Research Design, Fall 2021

03: elements of causal inference; experiments

Scott Spencer | Columbia University

goals of data science research

goals of data science research

descriptive

What do the data *describe* about the events that *already generated* that data?

associative

What do the data suggest about *correlations* between measured events?

predictive

What do the data suggest about the likelihood of what *may happen next*?

explicative

What do the data suggest about the *cause(s)* of measured events?

What is causation?

CAUSE, N. | That which produces an effect; that which gives rise to any action, phenomenon, or condition. *Cause* and *effect* are correlative terms.

How can we learn or test if thing A causes thing B?

causal inference and experiments

Causal effects involve the comparison of the outcome actually observed with other potential outcomes that could have been observed had the treatment taken on a different level, but that are not, in fact, observed. Causal inference is therefore fundamentally a missing data problem.

— Imbens & Rubin

causal inference, which concerns what *would happen* to an outcome y as a result of a treatment, intervention, or exposure z, given pre-treatment information x.

— Gelman, Hill, Ventari

What's a *treatment*? Why can't we observe these *potential* outcomes, these *missing* data?

The Road Not Taken

Two roads diverged in a yellow wood,
And *sorry I could not travel both*
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

— Robert Frost

 common notation for causal inference in experiments

i, an experimental unit

$z = 0$, the control group

$z = 1$, the treatment group

y_i^0, the potential outcome of unit i if no treatment

y_i^1, the potential outcome of unit i if treatment

$y_i = y_i^0 \cdot (1 - z_i) + y_i^1 \cdot z_i$, the observed outcome of unit i

$\tau_i = y_i^1 - y_i^0$, causal effect for unit i

$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} (y_i^1) - \frac{1}{m} \sum_{i=1}^{m} (y_i^0)$, sample average treatment effect

The fundamental problem of causal inference: we can never observe both y_i^0 *and* y_i^1. And we can only attribute an average treatment effect $\hat{\tau}$ to a unit if we assume that effects are constant across units.

$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} (y_i^1 - y_i^0)$, population average treatment effect

the potential outcomes approach, hypothetical data — *balanced* treatment and control groups?

Unit i	Female, x_{1i}	Age, x_{2i}	Treatment, z_i	Potential outcomes if $z_i = 0$, y_i^0	Potential outcomes if $z_i = 1$, y_i^1	Observed outcome, y_i
Audrey	1	40	0	**140**	135	140
Anna	1	40	0	**140**	135	140
Bob	0	50	0	**150**	140	150
Bill	0	50	0	**150**	140	150
Caitlin	1	60	1	160	**155**	155
Cara	1	60	1	160	**155**	155
Dave	0	70	1	170	**160**	160
Doug	0	70	1	170	**160**	160

Of note, with just 8 units, split equally between treatment and control groups, there are

$$\binom{n}{k} = 70$$

unique possible experiments!

Do you think this treatment assignment *balances* the treatment and control groups, or is it *biased*?

What's the sample average treatment effect $\hat{\tau}$ for this particular treatment assignment?

How does $\hat{\tau}$ compare with the *unknown true* average treatment effect?

Now re-assign the units to treatment and control groups *randomly* where $z \perp y^0, y^1$ and repeat. What do you get?

```
set.seed(3)
z <- sample(x = c(0,0,0,0,1,1,1,1), size = 8)
```

