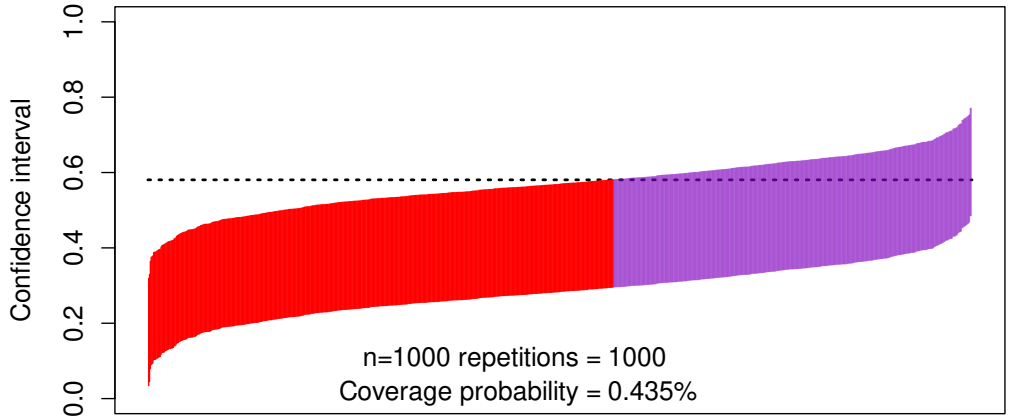
```
estimate <- do.call(c, estimate)

hist(estimate)
abline(v = trueval, col = "red", lwd = 4)
```



- The bias also affects the confidence intervals:

```
cis <- cbind(
  estimate - qnorm(0.975) * sd(estimate),
  estimate + qnorm(0.975) * sd(estimate)
)

ord <- order(rowSums(cis))
lower <- cis[ord, 1]
upper <- cis[ord, 2]
curve(trueval + 0 * x,
  ylim = c(0, 1), xlim = c(0, 1001), lwd = 2, lty = 3, xaxt = "n",
  xlab = "", ylab = "Confidence interval", cex.axis = 1.2, cex.lab = 1.2
)
for (i in 1:1000) {
  clr <- rgb(0.5, 0, 0.75, 0.5)
  if (upper[i] < trueval || lower[i] > trueval) clr <- rgb(1, 0, 0, 1)
  points(rep(i, 2), c(lower[i], upper[i]), type = "l", lty = 1, col = clr)
}
text(450, 0.10, "n=1000 repetitions = 1000 ", cex = 1.2)
text(450, 0.01, paste0(
  "Coverage probability = ",
```

```
  mean(lower < trueval & trueval < upper), "%"
), cex = 1.2)
```



5.10 Pros and cons of G-computation or weighting with data-adaptive regression

- Pros:
  - Easy to understand.
  - Alleviate model-misspecification bias.
- Cons:
  - Might be harder to implement depending on the regression procedures used.
  - No general approaches for computation of standard errors and confidence intervals.
  - For example, the bootstrap is not guaranteed to work, and it is known to fail in some cases.

5.11 Solution to these problems: robust semiparametric efficient estimation

- Intuitively, it combines the three above estimators to obtain an estimator with improved robustness properties
- It offers a way to use data-adaptive regression to
  - avoid model misspecification bias,
  - endow the estimators with additional robustness (e.g., multiple robustness), while
  - allowing the computation of correct standard errors and confidence intervals using Gaussian approximations

Chapter 6

Construction of a semiparametric efficient estimator for the NDE (a.k.a. the one-step estimator)

- Here we show the detail of how to construct an estimator for the NDE for illustration, but the construction of this estimator is a bit involved and may be complex in daily research practice
- For practice, we will teach you how to use our packages *medoutcon* (and *medshift*, as detailed in the extra material) for automatic implementation of these estimators of the NDE and other parameters

First, we need to introduce some notation to describe the EIF for the NDE

- Let $Q(M, W)$ denote $\mathbb{E}(Y \mid A = 1, M, W) - \mathbb{E}(Y \mid A = 0, M, W)$
- We can now introduce the semiparametric efficient estimator:

$$\begin{aligned}\hat{\psi} = &\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(A_i = 1)}{\hat{\mathbb{P}}(A_i = 1 \mid W_i)}\frac{\hat{\mathbb{P}}(M_i \mid A_i = 0, W)_i}{\hat{\mathbb{P}}(M_i \mid A_i = 1, W_i)} - \frac{\mathbb{I}(A = 0)}{\hat{\mathbb{P}}(A_i = 0 \mid W_i)}\right\}[Y_i - \hat{\mathbb{E}}(Y \mid A_i, M_i, W_i)] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}(A = 0)}{\mathbb{P}(A = 0 \mid W)}\left\{\hat{Q}(M_i, W_i) - \hat{\mathbb{E}}[\hat{Q}(M_i, W_i) \mid W_i, A_i = 0]\right\} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbb{E}}[\hat{Q}(M_i, W_i) \mid W_i, A_i = 0]\end{aligned}$$

- In this estimator, you can recognize elements from the G-computation estimator and the weighted estimators:
  - The third line is the G-computation estimator
  - The second line is a centered version of the second weighted estimator
  - The first line is a centered version of the first weighted estimator

- Estimating $\mathbb{P}(M \mid A, W)$ is a very challenging problem when M is high-dimensional. But, since we have the ratio of these conditional densities, we can re-paramterize using Bayes' rule to get something that is easier to compute:

$$\frac{\mathbb{P}(M \mid A=0, W)}{\mathbb{P}(M \mid A=1, W)} = \frac{\mathbb{P}(A=0 \mid M, W)\mathbb{P}(A=1 \mid W)}{\mathbb{P}(A=1 \mid M, W)\mathbb{P}(A=0 \mid W)}.$$

Thus we can change the expression of the estimator a bit as follows. First, some more notation that will be useful later:

