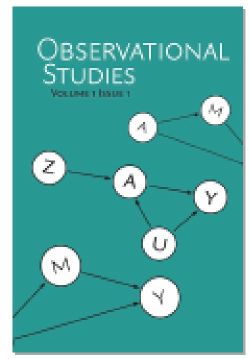
Cochran's Causal Crossword

Paul R. Rosenbaum

Observational Studies, Volume 1, Number 1, 2015, pp. 205-211 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2015.0021*



Observational Studies 1 (2015) 205-211 Submitted 5/15; Published 8/15

Cochran's Causal Crossword

Paul R. Rosenbaum rosenbaum@stat.wharton.upenn.edu
Department of Statistics, Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6340 US

Abstract

In discussing the "step from association to causation," Cochran described a certain "multiphasic attack" as "one of the most potent weapons in observational studies." This method emphasized assembling several weak strands of evidence that become stronger through mutual support by virtue of intersecting in appropriate ways.

Keywords: Differential effects; elaborate theories; evidence factors; intersecting strands of evidence; quasi-experiments.

1. Introduction

Cochran's "Observational Studies" is the less familiar, less readily accessible member of a pair of papers in which Cochran (1965, 1972) outlined the general structure of observational studies as a type of statistical investigation. Studies of this type were not new in 1965, nor was the attempt to think systematically about them (e.g., Campbell and Stanley 1963, Hill 1965), but Cochran was the first person to define the subject abstractly, that is, as a subject applicable to and informed by many academic disciplines. It is fitting that Dylan Small's new interdisciplinary journal *Observational Studies* makes Cochran's (1972) paper easily accessible once again.

Cochran's papers have many interesting aspects, but I will focus on just one aspect that appears in different forms near the end of both papers. The final sections of Cochran (1965, 1972) are entitled "The step from association to causation," and "Judgment about causality." These sections make several distinct and useful points, but I would like to focus on one of these, first in §2 by quoting what Cochran says, then in §3 by adding some interpretation.

2. What Cochran says

In discussing judgments about causality, in going beyond modelling empirical associations to reach causal conclusions, Cochran (1965, 1972) speaks again and again about "many different consequences," "variety of consequences," the "mechanism by which the effect is produced," and "completely different type[s] of research." How can many individually weak strands of evidence combine to become strong evidence by considering many different and varied consequences of a causal mechanism?

In one of the more often quoted remarks about causal inference, Cochran (1965, p. 252) wrote:

> First, as regards planning. About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: "Make your theories elaborate." The reply puzzled me at first, since by Occam's razor the advice usually given is to make theories as simple as is consistent with the known data. What Sir Ronald meant, as the subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many *different* consequences of its truth as possible, and plan observational studies to discover whether each of these consequences is found to hold.

After presenting a few illustrations of "many different consequences," Cochran (1965, p. 252) continues:

> Of course, the number and variety of consequences depends on the nature of the causal hypothesis, but imaginative thinking will sometimes reveal consequences that were not at first realized, and this multi-phasic attack is one of the most potent weapons in observational studies. In particular, the task of deciding between alternative hypotheses is made easier, since they may agree in predicting some consequences but will differ in others.

The second paper repeats similar points in different words and adds (Cochran 1972, p. 89):

> A claim of proof of cause and effect must carry with it an explanation of the mechanism by which the effect is produced. Except in cases where the mechanism is obvious and undisputed, this may require a completely different type of research from the observational study that is being summarized.

3. Analogies and methods

3.1 A limited analogy: The cable of many slender fibers

That individually weak strands of evidence may combine to form strong evidence was most famously suggested by Charles Sanders Peirce (1868):

> [We should] trust rather to the multitude and variety of ... arguments than to the conclusiveness of any one. [Our] reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be ever so slender, provided they are sufficiently numerous and intimately connected.

Although memorable, the cable analogy distorts in a key respect: the many fibers of a cable play identical roles in forming a strong cable, but strands of evidence must exhibit variety, must speak to different consequences of a theory, because, as Cochran says, "deciding between alternative hypotheses is made easier, since they may agree in predicting some consequences but will differ in others."

