International Clinical and Medical Case Reports Journal
Case Report (ISSN: 2832-5788)

The Illusion of Association: Simpson's Paradox in Clinical and Epidemiological Studies

Stephnie Hon[1], John Croucher[2*]

[1]International College of Management Sydney, Sydney, Australia

[2]Macquarie Business School, Macquarie University, Sydney, Australia

Citation: *Stephnie Hon, John Croucher. The Illusion of Association: Simpson's Paradox in Clinical and Epidemiological Studies. Int Clinc Med Case Rep Jour. 2026;5(11):1-7.*

Received Date: 11 November, 2026; **Accepted Date:** 12 November, 2026; **Published Date:** 13 November, 2026

***Corresponding author:** John Croucher, Macquarie Business School, Macquarie University, Sydney, Australia

Copyright: © John Croucher, Open Access 2026. This article, published in Int Clinc Med Case Rep Jour (ICMCRJ) (Attribution 4.0 International), as described by http://creativecommons.org/licenses/by/4.0/

ABSTRACT

Simpson's paradox is a statistical phenomenon in which a trend that appears in several separate groups of data either disappears or reverses when the groups are combined. That is, the relationship between two variables can look one way within each subgroup but look completely different when all the data are aggregated. Although widely reported in epidemiology and other branches of health care, the paradox remains a major source of misinterpretation in medical research. This paper examines Simpson's paradox and its consequences for analysing clinical outcomes and treatment effectiveness. It outlines the paradox's mathematical basis and shows how confounding, selection bias, and inappropriate aggregation can produce misleading conclusions. If Simpson's paradox is not recognised, it may lead to incorrect treatment recommendations and incorrect policy decisions. However, understanding the paradox helps identify hidden sources of bias and interpret statistical results within their clinical context. Accounting for Simpson's paradox is essential for all healthcare researchers, as it emphasises the need to look for subgroup effects before drawing conclusions from aggregated medical data. Another source of confounding variables is spurious correlation, where two variables may seem related only because they are each related to another variable that may not have been considered. For both Simpson's paradox and spurious correlation, case studies are provided.

Keywords: Simpson's paradox; Health care; Spurious correlation; Epidemiology; Confounding variables

INTRODUCTION

Simpson's paradox occurs when a trend observed in aggregated data disappears or reverses when the data are divided into meaningful subgroups. Although it was first discussed by the eminent statistician Karl Pearson[1] and several years later by George Udny Yule, who referred to it as "illusory association",[2] it was a paper nearly half a century later by Edward Hugh Simpson that illustrated how combining contingency tables can lead to paradoxical conclusions.[3]

Understanding Simpson's paradox is especially important in medical research because clinical decisions, treatment recommendations, and health policies are often based on statistical analyses of clinical results. If the paradox is overlooked or misinterpreted, it can lead to wrong conclusions and erroneous clinical decisions.[4]

The paradox appears in many areas of medical science, including epidemiology, where understanding it is essential to draw correct conclusions from the data.[5]

CASE STUDY 1

A prominent example of Simpson's paradox involves the treatment of kidney stones. In a landmark study, researchers found that one treatment appeared superior when all patients were analysed together, yet the alternative treatment demonstrated higher success rates within both small-stone and large-stone patient groups.[6] This study was designed to compare different methods of treating renal calculi to establish which was the most cost-effective and successful. Of the patients with renal calculi, 350 underwent open surgery (Treatment A), and 350 underwent percutaneous nephrolithotomy (Treatment B), which was less invasive. Treatment was defined as successful if stones were eliminated or reduced to less than 2.0 mm after three months. The overall success rate is shown in Table 1.

Table 1: Aggregated data for the kidney stone experiment

Treatment	No. successful	No. unsuccessful	Total	Success rate
A	273	77	350	78.0%
B	289	61	350	82.6%

The data in Table 1 clearly suggest that the less invasive Treatment B was better overall. Now the data are broken down according to kidney stone size. That is, the data were stratified into small vs. large stone subcategories and the success rate then computed according to the treatment. The result for small stones (< 2.0 cm) are shown in Table 2 and for large stones (≥ 2.0 cm) are shown in Table 3.

Table 2: Results for the kidney stone experiment with small stones (< 2.0 cm)

Treatment	No. successful	No. unsuccessful	Total	Success rate
A	81	6	87	93.1%
B	234	36	270	86.7%

Table 3: Results for the kidney stone experiment with large stones (≥ 2.0 cm)

Treatment	No. successful	No. unsuccessful	Total	Success rate
A	192	71	263	73.0%
B	55	25	80	68.8%

From Tables 2 and 3, for both small stones and large stones, Treatment A (surgery) had the higher success rate, completely reversing the conclusion from the aggregate data in Table 1. In this case, the same data produced a counterintuitive result. The reason for this was the presence of a confounding variable (stone size) that led to Simpson's paradox at work.