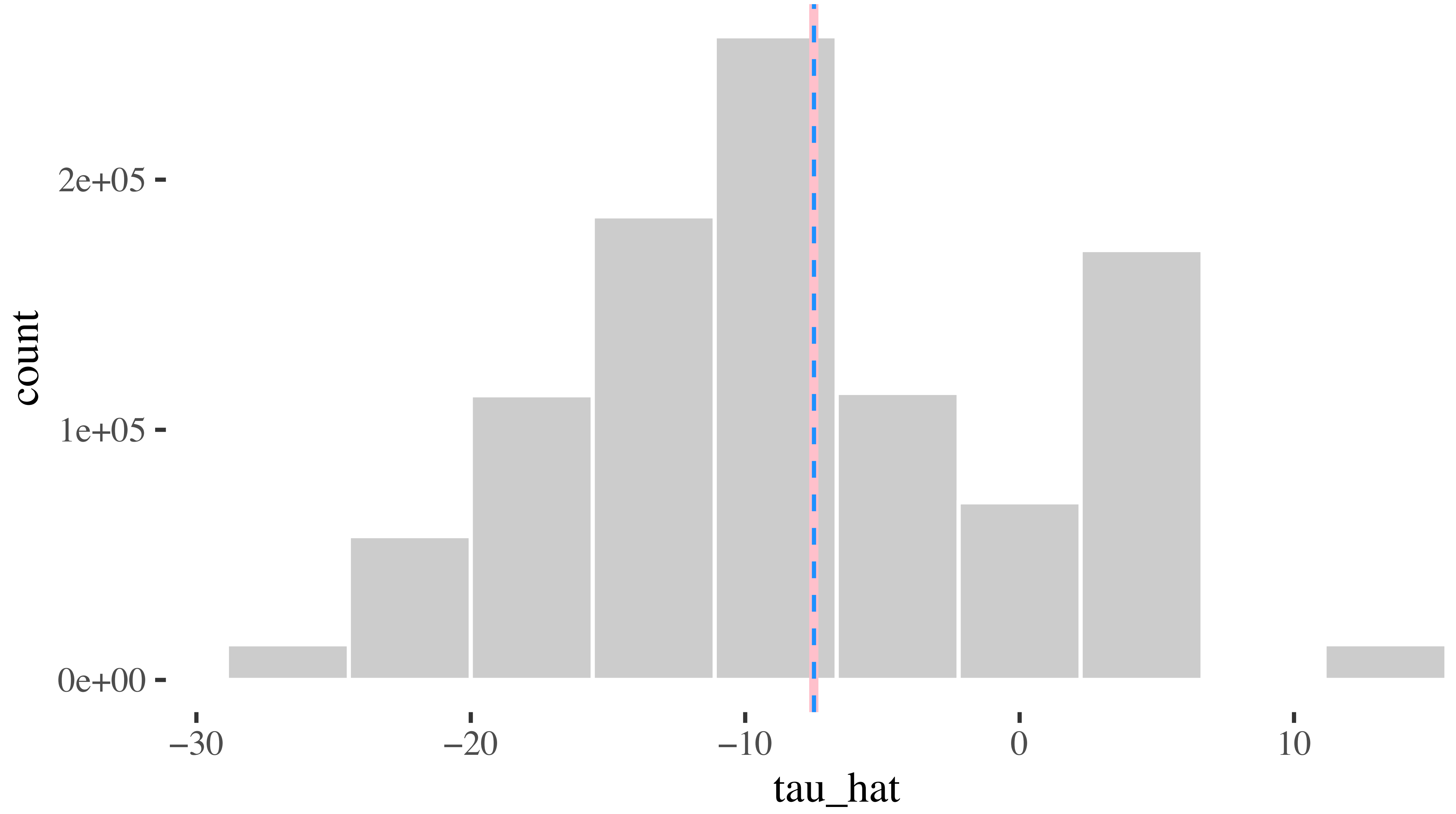
```
d <-
  read.table(text = '
  Unit     Female Age  z yi0 yi1
  Audrey     1     40   0 140 135
  Anna       1     40   0 140 135
  Bob        0     50   0 150 140
  Bill       0     50   0 150 140
  Caitlin    1     60   1 160 155
  Cara       1     60   1 160 155
  Dave       0     70   1 170 160
  Doug       0     70   1 170 160
', header = TRUE)

tau_tru <- with(d, mean(yi1 - yi0) )

d$yi    <- with(d, yi0 * (1 - z) + yi1 * z)
y1      <- with(d, mean(yi[z == 1]) )
y0      <- with(d, mean(yi[z == 0]) )
tau_hat <- y1 - y0

set.seed(123)

d$z     <- sample(c(0, 0, 0, 0, 1, 1, 1, 1), 8)
d$yi    <- with(d, yi0 * (1 - z) + yi1 * z)
y1      <- with(d, mean(yi[z == 1]) )
y0      <- with(d, mean(yi[z == 0]) )
tau_hat <- y1 - y0
```

No *single* randomized experiment guarantees that $\hat{\tau}$ will be close to the *unknown* true average treatment effect.

Try experimenting with different seeds in this code, and re-run to see how individual $\hat{\tau}$ is affected by the sample.

