The Battle for The Soul of Causal Inference

Presented to *PBHLT 7115 - Causal Methods in Public Health* at University of Utah (on invitation by Andy Wilson, PhD)

Justin Belair | Statistical Consultant | Author of Causal Inference in Statistics

2025-03-20

A Clash of Titans

A Clash of Titans

- **Donald Rubin**, a statistician: Rediscovered the potential outcomes notation $(Y_i(1), Y_i(0))$ in 1974 and sees the core task of causal inference as the contrast between these two potential outcomes.
  - "To avoid conditioning on some observed covariates [as Pearl suggests] in the hope of obtaining an unbiased estimator because of phantom but complementary imbalances on unobserved covariates, is neither Bayesian nor scientifically sound but rather it is **distinctly frequentist and nonscientific ad hocery.**"
- **Judea Pearl**, a computer scientist: Developed Structural Causal Models (SCMs) and sees the core task of causal inference as estimating what happens to systems of variables when they are perturbed, through an intervention.
  - "**Rubin will do well to expand the horizons of his students** with some of the tools that his admirers now deem illuminating."

A Brief Review of Potential Outcomes Theory

Rubin Causal Model

- Population of $i = 1, \ldots, N$ units (we omit sampling issues for now)
- Each unit is characterized by a set of covariates, X_i
- Two possible treatments, $T \in \{0, 1\}$
- The treatment actually taken by unit i is W_i, and the counterfactual treatment is $1 - W_i$

Rubin Causal Model

- Each unit has two potential outcomes: $Y_i(0)$ and $Y_i(1)$
- The observed outcome is

$$Y_i^{obs} = Y_i(1) \cdot W_i + Y_i(0) \cdot (1 - W_i) = Y(W_i)$$

- The unobserved (missing) outcome is

$$Y_i^{mis} = Y_i(1) \cdot (1 - W_i) + Y_i(0) \cdot W_i = Y(1 - W_i)$$

- The individual treatment effect is a contrast between the two potential outcomes: $Y_i(1) - Y_i(0)$, or $Y_i(1)/Y_i(0)$, or $\log(Y_i(1) - \log(Y_i(0))$, etc.

An Example

i	$Y_i(0)$	$Y_i(1)$	W_i	Y^{obs}	Individual Treatment Effect
1	1	?	0	1	?
2	?	0	1	0	?
3	?	1	1	1	?
4	0	?	0	0	?
5	?	0	1	0	?
6	1	?	0	1	?

Table 1: Potential outcomes example with 6 units

The assignement mechanism

In Rubin's framework, the central task of causal inference is to model the **treatment assignment mechanism**

- A probability distribution $P(W|X, Y(0), Y(1))$ which can depend on the covariates **and** the potential outcomes
- The difference between experimental and observational studies is in the treatment assignment mechanism
- The goal of **design** of a study (experimental or observational) is to make the treatment assignment independent of the potential outcomes
  - Formally, $P(W|X, Y(0), Y(1)) = P(W|X)$.

Dependance of treatment assignement on the potential outcomes

- The perfect doctor: gives the treatment to those who will benefit from it
- Assuming we know the true potential outcomes (this never happens) :

i	$Y_i(0)$	$Y_i(1)$	W_i	Y^{obs}	Individual Treatment Effect
1	1	0	1	0	-1
2	1	0	1	0	-1
3	1	1	0	1	0
4	0	0	0	0	0
5	1	0	1	0	-1
6	1	1	0	1	0

Table 2: The treatment is assigned by a perfect doctor

Computation of ATE

- The Average Treatment Effect (ATE) can be computed perfectly in this hypothetical scenario:

$$\text{ATE} = \frac{(-1) + (-1) + 0 + 0 + (-1) + 0}{6} = -\frac{1}{2}$$

A Brief Review of Potential Outcomes Theory

- In practice, we can only compare the average observed outcomes between treated and control groups

$$\hat{ATE}_{\text{naive}} = \bar{x}_1 - \bar{x}_0 = \frac{0+0+0}{3} - \frac{1+1+0}{3} = -\frac{2}{3}$$

- The key insight is that this overestimation of the treatment effect $-\frac{2}{3} < -\frac{1}{2}$ is not an artifact, it is a systematic bias in the way the treatment was assigned!