- Let $g(a \mid w)$ denote $\mathbb{P}(A=a \mid W=w)$
- Let $e(a \mid m, w)$ denote $\mathbb{P}(A=a \mid M=m, W=w)$
- Let $b(a, m, w)$ denote $\mathbb{E}(Y \mid A=a, M=m, W=w)$
- The quantity being averaged can be re-expressed as follows

$$\begin{aligned}
&\left\{\frac{\mathbb{I}(A=1)}{g(0 \mid W)}\frac{e(0 \mid M, W)}{e(1 \mid M, W)} - \frac{\mathbb{I}(A=0)}{g(0 \mid W)}\right\} \times [Y - b(A, M, W)] \\
&+ \frac{\mathbb{I}(A=0)}{g(0 \mid W)}\Big\{Q(M, W) - \mathbb{E}[Q(M, W) \mid W, A=0]\Big\} \\
&+ \mathbb{E}[Q(M, W) \mid W, A=0]
\end{aligned}$$

6.1 How to compute the one-step estimator (akin to Augmented IPW)

First we will generate some data:

```
mean_y <- function(m, a, w) abs(w) + a * m
mean_m <- function(a, w) plogis(w^2 - a)
pscore <- function(w) plogis(1 - abs(w))

w_big <- runif(1e6, -1, 1)
trueval <- mean((mean_y(1, 1, w_big) - mean_y(1, 0, w_big)) * mean_m(0, w_b
                + (mean_y(0, 1, w_big) - mean_y(0, 0, w_big)) *
                  (1 - mean_m(0, w_big)))

n <- 1000
w <- runif(n, -1, 1)
a <- rbinom(n, 1, pscore(w))
m <- rbinom(n, 1, mean_m(a, w))
y <- rnorm(n, mean_y(m, a, w))
```

Recall that the semiparametric efficient estimator can be computed in the following steps:

1. Fit models for $g(a \mid w)$, $e(a \mid m, w)$, and $b(a, m, w)$

- In this example we will use Generalized Additive Models for tractability
- In applied settings we recommend using an ensemble of data-adaptive regression algorithms, such as the Super Learner (van der Laan et al., 2007)

```
library(mgcv)
## fit model for E(Y | A, W)
b_fit <- gam(y ~ m:a + s(w, by = a))
## fit model for P(A = 1 | M, W)
e_fit <- gam(a ~ m + w + s(w, by = m), family = binomial)
## fit model for P(A = 1 | W)
g_fit <- gam(a ~ w, family = binomial)
```

2. Compute predictions $g(1 \mid w)$, $g(0 \mid w)$, $e(1 \mid m, w)$, $e(0 \mid m, w)$,$b(1, m, w)$, $b(0, m, w)$, and $b(a, m, w)$

```
## Compute P(A = 1 | W)
g1_pred <- predict(g_fit, type = 'response')
## Compute P(A = 0 | W)
g0_pred <- 1 - g1_pred
## Compute P(A = 1 | M, W)
e1_pred <- predict(e_fit, type = 'response')
## Compute P(A = 0 | M, W)
e0_pred <- 1 - e1_pred
## Compute E(Y | A = 1, M, W)
b1_pred <- predict(b_fit, newdata = data.frame(a = 1, m, w))
## Compute E(Y | A = 0, M, W)
b0_pred <- predict(b_fit, newdata = data.frame(a = 0, m, w))
## Compute E(Y | A, M, W)
b_pred  <- predict(b_fit)
```

3. Compute $Q(M, W)$, fit a model for $\mathbb{E}[Q(M, W) \mid W, A]$, and predict at $A = 0$

```
## Compute Q(M, W)
pseudo <- b1_pred - b0_pred
## Fit model for E[Q(M, W) | A, W]
q_fit <- gam(pseudo ~ a + w + s(w, by = a))
## Compute E[Q(M, W) | A = 0, W]
q_pred <- predict(q_fit, newdata = data.frame(a = 0, w = w))
```

4. Estimate the weights

$$\left\{ \frac{\mathbb{I}(A = 1)}{g(0 \mid W)} \frac{e(0 \mid M, W)}{e(1 \mid M, W)} - \frac{\mathbb{I}(A = 0)}{g(0 \mid W)} \right\}$$

using the above predictions:

```
ip_weights <- a / g0_pred * e0_pred / e1_pred - (1 - a) / g0_pred
```

5. Compute the uncentered EIF:

```
eif <- ip_weights * (y - b_pred) + (1 - a) / g0_pred * (pseudo - q_pred) +
  q_pred
```

6. The one step estimator is the mean of the uncentered EIF

```
## One-step estimator
mean(eif)
#> [1] 0.50127
```

6.2 Performance of the one-step estimator in a small simulation study

First, we create a wrapper around the estimator

```
one_step <- function(y, m, a, w) {
  b_fit <- gam(y ~ m:a + s(w, by = a))
  e_fit <- gam(a ~ m + w + s(w, by = m), family = binomial)
  g_fit <- gam(a ~ w, family = binomial)
  g1_pred <- predict(g_fit, type = 'response')
  g0_pred <- 1 - g1_pred
  e1_pred <- predict(e_fit, type = 'response')
  e0_pred <- 1 - e1_pred
  b1_pred <- predict(
    b_fit, newdata = data.frame(a = 1, m, w), type = 'response'
  )
  b0_pred <- predict(
    b_fit, newdata = data.frame(a = 0, m, w), type = 'response'
  )
  b_pred  <- predict(b_fit, type = 'response')
  pseudo <- b1_pred - b0_pred
  q_fit <- gam(pseudo ~ a + w + s(w, by = a))
  q_pred <- predict(q_fit, newdata = data.frame(a = 0, w = w))
  ip_weights <- a / g0_pred * e0_pred / e1_pred - (1 - a) / g0_pred
  eif <- ip_weights * (y - b_pred) + (1 - a) / g0_pred *
    (pseudo - q_pred) + q_pred
  return(mean(eif))
}
```

