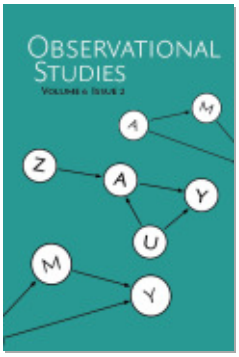
The Wrong Message from the Wrong Talk

Kenneth J. Rothman



Observational Studies, Volume 6, Issue 2, 2020, pp. 30-32 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2020.0006*

Observational Studies 6 (2020) 30-32 Submitted 7/19; Published 1/20

The Wrong Message from the Wrong Talk

Kenneth J. Rothman **krothman@rti.org**
Boston University and
Research Triangle Institute
3040 East Cornwallis Road
Research Triangle Park, NC USA 27709

Hill's 1965 address (Hill, 1965) on causal inference to the newly formed section on Occupational Medicine of the Royal Society is considered hallowed scripture, not merely revered by researchers, but even used as a template for adjudication of causal questions by courts of law. It certainly was a provocative and thoughtful after-dinner talk, but it does not merit the status of scripture, especially not for the reasons it is so revered. The message for which it is remembered was not the most important advice in the talk, and the talk itself was less important than another lecture, now largely forgotten, that Hill gave some years earlier.

The 1965 talk, which has been cited over 9000 times, gained prominence because it seems to offer a checklist of criteria for causal inference. Scientists, like anyone else, are inclined to follow a recipe if one exists. Apparently, few are bothered by how well the recipe works, or even whether it works. Consequently, numerous papers have attempted to infer causation through the application of the "Hill criteria," on topics ranging from the zika virus and microcephaly (Awadh et al., 2017) to neuropsychiatry (van Reekum et al., 2001) to climate change (Science in Society Archive). In one recent review of dietary sodium and cardiovascular risk, the authors divided the paper into sections that corresponded to each of Hill's nine criteria (Cogswell et al., 2016). In other papers, the authors employed scoring systems to quantify how well the criteria are met and from that derived an overall number which was intended to be a measure of causality (Mente et al., 2009). In one such approach, a discriminant analysis was used to generate weights for each of the nine criteria, which were used to derive an overall score that was interpreted as a probability indicating whether the association is causal (Swaen and Amelsvoort, 2009).

Philosophers of science consider these to be futile efforts, because they agree that there cannot be a logical basis for a checklist approach to scientific inference. It comes as no surprise, then, that critics have steadily questioned the legitimacy of inferential checklists, as well as the origin of Hill's list, and the applicability of specific points that Hill included in the list (Weed, 1988; Haack, 2014; Ward, 2009; Rothman and Greenland, 2005; Morabia, 1991; Philips and Goodman, 2004; Blackburn and Labarthe, 2012; Thygesen et al., 2005; Weiss, 2002). But we can hardly criticize Hill for originating the checklist, because he asked to be counted among the critics of a checklist approach to inference. He studiously avoided the word "criteria" in his talk, and he cautioned against using his "viewpoints" as if they were criteria, claiming that none of them was a *sine qua non* for causal inference.

In 2004, Philips and Goodman elegantly laid out the thesis that Hill's 1965 paper is esteemed for the wrong reason (Philips and Goodman, 2004): *"We will say only that Hill's list seems to have been a useful contribution to a young science that surely needed systematic thinking, but it long since should have been relegated to part of the historical foundation, as an early rough cut. Yet it is still being recited by many as something like natural law. Appealing in our teaching and epistemology to the untested "criteria" of a great luminary from the past is reminiscent of the "scientific" methods of the Dark Ages."* In their essay, they suggested two alternative messages that they deemed to be the "missed lessons" of Hill's talk. One of these was Hill's warning about reliance on statistical significance testing for inference: *"(W)e waste a deal of time, we grasp the shadow and lose the substance, we weaken our capacity to interpret data and to take reasonable decisions whatever the value of P. And far too often we deduce 'no difference' from 'no significant difference'"* (Hill, 1965). Hill continued, *"I wonder whether the pendulum has not swung too far - not only with the attentive pupils but even with the statisticians themselves."* This issue, which has been mostly submerged for decades, has finally surfaced to spawn serious debate and reflection (Wasserstein and Lazar, 2016; Amrhein et al., 2019). Reliance on significance testing for inference has led to countless misinterpretations of data, and has exacerbated the problem of replication of results. If that message had been the legacy of Hill's talk instead of the checklist, untold errors might have been avoided.

