



Causal criteria: time has come for a revision

Jørn Olsen¹ · Uffe Juul Jensen²

Received: 31 July 2018 / Accepted: 26 December 2018 / Published online: 16 January 2019
© Springer Nature B.V. 2019

Abstract

Epidemiologists study associations but they are usually interested in causation that could lead to disease prevention. Experience show, however, that many of the associations we identify are not the causes we take an interest in (correlation is not causation). In order to properly translate association into causes, a set of causal criteria was developed 50–60 years ago and they became important tools guiding this translational process (sometimes correlation is causation). Best known of these are the Bradford Hill ‘criteria’. In these last 50 years, epidemiologic theory and infrastructure have advanced rapidly without changes in these causal criteria. We think time has come to revisit the ‘old’ criteria to see which ones we should keep and which ones should be taken out or be replaced by new measures of association. Robustness of these criteria in attempts to make the association go away should have high priority. A group of renowned internationally recognized researchers should have this task. Since classifying associations as causes is often done in order to reduce or eliminate the exposures of concern results from conditional outcome research should also be used. We therefore suggest to add a ‘consequence’ criterion. We argue that a consequence criterion that provides a framework for assessing or prescribing action worthy or right in social contexts is needed. A consequence criterion will also influence how strict our causal criteria need to be before leading to action and will help in separating the ‘causal discussion’ and the discussion on what to do about it. A consequence criterion will be a tool in handling dilemmas over values (as social solidarity, fairness, autonomy). It will have implications for the interpretation and use of the procedural criteria of causality. Establishing interconnected procedural and consequence criteria should be a task for institutions representing and being recognized by experts, civil society and the state.

Keywords Causation · Criteria · Hill · Prevention · Translational research

Introduction

Our current, best-known causal criteria were presented more than 50 years ago [1–3] by, among others, Susser, the Surgeon General and Bradford Hill. They have served us well in epidemiology as a toolbox for a structured debate on causation. They were not labelled as criteria by Hill [1] but they have earned their status as criteria over the years. In spite of the rapid development in theoretical epidemiology, they have remained at least as a reference point for causal thinking in review committees and for decision makers.

Many advocate a more frequent use of causal terminology [4] even when reporting from single studies, but in spite of better research tools, we will usually not be in a position where single studies justify a causal label. Being able to identify all causal links in the process from exposure to disease does not mean we are able to study these links in an unbiased fashion under real life conditions. One thing is knowing what can go wrong in causal inference; another thing is avoiding these pitfalls in praxis.

Often we study problems in public health or clinical epidemiology where we have to take a stand on recommending acting or doing nothing. The consequence of choosing one or the other option should not determine our belief in causality but must be taken into consideration when we make public health decisions. We suggest therefore to add a consequence criterion.

We may be studying “Laws of Nature” and not just associations in specific populations [5], but we do it with imperfect tools [6] although new tools provide better and more

✉ Jørn Olsen
jo@ph.au.dk

¹ Department of Clinical Epidemiology, Aarhus University Hospital, Olof Palmes Alle 43-45, 8200 Aarhus N, Denmark

² Department of Philosophy and History of Ideas, School of Culture and Society, Aarhus University, Building 1465, Jens Chr. Skous Vej 7, 8000 Aarhus C, Denmark

valid designs. Time has come to implement these new methodologies and concepts in formal causal inference.

Epidemiology is a scientific discipline that aims at identifying preventable causes of diseases in order to reduce the burden of disease. If causes of diseases are eliminated or reduced, we will expect their effects, the disease occurrence, to shrink or to diminish. If E is causing D, eliminating E will at least reduce the incidence in the studied population with one case, often more [7]. If the exposure is not a cause, eliminating the exposure need not reduce the disease occurrence, except when other causal factors in the pathways linking the exposure to the endpoint of interest are also changed.

Studying causation requires a concept of causation, which is not only a technical or a philosophical concept but is part of everyday language. We learn about it in standard situations of causal interventions beginning in childhood (as when we turn on the light). If we had no concept of causation, we would be left with a very primitive language [6]. According to Hill [1], preventive medicine (including occupational medicine) is an intervention practice governed by a “decisive... question whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A” (p. 29).

Hill's list of 'conditions'

Hill avoided the term 'causal criteria' and talked instead about 'viewpoints and 'guidelines'. His guidelines are now widely referred to as criteria, but there have been almost no attempts at clarifying in what sense, Hill's guidelines are, in fact, criteria.

Using Feinstein's account of the role of criteria in clinical research and practice [8], we argue that Hill's guidelines are *procedural criteria*, i.e. criteria used to outline the performance of intervention procedures. Criteria for good preventive practice have undergone changes as epidemiological methods, disease-patterns, working-conditions, technology artefacts, culture and economic conditions have changed. Causal criteria cannot be used in the way they are often used in clinical practise where a certain number of criteria will lead to action since doing nothing may not be an option.

