Acceptance Code #

carelonresearch.com

Using the potential outcomes model to inform the design and interpretation of causal studies of the effectiveness and safety of healthcare interventions: an interdisciplinary perspective

Katherine M. Harris, PhD, MA[1]; Michael Grabner, PhD[1]; Shelly-Ann M. Love, PhD, MS[1]; Ruth Dixon, PhD[1]; Sarah Hoffman, PhD, MS[1]; Anna Wentz, PhD, MPH[1]

[1] Carelon Research, Wilmington, DE

Objectives

To describe the potential outcomes model (POM) as a framework for comparing the features and applications of three causal study designs that are commonly used to measure the effects of healthcare interventions on health-related outcomes.

Background

Healthcare decision makers need accurate, reliable information about the effects of medical treatments on health and cost outcomes. Causal effect estimation strives to combine data, statistical modeling, and subject matter expertise to measure the impacts of exposure to treatment on outcomes while reducing the potential for bias due to factors that are associated with both treatment assignment and outcomes.

The POM provides a common framework for the transparent design and testing of hypotheses about causal effects using randomized experiments and observational data.[1] It was introduced to the medical and social sciences by Donald Rubin in the 1970s; and its influence grew rapidly during the 2010s.[2-4] The POM defines a causal effect for a given individual as a comparison between two states of the world, referred to as "**potential outcomes**": one had they been treated and the other had they been untreated.[5] Because an individual cannot be both treated and untreated at the same time, only the outcome observed after treatment assignment is observable and the other is hypothetical and referred to as their "**counterfactual**" outcome. This concept informs the design, interpretation, and validation of causal estimates in ways that apply to both randomized and observational study designs.

The POM defines a set of three causal parameters or "estimands"[6]:

1) **Average treatment effect on the treated** (ATT) measures the expected treatment effect for the subgroup of treated individuals.
2) **Average treatment effect on the untreated** (ATU) measures the expected treatment effect for the subgroup of untreated individuals.
3) **Average treatment effect** (ATE) is the sum of the ATT and ATU weighted by the shares of treated and untreated individuals.

The ATE is equivalent to the expected value of the "simple difference in outcomes" (SDO) for treated and untreated groups, under three conditions. The first, "**exchangeability**," is satisfied when treatment assignment is unrelated to outcomes. [7,8] The second is referred to as the "**Stable Unit Treatment Value Assumption**" and is satisfied when (1) the effect of treatment on one individual does not affect the effect of treatment in others and (2) the effect of treatment is unrelated to the circumstances surrounding treatment assignment.[6] When using observational data to measure treatment effects, investigators can promote this consistency by defining treatment assignment in a way that is both conceptually appropriate given the specific causal question and measurable.

The third condition is "**positivity**." It requires that all study members have a non-zero chance of receiving the treatment or control condition. RCTs promote positivity by design.

In the absence of random assignment, it is reasonable to expect treatment and assignment to be correlated. The SDO in observational studies equals the ATE, plus two types of bias capturing associations between treatment assignment and outcomes related to violations of exchangeability[9,10]: (1) **Selection bias** is the difference in outcomes in the absence of treatment for treated and untreated individuals. It measures the presence of systematic differences between treatment and comparison groups that may be related to outcomes.[11] (2) **Selection on benefits of treatment** occurs when individuals treated respond differently to treatment than do untreated individuals causing treatment to appear more beneficial than it is.

Exhibit 1: Data from a hypothetical randomized experiment

Patient ID	Treatment assignment (D)	Actual and potential outcomes		Individual treatment effect	Observed outcome
		Treated outcome (Y^1)	Untreated outcome (Y^0)		
1	1	7	7	0	7
1	1	5	5	0	5
3	1	4	3	1	4
4	1	10	10	0	10
5	1	9	9	0	9
6	0	7	7	0	7
7	0	5	5	0	5
8	0	4	3	1	3
9	0	10	10	0	10
10	0	9	9	0	9

Blue shaded areas indicate unobservable values. Y^1 and Y^0 represent counterfactual outcomes in treated and untreated states; D is a dichotomous indicator of treatment (0-not treated, 1-treated).