A Brief Review of Potential Outcomes Theory

- The doctor does not need to be perfect
  - By simply favoring those that are more likely to heal (as any good doctor would), there is a *systematic* overestimation of the treatment effect
  - In formal terms, the probability of being assigned to treatment depends on the potential outcomes, i.e. higher potential outcomes lead to higher probability of being treated, thus
    - $P(W|X, Y(0), Y(1)) \neq P(W|X)$
- ***Can you think of other such examples?***

The Solution

- The simple solution to this problem is to make the assignment of treatment completely independent of anything relevant, namely by literally using a physical chance mechanism (e.g. a coin toss) to assign it randomly
  - Fisher first made this idea explicit and ushered in a new era of scientific thinking
  - Indeed, randomization ensures that $P(W|X, Y(0), Y(1)) = P(W|X)$

For observational studies

- In observational studies, we do not have the luxury of randomization
- We can imagine that the perfect doctor does not *know* the potential outcomes per se, but looks at patient's characteristics (e.g. age, sex, medical history, etc.)
- For a given patient profile X_i, if the doctor essentially assigns treatment randomly with a certain probability, then the treatment assignment is independent of the potential outcomes $P(W_i|X_i, Y_i(0), Y_i(1)) = P(W_i|X_i)$
  - This probability need not be 0.5, it can be any value *strictly* between 0 and 1
  - Thus, the doctor is essentially performing mini-randomized experiments for patients characterized by their covariates X

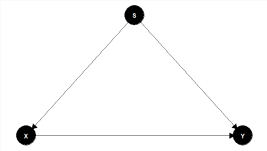
The Key Insight

- This is the key insight of the Rubin Causal Model:
  - If, conditional on a proper set of covariates, the treatment assignment is essentially random, then it is independent of the potential outcomes and for each value of X we can estimate the (conditional) treatment effect without bias by simply comparing the treated to the non-treated and then averaging the results over the X
  - Thus, in observational studies, we wish to compare patients that are similar in the covariates that we believe influence the fact that they receive treatment or control, this is why we speak of *balancing the groups on the covariates*
  - We can use matching, weighting, stratification, regression, etc. to achieve this balance
  - The statistical properties of these different techniques are continually being studied and refined

A Brief Review of Structural Causal Models

A Brief Review of Structural Causal Models

A Pearlian Causal Structural Model (SCM) uses a DAG and a set of functional relationships:



$$X \leftarrow f_X(S + U_X)$$
$$Y \leftarrow f_Y(X + S + U_Y)$$
$$S \leftarrow f_S(U_S),$$

where U are random error terms assumed independent of one another.

A Brief Review of Structural Causal Models

- The causal effect of X on Y is defined as the distribution of Y under a given intervention, denoted

$$P(Y|do(X = x)),$$

  where $do(X = x)$ is an intervention that sets X to x, a concept that Pearl invented.
- $do(X = x)$ is modelled by severing all the arrows entering into X in the DAG

- The intervention at X, i.e. $do(X = x)$, is propagated through the system of equations as follows:

$$
\begin{aligned}
X &\leftarrow x \\
Y &\leftarrow f_Y(x + S + U_Y) \\
S &\leftarrow f_S(U_S)
\end{aligned}
$$

A Brief Review of Structural Causal Models

- In observational settings, it is impossible to observe the effect of X on Y directly, as their is no $do(\cdot)$ operation
- In general, *observing* the units that have $X = x$ is possible, but will differ from intervening on X, i.e.

$$P(Y|X = x) \neq P(Y|do(X = x)).$$

The Key Insight

- Pearl developed the **backdoor criterion**, a set of rules that can be applied algorithmically to find an adjustment set Z that will **identify** the causal effect
- Namely, if Z respects certain conditions, then

$$P(Y|do(X = x), Z) = P(Y|X = x, Z = z).$$

  - That is, within units who have equal values of Z, *intervening* to fix X at x or *observing* that X is at x are the same.