Many have discussed causal criteria, also before Hill published his landmark paper in 1965. Causal criteria were also presented in the text by the Surgeon General's report on Smoking and Health [2] and later the International Centre for Cancer Research (IARC) added a probabilistic component to their classification of potential carcinogens [9].

Hill's 9 criteria were: (1) strength, (2) consistency, (3) specificity, (4) temporality, (5) biological gradient, (6) plausibility, (7) coherence, (8) experiment, and (9) analogy.

Before that, Hume had stipulated some of the criteria of our everyday use of the word cause (e.g., constant

conjunction). So we could say that the 'criteria go back'—in the sense that Hill exemplars from occupational medicine are in accordance with our everyday use of the word 'causation'. Hill's criteria, however, comprises more 'viewpoints' based on examples from a particular expert field.

Causation is often a delayed effect with a probabilistic outcome

Many discussions on causation often refer to Hume's "strong criteria" [10, 11].

For E to be a cause of D it must be true that:

1. E will always be followed by D—comment; E is a sufficient cause of D.
2. If E does not occur, D will not follow; comment; but for this to hold E has to be a necessary cause of D and the only necessary cause of D.

Hume believed these two statements to be alike, but they are not. The counterfactual condition in 2 will only be true if E is both a necessary and sufficient cause of D and is the only cause. The idea of a cause as a necessary and sufficient condition makes sense but is hardly ever seen in epidemiology, not even for infectious diseases although they were used in the Koch's postulates [12]. We see necessary causes but they are often the result of how we define the disease. If we include E in the definition of D, E will become a necessary cause by circular reasoning, as when we defined AIDS as a disease following HIV exposure. In the practice of epidemiology, we need other criteria to identify associations that may be likely causal candidates and thus targets for prevention [13]. We cannot limit our research to 'strong' sufficient and necessary causes but have to target component causes that act in concert (in causal fields) to onset an effect. Mackie and Rothman were the ones who linked Hume's causal criteria to a concept that often works in practice and explains its probabilistic nature and delayed effects; the component causal field model [14, 15].

On top of the list of causal criteria, we often placed strength of the association. The stronger the association is, the more likely it represents a causal link, although there is no estimate of strength in a standard directed causal graphs (DAGs). The strength of an association, we now think, is related to how common the other component causes in the causal field are in the population under study. According to Mackie's causal field theory causes will follow the INUS conditions; causes are 'insufficient but non-redundant parts of a condition which is itself unnecessary but sufficient for their effects'. However, strength is an important criterion because it makes other non-causal explanations less likely.

Consistency or reliability is also considered an important criterion, but the effect sizes are expected to depend upon the frequency of other component causes.

Evidence from randomized controlled trials, as illustrated by DAGs, will under perfect conditions reduce interpretation of a positive result to causation or chance. Randomized controlled trials are important tools, especially in clinical epidemiology, since confounding by indication may not be avoidable without randomization, especially when the treating doctors are good. Evidence from trials support causality but trials are also subject to error, especially if they need to be large and run for a long time.

Another important criterion is the dose–response association because it is hard to ‘explain away’ by confounding unless the confounder mimics the same dose–response effect; the higher the exposure, the more frequent is the outcome.

None of the mentioned criteria are *sine qua non*-criteria, except that a cause has to precede the effect; the association may not be a consequence of reverse causation.

It should also be a causal criterion—and perhaps the most important one—to have an association that remains after comprehensive attempts to remove it. “When you have eliminated the impossible, whatever remains, however improbable, must be the truth” (Sherlock Holmes). The task of the investigator should be to see if he/she can make an association go away, not just to add new data and repeat what has been done already. Repeating the same design does not bypass the verification problem and identifying observations that would not be compatible with the hypothesis may be more informative [16]. The new method development provides much better tools to see how robust an association is to falsification [17–32].

New method development should lead to new criteria

Many things have changed since Hill’s paper came out in 1965. Thinking of causal exposures with component causes that enter or exit these fields over time provides a concept that at least does not contradict empirical findings. Graphical presentations of causation based upon mathematical rules make us better prepared to decide on which data to collect and how they should be analyzed. Focus has changed from making inference based on *P* values [33] to bias analyses, for example by using instrumental variables, negative controls, triangulations, sibling comparison, use of cases as their own controls, use of marginal structural models, or use of inverse probability weighting to adjust for selection or confounding etc. [18–21, 23–32, 34]. The rapid improvement of computer technology has opened a whole set of new ways to analyze data and to better learn from simulation studies [17].

