Causation, Correlation and Association

By Dr Slava Vaisman, University of Queensland School of Mathematics and Physics

In the natural sciences, we are constantly observing variables that move in tandem. For example, when temperatures rise, retail sales often surge, or when a new corporate policy is rolled out, employee absenteeism might spike. These patterns are tempting to simplify, but they pose a fundamental question: Is the relationship causal, or are both variables being driven by a separate, external force? Confusing correlation (or, more generally, association) with causation is not just a statistical slip; it is one of the most pervasive and expensive logical fallacies in scientific research, business, and public policy. Misinterpreting these relationships can lead to deeply flawed scientific conclusions, wasted resources, ineffective laws, and fundamentally unsound strategic decisions.

Why is it relevant?

Mistaking correlation for causation leads to erroneous scientific conclusions, a breakdown of public trust, and wasted investment in strategies that target the wrong variables. For instance, a city might invest heavily in more police patrols in neighbourhoods with high ice cream sales, failing to realize that both are simply correlated with rising summer temperatures. By targeting the symptom rather than the driver, organizations often fail to address the real root cause of a problem.

The stakes involved make the causality subject essential not only for researchers, journal editors, and reviewers who safeguard scientific rigor, but also for the policymakers and analysts who translate data into real-world action.

Practical Steps

1. Before assuming A causes B, ask if there exists a hidden factor C that influences both A and B? A classic example is the correlation between ice cream sales and drowning incidents. The confounder here is hot weather: it causes people both to buy ice cream and to go swimming. More swimmers, in turn, lead to more drownings.

2. Ask if A causes B, or maybe B causes A, or could the causation be circular? For example, consider the link between happiness and productivity: are happy

employees more productive, or does being productive (receiving praise) cause happiness?

3. Finally, to verify that A causes B, perform a controlled test by randomly splitting your subjects into two groups. Change A for one group (the treatment) and leave it unchanged for the other (the control). Because the groups were assigned randomly, any difference in B must be caused by the change in A.

Case Study

Case Study 1: A city observes that neighbourhoods with the most public parks also have the highest household incomes and the lowest crime rates. This leads to the (false) conclusion that building a new park in a distressed area will directly create local prosperity.

Applying the three rules helps to expose the flaw.

1. Identify the confounder. The initial conclusion ignores the role of the municipal tax base. Wealthier neighbourhoods generate more tax revenue, which the city then uses to build and maintain parks. Prosperity causes the parks, not the other way around.

2. Examine the Direction. The assumed causal direction is also questionable. While parks may attract some residents, it is equally plausible that wealthy residents possess the political capital to demand more parks in their neighbourhoods, creating a circular relationship.

3. To properly justify the investment, the council would need to conduct a controlled experiment. This would involve introducing parks in a selected group of neighbourhoods while leaving a comparable control group unchanged. After several years, the council could then analyse the differences in household incomes and crime rates between the two sets of neighbourhoods. Without this rigorous comparison over time, the city cannot validate the expense. By treating a symptom of wealth (green space) rather than the root causes of poverty, it risks wasting public funds.

Case study 2: In orthopaedic research, observational studies may show that patients with very low BMI experience higher mortality rates after total joint arthroplasty (TJA). This suggests an observable association, and statistical analyses may simply indicate a clear correlation between low BMI and mortality. However, more comprehensive investigations often reveal a more complex picture than simply concluding that low BMI causes higher mortality. This example helps clarify the different possible relationships between these variables:

- *Association:* Patients with very low BMI were observed to have higher mortality rates after TJA.
- *Correlation:* Statistical analyses showed that lower BMI consistently coincided with increased postoperative mortality.
- *Causation:* At first glance, the findings may suggest that low BMI causes higher mortality risk. However, many low-BMI patients also had serious underlying illnesses such as cancer, frailty, or heart failure, which both reduced BMI and increased mortality risk. In this case, the observed relationship was partly driven by confounding and reverse causality rather than a direct causal effect of BMI itself.

This example shows why statistical relationships alone are insufficient to establish causation and why researchers must carefully account for confounding variables and reverse causality when interpreting observational findings. [1]

References

1. Zaniletti, I., Larson, D. R., Lewallen, D. G., Berry, D. J., & Kremers, H. M. (2023). How to distinguish correlation from causation in orthopaedic research. The Journal of arthroplasty, 38(4), 634-637.
2. Pearl, J. Glymour, M., and Jewell, N. P. (2016). *Causal inference in statistics: a primer*. Wiley.
3. Pearl, J. (2009). *Causality* (2nd edition). Cambridge University Press. Accessible at: https://doi.org/10.1017/CBO9780511803161
4. Dawid, P. (2021) 'Decision-theoretic foundations for statistical causality. Journal of Causal Inference, 9(1), 39–77. Accessible at: https://doi.org/10.1515/jci-2020-0008
5. Pearl, J. and D. Mackenzie (2018) *The Book of Why: The New Science of Cause and Effect*, Basic Books.
6. Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
7. Tsoumas, I., Sitokonstantinou, V., Giannarakis, G., Lampiri, E., Athanassiou, C., Camps-Valls, G., Kontoes, C. and Athanasiadis, I.N. (2025) 'Leveraging causality and explainability in digital agriculture', *Environmental Data Science*, *4*, e23. Accessible at: https://doi.org/10.1017/eds.2025.14

8. Abell, P. and Engel, O. (2021) ‘Subjective Causality and Counterfactuals in the Social Sciences: Toward an Ethnographic Causality?’, *Sociological Methods & Research*, *50*(4), 1842–1862. Accessible at: https://doi.org/10.1177/0049124119852373

9. Causality in digital medicine. (2021). *Nature Communications*, *12*(1), article 5471. Accessible at: https://doi.org/10.1038/s41467-021-25743-9