



Association, Correlation, and Causation: Three Important and Misleading Terms in Research

Farzan Madadzadeh^{1,2} and Sajjad Bahariniya^{3,*}

¹Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

²Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

³Department of Healthcare Services Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Abstract:

This study investigates three key ideas in statistics: association, correlation, and causation. When two variables change together, it does not necessarily mean that one variable causes the other to change. An association indicates that certain variables tend to appear or vary together, without confirming a direct link. Correlation goes a step further by measuring how strongly and in what direction two variables are linearly related, yet it still does not establish a cause-and-effect connection. Causation is the strongest form of relationship, yet proving it demands well-planned studies that account for potential confounding factors. Understanding the distinctions between these concepts is crucial for accurate data interpretation, sound study design, and the prevention of misleading conclusions.

Keywords: Causation, Statistics, Research interpretation, Directed acyclic graphs, Misleading terms.

© 2026 The Author(s). Published by Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Address correspondence to this author at the Department of Healthcare Services Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran;
E-mail: sajjadbahari98@gmail.com

Cite as: Madadzadeh F, Bahariniya S. Association, Correlation, and Causation: Three Important and Misleading Terms in Research. Open Public Health J, 2026; 19: e187494451232. <http://dx.doi.org/10.2174/01187494451232260409114348>



CrossMark

Received: October 27, 2025

Revised: December 02, 2025

Accepted: January 27, 2026

Published: June 03, 2026



Send Orders for Reprints to
reprints@benthamscience.net

Dear editor,

Association, correlation, and causation are fundamental ideas in statistics, but they are often misinterpreted or conflated. Association refers to a relationship between two variables in which a change in one tends to coincide with a change in the other. However, this association does not necessarily mean that one variable directly causes a change in the other [1]. In contrast, correlation is a more precise way of assessing the relationship between two variables. This relationship is usually expressed using a correlation coefficient between -1 and 1, which indicates the strength of the relationship in a linear fashion [2, 3]. It simply shows that two variables are connected or tend to vary together, without

demonstrating that one directly produces changes in the other [3]. The third, and most complex, of the three concepts is causality. This concept indicates that a change in one variable directly causes a change in the other. Demonstrating this type of relationship usually requires carefully designed experiments to ensure that the observed effect is actually caused by the variable under investigation [4].

Researchers use validated methodological approaches to determine whether a causal claim is justified. In fields such as epidemiology, the Austin-Bradford-Hill criteria provide a structured framework for assessing whether an observed relationship is likely to be causal. These criteria consider aspects such as the timing of events, consistency

of findings across studies, biological plausibility, and the existence of dose-response relationships. Modern approaches to causal inference also use counterfactual models, which assess causality by comparing actual outcomes with hypothetical outcomes under different conditions. Directed acyclic graphs (DAGs) are often used to map hypothetical causal pathways and identify potential confounding variables. Using these tools allows researchers to draw more robust conclusions and avoid mistaking simple associations or correlations for definitive evidence of cause and effect.

Research methodology highlights the importance of distinguishing between association, correlation, and causality. If these concepts are not carefully separated, study findings can easily be misinterpreted [5]. A careful analysis of rejected studies in biomedical research shows that a large number of these rejections are due to incorrect claims about causality, misuse of analytical techniques, or misunderstanding of statistical correlations [6]. These observations highlight the importance of distinguishing true causal relationships from descriptive associations, as well as the broader consequences that can arise when research concepts are misinterpreted.

This study highlights the importance of distinguishing between association, correlation, and causation. For example, careful reviews of randomized controlled trials conducted during the COVID-19 pandemic have shown that authors sometimes overstate the causal significance of their results, even when the trial data do not fully support such claims [7]. This pattern highlights the ongoing risk of confusing correlations with true causal effects, particularly in high-pressure research environments. Similarly, long-term assessments of methodological quality in fields such as preclinical urology have revealed significant inconsistencies, including inconsistent adherence to the standards required to make reliable causal claims [8].

Conducting sound research requires a clear understanding of the differences between association, correlation, and causation, as confusing them can lead to misleading scientific conclusions. For example, some influential observational studies on nutrition and chronic disease have mistakenly inferred causal relationships from observed associations. By carefully applying established causal frameworks, researchers can more accurately report their findings [9, 10].

In areas such as healthcare, where decisions based on statistical data can have real-world impacts on people's lives, confusing association, correlation, and causation can lead to incorrect conclusions [11]. Therefore, it is crucial to use sound judgment and critical thinking when evaluating statistical relationships in research. Researchers should be transparent about their methods and results. Ultimately, a proper understanding of these concepts enables better decision-making and contributes to a more informed society.

AUTHORS' CONTRIBUTIONS

The authors confirm contribution to the paper as follows: F.M.: Contributed to conceptualization, review, project management, writing the original draft, reviewing, and editing; S.B.: Contributed to writing the original draft, reviewing, and editing.

LIST OF ABBREVIATIONS

DAGs = Directed acyclic graphs
COVID-19 = Corona Virus Disease 2019

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Heumann C, Schomaker M. Association of two variables. Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R. Springer 2023; pp. 69-97.
- [2] Asuero AG, Sayago A, González AG. The correlation coefficient: An overview. Crit Rev Anal Chem 2006; 36(1): 41-59. <http://dx.doi.org/10.1080/10408340500526766>
- [3] Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg 2018; 126(5): 1763-8. <http://dx.doi.org/10.1213/ANE.0000000000002864> PMID: 29481436
- [4] Pearl J, Verma TS. A theory of inferred causation. Studies in Logic and the Foundations of Mathematics. Elsevier 1995; pp. 789-811. [http://dx.doi.org/10.1016/S0049-237X\(06\)80074-1](http://dx.doi.org/10.1016/S0049-237X(06)80074-1)
- [5] Kim MS, Yon DK, Lee SW, *et al.* Quantifying the impact of immortal time bias: empirical evidence from meta-analyses. J R Soc Med 2025; 118(9): 293-303. <http://dx.doi.org/10.1177/01410768251366880> PMID: 40842270
- [6] Hwang SY, Yon DK, Lee SW, *et al.* Causes for retraction in the biomedical literature: A systematic review of studies of retraction notices. J Korean Med Sci 2023; 38(41): e333. <http://dx.doi.org/10.3346/jkms.2023.38.e333> PMID: 37873630
- [7] Oh Y, Jung YJ, Sujata P, *et al.* Spin in randomized controlled trials of pharmacology in COVID-19: A systematic review. Account Res 2025; 32(3): 214-32. <http://dx.doi.org/10.1080/08989621.2023.2269083> PMID: 37818630
- [8] Park S, Lee S, Park S, Kim E, Pizzol D, Trott M. Methodological rigour in preclinical urology: A systematic review reporting research quality over a 14-year period. BJU Int 2023; 133(4): 387-99. <http://dx.doi.org/10.1111/bju.16171> PMID: 37667439
- [9] Murdock M. Causation, relationship, association, and correlation: Narrative review. EC Orthopaedics 2022; 13: 01-12.
- [10] Rezaee Z, Aliabadi S, Dorestani A, Rezaee NJ. Application of time series models in business research: correlation, association, causation. Sustainability 2020; 12(12): 4833. <http://dx.doi.org/10.3390/su12124833>
- [11] Altman N, Krzywinski M. Association, correlation and causation. Nat Methods 2015; 12(10): 899-900. <http://dx.doi.org/10.1038/nmeth.3587> PMID: 26688882