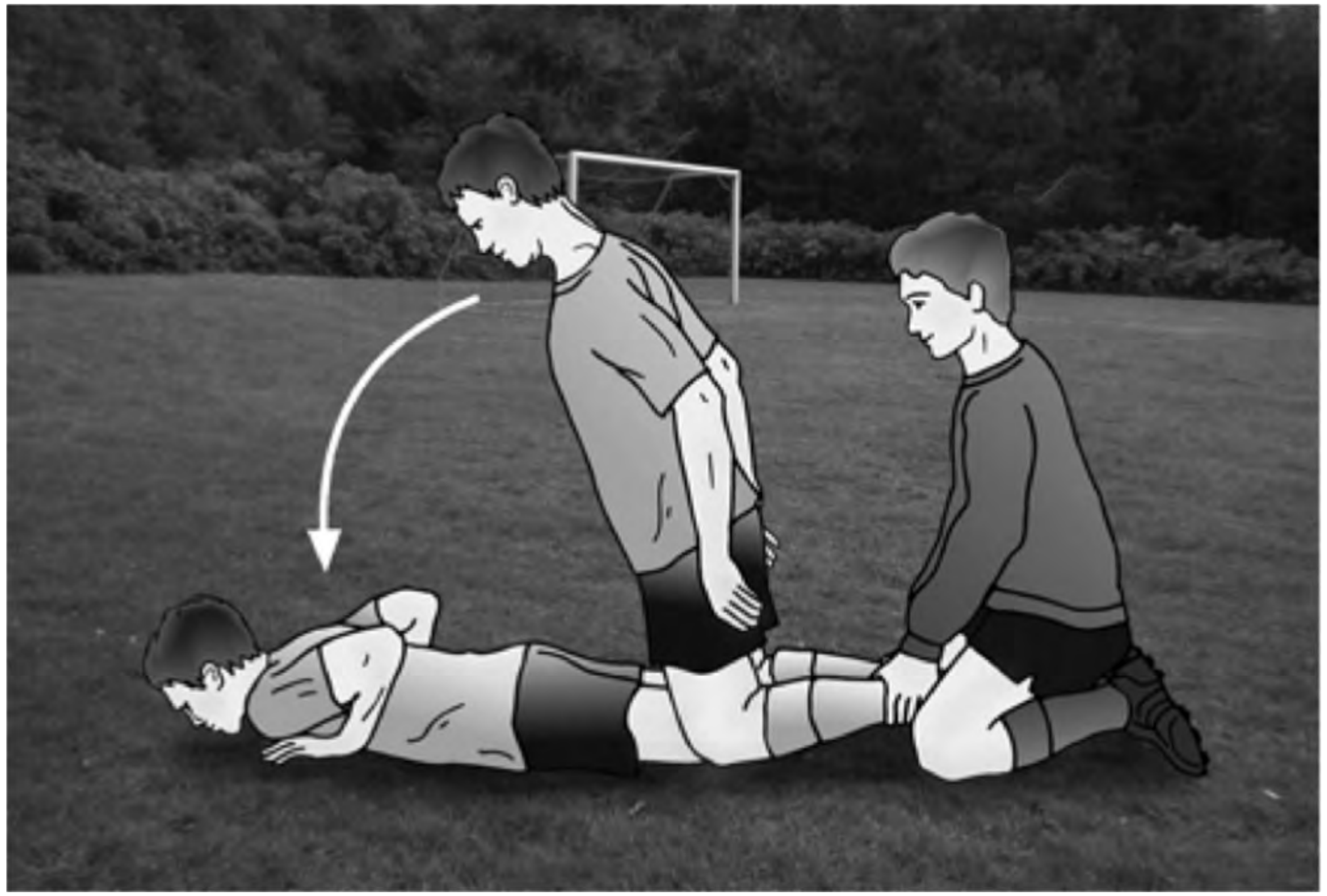
```
sim_experiment <- function(d) {

  d$z <- sample(c(0, 0, 0, 0, 1, 1, 1, 1), 8)
  y1  <- with(d, mean(yi1[z == 1]) )
  y0  <- with(d, mean(yi0[z == 0]) )

  return(y1 - y0)
}

tau_hat <- replicate( 1e6, sim_experiment(d) )

library(ggplot2)
library(ggthemes)

ggplot() +
  theme_tufte() +
  geom_histogram(aes(tau_hat),
                 bins  = 10,
                 fill  = "lightgray",
                 color = "white") +
  geom_vline(aes(xintercept = tau_tru),
             color = "pink",
             lwd = 1.1) +
  geom_vline(aes(xintercept = mean(tau_hat)),
             color    = "dodgerblue",
             linetype = "dashed")


E_tau_hat <- mean(tau_hat)
```

But randomly assigning units to treatment and control groups ensures that there are ***no differences in expectation in the distribution*** of potential outcomes between groups receiving different treatments — it's an *unbiased* estimator. In these simulations, $\mathbb{E}(\hat{\tau}) = -7.497 \simeq -7.5$



By collecting *more units*, we can improve balance in single experiments, and by collecting *pre-treatment* information, we can *adjust for imbalances* — techniques we cover later.

review of a published, randomized controlled experiment van der Horst, et al. *The Preventive Effect of the Nordic Hamstring Exercise on Hamstring Injuries in Amateur Soccer Players*



Purpose?

Population of interest?

Null hypothesis?

Alternative hypothesis?

Experimental design?

Results?

introducing your group projects

References

"cause, n.". OED Online. September 2020. *Oxford University Press*. https://www-oed-com.ezproxy.cul.columbia.edu/view/Entry/29147?rskey=AMcwBV&result=1&isAdvanced=false (accessed September 23, 2020).

Blitzstein, Joseph K., and Jessica Hwang. *Introduction to Probability*. Second edition. Boca Raton: Taylor & Francis, 2019.

Cox, D. R., and N. Reid. *The Theory of the Design of Experiments*. Monographs on Statistics and Applied Probability 86. Boca Raton: Chapman & Hall/CRC, 2000.

Gelman, Andrew, Jennifer Hill, and Aki Ventari. "Causal inference and randomized experiments, Chp. 18". In *Regression and Other Stories*. S.l.: Cambridge University Press, 2020.

Hernán, Miguel A, and James M Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

Imbens, Guido W, and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. 1st ed. An Introduction. Cambridge University Press, 2015.

Pearl, Judea. *CAUSALITY: Models, Reasoning, and Inference* Second Edition. Cambridge University Press, 2009.

Rosenbaum, Paul. "Randomized Experiments, Part I." In *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, 2017.