DAGs have provided a powerful tool to illustrate causation [20, 21, 34]. Presenting a plausible DAG with empirical support would argue for a causal association and should be one of the “causal criteria”. Putting the association through bias analyses will also make an important contribution by trying to quantify the potential role of selection and information bias, including confounding. If using ‘best bias analyses’ and reasonable assumptions will not make the association go away, it speaks in favor of causality. Use of counterfactual reasoning also made important contribution to causal understanding.

A consequence criterion

We have to evaluate the evidence we have in the light of the methods that were used to generate the findings. However, we often need more than procedural criteria for good epidemiological practice to be of use in real life.

Procedural criteria state prescriptions for doing something in a particular practice [35]. Hill’s procedural criteria give prescriptions for epidemiological research in the context of preventive medicine.

Acting in accordance with adequate procedural criteria does not always secure that adopted consequences will be accepted (‘in real life’) as appropriate (right, just, fair etc.). We have to consider what flows from decisions made in accordance with the procedural criteria. If society and its institutions acts upon it, it will have consequences in real life. However, if society and its institutions do not act it will often have consequences in real life as well.

In ‘real life’ procedural criteria for causal intervention are not sufficient. Hill gives examples to show how the strength of evidence demanded in a particular context of intervention should be determined in the light of human values as fairness, justice and autonomy. Before “we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and the sugar they do like” we should need ‘very strong evidence’.

Here Hill makes ‘human autonomy’ a criterion among others in preventive medicine. He is in accordance with Feinstein who also points to the need of combining procedural criteria with (what he labels) desirability criteria in medicine, criteria that prescribe actions that are considered worthy or right in social contexts [8].

The idea of coming to an agreement on causation is often related to action, to do nothing or something [36, 37]. This reflects back to our counterfactual consideration; what would have happened had the exposed not been exposed, but the question is now what would happen in the future if the people stop being exposed. A potential counterfactual future without the exposure may offer more benefits and less side effects than maintaining status quo.

The decision process need also to take into consideration if the ‘exposure’ is imposed from outside or a result of a personal choice.

Some may argue even for a moral obligation to action, i.e., to contribute to causal intervention in an environmental context. In the International Covenant on Economic, Social and Cultural Rights (1966) it is stated in Article 12 that the States recognize the right of everyone to the enjoyment of the highest attainable standard of physical and mental health [38]. Steps to be taken to achieve the full realization of this right include those necessary for the improvement of all aspects of environmental and industrial hygiene, and the prevention, treatment and control of epidemic, endemic, occupational and other diseases.

This emphasis on action was mentioned in Hill’s paper from 1965 [1]. “In occupational medicine, our object is usually to take action. If this be operative cause and that be deleterious effect, then we shall wish to intervene to abolish or reduce death or disease”. Paul Stolley further addressed our social responsibility in his talk to SER members [39] “This is not to say that all findings should not be scrutinized and challenged, but this should be done with a sense of responsibility”. We have no need for partly justified ‘opinions’.

It is not always the case that we have the luxury substantial evidence to evaluate effect risks in the light of causal criteria. Many drug trials are stopped at an early stage because the producer runs a high financial and ethical risk if they bring a harmful product on the market. A decision to implement a new vaccine—in spite of limited evidence—should be taken if the risk of doing nothing is considered to exceed the risk of using the vaccine. Other situations may call for decision making in situations where the risk of doing nothing is high, but the decision process is often heavily biased towards doing nothing. ‘Active’ mistakes are often more criticized than ‘passive’ mistakes.

Those who decide on this set of criteria should be driven by a wish to reach the truth—be like Kafka’s truth seeking dogs. They should have no conflicts of interest in the sense that they have no personal gain by the decisions they make. They should be familiar with epidemiologic research and the infrastructure and conditions for doing research.

By including a consequence criterion in a set of criteria for causal intervention we are confronted with dilemmas between different ethical and social values (e.g., between respecting individual autonomy and freedom and respecting social responsibility and solidarity). Here epidemiologists face problems and challenges they cannot solve alone.

Hill cleverly avoided simple checklists to classify research as good or bad. Such checklists may be of value in very standardized research protocols like RTCs but to think it is possible to navigate in the more complicated rivers of causation by only using predefined guidelines is naïve.

Conclusions

We still need causal criteria to summarize evidence, and we need to act on these criteria to preserve health, and to prevent diseases. These criteria should reflect the best knowledge from research, and much has changed in this field since Hill wrote his paper in 1965.

