Public Health Classics

This section looks back to some ground-breaking contributions to public health, reproducing them in their original form and adding a commentary on their significance from a modern-day perspective. Robyn M. Lucas and Anthony J. McMichael review *The environment and disease: association or causation?* by Sir Austin Bradford Hill on establishing relationships between illness and conditions of work or living. The original paper is reproduced by permission of The Royal Society of Medicine Press Limited (http://www.jrsm.org).

Association or causation: evaluating links between "environment and disease"

Robyn M. Lucas[1] & Anthony J. McMichael[2]

Epidemiological studies typically examine associations between an exposure variable and a health outcome. In assessing the causal nature of an observed association, the "Bradford Hill criteria" have long provided a background framework — in the words of one of Bradford Hill's closest colleagues, an "aid to thought" (1). First published exactly 40 years ago, these criteria also provided biomedical relevance to epidemiological research and quickly became a mainstay of epidemiological textbooks and data interpretation (2). Their checklist nature suited the study of simple, direct causation by disciplines characterized by classic scientific and mathematical training.

Most diseases have a multifactorial pathogenesis, but the conceptualization of their causation varies by discipline. While it is scientifically satisfying to elucidate the many component causes of an illness, in public health research the more important emphasis is on the discovery of necessary or sufficient causes that are amenable to intervention. Even so, over the four decades since Bradford Hill's paper appeared, the range of multivariate, multistage and multi-level research questions tackled by epidemiologists has evolved, as have their statistical methods and their engagement in wider-ranging interdisciplinary research. Within that context it is often not appropriate to seek the discrete cause or causes of a disease, but rather to identify a complex of interrelated and often interacting factors that influence the risk of disease (1). This complicates the assessment of causality.

The general context within which Bradford Hill developed his ideas about causal inference warrants brief review here. Most epidemiological research is non-experimental, being conducted in an inherently "noisy" environment in free-living populations. For example, the quality of the measurement of exposure and of health status is usually less than in controlled clinical trials or laboratory-based studies (measurement error); there are potential confounding variables that are statistically associated with the exposure variable of interest while also predictive of the health outcome in their own right, and these covariates must be controlled for; the sample of persons studied may not provide true information about the relationship between exposure and outcome in the source population, with respect either to the relationship that the sample actually displays (selection bias) or apparently displays (classification bias). Epidemiologists therefore seek research settings and study designs that maximize the signal-to-noise ratio.

These sources of noise, intrinsic to much epidemiological research, require one to proceed cautiously in making causal inference. Once sufficient studies have been done, in diverse settings, and adequately limiting random error (an intrinsic property of a stochastic universe), systematic error (bias) and logical error (confounding), then the causal nature of observed associations can reasonably be assessed.

Note, though, that particular phrase: "causal nature". Causation is an interpretation, not an entity; it should not be reified. The 18th-century Scottish philosopher David Hume pointed out that causation is induced logically, not observed empirically (3). Therefore we can never *know* absolutely that exposure X causes disease Y. There is no final *proof* of causation: it is merely an inference based on an observed conjunction of two variables (exposure and health status) in time and space. This limitation of inductive logic applies, of course, to both experimental and non-experimental research.

Around the mid-20th century, the philosopher Karl Popper offered a solution to this problem of reliance on induction. He stressed that science progresses by rejecting or modifying causal hypotheses, not by actually proving causation. While flirting briefly with Popper's ideas in the 1970s (4), epidemiologists have generally taken a practical data-based approach to the notion of causation, comfortably embracing Bradford Hill's criteria of causality. In general, these seem well suited to the mostly non-experimental, bias-prone, confounding-rich nature of epidemiological research. These nine criteria, or guidelines, lay particular emphasis upon the temporality of the relationship, its strength, the presence of a plausible dose–response relationship, the consistency of findings in diverse studies, and coherence with other disciplinary findings and biomedical theory.

[1] National Centre for Epidemiology and Population Health, The Australian National University, Canberra, ACT 0200, Australia. Correspondence should be sent to this author (email: robyn.lucas@anu.edu.au).

[2] National Centre for Epidemiology and Population Health, The Australian National University, Canberra, Australia.

