Epidemiology in History

Hume, Mill, Hill, and the Sui Generis Epidemiologic Approach to Causal Inference

Alfredo Morabia*

* Correspondence to Dr. Alfredo Morabia, Barry Commoner Center for the Biology of Natural Systems, Queens College, CUNY, 65-30 Kissena Boulevard, Flushing, NY 11367 (e-mail: alfredo.morabia@qc.cuny.edu).

Initially submitted December 4, 2012; accepted for publication August 21, 2013.

The epidemiologic approach to causal inference (i.e., Hill’s viewpoints) consists of evaluating potential causes from the following 2, noncumulative angles: 1) established results from comparative, observational, or experimental epidemiologic studies; and 2) reviews of nonepidemiologic evidence. It does not involve statements of statistical significance. The philosophical roots of Hill’s viewpoints are unknown. Superficially, they seem to descend from the ideas of Hume and Mill. Hill’s viewpoints, however, use a different kind of evidence and have different purposes than do Hume’s rules or Mill’s system of logic. In a nutshell, Hume ignores comparative evidence central to Hill’s viewpoints. Mill’s logic disqualifies as invalid nonexperimental evidence, which forms the bulk of epidemiologic findings reviewed from Hill’s viewpoints. The approaches by Hume and Mill cannot corroborate successful implementations of Hill’s viewpoints. Besides Hume and Mill, the epidemiologic literature is clueless about a plausible, pre-1965 philosophical origin of Hill’s viewpoints. Thus, Hill’s viewpoints may be philosophically novel, sui generis, still waiting to be validated and justified.

causality; David Hume; evidence; inference; John Stuart Mill; lung cancer; philosophy; tobacco

The epidemiologic approach to causal inference consists of evaluating potential causes from the following 2, noncumulative angles: 1) established results from comparative, observational, or experimental epidemiologic studies; and 2) reviews of nonepidemiologic evidence. It does not involve statements of statistical significance. The philosophical roots of Hill’s viewpoints are unknown. Superficially, they seem to descend from the ideas of English philosophers of the 18th and 19th centuries (David Hume, 1711–1776 (Figure 2), and John Stuart Mill, 1806–1873 (Figure 3)). The similarity between the set of “rules by which to judge of causes and effects” Hume published in his 1739 Treatise of Human Nature (16, p. 173) and Hill’s viewpoints (1) is intriguing. Hume’s rules (reproduced in the Appendix) comprise concepts present in Hill’s viewpoints, such as temporality, dose-response, consistency, analogy, and specificity (17, 18). Mill’s system of logic (19) relies on comparative designs and analytical concepts, such as confounding, effect modification, and bias relevant to the epidemiologic studies reviewed from Hill’s viewpoints. For instance, the Method of Difference (i.e., when all circumstances are in common but 1, the latter is the cause) (19) evokes cohort studies or controlled trials.

But do these similarities between Hume’s rules, Mill’s logic, and Hill’s viewpoints amount to a philosophical heritage? The immensity of the epistemological squabbles about these 2 giant philosophers precludes discussing in simple terms whether Hill’s viewpoints are consistent with Hume’s and Mill’s philosophical systems (20). Alternatively, Hume’s rules and Mill’s logic can be treated as practical approaches to causal inference similar to when epidemiologists apply each of Hill’s viewpoints to a causal statement.

Practically, Hume’s rules, Mill’s logic, and Hill’s viewpoints are different. In a nutshell, Hume ignores comparative evidence central to Hill’s viewpoints. Mill’s logic disqualifies as invalid nonexperimental evidence, which forms the bulk of epidemiologic findings reviewed from Hill’s viewpoints. The approaches by Hume and Mill cannot therefore corroborate successful implementation of Hill’s viewpoints (e.g., tobacco and lung cancer). Nor can they address the frustration...
expressed by epidemiologists who have proposed to replace, justify, or enrich Hill’s viewpoints by using well-characterized philosophies (21–24). A plausible explanation for the elusive origin of Hill’s viewpoints is that the epidemiologic approach to causal inference is sui generis (of its own genre), still needing to receive a proper philosophical justification.