From Table 2, most of the patients (75.6%) with small stones (easier to treat successfully) received Treatment B, while from Table 3, most of the patients with large stones (76.7%) received Treatment A, which were harder to

treat successfully. Within each stone-size category, Treatment A had the higher success rate, but because it was used predominantly on difficult cases, its overall success rate was dragged down.

Mathematical explanation for Case Study 1

The *overall* success rates are weighted averages:

For Treatment A:

$$0.25(93.1\%) + 0.75(73.0\%) = 78.0\%$$

For Treatment B:

$$0.77(86.7\%) + 0.23(68.8\%) = 82.6\%$$

Although Treatment A has the higher success rate in both subgroups, its much larger proportion of difficult cases produces a lower overall success rate. The discrepancy occurred because patients were unevenly distributed between treatment groups according to stone severity.

CASE STUDY 2

In a hypothetical scenario, two hospitals, A and B, are being compared with their mortality rates when treating pneumonia. A random sample of 1000 patients is taken from each hospital, with the overall mortality rate shown in Table 4.

Table 4: Aggregated data for the mortality rate for the two hospitals combined

Hospital	No. deaths	No. survivors	Total	Mortality rate
A	180	820	1000	18.0%
B	120	880	1000	12.0%

The aggregated data in Table 4 clearly suggest that the mortality rate is much higher in Hospital A when compared with Hospital B. The data are then broken down by pneumonia severity. That is, the data were stratified into mild vs. severe subcategories, and the mortality rate was then computed according to the type. The mortality rates for mild pneumonia patients are shown in Table 5 and severe pneumonia patients are shown in Table 6.

Table 5: Mortality rates for mild pneumonia patients

Hospital	No. deaths	No. survivors	Total	Mortality rate
A	10	490	500	2.0%
B	20	780	800	2.5%

Table 6: Mortality rates for severe pneumonia patients

Hospital	No. deaths	No. survivors	Total	Mortality rate
A	170	330	500	34.0%
B	100	100	200	50.0%

From Tables 5 and 6, for both mild and severe cases, Hospital A had the lower mortality rate, completely reversing the conclusion from the aggregate data in Table 4. Again, this is a counter-intuitive result using the same data. This was because a confounding variable (pneumonia severity) led to Simpson's paradox.

From Table 5, most of the patients (61.5%) with mild pneumonia (easier to treat successfully) received Treatment B, while from Table 6, most of the patients with severe pneumonia (71.4%) received Treatment A, which was harder to treat successfully. Within each severity category, Hospital A had a lower mortality rate when treating 50% of the severe cases, but because Hospital B predominantly treated the mild cases (80.0%), its overall mortality rate was dragged down.

It is important to consider where Simpson's paradox is relevant in many different scenarios, including, for example, evaluating surgical outcomes, assessing cancer survival rates, comparing COVID-19 mortality between healthcare systems, and ranking medical practitioners or medical centres. The two case studies shown here illustrate how healthcare assessments can be misleading if patient case-mix and disease severity are not properly adjusted. The overall result may tell the opposite story of the subgroup results.

SPURIOUS CORRELATION

In a similar situation to Simpson's paradox, another pitfall in statistical analysis occurs when a confounding (or lurking) variable can lead to misleading conclusions. A *spurious correlation* occurs when two variables appear statistically related, but the relationship is not causal. In other words, changes in one variable may seem associated with changes in another, even though neither variable directly influences the other. Such correlations can arise from coincidence or a hidden third confounding variable.[7]

It is important to recognise spurious correlations in medical and scientific research because it can lead to incorrect conclusions and poor decision-making. It is necessary to distinguish causation from correlation by conducting controlled experiments and using appropriate statistical methods to control for confounding variables.[8] Careful interpretation of statistical results is required, as relationships without a credible explanation may occur because of random variation or confounding variables.[7] Understanding spurious correlation helps ensure that conclusions are valid, reliable, and useful for decisions and policy development.[9]

CASE STUDY 3

A prominent example of spurious correlation in medicine involved the relationship between hormone replacement therapy (HRT) and heart disease. During the 1980s and 1990s, numerous observational studies suggested that postmenopausal women who used HRT had lower rates of coronary heart disease than women who did not. Researchers initially concluded that HRT protected women against cardiovascular disease.[7]

However, this conclusion was later shown to be largely false. Women who chose to take HRT tended to be wealthier, better educated, more health-conscious, and more likely to engage in healthy behaviours such as exercising, maintaining a nutritious diet, and obtaining regular medical care. These factors acted as confounding

variables, creating the illusion that HRT itself reduced heart disease risk[7] when in fact it was the lifestyle that was responsible.