Ref. No. **04-016048**

Rather than proposing absolute criteria, Bradford Hill considered these as aspects of the association between an exposure and an outcome that "we especially consider before deciding that the most likely interpretation of it is causation".

Bradford Hill's ideas about causal inference were formulated in the heady early years of the rise of noncommunicable disease epidemiology, which was essentially a post-Second World War phenomenon. His own experience included, in particular, the first definitive controlled clinical trial — of streptomycin in the treatment of tuberculosis, in the late 1940s (5) — and the early studies of cigarette smoking and lung cancer, principally the British doctors cohort study (6). Other early successes in non-experimental epidemiological studies of noncommunicable diseases included those that entailed substantial, quantifiable occupational exposures, for example to ionizing radiation (7), asbestos (8) and nickel (9). It is not surprising that, against that background, the challenge seemed not so much that of elucidating and apportioning complex causality but, more fundamentally, of inferring simple, relatively direct-acting causality.

Bradford Hill recognized the importance of moving from association to causation as a necessary step for taking preventive action against environmental causes of disease. But there are questions about the universal applicability of his classic criteria. How valid are they in the assessment of multifactorial causality? Are they useful in a widening research agenda within which, for example, we try to identify and quantify the effects of more distal, often indirectly acting determinants of health such as factors related to socioeconomic status, the effects of urban design on physical activity levels and the incidence of obesity, or the effects of ongoing climate change on risk of death from flooding? More subtly, does our reliance on causal criteria as an intellectual framework shape and direct our research questions and funding opportunities?

Ten years after Bradford Hill's classic paper, Rothman presented a model of causation that stressed the multifactorial pathogenesis of disease, with multiple component causes or factors that increase risk, and diverse causal pathways (10). He identified necessary elements and combinations of exposures sufficient to result in disease development. Causal inference, then, would focus more on how well the results of epidemiological studies fit with such a model. Rothman and Greenland note that none of Bradford Hill's criteria alone is sufficient to establish causality — for each criterion there are situations in which both lack of satisfaction of the criterion may be causal and satisfaction of the criterion may be non-causal. Temporality, the requirement that the exposure must precede the effect, is the only necessary criterion for a causal relationship between an exposure and an outcome (11).

In the following section we briefly review the Bradford Hill criteria and their contemporary use in epidemiology.

Strength. Bradford Hill suggested that strong associations were more likely to be causal than weak associations. The strong associations he cites (a 200-fold increase in mortality from scrotal cancer in chimney sweeps exposed to tar or mineral oils, and a 20-fold increased risk of lung cancer in smokers compared with non-smokers) have more credence, being less likely to be attributable solely to uncontrolled residual confounding. Relatively weak associations are common in contemporary epidemiology, so that we are reliant on strong study design and methodology, with minimization of bias, evaluation of the role of chance and comprehensive measurement of possible confounders for a valid measure of association. This is often difficult in the study of complex environmental influences on human health.

Consistency. Bradford Hill also felt more confidence in a causal explanation for an association if the same answer had been achieved in a variety of different situations — prospectively and retrospectively and in different populations. Conversely, the results of studies of the same phenomenon may vary because of difference in the methods, interaction with a third variable (including gene–environment interaction (12)) or chance (11). While similar results achieved by different methods and in different populations enhance confidence in a causal interpretation, consistency is not a necessary criterion for a causal interpretation. Indeed, lack of consistency may provide valuable insights into the component causes of an outcome (if there is interaction with a third factor that is variably present) and warrant further investigation, rather than a non-causal conclusion.

Specificity. This criterion is often stated to mean that any exposure may give rise to only a single outcome (13). While this may be true for some infectious diseases, for example only rubella virus causes rubella, it is clearly unlikely with respect to many environmental exposures. Bradford Hill recognized that diseases may have more than one cause and that one-to-one relationships are not frequent. However, if an association is limited to specific groups with a particular environmental exposure or is greatly increased in these groups, then the case for a causal association is strengthened. Weiss suggests resurrection of specificity as a useful concept in study design, particularly valuable in unravelling complex problems in causal attribution (14). He cites as valuable facets of study design and causal inference the examination of specificity of an outcome (do cycle helmet wearers experience a decrease in all types of injury, or just head injury?), specificity of exposure (is ovarian cancer caused by any type of endometriosis, or only ovarian endometriosis?) or specificity with regard to susceptibility (the association between a particular genotype and an outcome is manifest only under specific environmental conditions for which genetic susceptibility is important (15)).

Temporality. Temporality is a necessary criterion for a causal association between an exposure and an outcome, that is, the exposure must precede the outcome (although measurement of the exposure is not required to precede measurement of the outcome).

Biological gradient. It seems logical that the likelihood of a causal association is increased if a biological gradient or dose–response curve can be demonstrated. However, such quantitative relationships may be difficult to demonstrate or may be attributable to residual confounding where the confounder itself exhibits a biological gradient in relation to the outcome (11). In addition, it is clear that for many environmental exposures there is a threshold or non-linear association, for example the association between ambient temperature and disease (16, 17), exposure to ultraviolet radiation and disease (18), and alcohol consumption and mortality (19).

Plausibility. While it is reassuring if a causal association is biologically plausible, Bradford Hill notes that "this is a feature I am convinced we cannot demand. What is biologically plausible depends upon the biological knowledge of the day". Further, it is "too often based not on logic or data but only on prior beliefs" (11).

Coherence. Coherence and biological plausibility share a requirement that the cause-and-effect interpretation of an association should fit with the known facts of the natural history and biology of the disease. Do the temporal patterns of exposure and the known biological effects of the exposure fit with the observed disease patterns? For example, the "hygiene hypothesis" as a cause of some autoimmune and allergic diseases coheres with trends in developed countries to both fewer childhood infections and an increasing incidence of allergic and autoimmune disorders (20).

Experiment. Do preventive actions taken on the basis of a demonstrated cause-and-effect association alter the frequency of the outcome? With overtones of Koch's postulates, this criterion offered, in Bradford Hill's view, the strongest support of a causal interpretation. Laboratory experimentation and human clinical trials allow the manipulation of exposures in a controlled environment unlike human observational epidemiological studies. Laboratory animals are bred to simulate sensitivity to particular environmental exposures, exposed in a measured way, monitored for disease development and then sacrificed to examine pathological changes. The randomized clinical trial design aims to control bias and confounding in human studies to allow estimation of the true association between exposure and outcome. In practice, however, control of confounding and bias may be achieved only at the cost of representativeness or study power.

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. Clear-cut analogies, however, may add to the weight of evidence for otherwise weak associations. Consider the study of the association of passive smoking with lung cancer. Quantification of exposure and accurate measurement of all confounders may be difficult. However, by analogy to the known risk of lung cancer in active smokers, persons exposed to second-hand smoke plausibly have an increased lung cancer risk mediated by the same biological pathways.

Bradford Hill did not prescribe these criteria as rules that must be fulfilled before an association can be judged as causal, but as ways of examining if cause and effect is the reasonable inference. The difficulty of making causal inference in relation to more distal exposures centres on the difficulty of seeing the pure association of exposure and health effect — free from bias, confounding and interaction with other exposures. The research situations in which this can occur are limited mainly to clinical trials and perhaps large observational studies with impeccable design and execution. Contemporary environmental epidemiology confronts non-homogeneous health outcomes such as asthma, multiple sclerosis and suicide that are groupings of signs or symptoms likely to have multiple etiologies. Exposures can be difficult to quantify and even to define (e.g. socioeconomic status and urban design) as well as to link temporally and spatially to the disease outcome (e.g. air pollution and climate change).

How would Bradford Hill have dealt with some of the issues in contemporary social epidemiology? To take what is perhaps an extreme example, in its World Health Report of 1998 (21), WHO concluded that the world's greatest risk factor for disease was poverty. At about the same time, three of the world's prominent orthodox epidemiologists argued that it was not the task of epidemiology to focus on poverty as a cause of disease (22). This divergence of views bears on the question of how far upstream should the matter of cause — and therefore potential intervention — be pursued.