Hume’s Rules

For Hume, causal laws are derived from the co-occurrence of 2 “objects,” 1 deemed as the cause and the other its effect. In A Treatise of Human Nature, Hume explains that, “We remember to have seen that species of object we call flame, and to have felt that species of sensation we call heat. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one cause and the other effect” (16, p. 87).

The problem is that the causal process cannot be observed, and there is nothing in the cause that lets us suspect what the effect will be. Consider this modern example. Few will contest that aspirin relieves pain, and that “one [is] cause and the other effect” (16, p. 87). Still, the underlying causal process cannot be observed. Pain relief cannot be foretold by examining the pill and, vice versa, the actual mechanism by which aspirin relieves pain is intangible. Aspirin and pain relief simply occur sequentially.

A causal connection is therefore an idea that emerges from the repeated impression that 2 “objects” occur sequentially (16). The recurring experience of pain relief following consumption of aspirin generates the idea of their causal connection. These past observations do not guarantee, however, that the 2 objects will keep co-occurring in the future, as explained in this key passage of the Treatise: “Even after experience has informed us of their constant conjunction, ’tis impossible for us to satisfy ourselves by our reason, why we should extend that experience beyond those particular instances, which have fallen under our observation” (16, pp. 91–92).

From the impression that aspirin relieved our pain in the past, we cannot be reasonably certain that it will in the future and cannot therefore justify transferring “the past to the future, the known to the unknown” (16, p. 136). Hume’s tactic of using absurd examples to anchor his point is terribly effective; even solid, but not necessarily causal experiences, such as days inevitably following nights or all men being mortal, cannot be tagged as immutable (16, p. 124). A sudden extinction (if the sun imploded) or the future immortality of the human species cannot be ruled out. Hume deemphasizes these examples, however, as “ridiculous” to discuss the main class of co-occurring events for which we feel less confident, because their causal links are based on probable events, that is, “evidence that is still attended with uncertainty” (16, p. 124). This class of events, for which “all knowledge degenerates into probability” (16, p. 180), also interests epidemiologists.

Figure 1. Portrait of Sir Austin Bradford Hill. Reproduced with permission from Wellcome Library, London, United Kingdom.

Hume describes the process by which we separate “chance” from “causation” as equivalent to competing ideas battling in a mental arena to shift the balance onto their side: “That effect, which has been the most common, we always esteem the most likely [. . .] when we transfer the past to the future, the known to the unknown, every past experiment has the same weight, and that ’tis only a superior number of them, which can throw the ballance [sic] on any side” (16, pp. 133–136).

Thus, if aspirin relieves pain most but not all of the time, we will still view its relationship to pain relief as causal. Nonetheless, there is ample space for being mistaken; ideas having won battle after battle for centuries may still be defeated by new evidence. Consider the belief in the therapeutic properties of bloodletting to treat fevers. Doctors felt comfortable practicing it for almost 2,500 years, being under the impression that it worked more often than it did not, until the discoveries of bacteriology toward the end of the 19th century established that evacuating warm blood targeted the symptom but not the cause of fevers. Bloodletting rapidly lost ground thereafter.

Acknowledging the fact that observations based on probable evidence can be defeated by new evidence, Hume lists a set of rules, by which “we learn to distinguish the accidental circumstances from the efficacious causes” (16, p. 149). He goes on to say, “Since therefore ’tis possible for all objects to become causes or effects to each other, it may be proper to fix some general rules, by which we may know when they really are so” (16, p. 173).

