



PROJECT MUSE®

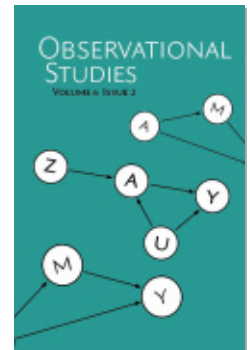
On the use and abuse of Hill's viewpoints on causality

Samantha Kleinberg

Observational Studies, Volume 6, Issue 2, 2020, pp. 17-19 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2020.0003>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/793344/summary>

On the use and abuse of Hill's viewpoints on causality

Samantha Kleinberg
Computer Science Department
Stevens Institute of Technology
Hoboken, NJ, USA 07030

samantha.kleinberg@stevens.edu

Here, then, are nine different viewpoints from all of which we should study association before we cry causation. **What I do not believe — and this has been suggested — is that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect.** None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question — is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect? (emphasis added, italics original) Hill (1965)

Not since Fisher¹ suggested $p < 0.05$ is often convenient has such a clear statement by a statistician been so misunderstood. Hill's sensible advice has been transformed like Samsa in Kafka's *Metamorphosis* into what his article warned against: a checklist. Google scholar returns over 100,000 articles using the phrase "Bradford Hill Criteria," it has been growing in usage in books since the 1990s (see figure 1),² and even the Wikipedia page on the topic is titled "Bradford Hill Criteria."³ And yet Hill wrote that there are no "hard-and-fast rules" for causality.

This is not just a marketing problem. How we talk influences how we think (Boroditsky, 2011) and the mutation of considerations into criteria is in fact part of their misuse. Hill referred to the pieces of evidence we may wish to examine as "aspects of [an] association [to] consider before deciding that the most likely interpretation of it is causation" (p. 295) and "viewpoints from [which] to study association before we cry causation" (p. 299). Considerations may influence our decisions, such as whether to believe a causal relationship exists, but they are also things we may evaluate and ignore if they're not relevant. Criteria, in contrast, are a benchmark against which we test something. In the case of causality, criteria provide a tantalizing yet misleading shortcut: check off these boxes and you can claim causality. Yet, there is no such checklist for causality and Hill's considerations are neither necessary nor sufficient to establish a causal relationship.⁴

1. Fisher (1925) said about a p -value threshold of 0.05 that "it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not." The message people heard appears to be " $p < 0.05$ or it didn't happen."
2. Variations of the phrase involving viewpoints and considerations are so rare that no Ngrams are found.
3. https://en.wikipedia.org/wiki/Bradford_Hill_criteria
4. See Kleinberg (2015) and (Rothman et al., 2008, p. 26) for a few examples detailing just why this is.

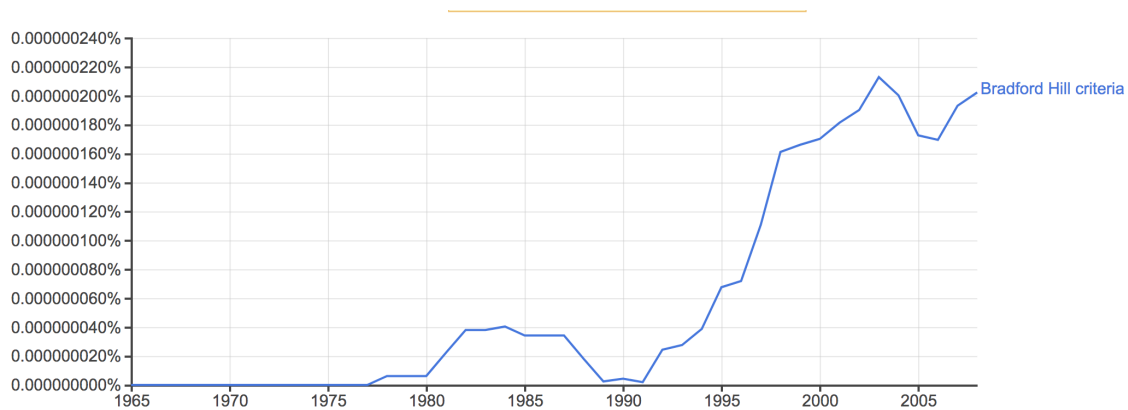


Figure 1: Google Ngram results for usage of “Bradford Hill Criteria” in books.

Checklists can be a powerful tool in safety critical domains where cognitive load is high and time is short (Gawande, 2010). However the settings where Hill’s views are most useful are not that. They are cases where experiments are difficult or impossible and we must cobble together piecemeal evidence for causal claims. These are cases where we also must assess whether a consideration is relevant to the topic. For example, Hill along with Richard Doll identified a link between smoking and lung cancer at a time when little was known about the etiology of lung cancer (Doll and Hill, 1950). It is not possible to conduct randomized experiments to test the hypothesis that smoking is responsible for cancer, but it is of great public health significance to know what causes cancer so it can be prevented. From this experience Hill distilled his views on how we can gain such causal knowledge into his famous article.

Yet rather than providing a starting point, Hill’s viewpoints have been widely and repeatedly used as a standard of evidence, the same way the majority of researchers use a p -value cutoff of 0.05. The precise danger of conventions is that one need not justify them, whereas a p -value threshold of 0.04 or 0.06 would invite significant scrutiny.⁵ However, many other factors such as effect size are important to determining whether a result is actually important or not. When Hill’s views are treated as criteria, they similarly become a causal inference figleaf. If these reasonable but still unvalidated pieces of evidence can be provided,⁶ then congratulations, you can claim causality.

So if I’m suggesting researchers quit the causal criteria cold turkey, what will replace them? It is perfectly fine to refer to Hill’s considerations as a starting point when thinking about what evidence one might gather when evaluating an association. The part that is not fine is making the leap from these pieces of evidence to a definitive claim of causality – and both failing to consider other types of evidence and forcing these considerations to fit scenarios where they do not apply. These are two critical areas for future research to

5. Editorial guidelines for the journal *Cognition* explicitly state that only effects with $p < 0.05$ can be described as statistically significant, stating that “for better or for worse, this is the current convention,” which it seems even journals are powerless to change.

6. This is all leaving aside the question of what it means to satisfy each criteria, which surely requires more nuance than present/absent.

explore. First, our methods and data have evolved in the years since Hill’s article, yet the considerations remain static. It is worth exploring whether there are other evidence types that may prove useful as well as updating how current methods might support the existing considerations (e.g. how do big data and simulations fit in?). Second, while Hill focused on epidemiology, the considerations have been used more broadly and it is important to examine how the needs for and standards of evidence vary across domains. By allowing them to evolve, Hill’s considerations will hopefully meet a better end than poor Samsa.

Acknowledgments

Thanks to Dylan Small for providing an outlet for these viewpoints.

References

- Boroditsky, L. (2011). How language shapes thought. *Scientific American*, 304(2):62–65.
- Doll, R. and Hill, A. B. (1950). Smoking and carcinoma of the lung. *British medical journal*, 2(4682):739.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Gawande, A. (2010). *The Checklist Manifesto*. Henry Holt and Company.
- Hill, A. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58(2):295–300.
- Kleinberg, S. (2015). *Why: A Guide to Finding and Using Causes*. O’Reilly Media.
- Rothman, K. J., Greenland, S., Lash, T. L., et al. (2008). *Modern epidemiology*, volume 3. Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia.