

Why Simulate?

Ashley I Naimi

Summer 2024

Contents

1	Why Simulate?	2
2	An Overview of Simulation Designs	2
2.1	Compartmental Models	3
2.2	Generative Modeling	3
2.3	Monte Carlo Estimation	4
2.4	Monte Carlo Simulation	5
2.5	The Monte Carlo Method	5
3	Example Simulation Questions	6
3.1	Can we estimate the value of π ?	6
3.2	Estimating the Central Tendency of a Random Variable	8
4	Some Example Simulation Research Questions	11
4.1	Inverse Probability Weighting versus G Computation in Time-Dependent Settings	11
4.2	The Impact of Random Measurement Error in Observational Data Analysis	12
4.3	Heterogeneous Treatment Effects in the Presence of Null Overall Effects	13
5	Some Motivating Questions	13
5.1	Example 1: Simple Regression in an RCT setting	13
5.2	Example 2: IPW vs Marginal Standardization	14
5.3	Example 3: Causal Mediation Analysis	14

1 Why Simulate?

The last two decades have seen tremendous growth in the quantitative sciences.¹ Statistical and analytic methods have become more nuanced, more complex, and more theoretical. The volume and complexity of these newer topics can make it difficult to gain a deeper understanding of basic concepts.

¹ Much of this section is based on [Rudolph et al. \(2021\)](#)

Even in cases where deep dives are possible, deep understanding of a complex concept is usually facilitated by breaking it down into simpler components and understanding the role that each component plays in the whole ([Gleick, 2011](#)). Tools that can be used to do this would therefore be of great benefit. Simulation is one such tool ([Burton et al., 2006](#); [Mooney, 1995](#); [Hodgson and Burke, 2000](#)).

Simulation is most commonly applied when assessing or comparing the performance of methods (e.g., estimators) in terms of bias or precision under known conditions. However, simulation can also be used to demonstrate many fundamental principles of data analysis, including study design, bias, and error, in a clear and systematic way. Additionally, simulation can be used to develop a deeper understanding of the scientific method, since theories, methods, or hypotheses can be subjected to experiments in a well-controlled, simulated environment.

Overall, developing an ability to use simulation methods can be invaluable in pursuing a deeper understanding of important scientific and quantitative concepts.

2 An Overview of Simulation Designs

[Morris et al. \(2019\)](#) define a Monte Carlo simulation study as “computer experiments that involve creating data by pseudo-random sampling from known probability distributions.” This is a good definition, but it captures at least four roughly distinct activities encountered in the quantitative sciences: compartmental modeling, generative modeling, Monte Carlo *estimation*, and Monte Carlo *simulation*.

2.1 Compartmental Models

Compartmental models are common in infectious disease modeling ([Vynnycky and White, 2010](#)) and the social sciences ([Epstein, 2006](#)). These models divide a population of individuals into sub-groups, or compartments, and creates relationships between them that are then tracked across time.

This models are quite common in infectious disease epidemiology, and include sub-variants such as the Susceptible, Infectious, and Recovered (SIR) model, or the Susceptible, Exposed, Infectious, and Recovered (SEIR) model. These are often used to understand how infections might spread in a population, where compartments are constructed to represent subsets of the population.

2.2 Generative Modeling

While the term “generative model” has been around for some time ([Epstein, 2006](#)), it is more recently being used to denote a specific type of *synthetic* data generation. Though there are multiple potential uses of “synthetic data” ([Figueira and Vaz, 2022](#)), in the context of simulation, synthetic data generation methods are being developed to solve one very specific but very important problem: the need to rely on parametric models to simulate data.

As we will see in this course, Monte Carlo simulations are conducted primarily by specifying a data generating mechanism with very specific forms. For example, the variables that we simulate might be normally distributed with constant variance. The relationships between them will often be linear, or deviate from linearity in very well behaved ways (e.g., using squared or cubic terms to define polynomial relationships).