Note the use of the adverb “really,” suggesting that the rules can validate causal statements. For philosophers Beau-champ and Rosenberg, the rules are “a set of individually necessary and jointly sufficient specifications of the truth of causal associations,” the “task” of which is to “specify the conditions that warrant causal statements” and “to determine the objective validity of such statements” (25, p. 23). Indeed, rules 1, 2, 3, 5, and 6 (Appendix) are “must” causal statements. Rules 4, 5, 6, and 7 include the term “always.”

Per Hume’s rules, aspirin and pain relief are viewed as causally associated because, most of the time, aspirin precedes pain relief, because analgesia is greater with 1 g of aspirin than with 500 mg, because extracts of bark from the willow tree (from which the natural active ingredient of aspirin is extracted) have analgesic effects, because other bark extracts (e.g., from the cinchona tree) have pharmaceutical properties, and so on.

The rules set a high threshold for causal validity, compatible with Hume’s skepticism about the potential success of our process of causal inference. Indeed, they work for aspirin and pain relief, but not for cigarette smoking and lung cancer. For the latter, we cannot pass the third rule (“There must be a constant union betwixt the cause and effect”) (Appendix) and its axiom (“that effect, which has been the most common, we always esteem the most likely”) (16, p. 133) because few heavy smokers (say, 10%) develop lung cancer.

HILL’S VIEWPOINTS

In Hill’s “The environment and disease: association or causation,” (1) some viewpoints refer to similar concepts as Hume’s rules (e.g., temporal relationship, analogy, specificity, dose-response, consistency). Other viewpoints are modern additions (e.g., strength of the association, biological plausibility, experimental or semiexperimental evidence, coherence) (17).

As do the rules, viewpoints list expected qualities of causal associations. But associations have different meaning in Hume’s rules and Hill’s viewpoints. Hume’s rules work to distinguish whether 1 recurrent sequence is causal (e.g., lung cancer occurrence among smokers). Smoking being less frequently followed by cancer (i.e., 10% of heavy smokers develop lung cancer in their lifetimes) than by nonsmoker (i.e., 90% of heavy smokers do not develop lung cancer in their lifetimes), smoking does not cause lung cancer per the principle of the most frequent co-occurrence. On the contrary, the associations reviewed from Hill’s viewpoints assume the comparison of (at least) 1 recurrent sequence involving the cause with (at least) another sequence of events not involving the cause. Each odds of lung cancer takes epidemiologists only halfway through the comparison; the co-occurrence of smoking and cancer needs to be contrasted with that of nonsmoking and noncancer (i.e., non-smokers rarely get lung cancer). Smoking can cause lung cancer because nonsmokers have lower odds of lung cancer than do smokers.

Thus, Hill’s viewpoints are informed by comparative evidence, whereas Hume’s rules are not. Hume’s “temporality” means that the cause preceded the outcome. Hill’s “temporality” means that in the 2 sequences of events, both the cause in

1 cohort, and its absence in another cohort, preceded their outcomes. Hume’s “dose-response” means that more cause corresponds to more effect, such as “a certain degree of heat gives pleasure; if you diminish that heat, the pleasure diminishes” (Appendix). Hill’s “dose-response” means that the contrast between the co-occurrence of exposure and disease and of non-exposure and nondisease becomes sharper when the intensity of exposure (e.g., nonsmoking, light smoking, moderate smoking, and heavy smoking) increases.

We can speculate that Hume’s reasoning lacks contrastive evidence because comparisons were less essential for the dominant sciences of the time, such as Newtonian physics, than they are in epidemiology. Moreover, Hume may not have been familiar with the single instance in which epidemiologic comparisons had been successfully applied before 1739, showing that people who were medically inoculated against smallpox had a lower lethality than those who were naturally infected (26, pp. 43–70) (27).

Hill’s viewpoints, as opposed to Hume’s rules, are not specifications of causal associations. Characteristically, and in contrast to Hume, Hill does not use the terms “must” or “alway$”s.” All the uses of “must” and the single use of “alway$”s” in Hill’s paper are about caveats not to use the viewpoints as a jointly sufficient and necessary validation set for causation. Hill’s classic line is, “What I do not believe 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” (1, p. 299).