There is a better analogy.

3.2 A better analogy: a crossword puzzle

"Generalization naturally starts from the simplest, the most transparent particular case," Georg Polya (1968, p. 60) wrote in discussing heuristic reasoning in mathematics. This simplest, most transparent case is not the important or general case that led us to be concerned with the topic in question; rather, it is the least cluttered, most immediately accessible and surveyable case, the example that perfectly exemplifies one issue in isolation from unneeded complications. Susan Haack (1995) suggests that the simplest, most transparent case of weak strands of evidence becoming stronger by virtue of mutual support is the case of a crossword puzzle. She writes (1995, pp. 81-82):

> The model is not ... how one determines the soundness or otherwise of a mathematical proof; it is, rather, how one determines the reasonableness or otherwise of entries in a crossword puzzle. ... [T]he crossword model permits pervasive mutual support, rather than, like the model of a mathematical proof, encouraging an essentially one-directional conception. ... How reasonable one's confidence is that a certain entry in a crossword is correct depends on: how much support is given to this entry by the clue and any intersecting entries that have already been filled in; how reasonable, independently of the entry in question, one's confidence is that those other already filled-in entries are correct; and how many of the intersecting entries have been filled in.

Haack is making two points here, the obvious one being that much of the conviction we develop that a crossword puzzle is filled-in correctly comes not from the individual clues, but from entries intersecting in appropriate ways. When we first pencil in an entry based on a clue, we may doubt that it is correct, but later, when other entries meet it in an appropriate way, we may be nearly certain it is correct, even though the direct evidence from the clue remains unconvincing on its own. It is important to recognize that, beyond this obvious point, there is a second point. Haack's second, subtle, point relates to her phrase above: "independently of the entry in question." She is concerned to exhibit mutual support without vicious circularity. If I can deduce B from assuming A, and if I can deduce A from assuming B, then the assertion of A-and-B based on these two deductions would be a logical error — vicious circularity — because both deductions are perfectly compatible with both A and B being false. In the crossword, two entries may meet appropriately yet both be incorrect entries. Haack is saying that B provides support for A only to the extent that we are confident about B not employing the support provided by its intersection with A, and A provides support to B only to the extent that we are confident about A not employing the support provided by its intersection with B; but, with this caveat, A and B may each support the other. The appropriate intersection of A and B provides support for both A and B, but we may reflect upon the evidence for B that does not derive from its appropriate intersection with A, and Haack refers to this as the "independent security" of B. Haack (1995, p. 84–86) continues:

> The idea of independent security is easiest to grasp in the context of the crossword analogy ... How reasonable one's confidence is that 4 across is correct depends, *inter alia*, on how reasonable one's confidence is that 2 down is correct. True, how reasonable one's confidence is that 2 down is correct in turn

> depends, *inter alia*, on how reasonable one's confidence is that 4 across is correct. But in judging how reasonable one's confidence is that 4 across is correct one need not, for fear of getting into a vicious circle, ignore the support given it by 2 down; it is enough that one judge how reasonable one's confidence is that 2 down is correct *leaving aside the support given it by 4 across.*

In a crossword puzzle, entries need not intersect to provide mutual support. If 2 down meets both 4 across and 6 across, then an entry in 6 across may support the entry in 2 down, and the entry in 2 down may support the entry in 4 across, so the entry in 6 across supports the entry in 4 across even though 6 across and 4 across do not intersect.

