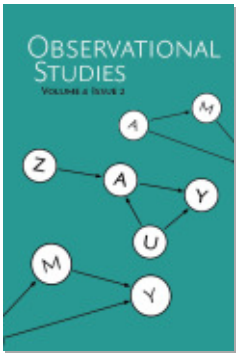
Another Ground Rule

Charles S. Reichardt



Observational Studies, Volume 4, Issue 2, 2018, pp. 57-60 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2018.0006*

Observational Studies 4 (2018) 57-60 Submitted 5/17; Published 1/18

Another Ground Rule

Charles S. Reichardt **Chip.Reichardt@du.edu**
Department of Psychology
University of Denver
Denver, CO 80208 USA

Bross (1960, p. 394) wrote, "If both proponents and critics have to watch their P's and Q's, we might hope that it would be easier to achieve broad agreement on scientific issues." Bross then went on to offer a ground rule (i.e., one of the P's and Q's) explicitly for critics of research hypotheses, though he emphasized (pp. 399-400) the "same ground rules should apply to both" proponents and critics. I have two purposes. First, to quibble with Bross about his ground rule. And second, to propose another ground rule.

Bross argues that alternative explanations should be judged tenable before they are allowed to see the light of day. And for an alternative explanation to be judged tenable, it must agree with available data. Certainly this is correct. Except that agreement with available data is not an infallible indicator of the adequacy of an explanation.

Consider Darwinian evolution. Significant data argued against Darwins theory at the time it was proposed (Bryson, 2003). For example, the best available evidence was that the earth was far too young, even according to Darwin's own account, for species to have evolved per natural selection. And the fossil record was too sparse providing too little evidence of the intermediate life forms that Darwin required. Plus Darwin's theory was at odds with well accepted contemporary thinking. Even T. X. Huxley, who was one of Darwin's staunchest supporters, believed Darwin was wrong about the rapidity with which evolution took place. Huxley, a saltationist, believed evolution happened suddenly rather than gradually. So was Darwin's alternative explanation for evolution in sufficient agreement with available data, and therefore sufficiently tenable, to permit publication according to Bross?

Or consider the work of Ingaz Semmelweis. When Ingaz Semmelweis was hired as a physician at the Vienna General Hospital in 1846, as many as twenty percent of the women giving childbirth in the hospital's First Obstetrical Clinic died from puerperal fever, also known as childbed fever. Semmelweis was determined to discover the cause. He uncovered a telling clue when a colleague cut his finger while performing an autopsy and died from symptoms similar to puerperal fever. Based on that evidence, Semmelweis hypothesized that the disease was caused by contamination from "cadaverous material" and found his hypothesis could explain another mystery. The First Obstetric Clinic at Vienna General had a much higher death rate from puerperal fever than the Second Clinic. The difference? The First Clinic was attended by physicians who often performed autopsies before serving on the obstetrics ward; the Second Clinic was attended by mid-wives, who did not perform autopsies. Acting on his hypothesis, Semmelweis instituted a policy that physicians wash their hands in a solution of chlorinated lime before examining patients. Occurrences of

puerperal fever declined dramatically. In April 1847, before the policy of hand washing was instituted, eighteen percent of the patients in the First Clinic died from puerperal fever. A month later, after hand washing was implemented, the mortality rate dropped to two percent. The same outcomes were obtained whenever hand washing was implemented by either Semmelweis or his students.

The results of Semmelweis' hand-washing experiments became widely known. But the findings ran counter to the conventional medical beliefs current at the time. Puerperal fever was thought to be due to multiple causes including effluviums that were thought to be spread through the air. In addition, the causes of illnesses were thought to be as unique as individuals themselves and determinable only on a case by case basis. As a result, Semmelweis' unconventional proposal that a simple cause such as a lack of cleanliness could be responsible for puerperal fever was rejected out of hand by the medical profession. Indeed, physicians believed their social status precluded them from having unclean hands. In spite of Semmelweis' demonstrations of the effectiveness of hand washing in reducing puerperal fever, most doctors did not adopt the practice until the germ theory of disease was advanced by Pasteur and Lister decades later.

So Semmelweis' alternative explanation for the cause of puerperal fever was not tenable in the eyes of the proponents of the current theories. Of course, Bross would have likely responded that Semmelweis' alternative explanation was highly tenable, given Semmelweis' data, and was rejected only because of what he calls tubular vision. Nonetheless, the acceptance of Semmelweis' theory was delayed because, for whatever reasons, it was judged unacceptable at the time and thousands subsequently died unnecessarily. Would Bross' ground rule have aided and abetted the rejection of Semmelweis' conclusions?

