Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: October 31, 2022

Citation: Weed DL, Alexander DD, et al., 2022. The Erosion of Causal Inference in Systematic Reviews in Epidemiology, Medical Research Archives, [online] 10(10). https://doi.org/10.18103/mra.v10i10.3252

DOI
https://doi.org/10.18103/mra.v10i10.3252

ISSN: 2375-1924

RESEARCH ARTICLE

The Erosion of Causal Inference in Systematic Reviews in Epidemiology

Douglas L. Weed, M.D., Ph.D.[1]; Dominik D. Alexander, Ph.D.[2]

[1] DLW Consulting Services LLC, 1302 N. Oak Forest Rd. Salt Lake City, UT 84103, USA
[2] MetaMethod, 16769 Bernardo Center Ct k1-144, San Diego, CA 92127, USA

* douglaslweed@aol.com

ABSTRACT
The assessment of disease causation is a complex process with a decades-long history of development and discussion. The family of methods involved had been in place for at least 30 years when meta-analysis and the systematic narrative review emerged to be added to study designs and statistical methods. More recently, methods to evaluate bias and quality have been added. Traditionally, near the end of a causal assessment, that is, after all the studies have been collected and described and sometimes meta-analyzed, investigators apply a set of conditions (or criteria or considerations) to evaluate whether an association observed in epidemiological studies supports a causal association. The criteria proposed by A.B. Hill—Hill's criteria—are arguably the best-known example. In this paper we describe and critically examine a trend in the epidemiological literature wherein some practitioners have been chipping away at this final step. In some instances, the use of these criteria-based methods has been totally rejected; in other situations, some of the traditional criteria (or considerations) have been eliminated while others remain. It is important to point out that these eliminations and exclusions are not replaced with some presumably better approach. Rather, there is a sense that these so-called "criteria" are no longer relevant. We see this process as eroding the reliability and validity of causal claims.

Keywords: Causality, Causal Criteria, Causal Inference, Epidemiology, Methods, Systematic Reviews

Introduction

The assessment of causation—i.e., what factors cause disease—is arguably the most important scientific issue facing epidemiologists. The methods used for this purpose have remained relatively stable with notable additions, the sum of which represent a family of interconnected methods. This "family" can be organized into several categories: the general scientific method and theoretical models, e.g., causal models, study design and statistical methods, methods used to evaluate bias and to assess the quality of studies, reviews, and meta-analyses, and the research synthesis methods, e.g., the systematic narrative review, meta- and pooled analyses, and criteria-based methods.

The use of these methods—their application—for a specific exposure-disease pair is a complex matter. Decisions about causality are made by investigators as well as by organizations such as the National Cancer Institute (NCI), the National Toxicology Program (NTP), the Agency for Toxicological Substances and Disease Registry (ATSDR), and the American Cancer Society (ACS) to name a few U.S. examples. The World Health Organization's International Agency for Research on Cancer (IARC) and other international organizations also participate in the process of making causal claims for cancer. Cancer is not the only outcome where causality is at issue. Indeed, the ATSDR examines the potential effects of chemicals and other agents across a wide range of disease outcomes. The European Food Safety Agency (EFSA) and "Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) are examples of similar agencies in Europe with an interest in causation.

A complete description of the many decisions involved in making causal claims is well beyond the scope of this paper. There are, however, some examples of these decisions in the peer-reviewed and the so-called "grey" literature that represent important variations from what can and should be considered typical practice, i.e., variations from the traditional approach.

Our concern here is that these divergent decisions make causal claims easier but not necessarily better. In short, we see these decisions as eroding the practice of causal inference. The examples presented here were not systematically identified and some represent a minority view. Others, however, appear to be accepted by many practitioners. The main reason we want to bring these to the attention of readers is that they appear to be reducing the rigor of causal claims and, as a result, potentially reducing the reliability and validity of the ensuing claims.

To keep our effort constrained given the wide variety of methods and methodological choices, we will focus primarily on what traditionally has been the last step in the process, the use of causal criteria. These criteria—sometimes referred to as "viewpoints" or "aspects" of causal associations have a long and storied history and remain widely used. The most famous—and likely the most well-known—set of criteria (or whatever else you want to call them) were proposed by Sir Austin Bradford Hill in what is considered a classic paper published in 1965 (1).