MILL’S LOGIC

Mill’s system of logic can usefully guide epidemiologic research about ways to “effectively test and eliminate many hypothetical causes” (28, p. 71). Indeed, Mill distinguishes cause-to-effect versus effect-to-cause approaches (19, p. 278); exclusive effects from “mixed and confounded” effects (19, p. 311); independent causes from those “modifying the effects of one another” (19, p. 315); and “effects of the same cause” (a→c←b) versus “effect of [independent] causes” (a→c←b) (19, p. 373). Susser has stressed, however, that Mill’s analytical logic cannot be used for epidemiologic causal inference (28, p. 71).

First, Mill excludes drawing valid conclusions from group comparisons of patients. Having discussed at length the evaluation of the health effects of mercury (“is, or is it not, some particular medicament (mercury, for instance) a remedy for a given disease”), Mill concludes that “we cannot tell that these other circumstances (“which we do know shall have effects susceptible of being confounded with those of the agents whose property we wish to study”) may not have produced the whole of the effect, independently or even in spite of the mercury. . . . Anything like a scientific use of the method of experiment in these complicated cases is, therefore, out of question . . . “ (19, pp. 320–323).

The experimental use of the “method of difference” (19, pp. 320–323) can establish a law of causation, but observational research cannot. Mill states, “Observation without experiment (supposing no aid from deduction) can ascertain sequences and coexistences but cannot prove causation” (19, p. 277). Thus, Mill’s logic of causal inference stops where Hill’s viewpoints start. The system of logic would not have allowed the Surgeon General’s Advisory Committee, assembled by US Surgeon General Luther Terry in 1963, to sense a causal link in the evidence about the health effects of tobacco. In contrast, Hill’s viewpoints decisively helped reach the conclusion that tobacco caused lung cancer in men (12).

DISCUSSION

A close look at their similarities and differences indicates that Hume’s rules and Hill’s viewpoints integrate different types of evidence (Hume ignores comparative evidence, which forms the core of Hill’s viewpoints) and have different purposes (Hume’s rules, but not Hill’s viewpoints, are conditions for causal statements). These differences explain the confidence of epidemiologists in the possibility of identifying valid causes as opposed to Hume’s notorious dubitative stance.

Mill’s system of logic deals with analytical aspects of comparisons, such as study design, unmeasured confounding, and statistical significance. These issues are not directly relevant to Hill’s viewpoints, which intervene beyond data analysis, when all the available evidence (i.e., epidemiologic, biological, experimental) is reviewed for lack of a jointly sufficient and necessary causal criterion.

Could there be other potential philosophical roots of Hill’s viewpoints besides Hume and Mill? The historical answer to this question is that, after a thorough review of the epidemiologic literature, no serious candidate emerges who critically assesses Hill’s viewpoints or discusses them from a philosophical perspective. Until 1965 (1, 3–15), Francis Bacon (1561–1626), Hume, Mill, Jakob Henle (1809–1885), Robert Koch (1843–1910), and Karl Popper (1902–1994) are mentioned (29). After 1965 (2, 17, 20–24, 28–68), we find Evans’s “unified concept” for infectious and noninfectious diseases integrating Henle-Koch postulates and Hill’s viewpoints (32); an unconcluded debate over the falsificationist approach of Karl Popper (1902–1994); and piecemeal references to Bertrand Russel (1872–1970), Rudolf Carnap (1891–1970), Carl G. Hempel (1905–1997), Mario Bunge, and other philosophers. However, the epidemiologic literature is clueless as to the origin of Hill’s viewpoints.