This reliance on well behaved parametric forms creates an inference problem: we might find that a particular method performs very well in data simulated with well-behaved parametric models. But can we expect this same performance to hold in settings where we collect our own data that is not guaranteed to follow some set of well-behaved parametric models?

Generative models are meant to solve this problem. They rely on, for example, neural networks ([Burkov, 2019](#), chapter 6), variational autoencoders ([Doersch, 2021](#)), generative adversarial networks ([Goodfellow et al., 2020](#)), or other algorithms that generally follow an “encoder-decoder” strategy. One

part of algorithm encodes the relationship in an empirical dataset that we are potentially using for a specific analysis, and simulates a dataset from this encoding. Some approaches try to simulate this data with a known true relationship between (say) an exposure and outcome of interest ([Athey et al., 2024](#), [Parikh et al. \(2022\)](#)).

The second part of the algorithm tries to “decode” the relationship, to see if it can tell whether the simulated data are “different” from the actual data. If certain parts of the data are, in fact, different, the initial encoder algorithm can re-simulate the data trying to improve the overall fidelity between the simulated and real data.

After several iterations of this, the hope is that the final simulated dataset should be indistinguishable from the real dataset. In doing so, these generative algorithms can then be used to evaluate the performance of analytic methods in datasets that look very much like real data, thus avoiding the problem that the simulated data are too artificial to be of general use.

2.3 Monte Carlo Estimation

Monte Carlo estimation (sometimes referred to as Monte Carlo integration) is a general procedure usually used in the context of an applied scientific question, such as when external data are used to address a particular question of interest. Monte Carlo estimation is usually deployed to solve for a parameter that answers this question using pseudo-random number generators. The most common use of Monte Carlo estimation in quantitative settings (to my knowledge) is in the context of causal inference to solve for the parametric g formula, first introduced by [Robins \(1986\)](#).

The g formula is a complex equation that (under some basic causal identifiability assumptions) allows us to quantify outcomes that would be observed under different exposure scenarios in a population of interest. The equation is often defined as:

$$E(Y^{\bar{a}_M}) = \int \cdots \int E(Y \mid \bar{a}_M, \bar{z}_M) \prod_{m=0}^M f(z_m \mid \bar{z}_{m-1}, \bar{a}_{m-1}) d\mu(z_m)$$

This complex equation does not usually have a closed-form solution, and thus Monte Carlo estimation is the standard procedure for solving this func-

tion. Several explanations of this are available in the literature (Robins and Hernán, 2009; Keil et al., 2014; Daniel et al., 2013; McGrath et al., 2020), as well as a large number of applied examples implementing the procedure (Naimi et al., 2021, Taubman et al. (2009); Westreich et al., 2012; Cole et al., 2013; Edwards et al., 2014; Young et al., 2011).

For example, we (Naimi et al., 2021) had a question about whether daily, low-dose aspirin can be used to prevent pregnancy loss in women trying to get pregnant, but who are at a higher risk of pregnancy loss and having a hard time getting and staying pregnant. We collected data on roughly 1200 women trying to get pregnant at a number of medical centers across the United States. These data were from the Effects of Aspirin for Gestation and Reproduction (EAGeR) Trial. One of our goals was to quantify the following parameter:

$$E(Y^{\bar{a}_M=1} - Y^{\bar{a}_M=0}),$$

which is the difference in outcomes that would be observed if all women took aspirin consistently over the course of follow-up, versus if all women took placebo consistently over the course of follow-up. This contrast is an answer to our underlying research question, and we used the g formula with Monte Carlo estimation to quantify this contrast of interest.

We will see a simpler example of Monte Carlo estimation below with an example of estimating π .

2.4 Monte Carlo Simulation

Monte Carlo simulation is an extension/generalization of Monte Carlo estimation. While Monte Carlo estimation is often used with data collected to answer a substantive question of interest, Monte Carlo simulation is often used to estimate a parameter that addresses a methodological question of interest.² This short course will be primarily about Monte Carlo simulation, and we will see many examples of this over the course of the next few days.