Consider the same ideas in a biological context. A high level of exposure to a toxin, such as cigarette smoke, is associated with a particular disease, say a particular cancer, in a human population, where experimentation with toxins is not ethical. A controlled randomized experiment shows that deliberate exposure to the toxin causes this same cancer in laboratory animals. A DNA-adduct is a chemical derived from the toxin that is covalently bound to DNA, perhaps disrupting or distorting DNA transcription. Exposure to the toxin is observed to be associated with DNA-adducts in lymphocytes in humans; e.g., Phillips (2002). A further controlled experiment shows that the toxin causes these DNA-adducts in cell cultures. A case-control study finds cases of this cancer have high levels of these DNA-adducts, whereas noncases (so-called "controls") have low levels. A pathology study finds these DNA-adducts in tumors of the particular cancer under study, but not in most tumors from other types of cancer. Certain genes are involved in repairing DNA, for instance, in removing DNA adducts; see Goode et al. (2002). In human populations, a rare genetic variant has a reduced ability to repair DNA, in particular a reduced ability to remove adducts, and people with this variant exhibit a higher frequency of this cancer, even without high levels of exposure to the toxin. Each of these entries in the larger puzzle is quite tentative as an indicator that the toxin causes cancer in humans, and some of the entries do not directly intersect; e.g., the rare genetic variant is not directly linked to the toxin. Yet, the filled in puzzle with its many intersections may be quite convincing.

Consider the same ideas in an economic context. Economic understanding depends, in part, on mathematical theories that derive predictions of economic actions from behavioral assumptions, and, in part, on empirical studies of how people or institutions do act in particular economic contexts. Taken in isolation, the assumptions in one mathematical theory may be quite speculative. Taken in isolation, the findings in one empirical study may be quite insecure, ambiguous and tentative. However, one mathematical theory may intersect with many empirical studies, and may also intersect with many other mathematical theories. Important economic facts – say, a high level of unemployment among recent high school graduates in a particular region at a particular time – may be compatible with several mutually incompatible economic theories – say, a theory that emphasizes rigidities in the labor market, or another that emphasizes the absence of a mechanism to provide adequate investments in human capital. But each theory intersects many particular facts, many empirical studies, and many other theories. Clarification comes, if it does, when an initially speculative theory has correctly met so many ambiguous facts or tentative empirical findings that the theory is no longer speculative, the facts no longer ambiguous, the findings no longer tentative.

3.3 What would it mean to take Cochran's advice seriously?

If you took Cochran's advice seriously, then you would ask of each new study what it contributes to the currently incomplete, partly penciled-in puzzle. You would welcome the completion of a new entry, even a small entry, compatible with the current tentative completion. You would also welcome a compelling new entry that challenged some current entries. You would welcome the suggestion that a particular entry is mistaken and constitutes a barrier to correct completion of the puzzle. A pencil and an eraser would be two tools of equal importance. You would agree with Sunstein (2005) in finding positive value in dissonance and dissent, and you would agree with Rescher (1995) in finding positive value in consensus only to the extent that this consensus has its origins in a rational appraisal of the evidence, whereupon the mere existence of consensus would have little importance beyond its important origins. You would tolerate inconsistency and uncertainty as necessary stepping stones on a path to greater consistency and greater certainty. You would welcome systematic attempts to take stock, to view the tentative completion as a whole, the appraisal of the gaps, the parts that appear secure, the other parts that are uncertain, needing work, in conflict, perhaps mistaken. You would welcome careful, patient, methodical scientific work. You would agree with Kafka (1917): "All human errors are impatience, the premature breaking off of what is methodical."

To take Cochran's advice seriously is to be skeptical of investigations that derive stout conclusions from slender evidence. It is to be skeptical of grand studies and grand conclusions, the suggestion that a single proposed entry settles a major issue, that consistent completion of the puzzle is inevitable given this one entry, and hence consistent completion is not needed and not worth the effort.

3.4 Methods

Several statistical methods cultivate varied strands of evidence within a single study, each strand being weak on its own, each strand vulnerable in a different way, but with the several strands gaining in strength if they agree in appropriate ways. Traditional methods are quasi-experimental designs; see Campbell and Stanley (1963), Shadish et al. (2002), West et al. (2008) and Wong et al. (2015). More recent methods include evidence factors (Rosenbaum 2010, 2015; Zhang et al. 2011), differential effects (Rosenbaum 2006, 2013, 2015; Zubizarreta et al. 2014) and attempts to integrate qualitative and quantitative causal inference (Rosenbaum and Silber 2001; Weller and Barnes 2014). Vanderweele (2015) expands on one of Cochran's themes, the role of mechanisms as evidence. Yang et al. (2014) encourage the tolerance of statistical inferences that terminate in dissonance, that is, inferences that demonstrate unresolved inconsistencies among intersecting strands of evidence.