Exhibit 2: Data from a hypothetical observational study

Patient ID	Treatment assignment (D)	Actual and potential outcomes		Individual treatment effect	Observed outcome
		Treated outcome (Y^1)	Untreated outcome (Y^0)		
1	1	7	1	6	7
1	1	5	1	4	5
3	1	4	2	2	4
4	1	10	1	9	10
5	1	9	8	1	9
6	0	5	6	-1	6
7	0	7	8	-1	8
8	0	1	7	-6	7
9	0	10	6	-1	6
10	0	9	7	-4	7

Blue shaded areas indicate unobservable values. Y^1 and Y^0 represent counterfactual outcomes in treated and untreated states; D is a dichotomous indicator of treatment (0-not treated, 1-treated).

Exhibit 3: Potential outcomes model (POM)-based causal analysis of simulated randomized experiment and observational study data

POM-based quantities	Mathematical formula	Study design	
		Randomized experiment	Observational
Simple difference in mean outcomes (SDO)	E[Y1 | D = 1] - E[Y0 | D = 0]	0.2	0.2
Difference in expected outcomes	E[Y1] - E[Y0]	0.2	0.9
Selection bias	E[Y0 | D = 1] - E[Y0 | D = 0]	0.0	-4.2
Average treatment effect on the treated (ATT)	E[Y1 | D = 1] - E[Y0 | D = 1]	0.2	4.4
Average treatment effect on the untreated (ATU)	E[Y1 | D = 0] - E[Y0 | D = 0]	0.2	-2.6
Average treatment effect (ATE)	p x ATT + (1-p) x ATU	0.2	0.9
Selection based on benefits	(1-p) x (ATT - ATU)	0.0	3.5
Average treatment effect on the treated + selection bias	ATT + selection bias	0.2	0.2

"E" represents expected outcomes; Y^1 and Y^0 represent counterfactual outcomes in treated and untreated states; D is dichotomous indicator of treatment (0 represents not treated, 1 represents treated) exposure; p is the probability of treatment and is equal to 0.5 in both datasets.

Exhibit 4: Side-by-side comparison of causal study designs

	Randomized experiments	Quasi-experiments	Observational studies
Description	Use physical randomization of treatment and experimental control over the administration of treatment to eliminate factors that confound the relationship between treatment and outcomes.	Leverage natural experiments that determine exposure to interventions of research interest, independent of patient characteristics (e.g., travel distance to emergency care to "randomize" assignment to time sensitive treatment).	Use observed confounders statistical modeling to identify the effects of interventions under the assumption of conditional exchangeability[a], often in large, diverse populations where physical randomization is impractical. Guided by graphic representations of relationships among exposures, outcomes, and measured and unmeasured confounders.
Empirical approaches	• Unadjusted and adjusted tests of group differences • Detection of heterogenous treatment effects	• Difference-in-differences • Instrumental variables • Regression discontinuity • Randomization inference	• Linear and non-linear regression models • Propensity score methods • Double robust methods
Strengths	• Absence of selection bias permits estimation of average treatment effects over treated and untreated subjects • Clear interpretation of effect estimates • Simple statistical tests • Strong internal validity • Transparency and traceability via pre-established protocols	• Potential to isolate causal effects by de-coupling treatment assignment and outcomes • Reflects real-world practice	• Overcomes pragmatic and ethical considerations associated with RCTs • Cost-effective use of real-world data • Flexibility to support investigation of emergent public health concerns and long-term outcomes
Opportunities	Consider pragmatic trials or pragmatic elements to traditional trials to promote understanding of the effectiveness of complex interventions in real-world settings, especially when treatment adherence is a concern.	Use care in generalizing the results of quasi-experiment. If treatment effects differ across population subgroups, treatment effect estimates, referred as local average treatment effects.[b] Supplemental analysis can help to characterize subpopulations contributing to the identification of LATEs.	Use double robust methods and ML to reduce the potential for biased causal effect estimates due to misspecification of outcome and/or treatment assignment models. Use Quantitative Bias Analysis to estimate magnitude and direction of potential remaining biases (e.g., unmeasured confounding). Use "target trial methods" to improve accuracy of causal effect estimates by aligning the timing of treatment exposure and outcomes measurement.