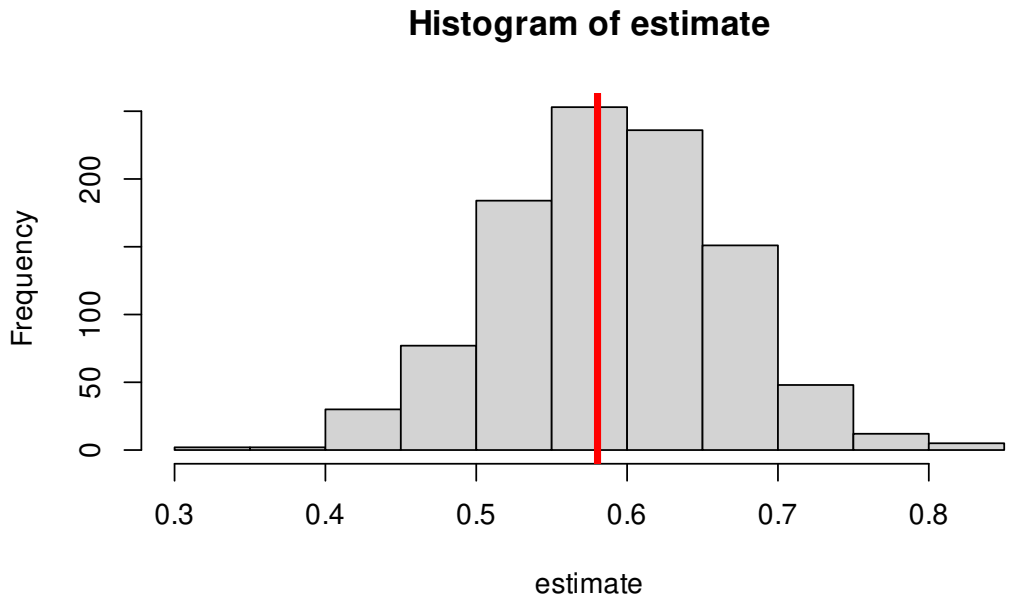
Let us first examine the bias

- The true value is:

```
w_big <- runif(1e6, -1, 1)
trueval <- mean((mean_y(1, 1, w_big) - mean_y(1, 0, w_big)) * mean_m(0, w_b
  + (mean_y(0, 1, w_big) - mean_y(0, 0, w_big)) * (1 - mean_m(0, w_big)))
print(trueval)
#> [1] 0.5804
```

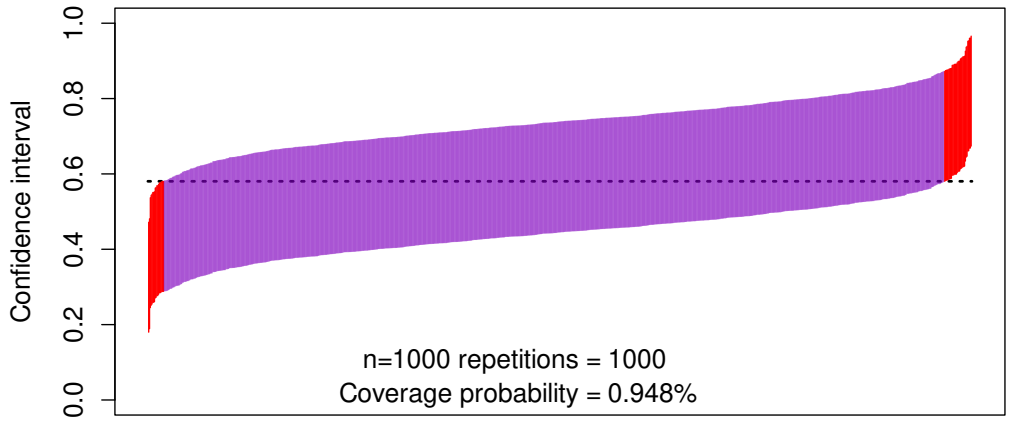
- Bias simulation

```
estimate <- lapply(seq_len(1000), function(iter) {
  n <- 1000
  w <- runif(n, -1, 1)
  a <- rbinom(n, 1, pscore(w))
  m <- rbinom(n, 1, mean_m(a, w))
  y <- rnorm(n, mean_y(m, a, w))
  estimate <- one_step(y, m, a, w)
  return(estimate)
})
estimate <- do.call(c, estimate)

hist(estimate)
abline(v = trueval, col = "red", lwd = 4)
```



- And now the confidence intervals:

```
cis <- cbind(
  estimate - qnorm(0.975) * sd(estimate),
  estimate + qnorm(0.975) * sd(estimate)
)

ord <- order(rowSums(cis))
lower <- cis[ord, 1]
upper <- cis[ord, 2]
curve(trueval + 0 * x,
  ylim = c(0, 1), xlim = c(0, 1001), lwd = 2, lty = 3, xaxt = "n",
  xlab = "", ylab = "Confidence interval", cex.axis = 1.2, cex.lab = 1.2
)
for (i in 1:1000) {
  clr <- rgb(0.5, 0, 0.75, 0.5)
  if (upper[i] < trueval || lower[i] > trueval) clr <- rgb(1, 0, 0, 1)
  points(rep(i, 2), c(lower[i], upper[i]), type = "l", lty = 1, col = clr)
}
text(450, 0.10, "n=1000 repetitions = 1000")
text(450, 0.01, paste0(
  "Coverage probability = ",
  mean(lower < trueval & trueval < upper), "%"
), cex = 1.2)
```



6.3 A note about targeted minimum loss-based estimation (TMLE)

- The above estimator is great because it allows us to use data-adaptive regression to avoid bias, while allowing the computation of correct standard errors
- This estimator has a problem, though:
  - It can yield answers outside of the bounds of the parameter space

