Exploring Causal Inference: Counterfactuals, RCTs, and Practical Pitfalls

Avinash Bunga

Master of Science in Information Systems and Business Analytics

Park University

CIS625HOS2P2025 Machine Learning for Business

Professor: Abdelmonaem Jornaz

May 11, 2025

Exploring Causal Inference: Counterfactuals, RCTs, and Practical Pitfalls

Causal inference moves from observing patterns to testing whether one factor truly causes an outcome. This assignment addresses four key tasks: illustrating the fundamental problem of causal inference, clarifying when associations can imply causation, designing a randomized control trial for a school-based oatmeal intervention, and identifying common RCT pitfalls. Practical, automotive-themed examples and structured tables support each concept. All numerical values in these examples and tables are hypothetical and provided solely for illustrative purposes (Pearl, 2010).

1. Fundamental Problem of Causal Inference

The fundamental problem of causal inference arises because the same unit cannot be observed simultaneously with and without treatment. For example, a logistics company installing speed limiters on half of its delivery vans to test accident claims cannot measure claims for a single van under both conditions simultaneously.

Because the missing outcome (counterfactual) cannot be observed directly, methods such as randomization, matching, or modeling are used to estimate what accident claims would look like under the opposite condition (Meghanath, 2024). This challenge underscores the need for careful study design in causal analysis.

Table 1

Accident claims and counterfactual estimates by limiter status

Van ID	Limiter Installed?	Accident Claims (2024)	Counterfactual Claims
101	Yes	2	?
102	No	5	?

Note. The question mark indicates the unobserved counterfactual outcome for each van.

2. When “Association Is Causation”

An observed association may suggest but does not confirm causation. Three strict criteria are required: the cause precedes the effect, plausible confounders are ruled out, and the effect disappears when the cause is removed.

A real world traffic safety case illustrates these principles. When a city reduced its speed limit from 60 mph to 50 mph without other changes, serious accident rates fell by 20%. Accident rates rose again when the limit returned to 60 mph, supporting a causal interpretation (Lee, Aronson, & Nunan, 2019).

3. RCT Design: Oatmeal for School Breakfast

A randomized control trial evaluated whether serving oatmeal with fruit toppings improved student focus during first-period classes. A total of 240 students in Grades 4–6 were randomly assigned stratified by grade and baseline focus score to an oatmeal breakfast group or a regular-menu group.

The five-week intervention included three measurement points. Baseline focus scores were recorded in Week 0. Daily teacher-rated focus scores were collected during Weeks 1–4, and a post-test quiz measured attention in Week 5. Confounders such as allergies, sleep hours, and socioeconomic background were balanced by stratified randomization. Parental consent and opt-out options ensured ethical compliance (Alexander, Lopes, Ricchetti-Masterson, & Yeatts, n.d.).

Table 2

RCT timeline for oatmeal intervention

Week	Treatment Group	Control Group	Measurement
0	Regular menu	Regular menu	Pre-test focus self-rating
1–4	Oatmeal + fruit topping	Regular menu	Daily teacher focus rating

5	None	None	Post-test quiz on attention

Note. Table 2 outlines randomized group assignments and corresponding focus measurements to evaluate the effect of the oatmeal intervention.

4. Pitfalls of RCTs: The following two pitfalls illustrate common challenges in applying RCTs:

- **Pitfall 1: Non-compliance and attrition.** Participants may deviate from assignments, diluting treatment effects. For example, if 15% of control students eat oatmeal outside the study or 10% of treatment students skip breakfast, intent-to-treat analysis preserves assignment but underestimates intervention impact.
- **Pitfall 2: Limited generalizability.** Findings from one setting may not apply elsewhere. A school with health-conscious families may not represent broader district demographics.

Table 3

Demographic comparison of pilot school and district average

Demographic Metric	Pilot School (%)	District Average (%)
Free/reduced lunch	10	45
Average parental education	70	40

Note. Table 3 compares pilot school demographics with district averages to highlight generalizability concerns.

Conclusion

The assignment shows how clear experimental design, strict criteria for interpreting associations, correct counterfactual estimation, and being aware of common mistakes are needed for making conclusions about causes. Examples from delivery logistics, traffic safety, and educational interventions reveal how structured methods enable valid cause-and-effect conclusions.

References

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K., & Yeatts, K. B. (n.d.). *Randomized controlled trials (experimental studies)* (2nd ed.). ERIC Notebook Series. Department of Epidemiology, UNC-Chapel Hill. Retrieved May 11, 2025, from https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC10.pdf

CodeEmporium. (n.d.). *Causal inference – EXPLAINED!* [Video]. YouTube. Retrieved May 8, 2025, from https://www.youtube.com/watch?v=Od6oAz1Op2k

Lee, H., Aronson, J. K., & Nunan, D. (2019, March 5). *Association or causation? How do we ever know?* Catalog of Bias. Retrieved May 11, 2025, from https://catalogofbias.org/2019/03/05/association-or-causation-how-do-we-ever-know/

Meghanath, G. (2024, November 5). *Causal analysis overview: Causal inference versus experimentation versus causal discovery*. Data Science at Microsoft. Medium. Retrieved May 11, 2025, from https://medium.com/data-science-at-microsoft/causal-analysis-overview-causal-inference-versus-experimentation-versus-causal-discovery-d7c4ca99e3e4

Pearl, J. (2010). *An introduction to causal inference*. *The International Journal of Biostatistics, 6*(2), Article 7. https://doi.org/10.2202/1557-4679.1203

Yaghi, S., Siegler, J. E., & Nguyen, T. N. (2023). *Pitfalls of randomized controlled trials in stroke: How can we do better? Stroke: Vascular and Interventional Neurology, 3*(4). https://doi.org/10.1161/SVIN.123.000807