Criteria-based Methods

These criteria should only be applied to a body of epidemiological and biologic evidence after it has been concluded that the observed association in epidemiological evidence is not due to chance (1-3). It will be helpful and important to describe these so-called criteria, their characteristics, and some of their strengths and limitations, as a prelude to examining how different investigators have either misapplied, reduced their number, or eliminated them altogether.

Temporality—the concept that a causal factor must precede its effect—is a true criterion (3-7). In the absence of temporality, i.e., if the presumed effect precedes its hypothetical cause in time, then causation can be ruled out.

Strength of Association: refers to the magnitude of the relative risk estimates observed in the epidemiology studies. Typically, the larger the relative risk (RR), the more likely the observed association is causal (3-7). Small magnitudes of association (sometimes called "weak" or "modest" associations), e.g., relative risks (RRs) of 2.0 or less, are less likely to represent causal associations. Bias (due especially to uncontrolled and residual confounding) can explain the presence of weak associations.

Consistency of Association: refers to the extent to which scientific results are similar (e.g., in direction and magnitude) across the entire body of epidemiological evidence. Typically, the more consistent are the results, the more likely the observed association is causal (3-7). One of the additional values of meta-analysis is that it

provides a quantitative assessment of consistency (8) through statistical tests of heterogeneity. Based on these results, a collection of epidemiologic studies may be considered homogeneous, which, by definition, means they are consistent. Moreover, between study review after heterogeneity testing may indicate methodological reasons why statistical associations are inconsistent.

Biologic Gradient (Exposure-Response): refers to the extent to which the relative risk estimates increase in magnitude as the measure of the exposure increases in the epidemiology studies. Typically, a regularly increasing relationship between exposure and risk estimate is more likely to represent a causal relationship than other patterns (4-7).

Biologic Plausibility: refers to the extent to which a mechanism of action has been proposed, studied, and demonstrated, typically in toxicological and other types of laboratory-based studies. It is generally accepted that as the evidence explaining the mechanism of action for a disease increases, the more likely the association is causal (9). A disease mechanism has many features, including but not limited to the many intracellular and extracellular changes that occur from the initiating causal event (e.g., an exposure or some unknown "idiopathic" event) to the subsequent disease event. Indeed, latency (discussed briefly above) can be considered one of many features of a disease mechanism. Assessing biological plausibility also involves distinguishing between what happens in humans and what happens in animals. Although animal testing (also called animal bioassay testing) has been used for many years as a component of assessing biological plausibility, its relevance to human health is under intense scrutiny in the scientific community. The primary concern has always been the extent to which the results of animal testing can be extrapolated to humans given that the experimental doses given to animals were often many times more potent than what humans experience.

Specificity: refers to two related ideas. First, it refers to the precision with which the exposure and the outcome can be defined and characterized. For example, studies of benzene and chronic lymphocytic leukemia have more specific measures of exposures and outcomes than, say, studies of solvents and leukemia (encompassing several sub-types of leukemia). As Hill noted, the more specific the exposure and disease involved, the stronger the argument in favor of causation. Secondly (and traditionally), specificity also refers to the extent that the disease (outcome) has one or more causes (4,6).

Coherence: refers to the extent to which the evidence and hypotheses for the results fit together into a reasonable and well-tested explanation (3). In the classic description of this so-called criterion, coherence was defined as the extent to which the causal hypothesis does not conflict with the available evidence. Coherence can be assessed in terms of the extent to which other causal criteria (or "guidelines") have been met. The more criteria that are satisfied, the more coherent the causal explanation.

Experimentation: refers to the extent to which a randomized clinical trial (e.g., a prevention trial) or an observational intervention study has been undertaken (7). This is an uncommon condition to be satisfied using randomized trials in the study of chronic diseases. Note that this criterion does not refer to animal experimentation.

Analogy: the extent to which the purported exposure-disease relationship under consideration is similar (in types and characteristics of evidence) to other relationships, known to be causal or not (10). Hill's version of this criterion (or consideration) is somewhat more nuanced. We will address that fact later in this paper.

Use and Uses of the Causal Criteria

These are widely used criteria with scientific justification; the method has been used for nearly sixty years in hundreds, if not thousands, of applications involving many different exposures and many different diseases and conditions. Furthermore, many causes of diseases have been identified using this methodology along with all the other methods described earlier. Research institutions and governmental regulatory agencies regularly use these criteria. Examples include: the International Agency for Research on Cancer and the U.S. Environmental Protection Agency (11,12). For the past 50 years, epidemiology textbooks have recommended and discussed the use of these criteria for causal inference (4-7, 11-22).