  - E.g., if Y is binary, it could yield direct and indirect effects outside of $[-1, 1]$
  - To solve this, you can compute a TMLE instead (implemented in the R packages, coming up)

6.4 A note about cross-fitting

- When using data-adaptive regression estimators, it is recommended to use cross-fitted estimators
- Cross-fitting is similar to cross-validation:
  - Randomly split the sample into K (e.g., K=10) subsets of equal size
  - For each of the 9/10ths of the sample, fit the regression models
  - Use the out-of-sample fit to predict in the remaining 1/10th of the sample
- Cross-fitting further reduces the bias of the estimators
- Cross-fitting aids in guaranteeing the correctness of the standard errors and confidence intervals
- Cross-fitting is implemented by default in the R packages that you will see next

Chapter 7

R packages for estimation of the causal (in)direct effects

We'll now turn to working through a few examples of estimating the natural, interventional, and stochastic direct and indirect effects. As our running example, we'll a simple data set from an observational study of the relationship between BMI and kids' behavior, freely distributed with the mma R package on CRAN[1]. First, let's load the packages we'll be using and set a seed; then, load this data set and take a quick look

```
library(tidyverse)
library(sl3)
library(medoutcon)
library(medshift)
library(mma)
set.seed(429153)

# load and examine data
data(weight_behavior)
dim(weight_behavior)
#> [1] 691  15

# drop missing values
weight_behavior <- weight_behavior %>%
  drop_na() %>%
  as_tibble()
weight_behavior
#> # A tibble: 567 x 15
#>      bmi   age sex   race  numpeople   car gotosch snack tvhours cmpthour
#>    <dbl> <dbl> <fct> <fct>     <int> <int> <fct>   <fct>   <dbl>     <dbl>
```

[1] https://CRAN.R-project.org/package=mma

```
#>  1   18.2   12.2  F       OTHER           5       3  2         1
4           0
#>  2   22.8   12.8  M       OTHER           4       3  2         1
4           2
#>  3   25.6   12.1  M       OTHER           2       3  2         1
0           2
#>  4   15.1   12.3  M       OTHER           4       1  2         1
2           1
#>  5   23.0   11.8  M       OTHER           4       1  1         1
4           3
#>  # ... with 562 more rows, and 5 more variables: cellhours <dbl>, sports <
#>  #   exercises <int>, sweat <int>, overweigh <dbl>
```

The documentation for the data set describes it as a "database obtained from the Louisiana State University Health Sciences Center, New Orleans, by Dr. Richard Scribner. He explored the relationship between BMI and kids' behavior through a survey at children, teachers and parents in Grenada in 2014. This data set includes 691 observations and 15 variables." Note that the data set contained several observations with missing values, which we removed above to simplify the demonstration of our analytic methods. In practice, we recommend instead using appropriate corrections (e.g., imputation, inverse weighting) to fully take advantage of the observed data.

Following the motivation of the original study, we focus on the causal effects of participating in a sports team (sports) on the BMI of children (bmi), taking into consideration several mediators (snack, exercises , overweigh); all other measured covariates are taken to be potential baseline confounders.

7.1 `medoutcon`: Natural and interventional (in)direct effects

The data on a single observational unit can be represented $O = (W, A, M, Y)$, with the data pooled across all participants denoted $O_1, \ldots, O_n$, for a of n i.i.d. observations of O. Recall the DAG from an earlier chapter, which represents the data-generating process:

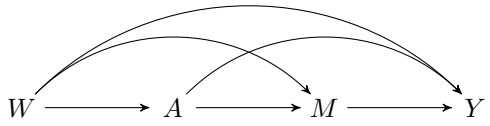


Figure 7.1: Directed acyclic graph under *no intermediate confounders* of the mediator-outcome relation affected by treatment

7.1.1 Natural (in)direct effects

To start, we will consider estimation of the *natural* direct and indirect effects, which, we recall, are defined as follows

$$\mathbb{E}[Y_{1,M_1} - Y_{0,M_0}] = \underbrace{\mathbb{E}[Y_{1,M_1} - Y_{1,M_0}]}_{\text{natural indirect effect}} + \underbrace{\mathbb{E}[Y_{1,M_0} - Y_{0,M_0}]}_{\text{natural direct effect}}.$$

- Our medoutcon R package[2] (Hejazi et al., 2022,?), which accompanies Díaz et al. (2020), implements one-step and TML estimators of both the natural and interventional (in)direct effects.
- Both types of estimators are capable of accommodating flexible modeling strategies (e.g., ensemble machine learning) for the initial estimation of nuisance parameters.
- The medoutcon R package uses cross-validation in initial estimation: this results in cross-validated (or "cross-fitted") one-step and TML estimators (Klaassen, 1987; Zheng and van der Laan, 2011; Chernozhukov et al., 2018), which exhibit greater robustness than their non-sample-splitting analogs.
- To this end, medoutcon integrates with the sl3 R package (Coyle et al., 2022), which is extensively documented in this book chapter[3] (Phillips, 2022; van der Laan et al., 2022).