² This distinction, however, is somewhat artificial, as the separation between Monte Carlo estimation versus simulation is not always straightforward.

2.5 The Monte Carlo Method

What's important is that all three of these approaches rely on the Monte Carlo method to be deployed. The Monte Carlo method was introduced in the early 20th century (Metropolis and Ulam, 1949). It was developed by John von Neu-

mann, Stanislaw Ulam, and Nicholas Metropolis to solve problems around the behavior of fissionable material under different conditions ([Metropolis, 1987](#)).

The basic idea behind this method is to use chance to solve problems that would either be intractable, or just too difficult to solve analytically. Because of its use of chance, it is named after the famed Monte Carlo Casino in Monaco ([Dunn and Shultis, 2022](#)).

3 Example Simulation Questions

Monte Carlo estimation proceeds by using pseudo-random number generation to quantify a parameter of interest. The number of questions that could be answered using Monte Carlo simulation is endless. More generally, Monte Carlo methods can be used to gain insights, experiment, and answer an enormously wide range of questions related to quantitative methodology. Here are a few examples:

3.1 Can we estimate the value of π ?

For thousands of years, the ratio of a circle's circumference to its diameter has played an important role in the unfoldment of human history. Equally fascinating is the history of how this constant π has been quantified over the millenia ([Beckmann, 1971](#)). Interestingly, we can use the Monte Carlo method to quantify π .³ For example, suppose you didn't know, but had to estimate π . Suppose further that you knew the area of a circle included π as a constant:

$$A_C = \pi \times r^2$$

We can construct an equation to solve for π by taking the ratio of a quarter of a unit circle to the area of the unit square:

$$\frac{A_C/4}{A_S} = \frac{(\pi \times r^2)/4}{L \times W}.$$

Because we're dealing with the unit circle and square, we can replace r , L , and W by 1, and we get:

$$\frac{A_C/4}{A_S} = \pi/4$$

If you chose $L = W = r = 1$, you could take the ratio of the area of the

³ Note that this is not nearly the most efficient way of calculating π . To my knowledge, the current "best" approach to quantifying π is based on a hexadecimal algorithm in base 16. See [Bailey et al. \(1997\)](#).

quarter circle to the area of the unit square to give:

$$\pi = 4 \times \frac{A_{C/4}}{A_S}$$

You could then randomly spread points over the entire unit square. Taking four times the proportion of points that fall in the quarter circle relative to the unit square will give you an estimate of π :

```
# estimating pi
pi_est <- function(simulation_n) {

  x <- runif(simulation_n, 0, 1)
  y <- runif(simulation_n, 0, 1)

  # if point lies within radius, set to one, otherwise zero
  rad <- as.numeric(x^2 + y^2 <= 1)

  # proportion of points within quarter circle (circle area) to
# proportion of points within unit square (square area)
  res <- c(simulation_n, (sum(rad)/simulation_n)*4)

  return(res)

}

n_list <- c(500, 1000, 5000, 50000, 500000, 100000, 10000000)

pi_data <- NULL
for(i in n_list){
  pi_data <- rbind(pi_data, pi_est(simulation_n = i))
}
```

We can present the results in a fancy-ish table using the `kable` function in the `knitr` package. Here we see that, as we increase the sample size used to estimate π , the estimated value gets closer and closer to the truth:

```
knitr::kable(pi_data, "simple",
             col.names = c("Simulation N",
                           "Estimated pi"))
```

Simulation N	Estimated pi
5e+02	3.096000
1e+03	3.164000
5e+03	3.181600
5e+04	3.143760
5e+05	3.141904
1e+05	3.141160
1e+07	3.142735

3.2 Estimating the Central Tendency of a Random Variable

Here, we'll use another example to illustrate the concept of simulation in a slightly more familiar setting. Suppose we have a dataset that contains information on the body mass index of 20 individuals sampled from the general population. We assume this BMI variable is distributed normally.

Suppose further that we'd like to estimate the central tendency of this distribution as accurately and efficiently as possible.