The Key Insight

Using the backdoor criterion, we can compute the causal effect as a weighted average

$$\begin{aligned} P(Y|do(X=x)) &= \sum_z P(Y|do(X=x), Z=z)P(Z=z) \\ &= \sum_z P(Y|X=x, Z=z)P(Z=z). \end{aligned}$$

The Debate - M-Bias

The debate

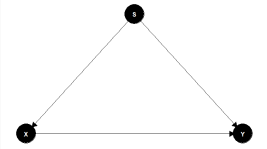
- Rubin today: Let's watch this 2024 video at 12min49
- Rubin's position has essentially been that Pearl's approach is not useful, at least not to him
  - In Imbens & Rubin (2015), he devotes $\sim$ 100 words to Pearl:
    - "Pearl's work is interesting [...] In our work, [...] we have not found this approach to aid drawing of causal inferences, and we do not discuss it further in this text."
- Pearl, on the other hand, repeatedly discusses Rubin's work and consistently argues that his own framework is more relevant
  - He criticizes Rubin's approach as being non-communicable, namely because ignorability conditions such as $P(W|X, Y(0), Y(1)) = P(W|X)$ are deemed too abstract, while DAGs communicate clear knowledge

Rubin Innoncently Publishes A Nice Stats Paper

- In a 2007 paper, *The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials*, Rubin shows how modelling the treatment assignment is a matter of **design** of the observational study, namely through propensity score matching methods
  - In it, he advocates that any any covariate should be controlled for, namely by ensuring balance between the treated and control groups
    - "[...] covariates are variables that take the same value for each treatment for each unit no matter which treatment is applied to the units, such as quantities measured before treatments are assigned (e.g. age, pre-treatment blood pressure or cholesterol level)."
- This point of view is largely adopted by the statistical community
  - Pearlian scholars sensed an opportunity to disrupt the statistical establishment...

Review of confounders in Pearl's framework

- A confounder is a variable that *creates* a backdoor path between treatment and outcome, as illustrated here

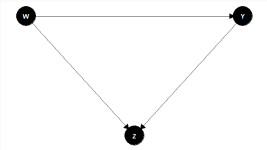


- Using the backdoor criterion, it can easily be shown that putting such a variable in the adustment set allows one to identify the causal effect, i.e.

$$P(Y|do(X), S) = P(Y|X, S).$$

Review of collliders in Pearl's framework

- A collider is a variable that *blocks* a backdoor path between treatment and outcome, as illustrated here



- Using the backdoor criterion, it can easily be shown that putting such a variable in the adjustment set would create a spurious association, i.e.

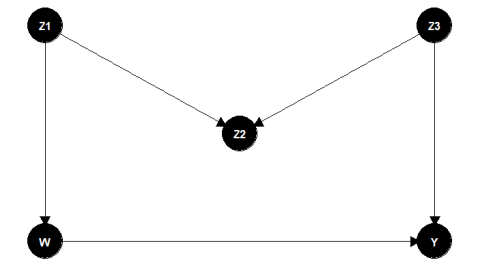
$$P(Y|do(X), S) \neq P(Y|X, S).$$

The Debate - M-Bias

- We can imagine statisticians dismissing this notion that to them was already very clear
  - Conditioning on something post-treatment can **obviously** create bias, e.g. differential loss-to-followup
  - As a matter of fact, a great paper Miguel A Hernán, Sonia Hernández-Díaz, and James M Robins, *A Structural Approach to Selection Bias*, published in 2004, argues that all forms of selection bias can be seen as bias induced by conditioning oncollider
  - Statisticians:
    - Before treatment : Good
    - After Treatment : Bad

The "Ah ha!" Moment

- Ian Shrier wrote a letter in 2008 asking if Rubin has considered the following DAG



Why This DAG Matters?