7.1.2 Interlude: `sl3` for nuisance parameter estimation

- To fully take advantage of the one-step and TML estimators, we'd like to rely on flexible, data adaptive strategies for nuisance parameter estimation.
- Doing so minimizes opportunities for model misspecification to compromise our analytic conclusions.
- Choosing among the diversity of available machine learning algorithms can be challenging, so we recommend using the Super Learner algorithm for ensemble machine learning (van der Laan et al., 2007), which is implemented in the sl3 R package[4] (Coyle et al., 2022).
- Below, we demonstrate the construction of an ensemble learner based on a limited library of algorithms, including n intercept model, a main terms GLM, Lasso (ℓ_1-penalized) regression, and random forest (ranger).

```
# instantiate learners
mean_lrnr <- Lrnr_mean$new()
fglm_lrnr <- Lrnr_glm_fast$new()
lasso_lrnr <- Lrnr_glmnet$new(alpha = 1, nfolds = 3)
rf_lrnr <- Lrnr_ranger$new(num.trees = 200)

# create learner library and instantiate super learner ensemble
lrnr_lib <- Stack$new(mean_lrnr, fglm_lrnr, lasso_lrnr, rf_lrnr)
sl_lrnr <- Lrnr_sl$new(learners = lrnr_lib, metalearner = Lrnr_nnls$new())
```

[2] https://github.com/nhejazi/medoutcon
[3] https://tlverse.org/tlverse-handbook/sl3
[4] https://github.com/tlverse/sl3

- Of course, there are many alternatives for learning algorithms to be included in such a modeling library. Feel free to explore!

7.1.3 Efficient estimation of the natural (in)direct effects

- Estimation of the natural direct and indirect effects requires estimation of a few nuisance parameters. Recall that these are
  - $g(a \mid w)$, which denotes $\mathbb{P}(A = a \mid W = w)$
  - $h(a \mid m, w)$, which denotes $\mathbb{P}(A = a \mid M = m, W = w)$
  - $b(a, m, w)$, which denotes $\mathbb{E}(Y \mid A = a, M = m, W = w)$
- While we recommend the use of Super Learning, we opt to instead estimate all nuisance parameters with Lasso regression below (to save computational time).
- Now, let's use the medoutcon() function to estimate the *natural direct effect*:

```
# compute one-step estimate of the natural direct effect
nde_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = NULL,
  M = weight_behavior[, c("snack", "exercises", "overweigh")],
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "direct",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
summary(nde_onestep)
#> # A tibble: 1 x 7
#>   lwr_ci param_est upr_ci var_est  eif_mean estimator param
#>    <dbl>     <dbl>  <dbl>   <dbl>     <dbl> <chr>     <chr>
#> 1 -0.600   -0.0880  0.424  0.0683 -5.25e-16 onestep   direct_natural
```

- We can similarly call medoutcon() to estimate the *natural indirect effect*:

```
# compute one-step estimate of the natural indirect effect
nie_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = NULL,
  M = weight_behavior[, c("snack", "exercises", "overweigh")],
```

```
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "indirect",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
summary(nie_onestep)
#> # A tibble: 1 x 7
#>   lwr_ci param_est upr_ci var_est eif_mean estimator param
#>    <dbl>     <dbl>  <dbl>   <dbl>    <dbl> <chr>     <chr>
#> 1  0.477      1.04   1.61  0.0838 3.48e-16 onestep   indirect_natural
```

- From the above, we can conclude that the effect of participation on a sports team on BMI is primarily mediated by the variables snack, exercises , and overweigh, as the natural indirect effect is several times larger than the natural direct effect.
- Note that we could have instead used the TML estimators, which have improved finite-sample performance, instead of the one-step estimators. Doing this is as simple as setting the estimator = "tmle" in the relevant argument.

7.1.4 Interventional (in)direct effects

Since our knowledge of the system under study is incomplete, we might worry that one (or more) of the measured variables are not mediators, but, in fact, intermediate confounders affected by treatment. While the natural (in)direct effects are not identified in this setting, their interventional (in)direct counterparts are, as we saw in an earlier section. Recall that both types of effects are defined by static interventions on the treatment. The interventional effects are distinguished by their use of a stochastic intervention on the mediator to aid in their identification.

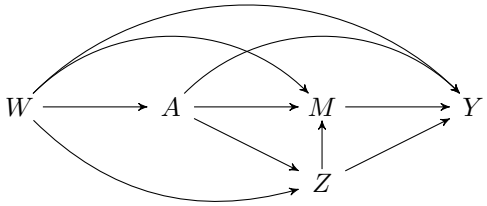


Figure 7.2: Directed acyclic graph under intermediate confounders of the mediator-outcome relation affected by treatment

Recall that the interventional (in)direct effects are defined via the decomposition:

$$\mathbb{E}[Y_{1,G_1} - Y_{0,G_0}] = \underbrace{\mathbb{E}[Y_{1,G_1} - Y_{1,G_0}]}_{\text{interventional indirect effect}} + \underbrace{\mathbb{E}[Y_{1,G_0} - Y_{0,G_0}]}_{\text{interventional direct effect}}$$

- In our data example, we'll consider the eating of snacks as a potential intermediate confounder, since one might reasonably hypothesize that participation on a sports team might

subsequently affect snacking, which then could affect mediators like the amount of exercises and overweight status.
- The interventional direct and indirect effects may also be easily estimated with the medoutcon R package[5] (Hejazi et al., 2022,?).
- Just as for the natural (in)direct effects, medoutcon implements cross-validated one-step and TML estimators of the interventional effects.

7.1.5 Efficient estimation of the interventional (in)direct effects

- Estimation of these effects is more complex, so a few additional nuisance parameters arise when expressing the (more general) EIF for these effects:
  - $q(z \mid a, w)$, the conditional density of the intermediate confounders, conditional only on treatment and baseline covariates;
  - $r(z \mid a, m, w)$, the conditional density of the intermediate confounders, conditional on mediators, treatment, and baseline covariates.
- To estimate the interventional effects, we only need to set the argument Z of medoutcon to a value other than NULL.
- Note that the implementation in medoutcon is currently limited to settings with only binary intermediate confounders, i.e., $Z \in \{0, 1\}$.
- Let's use medoutcon() to estimate the *interventional direct effect*:

```
# compute one-step estimate of the interventional direct effect
interv_de_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = (as.numeric(weight_behavior$snack) - 1),
  M = weight_behavior[, c("exercises", "overweigh")],
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "direct",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
summary(interv_de_onestep)
#> # A tibble: 1 x 7
#>   lwr_ci param_est upr_ci var_est eif_mean estimator param
#>    <dbl>     <dbl>  <dbl>   <dbl>    <dbl> <chr>     <chr>
#> 1 -0.309     0.239  0.788  0.0782 1.94e-15 onestep   direct_interventiona
```