We know that we can use either the mean, the median, or the mode value of this distribution as a measure of central tendency, and if BMI is normally distributed, then they should all be the same. So should we use the mean, the median, or the mode to summarize this distribution? Does it matter? Why or why not? We can use simulation to gain some insight into this question:

```
# write a function that computes the statistical
# mode. modified from: https://stackoverflow.com/a/8189441
mode_func <- function(x) {
  ux <- unique(y)
  ix <- which.max(tabulate(match(y, ux)))
  if(ix==1){
    warning("no duplicates in data, mode does not exist")
  }
}
```



```

    res <- sample(ux, size = 1)
  } else{
    res <- ux[ix]
  }
  return(res)
}

# set the seed value
set.seed(123)

# how many observations?
n = 200

# start the simulation loop
res <- NULL
for(i in 1:1e4){

  # simulate a variable from a normal
  # with mean 26.5 and SD 5.25
  y <- round(rnorm(n, mean = 26.5, sd = 5.25))

  # estimate mean, median, and mode
  mode_estimator <- mode_func(y)
  mean_estimator <- mean(y)
  median_estimator <- median(y)

  # store results
  res <- rbind(res,
               cbind(mode_estimator,
                     mean_estimator,
                     median_estimator)
  )
}

```

```
head(res)
```

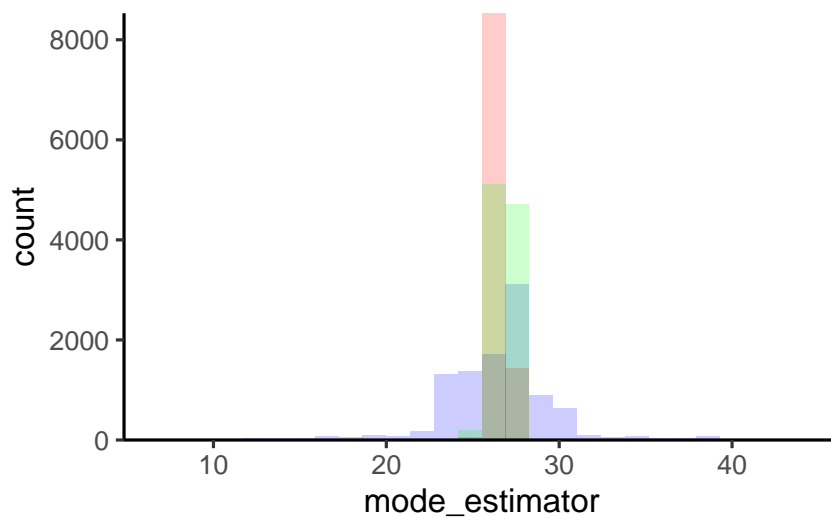
```
##      mode_estimator mean_estimator median_estimator
## [1,]           25         26.480             26
## [2,]           27         26.695             27
## [3,]           25         26.645             27
## [4,]           26         26.385             27
## [5,]           28         26.675             27
## [6,]           29         26.230             26
```

```
# convert results to data frame
```

```
res <- data.frame(res)
```

```
# plot results
```

```
ggplot(res) +
  geom_histogram(aes(mode_estimator),
                 alpha = .2, fill = "blue") +
  geom_histogram(aes(mean_estimator),
                 alpha = .2, fill = "red") +
  geom_histogram(aes(median_estimator),
                 alpha = .2, fill = "green") +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,0))
```



Is one better than the other? Why? Clearly, the figure shows us that the mode as an estimator is not great. It has a much wider spread than the other two. But what about the mean versus median?

```
res %>%
  summarize_all(list(mean = mean,
                     std_dev = sd))
```

	mode_estimator_mean	mean_estimator_mean	median_estimator_mean
## 1	26.5094	26.49731	26.5023

	mode_estimator_std_dev	mean_estimator_std_dev	median_estimator_std_dev
## 1	3.33467	0.3718288	0.5456872

This simple table shows that the average of all the values from the mean, median, and mode estimators are the same: roughly 26.5, which is what we set the true mean to in the simulation code above.