Acknowledgments

Supported by a grant from the Measurement, Methodology and Statistics Program of the US National Science Foundation.

References

Campbell, D. T., Stanley, J. C. (1963). *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally.

Cochran, W. G. (1965). The planning of observational studies of human populations (with Discussion). *Journal of the Royal Statistical Society* A, 234-266.

Cochran, W. G. (1972). Observational Studies. *Statistical Papers in Honor of George W. Snedecor*. Ames: Iowa State University Press, 70-90.

Goode, E. L., Ulrich, C. M. and Potter, J. D. (2002). Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiology Biomarkers and Prevention*, 11:1513-1530.

Haack, S. (1995). *Evidence and Inquiry*. Oxford: Blackwell.

Hill, A. B. (1965). The environment and disease: association or causation?. *Proceedings of the Royal Society of Medicine*, 58:295-300.

Kafka, F. (1917). *The Blue Octavo Notebooks*. Cambridge: Exact Change.

Peirce, C. S. (1868). Some consequences of four incapacities. *Journal of Speculative Philosophy*, 2, 140-157. Reprinted in: Talisse, R. B. and Aikin, S. F., eds. (2011) *The Pragmatism Reader: From Peirce through the Present*, Cambridge MA: Harvard University Press.

Phillips, D. H. (2002). Smoking-related DNA and protein adducts in human tissues. *Carcinogenesis*, 23:1979-2004.

Polya, G. (1968). *Mathematics and Plausible Reasoning*, Volume II, 2nd edition. Princeton, NJ: Princeton University Press.

Rescher, N. (1995). *Pluralism: Against the Demand for Consensus*. New York: Oxford University Press.

Rosenbaum, P. R. and Silber, J. H. (2001). Matching and thick description in an observational study of mortality after surgery. *Biostatistics*, 2, 217-232.

Rosenbaum, P. R. (2006). Differential effects and generic biases in observational studies. *Biometrika*, 93:573-586.

Rosenbaum, P. R. (2010). Evidence factors in observational studies. *Biometrika*, 97:333-345.

Rosenbaum, P. R. (2013). Using differential comparisons in observational studies. *Chance*, 26:18-25.

Rosenbaum, P. R. (2015). How to see more in observational studies: Some new quasi-experimental devices. *Annual Review of Statistics and Applications*, 2:21-48.

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Sunstein, C. R. (2005). *Why Societies Need Dissent*. Harvard University Press.

Vanderweele, T. J. (2015). *Explanation in Causal Inference*. New York: Oxford.

Weller, N. and Barnes, J. (2014). *Finding Pathways: Mixed-method Research for Studying Causal Mechanisms*. Cambridge University Press.

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M., Mullen, P. D. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98:1359-66.

Wong, M., Cook, T. D., and Steiner, P. M. (2015). Adding design elements to improve time series designs: No Child Left Behind as an example of causal pattern matching. *Journal of Research on Educational Effectiveness*, 8:245-279.

Yang, F., Zubizarreta, J. R., Small, D. S., Lorch, S., and Rosenbaum, P. R. (2014). Dissonant conclusions when testing the validity of an instrumental variable. *American Statistician*, 68:253-263.

Zhang, K., Small, D. S., Lorch, S., Srinivas, S., and Rosenbaum, P. R. (2011). Using split samples and evidence factors in an observational study of neonatal outcomes. *Journal of the American Statistical Association*, 106:511-524.

Zubizarreta, J. R., Small, D. S., Rosenbaum, P. R. (2014). Isolation in the construction of natural experiments. *Annals of Applied Statistics*, 8:2096-2121.