[5] https://github.com/nhejazi/medoutcon

- We can similarly estimate the *interventional indirect effect*:

```
# compute one-step estimate of the interventional indirect effect
interv_ie_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = (as.numeric(weight_behavior$snack) - 1),
  M = weight_behavior[, c("exercises", "overweigh")],
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "indirect",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
summary(interv_ie_onestep)
#> # A tibble: 1 x 7
#>   lwr_ci param_est upr_ci var_est eif_mean estimator param
#>    <dbl>     <dbl>  <dbl>   <dbl>    <dbl> <chr>     <chr>
#> 1  0.524      1.06   1.60  0.0758 1.69e-16 onestep   indirect_interventio
```

- From the above, we can conclude that the effect of participation on a sports team on BMI is largely through the interventional indirect effect (i.e., through the pathways involving the mediating variables) rather than via its direct effect.
- Just as before, we could have instead used the TML estimators, instead of the one-step estimators. Doing this is as simple as setting the estimator = "tmle" in the relevant argument.

Chapter 8

Appendix: Additional topics of interest

The literature on mediation analysis has grown considerably in the last few decades and there are now many novel methods to tackle important questions with complex data structures. While we are unable to cover all these interesting methods in this workshop, here we provide a few references for further reading.

This list is not meant to be comprehensive, just some of our own work and some work by others that we know and consider interesting.

8.1 Mediation with time-varying treatments, mediators, and covariates

The issue of intermediate confounding is exacerbated in a setting with multiple treatments and mediators measured at various time points. Here are a couple of papers on this topic that are interesting:

- Mediation analysis with time varying exposures and mediators[1] by Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen
- Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes[2] by Wenjing Zheng and Mark J. van der Laan
- Efficient and flexible causal mediation with time-varying mediators, treatments, and confounders[3] by Iván Díaz, Nicholas Williams, and Kara E. Rudolph

[1] https://rss.onlinelibrary.wiley.com/doi/full/10.1111/rssb.12194
[2] https://www.degruyter.com/document/doi/10.1515/jci-2016-0006/html
[3] https://arxiv.org/abs/2203.15085

8.2 Mediation with monotonicity of A-Z relationship

- On identification of natural direct effects when a confounder of the mediator is directly affected by exposure[4] by Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele
- Efficient and flexible estimation of natural mediation effects under intermediate confounding and monotonicity constraints[5] by Kara E. Rudolph and Iván Díaz

8.3 Mediation with instrumental variables

- Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables[6] by Markus Frolich and Martin Huber
- Causal mediation with instrumental variables[7] by Kara E. Rudolph, Nicholas Williams, and Iván Díaz

8.4 Mediation with separable effects

- An Interventionist Approach to Mediation Analysis[8] by James M. Robins, Thomas S. Richardson, and Ilya Shpitser
- Conditional Separable Effects[9] by Mats J. Stensrud, James M. Robins, Aaron Sarvet, Eric J. Tchetgen Tchetgen, and Jessica G. Young
- Separable Effects for Causal Inference in the Presence of Competing Events[10] by Mats J. Stensrud, Jessica G. Young, Vanessa Didelez, James M. Robins, and Miguel A. Hernán
- A Generalized Theory of Separable Effects in Competing Event Settings[11] by Mats J. Stensrud, Miguel A. Hernán, Eric J. Tchetgen Tchetgen, James M. Robins, Vanessa Didelez, and Jessica G. Young

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4230499/
[5] https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13850
[6] https://rss.onlinelibrary.wiley.com/doi/full/10.1111/rssb.12232
[7] https://arxiv.org/abs/2112.13898
[8] https://dl.acm.org/doi/abs/10.1145/3501714.3501754
[9] https://www.tandfonline.com/doi/abs/10.1080/01621459.2022.2071276
[10] https://www.tandfonline.com/doi/abs/10.1080/01621459.2020.1765783
[11] https://link.springer.com/article/10.1007/s10985-021-09530-8

Chapter 9

Appendix: Stochastic direct and indirect effects

9.1 Definition of the effects

Consider the following directed acyclic graph.

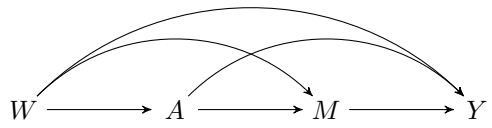


Figure 9.1: Directed acyclic graph under no intermediate confounders of the mediator-outcome relation affected by treatment

9.2 Motivation for stochastic interventions

- So far we have discussed controlled, natural, and interventional (in)direct effects
- These effects require that $0 < \mathbb{P}(A = 1 \mid W) < 1$
- They are defined only for binary exposures
- *What can we do when the positivity assumption does not hold or the exposure is continuous?*
- Solution: We can use stochastic effects

9.3 Definition of stochastic effects

There are two possible ways of defining stochastic effects:

- Consider the effect of an intervention where the exposure is drawn from a distribution

  - For example incremental propensity score interventions

- Consider the effect of an intervention where the post-intervention exposure is a function of the actually received exposure
  - For example modified treatment policies
- In both cases $A \mid W$ is a non-deterministic intervention, thus the name *stochastic intervention*