Hill's Causal Criteria and the Scientific Method

It is important to point out that these criteria reflect the application of the general scientific method and not a substitute for it. The criterion of consistency, for example, reflects the scientific principles of replicability and testability. The criterion of

strength (of association) reflects the scientific concept of the need to critically test alternative explanations including those not measured. Experimentation, likewise, reflects the need to test and control for alternative hypotheses. Temporality is a key feature of any causal hypothesis. Specificity reflects the need to test the hypothesis of interest and not some different hypothesis. Biological plausibility incorporates biological explanations with those explanations in human populations by examining the extent to which the basic causal hypothesis has been tested in cellular and animal systems. Exposure (or "dose") response reflects a basic toxicological principle: the greater the exposure to a causal agent, the greater the effect. Coherence and analogy, in turn, represent the scientific need for an explanation that "fits" with the various sources of evidence involved as well as with what has been decided in the past regarding causation. In sum, the criteria and the general scientific method are not only compatible but inseparable.

General Critiques of Hill's Causal Criteria

One of the most important critiques of Hill's criteria can be traced to the influential textbook by Rothman and Greenland (19, p. 27) who wrote that "the standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions." This critique has been repeated and expanded upon in more recent accounts to such an extent that, in our view, some authors have rejected the concept of causal criteria (or standards or viewpoints or guides) completely. To this apparent rejection of the Hill criteria another critique has emerged, namely, that the Hill criteria should not be considered a checklist for causal claims. A good example of how these two general critiques of the Hill criteria have affected subsequent discussions of the methodology of causal claims, is in a paper by Kundi (23, p.971). He writes that "it is a complete misinterpretation of the nine issues considered by Bradford Hill that they can be a type of checklist to establish causation. But it may turn out that they owe their popularity, still persisting after 40 years, exactly to this misconception." Kundi then proposes his own list of considerations—it is indeed a list despite his complaints about Hill—which includes strength of association, consistency of association, dose-response, temporality, coherence, experimental evidence, and biological plausibility. Put another way, Kundi's critiques are, in a sense, meaningless. Nothing has really changed from the version described by Hill other than the fact that Kundi removes the Hill criteria of analogy and specificity completely.

In the end, the exceptions and reservations mentioned by Rothman and Greenland have not disappeared and the structure of the method—as a list—remains. And yet, Kundi's major claim was that Hill's criteria are somehow inadequate and, worse yet, considered a list of conditions.

Some authors simply reject Hill's criteria. Ward (24, p.16), for example, writes that "satisfactions of the Bradford Hill criteria do not 'justify' causal claims" and that "satisfactions of the Bradford Hill criteria neither guarantee the truth of a causal conclusion nor make it improbable that a causal conclusion is false." Ward's conclusion emerged from a long and complex philosophical discussion too long to be examined here. The result, however, is that the author rejected Hill's criteria but did not replace it with another set of criteria or, as in the Kundi paper, the same criteria in a similar list. Rather, Ward finishes his argument by recommending that causal claims should emerge after "only careful and reflective analyses using the appropriate methodological safeguards and statistical tools." This cannot be much of an improvement given that "methodological safeguards and statistical tools" already exist and, more importantly, these vague recommendations give the practitioner no practical guidance on how to interpret epidemiological and biological evidence. Who's safeguards and which statistical tools should we apply? How is this an improvement on the existing process of causal claims that is not simply the application of Hill's criteria but rather the application of a family of methods the last step of which is Hill's list? Rejecting one method—Hill's criteria—and replacing it with a vague nod towards safeguards and tools cannot be considered progress.

Less Extreme Critiques of Hill's Criteria

Another less extreme version of rejecting Hill's criteria is provided by the National Toxicology Program at the National Institute of Environmental Health Sciences (NIEHS), within the National Institutes of Health (25). In their version, only 6 of the 9 Hill criteria remain. Rejected are specificity, coherence, and analogy. No justification is provided. The authors cite Hill (1) and yet leave out a third of the "list" as if they did not exist in the original paper. It does not take much to realize that by reducing the number of criteria, causal claims become easier, i.e., more likely. We will return to the analogy criterion below.