However, the standard deviation of each estimator is very different across the three. The estimator with the smallest standard deviation is the mean estimator, which suggests the mean is the best option in our setting.

As we proceed through this course, we'll see how these and other tools can be used to answer questions about the performance of quantitative methods in a wider range of settings.

4 Some Example Simulation Research Questions

The number of questions that can be addressed using simulation methods is nearly endless. However, to gain a more precise understanding of the kinds of questions asked and answered in actual research settings, let's consider a few published research papers that rely on simulation:

4.1 Inverse Probability Weighting versus G Computation in Time-Dependent Settings

In 2023, [Rudolph et al. \(2023\)](#) compared different variations of inverse probability weighting and different variations of the g computation estimator for estimating the average treatment effect in survival data with a time-dependent

treatment. Generally, there are few methods available to estimate effects for treatments that vary over time. The most commonly used techniques are IP weighting and g computation. However, researchers are often unsure about which approach to pick in a given setting. [Rudolph et al. \(2023\)](#) sought to generate insights about how to navigate this question using simulation. They found that methods performed similarly, except for a variation of g computation known as the iterative conditional expectation (ICE) g computation estimator. This latter estimator had the lowest bias, but also the lowest precision, in the settings explored.

4.2 The Impact of Random Measurement Error in Observational Data Analysis

Measurement error occurs when the tools used to collect data can yield incorrect information about the underlying feature being measured. Measurement error can be non-random, or differential, where the degree of error can depend on other variables; or it can be random, or non-differential, where the degree of error does not depend on any other variables in the system under study.

Researchers in epidemiology and other quantitative sciences often state that non-differential misclassification biases results towards the null effect. Practically, this means that if non-differential measurement error is present, and a study yields an effect that is non-null, then this effect represents a conservative estimate of the exposure under study. This implies the effect is stronger than estimated in the study. However, this is only true if non-differential misclassification biases towards the null, and only towards the null. If this type of error can bias away from the null, then all bets are off, and we cannot conclude that the effect estimated is conservative.

In 2018, [Brakenhoff et al. \(2018\)](#) conducted a simulation study that mimicked research in cardiovascular epidemiology to evaluate whether, in these more realistic settings, non-differential measurement error biases only towards the null. They showed that the direction of effect of random measurement error on the estimated exposure-outcome relations can be difficult to anticipate, suggesting that caution is warranted when concluding that estimated effects are conservative.

4.3 Heterogeneous Treatment Effects in the Presence of Null Overall Effects

Interest in heterogeneous treatment effects, a term meant to connote the fact that the effect of a treatment or exposure might differ widely across subgroups of a population, is increasing. There has been some theoretical work suggesting that, in a given sample, if there is no overall effect of a treatment under study, then the existence of strong sub-group effects is unlikely. This is important, since researchers often estimate “significant” subgroup effects in the absence of overall effects, suggesting that these estimates are the result of noise and variability.

In 2013, [Abrahamowicz et al. \(2013\)](#) conducted a simulation study to see how common it would be to encounter null overall effects even in the presence of strong sub-group effects. They also explored what factors affected the presence of such scenarios. They explored three study designs: small clinical studies, case-control studies, and large cohort studies, each under different total sample size (N), relative size of the affected subgroup, and true treatment effect. Overall, their simulation provided evidence against previous theoretical conceptualizations of the problem: if the treatment really only has an effect in one subgroup of the population, a null overall effect can often coincide with a non-null treatment-by-subgroup interaction.

5 Some Motivating Questions

This short course will be motivated by an actual simulation study. Throughout, we will demonstrate the concepts and tools needed to conduct a simulation study using some actual examples to motivate our study. We will focus on the following research questions:

5.1 Example 1: Simple Regression in an RCT setting

Estimating intention-to-treat effects in randomized trials is usually a simple procedure comparing the summary measure (e.g., risk) of the outcome between the treated and placebo groups. However, there is some evidence suggesting that if we further adjust for variables that explain variability in the outcome, we can increase the efficiency of the intention-to-treat effects estimator. We’ll conduct a simulation to address the following research question:

How much improvement in the performance of an ITT effect estimator do we gain when we adjust for variables that explain the outcome?