The task for such a revision of procedural criteria as well as a consequence criteria should be left for authorities who can speak on behalf of the scientific communities and who are recognized and trusted by stakeholders in states and civil society. National institutions should collaborate (e.g., in the context of WHO) aiming at formulating international standards and criteria to promote public health. It will be a useless exercise, unless the criteria are widely accepted and used (lead to action). If this is not done, many of our research findings will not be used in practice.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

References

1. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med.* 1965;58:295–300.
2. Smoking and health. Report of the advisory committee to the Surgeons General of the Public Health Service. Publication no. 1103. Washington; 1964.
3. Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol.* 1991;133:635–48.
4. Hernan MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health.* 2018;108:616–9.
5. Keiding N, Louis TA. Perils and potentials of self-selected entry to epidemiological studies and surveys. *J R Stat Soc A.* 2016;179:319–76.
6. von Wright GH. Explanation and understanding. New York: Cornell University Press; 1971.
7. Nohr EA, Olsen J. Commentary: Epidemiologists have debated representativeness for more than 40 years—has the time come to move on? *Int J Epidemiol.* 2013;42:1016–7.
8. Feinstein AR. Clinical biostatistics. XLV. The purposes and functions of criteria. *Clin Pharmacol Ther.* 1978;24:779–92.
9. IARC monographs on the evaluation of carcinogenic risk to humans, vol 100. Lyon; 2012. <https://monographs.iarc.fr/agents-classified-by-the-iarc/>.
10. Hume D. An enquiry concerning human understanding. London; 1748.
11. Morabia A. On the origin of Hill’s causal criteria. *Epidemiology.* 1991;2:367–9.
12. Evans AS. Causation and disease: the Henle–Koch postulates revisited. *Yale J Biol Med.* 1976;49:175–95.
13. Olsen J. What characterises a useful concept of causation in epidemiology? *J Epidemiol Community Health.* 2003;57:86–8.

14. Mackie JL. *The cement of the universe: a study of causation*. London: Clarendon Press; 1980.
15. Rothman KJ. Causes. *Am J Epidemiol*. 1976;104:587–92.
16. Popper K. *The logic of scientific discovery*. New York: Routledge; 1959.
17. Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. New York: Springer; 2009.
18. Lawlor DA, Tilling K, Davey SG. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45:1866–86.
19. VanderWeele T. *Explanation in causal inference—methods for mediation and interaction*. New York: Oxford University Press; 2015.
20. Pearl J. *Causality: models, reasoning and inference*. New York: Cambridge University Press; 2009.
21. Pearl J, Glymour M, Jewell NP. *Causal inference in statistics. A primer*. Chennai: Wiley; 2016.
22. Daniel RM, De Stavola BL, Vansteelandt S. Commentary: The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *Int J Epidemiol*. 2016;45:1817–29.
23. Broadbent A, Vandenbroucke JP, Pearce N. Response: formalism or pluralism? A reply to commentaries on ‘causality and causal inference in epidemiology’. *Int J Epidemiol*. 2016;45:1841–51.
24. Krieger N, Davey SG. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 2016;45:1787–808.
25. Krieger N, Davey SG. Response: FACEing reality: productive tensions between our epidemiological questions, methods and mission. *Int J Epidemiol*. 2016;45:1852–65.
26. Pearce N, Lawlor DA. Causal inference—so much more than statistics. *Int J Epidemiol*. 2016;45:1895–903.
27. Crislip M. Causation and Hill’s criteria. 2000. <https://sciencebas.edmedicine.org/causation-and-hills-criteria/>. Accessed 6 June 2018.
28. Ward AC. The role of causal criteria in causal inferences: Bradford Hill’s “aspects of association”. *Epidemiol Perspect Innov*. 2009;6:2.
29. Phillips CV, Goodman KJ. Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerg Themes Epidemiol*. 2006;3:5.
30. Hoffer M. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol*. 2005;2:11.
31. Breskin A, Cole SR, Westreich D. Exploring the subtleties of inverse probability weighting and marginal structural models. *Epidemiology*. 2018;29:352–5.
32. Haber N, Smith ER, Moscoe E, et al. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): a systematic review. *PLoS ONE*. 2018;13:e0196346.
33. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31:337–50.
34. Pearl J, Mackenzie D. *The book of why: the new science of cause and effect*. New York: Basil Books; 2018.
35. Jensen UJ. *Practice and progress: a theory for the modern health care system*. Oxford: Blackwell Scientific Publications; 1987.
36. Sorensen TIA. To see and then to act, that is the challenge. *Eur J Epidemiol*. 2017;32:737–9.
37. Hernan MA, Robins JM. Instruments for causal inference: An epidemiologist’s dream? *Epidemiology*. 2006;17:360–72.
38. Covenant on Economic, Social and Cultural Rights. Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A(XXI) of 16 December 1966 entry into force 3 January 1976, in accordance with article 27. United Nations Human Rights Office of the High Commissioner; 1966.
39. Outgoing SER President Addresses Group on Faith, Evidence and the Epidemiologist. 2017. <http://www.epimonitor.net/Stolley-1983-Speech.htm>; Paul Stolley speech.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.