The issue came to light with one of the largest and most influential medical studies ever conducted in the United States. Established in 1991 and funded by the U.S. National Institutes of Health, it was designed to investigate major causes of death, disability, and poor quality of life among postmenopausal women. The hormone replacement therapy (HRT) component of the study involved 16,608 healthy postmenopausal women, aged 50 to 79 years, who still had a uterus. Participants were randomly assigned to receive either a daily combination of oestrogen and progesterone or an identical placebo.[11] Because it was a randomised controlled trial, the study was able to eliminate many of the confounding factors that had affected earlier observational research.

The Women's Health Initiative trial found results that contradicted those previously reported in observational studies. It concluded that women who received HRT, compared with the placebo group, experienced:

- A 29% increase in coronary heart disease.
- A 41% increase in stroke.
- More than a twofold increase in venous thromboembolism (blood clots).
- A 26% increase in breast cancer.

That is, HRT did not provide the expected cardiovascular benefits and, in some cases, increased the risk of adverse health outcomes, including stroke and blood clots.[11] The earlier correlation between HRT use and lower heart disease rates had therefore led researchers to a false causal conclusion.

This case shows why correlation by itself is not enough to establish causation. Strong statistical relationships can arise from hidden, confounding variables rather than genuine cause and effect. It is important to control for these confounding factors which must be taken into account before drawing causal deductions.[7,8]

CONCLUSION

These case studies illustrate the importance of recognising confounding variables in medical research. Factors such as age, disease severity, sex, socioeconomic status, and comorbidities, among others, can modify outcomes and distort associations when data are analysed overall. In evidence-based medicine, understanding Simpson's paradox helps researchers account for confounding factors and estimate more accurate treatment effects.[12] These approaches allow clinicians to uncover genuine causal relationships rather than misleading conclusions.

The paradox also has implications for public health and healthcare management, where policymakers may rely on aggregated data to compare hospitals and allocate resources. However, without considering subgroup differences, decisions may be based on erroneous interpretations of healthcare performance. Consequently, understanding Simpson's paradox improves the quality of evidence used in policy development.[13]

A similar situation arises when using correlation to analyse the possible relationship between two variables. Confounding factors are not always easy to spot, but if in doubt, you can calculate a *partial* correlation coefficient that filters out the confounding factor. For example, an experiment may show that salary and systolic blood pressure have a strong positive correlation. On the surface, it may seem that the more money you earn (possibly in a more stressful job), the higher your systolic blood pressure will be. A confounding factor here is age. As you get older, you usually progress in your career and earn a higher salary. Also, as you age, your systolic blood pressure is likely to become higher. In fact, systolic blood pressure and salary may not be correlated at all.

When Simpson's paradox occurs in a medical context, any conclusion should be treated with caution. Aggregated data can produce deceptive conclusions if important confounding variables are ignored, but carefully examining subgroup effects can lead to more accurate clinical decisions and improve the reliability of the research.[14]

REFERENCES

1. Pearson, K. (1899). On the Theory of Genetic (Reproductive) Selection. *Phil. Trans. R. Soc. Ser. A.* 1899;192:257-330.
2. Yule, G.U. Notes on the Theory of Association of Attributes in Statistics. *Biometrika*. 1903;2:121–13.
3. Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B* (Methodological). 1951;13(2):238-241.
4. Bonovas, S, Piovani, D. Simpson's paradox in clinical research: A cautionary tale. *Journal of Clinical Medicine*, 2023;12(4):1633.
5. Sprenger J, Weinberger N. (2021). Simpson's Paradox. In *The Stanford Encyclopedia of Philosophy*, Summer 2021 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA.
6. Charig, C. R., Webb, D. R., Payne, S. R., & Wickham, J. E. A. (1986). Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *British Medical Journal*, 292(6524), 879-882.
7. Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.
8. Katz, M. H. (2022). *Multivariable analysis: A practical guide for clinicians and public health researchers* (4th ed.). Cambridge University Press.
9. Shepherd, D. A., & Suddaby, R. (2017). Theory building: A review and integration. *Journal of Management*, *43*(1), 59-86.
10. Stampfer, M. J., & Colditz, G. A. (1991). Estrogen replacement therapy and coronary heart disease: A quantitative assessment of the epidemiologic evidence. *Preventive Medicine*, 20(1), 47-63.
11. Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A. A., Howard, B. V., Johnson, K. C., Kotchen, J. M., & Ockene, J. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association*, 288(3), 321-333.

12. Sprenger, J., & Weinberger, N. (2021). Simpson's paradox. *Stanford Encyclopedia of Philosophy*
13. Sanders, S., Ehrlich, J., & Boudreau, J. (2023). Simpson's aggregation paradox in nonparametric statistical analysis: Theory, computation, and susceptibility in public health data. *Frontiers in Applied Mathematics and Statistics*, 9, Article 1169164.
14. Rojanaworarit, C. (2020). Misleading epidemiological and statistical evidence in the presence of Simpson's paradox: An illustrative study using simulated scenarios of observational study designs. *Journal of Medicine and Life*, 13(1), 37-44.