- In this DAG, Z_2 is a pretreatment variable, and Z_1 and Z_3 are unmeasured
- Rubin would probably argue that we should adjust for Z_2, but the backdoor criterion tells us that this would create collider bias, which would be impossible to remedy due to Z_1 and Z_3 being unmeasured
- We got them! Statisticians are wrong, and Pearl's thinking has revelead the truth to all!

Rubin's Reply

- In Rubin's reply, he remains puzzled and does not really address Shrier's question
  - "[...] the letter uses notation and terminology entirely different from what was used in the target article, which puzzles and confuses me."

Pearl Himself Joins The Fray

- Pearl reiterates why the M-bias example contradicts Rubin's principle of always controlling for covariates, and he does an effort to say so in Rubin's own language (e.g. ignorability, bias, asymptotics, etc.)

The Mediator (This Has Nothing To Do With Mediation)

- Sjölander writes in 2009 trying to mediate the gap between the two, briefly describing both frameworks and putting down mathematical notation to help the debate move forward

Rubin's Final Form

- Rubin writes his final reply in 2009, stating that this example is far too contrived to be of any importance
- You can read through the page how annoyed he is, and he even sounds quite arrogant.

Spicy Quotes

- "The only design advice that I can garner from these letters is that we should ponder the possible existence of unobserved imbalances–OK, and then do nothing about some observed imbalances–not OK."
- "The rationale for this implicit advice is the well-known mathematical fact that marginal independence between two random variables does not necessarily imply conditional independence between them given a third random variable (e.g. the sum of the first two). The letters seem to suggest that this fact has only been recently discovered and is important for causal inference in observational studies; I think that neither is accurate."

(More) Spicy Quotes

- "Time spent designing observational studies to have observed covariate imbalances because if hoped-for compensating imbalances in unobserved covariates is neither practically nor theoretically justifiable."
- "To avoid conditioning on some observed covariates in the hope of obtaining an unbiased estimator because of phantom but complementary imbalances on unobserved covariates, is neither Bayesian nor scientifically sound but rather it is distinctly frequentist and nonscientific ad hocery."

Let's Talk About It

- A DAG is essentially a tool to economically encode a set of **conditional independence relationships** between variables
  - The *absence* of an arrow between two variables A and B means that A and B are conditionally independent given a set of variables, say C
- See online Daggity tool

The Real Meaning of M-Bias

- The M-bias DAG implies the following set of conditional independencies
  - $W \perp Z_2 | Z_1$
  - $W \perp Z_3$
  - $Z_1 \perp Z_3$
  - $Z_1 \perp Y | W$
  - $Z_2 \perp Y | W, Z_3$
  - $Z_2 \perp Y | Z_1, Z_3$
- If we give Rubin the benefit of the doubt, the main thrust of his argument is that assuming this exact set of conditional independencies is not a good reason to believe that we should be wary of adjusting for a collider
  - In other words, M-Bias is a theoretical curiosity, but not a practical concern

Isn't It All Just Conditional Independence?

In an unpublished technical report *Myth, Confusion, and Science in Causal Analysis* by Pearl in 2009

- "While finding a pure M-structure, totally free of bias, may indeed be rare in practical studies (not unlike the rarity of finding any conditional independence,) cases containing local M-structures are abound."
  - We just saw that M-structures in fact encode a large set of conditional independencies, this statement is plainly contradictory, right?

More Spicy Quotes

In *Causal Inference: History, Perspectives, Adventures, and Unification (An Interview with Judea Pearl)* (2022), Pearl continues to maintain that Rubin refuses to acknowledge M-structures, that he is dogmatic, etc.

- "[...] Rubin's potential outcome framework became popular in several segments of the research community [...]. These researchers talked "conditional ignorability" to justify their methods, though **they could not tell whether it was true or not**. Conditional ignorability gave them a formal notation to state a license to use their favorite estimation procedure even though they could not defend the assumptions behind the license. [...] It is hard to believe that something so simple as a graph could replace the opaque concept of "conditional ignorability" that people find agonizing and incomprehensible. The back-door criterion made it possible [...]"