9.3.1 Example: incremental propensity score interventions (IPSI) (Kennedy, 2018)

Definition of the intervention

- Assume A is binary, and $\mathbb{P}(A = 1 \mid W = w) = g(1 \mid w)$ is the propensity score
- Consider an intervention in which each individual receives the intervention with probability $g_\delta(1 \mid w)$, equal to
$$g_\delta(1 \mid w) = \frac{\delta g(1 \mid w)}{\delta g(1 \mid w) + 1 - g(1 \mid w)}$$
- e.g., draw the post-intervention exposure from a Bernoulli variable with probability $g_\delta(1 \mid w)$
- The value δ is user given
- Let A_δ denote the post-intervention exposure distribution
- Some algebra shows that δ is an odds ratio comparing the pre- and post-intervention exposure distributions
$$\delta = \frac{\text{odds}(A_\delta = 1 \mid W = w)}{\text{odds}(A = 1 \mid W = w)}$$
- Interpretation: *what would happen in a world where the odds of receiving treatment is increased by* δ
- Let Y_{A_δ} denote the outcome in this hypothetical world

9.3.1.1 Illustrative application for IPSIs

- Consider the effect of participation in sports on children's BMI
- Mediation through snacking, exercising, etc.
- Intervention: for each individual, increase the odds of participating in sports by $\delta = 2$
- The post-intervention exposure is a draw A_δ from a Bernoulli distribution with probability $g_\delta(1 \mid w)$

Example: modified treatment policies (MTP) (Díaz and Hejazi, 2020)

Definition of the intervention

- Consider a continuous exposure A taking values in the real numbers

- Consider an intervention that assigns exposure as $A_\delta = A - \delta$
- Example: A is pollution measured as $PM_{2.5}$ and you are interested in an intervention that reduces $PM_{2.5}$ concentration by some amount δ

9.3.2 Mediation analysis for stochastic interventions

- The total effect of an IPSI can be computed as a contrast of the outcome under intervention vs no intervention:
$$\psi = \mathbb{E}[Y_{A_\delta} - Y]$$
- Recall the NPSEM
$$W = f_W(U_W) \tag{9.1}$$
$$A = f_A(W, U_A) \tag{9.2}$$
$$M = f_M(W, A, U_M) \tag{9.3}$$
$$Y = f_Y(W, A, M, U_Y) \tag{9.4}$$
- From this we have
$$M_{A_\delta} = f_M(W, A_\delta, U_M)$$
$$Y_{A_\delta} = f_Y(W, A_\delta, M_{A_\delta}, U_Y)$$
- Thus, we have $Y_{A_\delta} = Y_{A_\delta, M_{A_\delta}}$ and $Y = Y_{A, M_A}$
- Let us introduce the counterfactual $Y_{A_\delta, M}$, interpreted as the outcome observed in a world where the intervention on A is performed but the mediator is fixed at the value it would have taken under no intervention:
$$Y_{A_\delta, M} = f_Y(W, A_\delta, M, U_Y)$$
- Then we can decompose the total effect into:
$$\mathbb{E}[Y_{A_\delta, M_{A_\delta}} - Y_{A, M_A}] = \underbrace{\mathbb{E}[Y_{A_\delta, M_{A_\delta}} - Y_{A_\delta, M}]}_{\text{stochastic natural indirect effect}} + \underbrace{\mathbb{E}[Y_{A_\delta, M} - Y_{A, M}]}_{\text{stochastic natural direct effect}}$$

9.4 Identification assumptions

- Confounder assumptions:
  - $A \perp\!\!\!\perp Y_{a,m} \mid W$
  - $M \perp\!\!\!\perp Y_{a,m} \mid W, A$
- No confounder of $M \to Y$ affected by A
- Positivity assumptions:
  - If $g_\delta(a \mid w) > 0$ then $g(a \mid w) > 0$
  - If $\mathbb{P}(M = m \mid W = w) > 0$ then $\mathbb{P}(M = m \mid A = a, W = w) > 0$

Under these assumptions, stochastic effects are identified as follows

- The indirect effect can be identified as follows

$$\mathbb{E}(Y_{A_\delta} - Y_{A_\delta,M}) =$$
$$\mathbb{E}\left[\sum_a \{\mathbb{E}(Y \mid A = a, W) - \mathbb{E}(Y \mid A = a, M, W)\} g_\delta(a \mid W)\right]$$

- The direct effect can be identified as follows

$$\mathbb{E}(Y_{A_\delta} - Y_{A_\delta,M}) =$$
$$\mathbb{E}\left[\sum_a \{\mathbb{E}(Y \mid A = a, M, W) - Y\} g_\delta(a \mid W)\right]$$

- Let's dissect the formula for the indirect effect in R:

```
n <- 1e6
w <- rnorm(n)
a <- rbinom(n, 1, plogis(1 + w))
m <- rnorm(n, w + a)
y <- rnorm(n, w + a + m)
```

- First, fit regressions of the outcome on (A, W) and (M, A, W):

```
fit_y1 <- lm(y ~ m + a + w)
fit_y2 <- lm(y ~ a + w)
```

- Get predictions fixing $A = a$ for all possible values a

```
pred_y1_a1 <- predict(fit_y1, newdata = data.frame(a = 1, m, w))
pred_y1_a0 <- predict(fit_y1, newdata = data.frame(a = 0, m, w))
pred_y2_a1 <- predict(fit_y2, newdata = data.frame(a = 1, w))
pred_y2_a0 <- predict(fit_y2, newdata = data.frame(a = 0, w))
```