Rejecting some aspect of Hill's approach can take many forms. Consider a recent discussion by Fedak et al. (26, p. 2) who argue that "statistical significance—not the magnitude of association—is the accepted benchmark for judging the strength of an observed association, and thus its potential causality." Simply put, these authors are rejecting in full the reason the magnitude of association matters, namely, that unmeasured confounders are less likely to affect a presumed causal association if the magnitude of association is "large," i.e., greater than 2.0 or some other threshold. The Fedak et al. (26) view ignores the fact that statistical significance—i.e., the role of chance—is already an integral part of the original Hill paper (1). Hill wrote that his considerations (or viewpoints or criteria) were to be applied when the "observed association is perfectly clear-cut and beyond what we would care to attribute to chance" (p. 295). Rejecting the magnitude argument by replacing it with statistical significance is another example of making causal claims easier. By doing so, the authors are ignoring a key scientific problem: the existence of unmeasured alternative explanations in the form of unmeasured confounders.

There is another example of the rejection of Hill's criteria that requires attention. As noted above, the NIEHS view of Hill's criteria eliminates the criterion (or consideration) of analogy. Indeed, if one were to carefully review current epidemiological textbooks and other written commentary on Hill's criteria, it would be relatively easy to find many authors who believe analogy—as it appeared in Hill's paper—to be, for want of a better word, worthless. Some write that analogy has "not stood the test of time" and "cannot be considered essential" (27, p. 5519). Others call it "vague" or "weak" (28) In a recent commentary, one author claims that analogy "doesn't work" and is "possibly detrimental" (29). Others write that analogy has major limitations (30).

As described elsewhere (10), "many epidemiology textbook and books published between 1970 and the present do not mention analogy when they discuss causation (5,6.13-15,20,22,32-34). In the other texts and in published articles, analogy is ignored (3,31,35,36) or equated with (and thus replaced by) biological plausibility or coherence (7,11,37,38). In many influential accounts, analogy has been downgraded to an unessential manifestation of a scientist's imagination (4,16,19,24,39,40). Consider, for example, Lucas and McMichael's (39) view of analogy: 'Analogy. Bradford Hill and other epidemiologists recognized that the notion of analogy can be taken to impractical extremes and may depend on the imagination of scientists to see analogies.'"

The problem with these overwhelmingly negative accounts of analogy is that they ignore what Hill wrote. In short, Hill noted (1, p. 299) that: "In some circumstances, it would be fair to judge by analogy. With the effects of thalidomide and rubella before us, we would surely be ready to accept slighter but similar evidence with another drug or another virus."

Putting aside the implication that analogy only applies in some circumstances, the word "slighter," as well as the role of judgment (41), Hill's description suggests a very different interpretation of analogy than that provided in the examples above. A key concept found in his description is a comparison of two bodies of evidence. In essence, Hill writes that a causal claim can emerge by comparing the evidence collected for one accepted causal relationship to that collected for another association not yet determined to be causal. Comparing bodies of evidence in a systematic manner and assessing their similarity (or lack thereof) is a completely different and much more involved issue than what the scientific community has believed. We will not pursue this further except to point out that if Hill is to be taken at his word, analogy should be considered a critically important consideration and one that can be used to test the reliability of causal claims (10,42).

There are at least two places in Hill's original account that require reconsideration. He wrote that none of his considerations were required for causation although temporality—that the exposure precede the onset of disease—does seem to be a necessary condition. Furthermore, Hill wrote that biologic plausibility cannot be required for causation which seems dated given the explosion in our understanding of the biological and genetic origin of many diseases that has emerged in the past 60 years. Nevertheless, there could be situations in which the epidemiologic evidence was so convincing that a biologic explanation could be considered unnecessary.

Other Examples of the Erosion of Causal Inference

The use of causal criteria is not the only place where causal inference appears to be eroding. One of the methodologic developments that occurred twenty years after Hill's seminal paper is the appearance and acceptance of the systematic

narrative review as a necessary approach for causal assessments. Making causal claims in the absence of a systematic approach to collecting, describing, and interpreting evidence—both epidemiological studies and laboratory studies—is another example of the erosion phenomenon. There are, in turn, examples of the misuse of meta-analysis in causal inference. It should be considered a method that, if "positive," means that the relationship under consideration is therefore, causal. Meta-analysis is an important and powerful tool but it is not a sufficient method for causation.