More Spicy Quotes

In *Causal Inference: History, Perspectives, Adventures, and Unification (An Interview with Judea Pearl)* (2022), Pearl continues to maintain that Rubin refuses to acknowledge M-structures, that he is dogmatic, etc.

- It is not clear to me how we can tell whether a DAG is true or not, any differently than a set of conditional independencies between counterfactuals

More Spicy Quotes

- " In the potential outcome framework, problems are defined algebraically as assumptions about counterfactual independencies, also known as "ignorability assumptions." These **types of assumptions are too complicated to interpret or verify by unaided judgment**. In the structural framework, on the other hand, **problems are defined in the language in which scientific knowledge is stored – causal graphs**. Dependencies of counterfactuals, if truly needed, can be deduced from the graph, but in almost all cases they can be replaced by causal dependencies among observables, which are vividly displayed in the graphs."
  - On one hand, whole swaths of scientists refuse to acknowledge causal graphs, but on the other, this is the language in which scientific knowledge is stored...

More Spicy Quotes

- "Rubin will do well to expand the horizons of his students with some of the tools that his admirers now deem illuminating", Pearl (2009)

My Take On The Matter

- Rubin and Pearl are both, to some degree, preaching for the choir
- DAGs are (extremely) useful to help encode complex conditional independencies, but not without thinking in terms of counterfactuals
  - In a sense, they are so useful and intuitive that their apparent simplicity can be extremely misleading in practice
  - I always use them when approaching a causal inference problem
  - I am sympathetic with Rubin's formalism, as it is much more elegant and does not require any new tools (except counterfactuals), whereas SCMs require a whole new set of formal concepts (e.g. Markov compatibilty, d-separation, backdoor criterion, do-calculus, etc.) and rely on a much richer set of assumptions (i.e. structural equations with uncorrelated errors)
  - Econometrics used to rely heavily on Structural Equations, but shifted to potential outcomes due to a crisis that led to a "credibility revolution", where the focus was more on designing studies than estimating parameters in structural models

My Take On The Matter

A proper formalization of graphical models with counterfactuals is possible, but complicated - See Richardson and Robins' 148 page working paper, *Single World Intervention Graphs (SWIGs):A Unification of the Counterfactual and Graphical Approaches to Causality)*

My Take On The Matter

- To me, the key idea of causal inference still boils down to comparison of counterfactual values
- DAGs help with the problem of *identification*, but not with everything statisticians care about: *estimation*, *sampling*, ruling-out competing explanations through *design*, clearly defining *interventions*, etc.
- SCMs are likely to be very useful in highly-structured, tightly-controlled environments (e.g. online settings, industrial applications, etc.)
- Potential outcomes are likely to be very useful in more complex, less-controlled environments (e.g. social sciences, health sciences, etc.) where we use design aspects to eliminate competing explanations for putative causal effects

My Take On The Matter

- The two frameworks are not mutually exclusive, and in fact, they are complementary
- This is the approach I am taking in my book, *Causal Inference in Statistics*, where I introduce the reader to both frameworks and show how they can be used together (To learn more, see links on last slide)
  - My assumption is that most researchers care about the practical aspects of causal inference, and not so much about the philosophical underpinnings of the two frameworks, much like the bayesian-frequentist debate

Thank You :) - Questions?

Contact Information and other clickable stuff

- **My Personal Website**
- If you are interested in Consulting, Training, Workshops, Public Speaking, Guest Lectures, and Other Custom Services in statistics and causal inference, please do not hesitate to contact me **belairjustin@gmail.com**
- The first chapter of my book, *Causal Inference in Statistics*, is available for free on my website by **clicking here**
- I share a montly newsletter to stats nerds, causal inference enthusiasts, and data scientists, **subscribe here**
- I founded **biostatistics.ca**, a blog and biostatistics community.
- I share content daily on **Linkedin**
- I am working online an online course, **Introduction to Biostatistics**
- I created a **Linkedin Causal Inference Group**
- If you would just like to chat about science, stats, causal inference, or any other topic, please send me a DM on Linkedin or reach out via **belairjustin@gmail.com**.