A dozen years before the 1965 lecture for the Royal Society, Hill gave the Cutter Lecture at Harvard. Some background is relevant: in 1951, Prof. Hugh Sinclair of Oxford University was the Cutter Lecturer (Sinclair, 1951). Sinclair's theme was to exalt the experimental method, while attacking nonexperimental science, at least regarding nutritional research: *"The use of the experimental method has brilliant discoveries to its credit, whereas the method of observation has achieved little...The observer must await the occurrence of the natural succession of events he wishes to study, and he is very apt to be misled by the fallacy of post hoc ergo propter hoc or by the existence of a correlation without causality"* (Sinclair, 1951). When Hill delivered his Cutter Lecture two years later, he pointedly titled it "Observation and Experiment" (Hill, 1953). He made the case for nonexperimental epidemiology, starting with the example of Snow's work on cholera, moving on to studies of rubella, and then to his own work on smoking and lung cancer. He made it clear that he was not criticizing experimentation; indeed he preferred when possible to get experimental evidence, a point reiterated in his 1965 Royal Society talk. But he emphasized that *"...my preference in preventive medicine for the experimental approach...does not lead me to repudiate or even, I hope, to underrate the claims of accurate and designed observations"* (Hill, 1953). Hill's lecture was clearly a response to Sinclair's attack on nonexperimental science. He repeated Sinclair's phrases: *"The observer may well have to be more patient than the experimenter – awaiting the occurrence of the natural succession of events he desires to study; he may well have to be more imaginative – sensing the correlations that lie below the surface of his observations; and he may well have to be more logical and less dogmatic – avoiding as the evil eye the fallacy of post hoc ergo propter hoc, the mistaking of correlation for causation"* (Hill, 1953).

Hill's inspirational rebuttal to Sinclair, along with his work on smoking and lung cancer, helped to fortify the philosophic foundation for nonexperimental epidemiologic research. He asked whether we can draw inferences at all without experimentation, and he answered affir-

matively. In comparison, the issue that became the focus of his 1965 talk was misconstrued and far less compelling.

References

Amrhein V, Greenland S, McShane B (2019). Scientists rise up against statistical significance. *Nature*, 567: 305-307. doi: 10.1038/d41586-019-00857-9.

Awadh A, Chughtai AA, Dyda A, Sheikh M, Heslop DJ, MacIntyre CR. (2017). Does zika virus cause microcephaly – applying the Bradford Hill viewpoints. *PLOS Currents Outbreaks*, Feb 22.

Blackburn H, Labarthe D (2012). Stories from the evolution of guidelines for causal inference in epidemiologic associations: 1953–1965. *Am J Epidemiol*, 176: 1071–1077.

Cogswell ME, Mugavero K, Bowman BA, Frieden TR (2016). Dietary sodium and cardiovascular disease risk – measurement matters. *N Engl J Med*, 375: 580-586.'

Haack S (2014). *Evidence matters; science, proof, and truth in the law.* Cambridge University Press, pp 258-263.

Hill, AB (1953). Observation and experiment. *N Engl J Med*, 248: 995-1001.

Hill, AB (1965). The environment and disease: association or causation? *Proc R Soc Med.*, 58:295–300.

Mente A, de Koning L, Shannon HS, Anand SS. (2009). A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease. *Arch Intern Med.*, 169(7):659–669.

Morabia A (1991). On the origin of Hill's criteria. *Epidemiology*,2: 367-369.

Philips CV, Goodman KJ (2004). The missed lessons of Sir Austin Bradford Hill. *Epidemiologic Perspectives & Innovations*, 1:3. doi:10.1186/1742-5573-1-3

Rothman KJ and Greenland S (2005). Causation and causal inference in epidemiology. *Am J Public Health*, 95: S144–S150. doi:10.2105/AJPH.2004.059204

Science in Society Archive: http://www.i-sis.org.uk/TheBradfordHillCriteria.php

Sinclair HM (1951). Nutritional surveys of population groups. *N Engl J Med*, 145, 39-47.

Swaen G, Amelsvoort L (2009). A weight of evidence approach to causal inference. *J Clin Epidemiol*, 62: 270-277

Thygesen LC, Andersen GS, Andersen H. (2005). A philosophical analysis of the Hill criteria. *J Epidemiol Community Health*, 59: 512–516. doi: 10.1136/jech.2004.027524

van Reekum R, Streiner DL, Conn DK (2001). Applying Bradford Hill's criteria for causation to neuropsychiatry: challenges and opportunities. *J Neuropsychiatry Clin Neurosci.*, 13:318-325.

Ward AC (2009). The role of causal criteria in causal inferences: Bradford Hill's "aspects of association." Epidemiologic Perspectives & Innovations , 6:2 doi:10.1186/1742-5573-6-2

Wasserstein RL, Lazar NA (2016). The ASA's Statementon p-Values: Context, Process, and Purpose. *The American Statistician*, 70: 129-133, doi:10.1080/00031305.2016.1154108

Weed DL (1988). Causal criteria and Popperian refutation. In *Causal Inference*, Rothman KJ, ed., Epidemiology Resources, Inc.

Weiss NS. (2002). Can the "specificity" of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology*, 13: 6–8.