- Compute

$$\{\mathbb{E}(Y \mid A = a, W) - \mathbb{E}(Y \mid A = a, M, W)\}$$

  for each value a

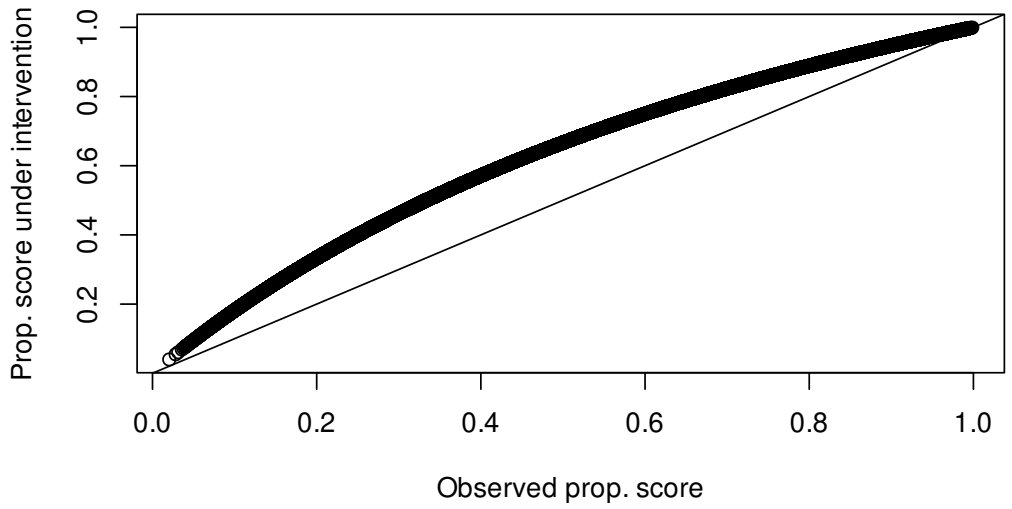
```
pseudo_a1 <- pred_y2_a1 - pred_y1_a1
pseudo_a0 <- pred_y2_a0 - pred_y1_a0
```

- Estimate the propensity score $g(1 \mid w)$ and evaluate the post-intervention propensity score $g_\delta(1 \mid w)$

```
pscore_fit <- glm(a ~ w, family = binomial())
pscore <- predict(pscore_fit, type = 'response')
## How do the intervention vs observed propensity score compare
pscore_delta <- 2 * pscore / (2 * pscore + 1 - pscore)
```

- What do the post-intervention propensity scores look like?

```
plot(pscore, pscore_delta, xlab = 'Observed prop. score',
     ylab = 'Prop. score under intervention')
abline(0, 1)
```



9.5 What are the odds of exposure under intervention vs real world?

```
odds <- (pscore_delta / (1 - pscore_delta)) / (pscore / (1 - pscore))
summary(odds)
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>       2       2       2       2       2       2
```

- Compute the sum $\sum_a \{\mathbb{E}(Y \mid A = a, W) - \mathbb{E}(Y \mid A = a, M, W)\} g_\delta(a \mid W)$

```
indirect <- pseudo_a1 * pscore_delta + pseudo_a0 * (1 - pscore_delta)
```

- The average of this value is the indirect effect

```
## E[Y(Adelta) - Y(Adelta, M)]
mean(indirect)
#> [1] 0.1091
```

- The direct effect is

$$\mathbb{E}(Y_{A_\delta} - Y_{A_\delta,M}) = \mathbb{E}\left[\sum_a \{\mathbb{E}(Y \mid A = a, M, W) - Y\} g_\delta(a \mid W)\right]$$

- Which can be computed as

```
direct <- (pred_y1_a1 - y) * pscore_delta +
      (pred_y1_a0 - y) * (1 - pscore_delta)
mean(direct)
#> [1] 0.10934
```

9.6 Summary

- Stochastic (in)direct effects
  - Relax the positivity assumption
  - Can be defined for non-binary exposures
  - Do not require a cross-world assumption
- Still require the absence of intermediate confounders
  - But, compared to the NDE and NIE, we can design a randomized study where identifiability assumptions hold, at least in principle
  - There is a version of these effects that can accommodate intermediate confounders (Hejazi et al., 2022)
  - R implementation to be released soon... stay tuned!

Bibliography

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21,**.

Coyle, J. R., Hejazi, N. S., Malenica, I., Phillips, R. V., and Sofrygin, O. (2022). *sl3: Modern Pipelines for Machine Learning and Super Learning*. R package version 1.4.5.

Díaz, I. and Hejazi, N. S. (2020). Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82,** 661–683.

Díaz, I., Hejazi, N. S., Rudolph, K. E., and van der Laan, M. J. (2020). Non-parametric efficient causal mediation with intermediate confounders. *Biometrika* .

Hejazi, N. S., Díaz, I., and Rudolph, K. E. (2022). medoutcon: Efficient natural and interventional causal mediation analysis. R package version 0.1.6.

Hejazi, N. S., Rudolph, K. E., and Díaz, I. (2022). medoutcon: Nonparametric efficient causal mediation analysis with machine learning in R. *Journal of Open Source Software* .

Hejazi, N. S., Rudolph, K. E., van der Laan, M. J., and Díaz, I. (2022). Nonparametric causal mediation analysis for stochastic interventional (in) direct effects. *Biostatistics* **(in press),**.

Kennedy, E. H. (2018). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association* .

Klaassen, C. A. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *The Annals of Statistics* pages 1548–1562.

Miles, C. H. (2022). On the causal interpretation of randomized interventional indirect effects. *arXiv preprint arXiv:2203.00245* .

Phillips, R. V. (2022). Super (machine) learning. In *Targeted Learning in* `R`*: Causal Data Science with the* `tlverse` *Software Ecosystem*. Springer.

Rudolph, K., Diaz, I., Hejazi, N., van der Laan, M., Luo, S., Shulman, M., Campbell, A., Rotrosen, J., and Nunes, E. (2020). Explaining differential effects of medication for opioid use disorder using a novel approach incorporating mediating variables. *Addiction* .

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2014). On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology* **25,** 282.

van der Laan, M. J., Coyle, J. R., Hejazi, N. S., Malenica, I., Phillips, R. V., and Hubbard, A. E. (2022). *Targeted Learning in `R`: Causal Data Science with the `tlverse` Software Ecosystem*. CRC Press. in preparation.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology* **6,**.

Zheng, W. and van der Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pages 459–474. Springer.