[a]Conditional exchangeability refers to the independence of potential outcomes from the treatment assignment, given a set of observed covariates (Hernan 2020, Fine Point 4.1). [b]Local average treatment effects (LATEs) represent only unobserved subpopulations whose treatment status was affected by the quasi-randomizing mechanism (Cunningham 2021, pp. 176).

Methods

- Data Sources: Simulated data representing a randomized experiment and an observational study with the same observable mean difference in outcomes (SDO) **(Exhibits 1 and 2).**
- Analysis: Calculated using formulas listed in Exhibit 3, including simple difference in mean outcomes (SDO), difference in expected outcomes; selection bias, average treatment effect on the treated (ATT), average treatment effect on the untreated (ATU), average treatment effect (ATE), selection based on benefits, and average treatment effect on the treated plus selection bias **(Exhibit 3).**
- Compared the key features of three POM-based causal study designs: (1) randomized experiments, (2) quasi-experiments, and (3) observational studies **(Exhibit 4).**

Results

In the absence of selection bias and selection based on benefits of treatment in the randomized design, SDO = ATE = ATT = ATU = 0.2. By contrast, the ATE from the observational study was 0.9 reflecting presence of selection bias (-4.2) and selection on benefits of treatment (1.8) that resulted in the non-equivalence of average treatment effects in treated and untreated groups (ATT = 4.4 vs ATU = -2.6. Finally, the SDO was equal to ATT + selection bias, which does not have a clear causal interpretation.

While randomized studies can yield unbiased causal estimates, these estimates may not reflect real-world clinical practice or be generalizable. Quasi-experiments and observational studies can improve generalizability and realism, but require subject matter expertise, complex models, and assumptions to address sources of bias. POM concepts can strengthen causal inference even in the absence of randomization.

Conclusions

The POM provides a concrete framework assessing the trade-offs associated with alternative approaches for measuring the effectiveness of healthcare interventions on health outcomes. The POM also clarifies the implications of study design decisions for causal inference with observational data more transparent.

References

1. Rubin, DB For Objective Causal Inference, Design Trumps Analysis. The Annals of Applied Statistics 2008, Vol. 2, No. 3, 808–840.
2. Splawa-Neyman, Jerzy. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Annals of Agricultural Sciences, 1923, 1–51.
3. Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688–701.
4. Scott Cunningham, Causal Inference Mixtape. Yale University Press, 2021, Chapter 5, Section 5.3.3.3
5. Donald Rubin, For Objective Causal inference, Design Trumps Analysis. Annals of Applied Statistics 2008, 2(3), 808-840
6. Schwartz S, Gatto NM, Campbell UB. Extending the sufficient component cause model to describe the Stable Unit Treatment Value Assumption (SUTVA). Epidemiol Perspectives and Innovations. 2012, 9:3.
7. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, p. 26.
8. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, Chapter 1.
9. Angrist, J. D., & Pischke, J.-S. (2008). Mostly harmless econometrics. Princeton University Press. Page 54.
10. Cunningham, S. (2021). Causal Inference: The Mixtape. New Haven, CT: Yale University Press. Pp.133.
11. See Varga, A., et al. Bias? Health Serv Outcomes Res Method 23, 354–375 (2023).; and Lu H, et al. Epidemiology. 2022 Sep 1;33(5):699-706. for discussion of alternative definitions of the term "selection bias" in the context of regression analysis and measurement error.

Funding and disclosures

No funding was received for the conduct of this study. All authors are employees of Carelon Research, which conducts health outcomes, safety, and epidemiological research with both internal and external funding, including a variety of private and public entities.

QR Code

Katherine Harris | Katherine.Harris@carelon.com