Hill’s viewpoints supplement a distinct response to a topic that has agitated philosophers since Bacon (69), that is, the causal inference based on observational but nonexperimental (strictly defined) evidence. The approach encapsulated in Hill’s viewpoints successfully passed a major challenge when the committee assembled by the US Surgeon General, in which half of the participants represented the tobacco industry, unanimously admitted the carcinogenicity of tobacco (70). Hill’s viewpoints are widely used in epidemiology and public health (24, 46, 52, 66). Still, an explanation of why Hill’s viewpoints work is unsurprisingly wanting, because philosophy usually follows rather than leads scientific innovation. The method called “inference to the best explanation” (71, 72) appears promising in explaining the sometimes intractable polarization among epidemiologists between Popperians and non-Popperians (see various contributions in Rothman (23)), refutationists versus verificationists (38), inductivists versus deductivists (22),
and Bayesians versus non-Bayesians (71, pp. 103–120). It can possibly be traced to Charles Sanders Peirce (1839–1914) (73), but its extension to epidemiology is recent (71, pp. 71–90) (66, 74).

Thus, epidemiologists may have developed a novel, sui generis approach to discover causes “however tangled the skein of causation is” (15, p. 1000), which still awaits its philosophers. If this statement is too bold for my philosophical expertise, it would suffice to trace Hill’s viewpoints to the ideas of a single philosopher to prove it wrong.

ACKNOWLEDGMENTS

Author affiliations: Barry Commoner Center for the Biology of Natural Systems, Queens College, City University of New York, New York, New York (Alfredo Morabia); and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York (Alfredo Morabia).

A.M. received funding from the National Library of Medicine (grant 1G13LM010884-01A1).

Conflict of interest: none declared.

REFERENCES

The following material is excerpted from Hume’s A Treatise of Human Nature, published in 1739 (16, pp. 173–175).

Section XV: Rules by which to judge of causes and effects.

Since therefore ‘tis possible for all objects to become causes or effects to each other, it may be proper to fix some general rules, by which we may know when they really are so.

1. The cause and effect must be contiguous in space and time.
2. The cause must be prior to the effect.
3. There must be a constant union betwixt the cause and effect. ‘Tis chiefly this quality, that constitutes the relation.
4. The same cause always produces the same effect, and the same effect never arises but from the same cause. This principle we derive from experience, and is the source of most of our philosophical reasonings. For when by any clear experiment we have discover’d the causes or effects of any phenomenon, we immediately extend our observation to every phenomenon of the same kind, without waiting for that constant repetition, from which the first idea of this relation is deriv’d.
5. There is another principle, which hangs upon this, viz. that where several different objects produce the same effect, it must be by means of some quality, which we discover to be common amongst them. For as like effects imply like causes, we must always ascribe the causation to the circumstance, wherein we discover the resemblance.
6. The following principle is founded on the same reason. The difference in the effects of two resembling objects must proceed from that particular, in which they differ. For as like causes always produce like effects, when in any instance we find our expectation to be disappointed, we must conclude that this irregularity proceeds from some difference in the causes.

7. When any object increases or diminishes with the increase or diminution of its cause, 'tis to be regarded as a compounded effect, deriv'd from the union of the several different effects, which arise from the several different parts of the cause. The absence or presence of one part of the cause is here suppos'd to be always attended with the absence or presence of a proportionable part of the effect. This constant conjunction sufficiently proves, that the one part is the cause of the other. We must, however, beware not to draw such a conclusion from a few experiments. A certain degree of heat gives pleasure; if you diminish that heat, the pleasure diminishes; but it does not follow, that if you augment it beyond a certain degree, the pleasure will likewise augment; for we find that it degenerates into pain.

8. The eighth and last rule I shall take notice of is, that an object, which exists for any time in its full perfection without any effect, is not the sole cause of that effect, but requires to be assisted by some other principle, which may forward its influence and operation. For as like effects necessarily follow from like causes, and in a contiguous time and place, their separation for a moment shews, that these causes are not compleat ones.

Here is all the Logic I think proper to employ in my reasoning; and perhaps even this was not very necessary, but might have been supply'd by the natural principles of our understanding.