5.2 Example 2: IPW vs Marginal Standardization

Similar to the study by [Rudolph et al. \(2023\)](#), we can ask whether there is a difference in the performance of an IP weighted estimator versus a g computation (i.e., marginal standardization) estimator in settings where the exposure does not vary over time. We can use simulation to address the following research question:

How does inverse probability weighting compare to marginal standardization when used to adjust for confounding in a treatment effect estimation setting?

5.3 Example 3: Causal Mediation Analysis

Mediation analysis is an extremely popular form of analyzing data to determine whether an estimated association between an exposure and outcome of interest is due entirely, partially, or not at all to a third potentially mediating variable. Several methods have been developed to estimate direct and indirect effects due to mediation in observational data. Most require that certain very strong assumptions hold. For example, one usually has to adjust for *confounders* of the relationship between the mediator and outcome of interest to avoid bias. However, many methods also require that these confounders are not affected by the exposure under study. We can use simulation to answer the following research question:

If we don't use methods that account for mediator-outcome confounders affected by the exposure, how misleading can the results of a mediation analysis be?

Over the next few days, we will explore concepts, tools, and strategies to enable us to more clearly define what we mean by this question, and how we can construct a simulation study to provide us with some answers to these questions.

References

- Michal Abrahamowicz, Marie-Eve Beauchamp, Pierre Fournier, and Alexandre Dumont. Evidence of subgroup-specific treatment effect in the absence of an overall effect: is there really a contradiction? *Pharmacoepidemiology and Drug Safety*, 22(11):1178–1188, 2013.
- Susan Athey, Guido W. Imbens, Jonas Metzger, and Evan Munro. Using wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*, 240(2):105076, 2024. ISSN 0304-4076. doi: <https://doi.org/10.1016/j.jeconom.2020.09.013>. URL <https://www.sciencedirect.com/science/article/pii/S0304407621000440>.
- David Bailey, Peter Borwein, and Simon Plouff. On the rapid computation of various polylogarithmic constants. *Math. Comp*, pages 903–13, 1997.
- P Beckmann. *A History of Pi*. St. Martin's Publishing Group, 1971.
- Timo B. Brakenhoff, Maarten van Smeden, Frank L. J. Visseren, and Rolf H. H. Groenwold. Random measurement error: Why worry? an example of cardiovascular risk factors. *PLOS ONE*, 13(2):1–8, 2018.
- A Burkov. *The Hundred Page Machine Learning Book*. Andriy Burkov, 2019.
- Andrea Burton, Douglas G Altman, Patrick Royston, and Roger L Holder. The design of simulation studies in medical statistics. *Stat Med*, 25(24):4279–4292, 2006.
- Stephen R. Cole, David B. Richardson, Haitao Chu, and Ashley I. Naimi. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol*, 177(9):989–996, 2013.
- R.M. Daniel, S.N. Cousens, B.L. De Stavola, M. G. Kenward, and J. A. C. Sterne. Methods for dealing with time-dependent confounding. *Stat Med*, 32(9):1584–618, 2013.
- Carl Doersch. Tutorial on variational autoencoders. *arXiv:1606.05908*, <https://arxiv.org/abs/1606.05908>, 2021.
- W.L. Dunn and J.K. Shultis. *Exploring Monte Carlo Methods*. Elsevier Science, 2022.

Jessie K Edwards, LJ McGrath, Buckley JP, MK Schubauer-Berigan, SR Cole, and Richardson DB. Occupational radon exposure and lung cancer mortality: Estimating intervention effects using the parametric g-formula. *Epidemiol*, 25(6):829–34, 2014.

JM Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, Princeton, NJ, 2006.