Summary and Recommendations

We have described several efforts to reject causal criteria in toto or piecemeal. These we believe are examples of a broader effort to make causal claims easier, that is, less rigorous. What we have mentioned but not discussed are ways to evaluate the extent to which any change in the status quo makes causal claims better. It is relatively easy, as shown in this paper, to reveal efforts by authors and institutions to chip away at established methodologies, like Hill's criteria. It is another matter altogether to demonstrate how any of the proposed changes improve the practice of causal inference. Nevertheless, there is an implicit assumption made by these authors that they are improving the practice of causal inference by eliminating all or some of Hill's criteria. We are not convinced.

One of the central concerns of any form of scientific measurement—or more generally, scientific assessment—is the extent to which the method used for that assessment is reliable. We suggest that reducing the criteria for making causal claims, if anything, would appear to make assessments of causation less rather than more reliable. We have not demonstrated much less proved this statement but the fact that the criteria, largely as they were originally described, can be easily linked to basic scientific principles indicates to us that rejecting them is a bad idea.

Put another way, if a method is simply rejected—as some have advocated for the causal criteria—then how do we know that what remains is at least as good or better than what existed before the purge? Similarly, if you reject the criteria and replace them with something else, how do we know that the new method is better than the old? These questions are important and are not only unanswered but unasked in most accounts of causal inference. We leave these for another day.

The assessment of causation is a central problem in epidemiology, preventive medicine, therapeutic medicine, and public health. When causation is the issue, there is a consensus in the epidemiological and biomedical communities that a systematic approach to identifying, describing, and interpreting evidence is appropriate and should include the use of causal criteria. The criteria that have been used for decades—as described in this paper—are consistent with scientific principles and the basic scientific method. It follows that removing or simply changing these criteria should not be adopted without a careful assessment of the potential impact on causal inference of such changes. We recommend that investigators involved in the process of causal inference use methods consistent with the consensus view at least for now. We also recommend that research into the practice of causal inference seems prudent.

References

1. Hill AB. The environment and disease: Association or causation? *Proc Roy Soc Med*. 1965;58:295-300.

2. Hill AB. Statistical evidence and inference. Chapter 24 in: *A Short Textbook of Medical Statistics*. London:Hodder and Stoughton 1971:283-296.

3. Susser M. Rules of inference in epidemiology. *Reg Tox Pharm*. 1986;6:116-28.

4. Aschengrau A, Seage GR. The epidemiologic approach to causation. Chapter 15 in: *Essentials of Epidemiology in Public Health*. Sudbury, MA:Jones and Bartlett, 2003;375-401.

5. Beaglehole R, Bonita R, Kjellstrom T. Causation in epidemiology. Chapter 5 in: *Basic Epidemiology*. Geneva:World Health Organization, 1993;71-81.

6. Gordis L. From association to causation: Deriving inferences from epidemiologic studies. Chapter 13 in: *Epidemiology*. 2nd ed. Philadelphia:W.B. Saunders, 2000;184-203.

7. Goodman SN, Samet JM. Causation and causal inference. Chapter 1 in: Schottenfeld D, Fraumeni Jr, JF. *Cancer Epidemiology and Prevention*. 3rd ed. New York:Oxford, 2006;3-9.

8. Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol*. 2000;29:387-390.

9. Weed DL, Hursting SD. Biologic plausibility in causal inference: current method and practice. *Am J Epidemiol*. 1998;147:415-425.

10. Weed DL. Analogy in causal inference: rethinking Austin Bradford Hill's neglected consideration. *Ann Epidemiol*. 2018;28:343-346.

11. Cogliano VJ, Baan RA, Straif K, et al. The science and practice of carcinogen identification and evaluation. *Environ Health Perspect*. 2004;112:1269-1274.

12. United States Environmental Protection Agency (EPA). *Guidelines for Carcinogen Risk Assessment*. Risk Assessment Forum. U.S. Environmental Protection Agency, Washington, DC. EPA/630/P-03/001F March, 2005.

13. MacMahon B, Pugh TF. Concepts of cause. Chapter 2 in: *Epidemiology: Principles and Methods*. Boston:Little, Brown, 1970.

14. Mausner JD, Bahn AK. The search for causal relations: Observational studies. Chapter 5 in: *Epidemiology: An Introductory Text*. Philadelphia:W.B. Saunders. 1974;91-111.