Bross also notes that he allows speculation in a research report, just not as part of a conclusion section. Because Einstein's general theory of relativity had little data in its support at the time it was published, it was a highly speculative conclusion. Indeed, evidence that light bends in the presence of gravity, as predicted by the theory, was not provided for four years after Einstein's theory was published. And one of the theory's main tenets, the existence of gravity waves, was not confirmed until 100 years later. So would Bross have stood in the way of publishing that largely unsupported speculation?

The point is the following. Bross is correct that those who propose alternative explanations, whether proponent or critic, should be required to check their explanations against available data. But data is always incomplete and can be all too easily misinterpreted. So judging what is and is not in agreement with data can be tricky. Assessing tenability is an inexact science. As a result, researchers and editors must be circumspect in applying Bross' rule.

I might note one other caution with regard to Bross' rule. To be tenable, any one alternative explanation need not, by itself, account for all of the observed results. I've seen instances where an estimate of a treatment effect, for example, is said to be immune to a rival hypothesis because the alternative explanation was insufficient to account for the entirety of the estimate. But more than one bias can be present. And perhaps together they could account for the whole treatment effect estimate. So the tenability of alternative explanations needs to be considered en masse rather than one at a time.

Now let me consider my second purpose which is to provide another ground rule – especially for proponents, rather than critics, of explanations. Like Bross (1960), Doll and

Hill (1952) were concerned with the theory that smoking causes lung cancer. And Doll and Hill presented data that well supported that theory. The critical point, for my presentation, is that Doll and Hill included a section of their article entitled "Validity of the Results." Cochran (1965, pp. 252-253) explains the importance of such a section:

> When summarizing the results of a study that shows an association consistent with the causal hypothesis, the investigator should always list and discuss all alternative explanations of his [sic] results (including different hypotheses and biases in the results) that occur to him. This advice may sound trite, but in practice is often neglected. A model is the section "Validity of the results" by Doll and Hill (1952), in which they present and discuss six alternative explanations of their results in a study.

Certainly it is well accepted by now that proponents of scientific hypotheses should openly report weaknesses of their research and alternative explanations for their findings. For example, the American Psychological Association's (2010, p. 35) widely used Publication Manual states:

> Your interpretation of the results should take into account (a) sources of potential bias and other threats to internal validity, (b) the imprecision of measures, (c) the overall number of tests or overlap among tests, (d) the effect sizes observed, and (e) other limitations or weaknesses of the study.

But according to the APA Manual, such an accounting of limitations and weaknesses is to be placed in the discussion section of an article. In contrast, I believe, as Cochran recommends, that such critical reflections on a research study deserve their own dedicated section, as in Doll and Hill.

And not only that, reviewers and editors should insist that such a section goes to the heart of limitations and weaknesses in a research study rather than just skimming the surface. I've read reports where the limitations of a study include such obvious reflections as that the results should not be generalized beyond the population of participants and the outcome measures that were used. But the same reports ignore warnings of much more insidious concerns such as omitted variables and hidden biases. Perhaps the authors are unaware of such difficulties. If so, that is all the more reason for reviewers and editors to insist they be acknowledged forthrightly. Or perhaps the authors are aware but afraid that acknowledging such weaknesses would disqualify their research from publication. If so, reviewers and editors must explicitly disavow such disqualification when the research is otherwise of high quality – because such weaknesses are simply an inherent feature in some realms of research, such as observational studies. Even sensitivity analyses are not guaranteed to bracket the true sizes of treatment effects.

Are we doing what we should to encourage substantive researchers to abide by the practice endorsed by Cochran – one of the founders of statistical methods in observational studies? Honesty and integrity in science require that researchers be their own worst critics. We should insist that a discussion of limitations and weaknesses be the whole truth and that it appears in its own section and not as a brief mention of superficial weaknesses hidden amid substantive summarizing in a discussion section.

Acknowledgments

Thanks to Kris Preacher, Bill Revelle, and Anne Strobridge for their helpful comments on the manuscript.

References

American Psychological Association (2010). *Publication Manual* (6th edition), Washington, DC: American Psychological Association.

Bross, I. D. J. (1960). Statistical criticism. *Cancer*, 13, 394-400.

Bryson, B. (2003). *A Short History of Nearly Everything.* NY: Broadway Books.

Cochran W. G. (1965). The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society, Series A*, 128, 234-266.