Alvaro Figueira and Bruno Vaz. Survey on synthetic data generation, evaluation methods and gans. *Mathematics*, 10(15), 2022. ISSN 2227-7390. DOI: 10.3390/math10152733. URL <https://www.mdpi.com/2227-7390/10/15/2733>.

James Gleick. *Genius: The Life and Science of Richard Feynman*. Open Road Media, New York, NY, 2011.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, oct 2020. ISSN 0001-0782. DOI: 10.1145/3422622. URL <https://doi.org/10.1145/3422622>.

Ted Hodgson and Maurice Burke. On simulation and the teaching of statistics. *Teaching Statistics*, 22(3):91–96, 2000.

Alex Keil, Jessie K Edwards, David B. Richardson, Ashley I. Naimi, and Stephen R. Cole. The parametric g-formula for time-to-event data: towards intuition with a worked example. *Epidemiol*, 25(6):889–97, 2014.

Sean McGrath, Victoria Lin, Zilu Zhang, Lucia C. Petito, Roger W. Logan, Miguel A. Hernán, and Jessica G. Young. gformula: An r package for estimating the effects of sustained treatment strategies via the parametric g-formula. *Patterns*, 1(3):100008, 2020.

N Metropolis. The beginning of the monte carlo method. *Los Alamos Science Special Issue*, 15:125–30, 1987.

N Metropolis and S Ulam. The Monte Carlo method. *J Am Stat Assoc*, 44(247): 335–341, 1949.

C Mooney. Conveying truth with the artificial: using simulated data to teach statistics in social sciences. *SocInfo Journal*, 1(Part 7):1–5, 1995.

Tim P. Morris, Ian R. White, and Michael J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102, 2019.

Ashley I Naimi, Neil J Perkins, Lindsey A Sjaarda, Sunni L Mumford, Robert W Platt, Robert M Silver, and Enrique F Schisterman. The effect of preconception-initiated low-dose aspirin on human chorionic gonadotropin-detected pregnancy, pregnancy loss, and live birth : Per protocol analysis of a randomized trial. *Ann Intern Med*, 174(5):595–601, 2021.

Harsh Parikh, Carlos Varjao, Louise Xu, and Eric Tchetgen Tchetgen. Validating causal inference methods. *arXiv:2202.04208*, <https://arxiv.org/abs/2202.04208>, 2022.

J. M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.

James M Robins and Miguel Á Hernán. Estimation of the causal effects of time-varying exposures. In G Fitzmaurice, M. Davidian, G Verbeke, and G Molenberghs, editors, *Advances in Longitudinal Data Analysis*, pages 553–599. Chapman & Hall, Boca Raton, FL, 2009.

Jacqueline E Rudolph, Matthew P Fox, and Ashley I Naimi. Simulation as a tool for teaching and learning epidemiologic methods. *American Journal of Epidemiology*, 190(5):900–907, 2021.

Jacqueline E Rudolph, Enrique F Schisterman, and Ashley I Naimi. A simulation study comparing the performance of time-varying inverse probability weighting and g-computation in survival analysis. *Am J Epidemiol*, 192(1):102–110, 2023.

S. L. Taubman, J. M. Robins, M. A. Mittleman, and M. A. Hernán. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol*, 38(6):1599–611, 2009.

E. Vynnycky and R. White. *An Introduction to Infectious Disease Modelling*. OUP Oxford, New York, NY, 2010.

Daniel Westreich, Stephen R. Cole, Jessica G. Young, Frank Palella, Phyllis C. Tien, Lawrence Kingsley, Stephen J. Gange, and Miguel Á Hernán. The

parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident aids or death. *Stat Med*, 31(18):2000–2009, 2012.

Jessica G Young, Lauren E Cain, James M Robins, Eilis J O'Reilly, and Miguel A Hernán. Comparative effectiveness of dynamic treatment regimes: An application of the parametric g-formula. *Statistics in Biosciences*, 3(1): 119–143, 2011.