15. Kleinbaum DG, Kupper LL, Morgenstern H. Fundamentals of Epidemiologic Research. Chapter 2 in: *Epidemiologic Research*. Belmont, CA:Lifetime Learning. 1982;19-39.

16. Rothman KJ. Causal Inference in Epidemiology. Chapter 2 in: *Modern Epidemiology*. Boston: Little, Brown. 1986;7-21.

17. Weed DL. Causal and Preventive Inference. Chapter 17 in: Greenwald P, Kramer BS, Weed DL. *Cancer Prevention and Control*. New York:Marcel Dekker, 1995;285-302.

18. Kelsey JL, Pettiti DB, King AC. Key Methodologic Concepts and Issues. Chapter 2 in: Brownson RR, Pettiti DB. *Applied Epidemiology: Theory to Practice*. New York:Oxford University Press, 1998;35-69.

19. Rothman KJ and Greenland S. Causation and Causal Inference. Chapter 2 in: *Modern Epidemiology*. 2nd ed. Philadelphia:Lippincott, Raven. 1998.

20. Vetter N, Matthews I. Causation. Chapter 3 in: *Epidemiology and Public Health Medicine*. London: Churchill, Livingstone, 1999;23-30.

21. Rothman KJ. What is Causation? Chapter 2 in: *Epidemiology: An Introduction*. New York:Oxford University Press, 2002;8-23.

22. Bhopal R. Cause and effect: The epidemiological approach. Chapter 5 in: *Concepts of Epidemiology*. New York:Oxford University Press, 2002;98-132.

23. Kundi M. Causality and the interpretation of epidemiologic evidence. *Environ Health Perspect*. 2006;114:969–974.

24. Ward AC. The role of causal criteria in causal inferences: Bradford Hill's "aspects of association." *Epidemiol Perspect Innovat*. 2009, 6:2.

25. National Toxicology Program (NTP). *Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration*. National Institute for Environmental Health Sciences (NIEHS), 2019.

26. Fedak KM, Bernal A, Capshaw ZA, et al. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol*. 2015;12:14 DOI 10.1186/s12982-015-0037-4.

27. Carbone M, Klein G, Gruber J, et al. Modern criteria to establish human cancer etiology. *Ca Res*. 2004;64: 5518–5524.

28. Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy. J *Fam Plann Reprod Health Care*. 2008;34:261-264.

29. Ioannidis JPA. Exposure-wide epidemiology: revisiting Bradford-Hill. *Stat Med*. 2015;doi:10.1002/sim.6825.

30. Van Reekum R, Streiner EL, Conn DK. Applying Bradford-Hill's criteria for causation to neuropsychiatry. *J Neuropsych Clin Neurosci*. 2001;13:318-325.

31. International Agency for Research on Cancer (IARC). Preamble. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. 2006. World Health Organization: Lyon, France.

32. Susser M. Criteria of judgment. Chap. 11 in *Causal Thinking in the Health Sciences*. Oxford:University Press, 1973:140-162.

33. Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. 2nd ed. Oxford:University Press; 2004.

34. Baker D. Review of environmental health and epidemiological principles. In: Baker D, Nieuwenhuijsen MJ. *Environmental Epidemiology: Study methods and applications*. Oxford:University Press; 2008:15-40.

35. Hall W. A simplified logic of causal inference. *Aust NZ J Psychiatry*. 1987;21:507-513.

36. Adami HO, Berry CL, Breckenridge CB, et al. Toxicology and epidemiology: Improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. *Tox Sci*. 2011;122:223-234.

37. Schlesselman JJ. "Proof" of cause and effect in epidemiological studies: criteria for judgment. Prev Med 1987;16:195-210.

38. Thygesen LC, Andersen GS, Andersen H. A philosophical analysis of the Hill criteria. *J Epidemiol Commun Health*. *2005*;59:512-516.

39. Lucas RM, McMichael AJ. Association or causation: evaluating links between "environment and disease." *Bull World Health Org*. 2005;83:792-795.

40. Rothman KJ and Greenland S. Causation and causal inference. *Am J Pub Health*. 2005;95(Suppl 1):S144-150.

41. Weed DL. The nature and necessity of scientific judgment. *J Law Policy*. 2007;15:135-164.

42. Weed DL. Commentary: On the reliability of causal claims. *Global Epidemiol*. 2022;(in press).