In Australia, is the shockingly low life expectancy of Aboriginals to be attributed to their high prevalence of health-endangering behaviour at the individual level, including unbalanced diets, excessive alcohol consumption, cigarette smoking, sedentary behaviour, poor hygiene and dangerous driving? Is it attributable to population-level factors such as poor education, lack of primary health care, levels of access to processed foods and alcohol, and so on? Or is it caused by the social context of cultural disintegration, low self-esteem and poverty? Whatever the answer (and, in fact, all levels of causation are relevant), it is clear that causal relationships become more complex, less quantifiable and less amenable to formal causal inference as one moves from proximal to distal determinants of health outcomes (23).

The existence of formal criteria for causal inference may steer current research towards comfortable, tightly specified, research questions, and thus deter us from consideration of the "big picture" where the data are often fuzzy and residual confounding is likely. Indeed, funding bodies may prefer to award research grants to studies with clear delineation of both exposure and health outcome and a study design conducive to causal inference.

In conclusion, epidemiological studies seek understanding of the links between environment and health, and thus provide support for evidence-based practice. Whether such links can be considered causal can only be assessed with confidence once full consideration has been taken of epidemiological noise — chance, bias and confounding. Both practical and ethical considerations mean that causality cannot, in general, be proved in human studies. Rather, it must be induced from demonstrated associations between an exposure and a health outcome. Characteristics of that association, judged against some framework, then help us to assess whether that association is or is not causal.

In modern times, epidemiologists have extended their research horizons to encompass the domains of social epidemiology, of population-level relationships not reducible to individual-level study, and of the health consequences of complex environmental and social change processes. The notion of cause has become more complex, with most health outcomes having multiple component causes. Distinguishing which of these are necessary or sufficient is central to preventive efforts. Bradford Hill's criteria provide a framework against which exposures can be tested as component causes, but they are not absolute. As with statistical *P*-tests, the criteria of causality must be viewed as aids to judgement, not as arbiters of reality. ■

Acknowledgements

We thank Alistair Woodward, University of Auckland, for helpful suggestions during the drafting of this paper.

Competing interests: none declared.

References

1. Doll R. Proof of causality: deduction from epidemiological observation. Perspect Biol Med 2002;45:499-515.
2. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295-300.
3. Hume D. *A treatise of human nature.* Norton D, Norton M, eds. (reprinted from the original of 1740). Oxford and New York: Oxford University Press; 2000.
4. Buck C. Popper's philosophy for epidemiologists. Int J Epidemiol 1975;4:159-68.
5. Medical Research Council Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. BMJ 1948;2:769-82.
6. Doll R, Hill AB. Smoking and carcinoma of the lung. Preliminary report. BMJ 1950;739-48.
7. Folley J, Borges W, Yamaskai T. Incidence of leukemia in survivors of the atomic bomb in Hiroshima and Nagasaki. Am J Med 1952;13:311-21.
8. Wagner J, Sleggs C, Marchand P. Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province. British J Ind Med 1960;17:260-71.
9. Hill AB. Principles of medical statistics. London: The Lancet Ltd; 1967:308-9.
10. Rothman KJ. Causes. Am J Epidemiol 1976;104:587-92.
11. Rothman KJ, Greenland S. Modern epidemiology, 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998.
12. MacMahon B. Gene-environment interaction in human disease. J Psych Res 1968;6(Suppl 1):393-402.
13. Lemen R. Chrysotile asbestos as a cause of mesothelioma. Int J Occup Environ Health 2004;10:233-9.
14. Weiss NS. Can the 'specificity' of an association be rehabilitated as a basis for supporting a causal hypothesis? Epidemiology 2002;13:6-8.
15. Dwyer T, Ponsonby A-L, Stankovich J, Blizzard L, Easteal S. Measuring environmental factors can enhance the search for disease causing genes? J Epidemiol Community Health 2004;58:613-5.
16. Wilkinson P, Pattenden S, Armstrong B, Fletcher A, Kovats RS, Mangtani P, et al. Vulnerability to winter mortality in elderly people in Britain: population based study. BMJ 2004;329:647.
17. Rooney C, McMichael AJ, Kovats RS, Coleman MP. Excess mortality in England and Wales, and in Greater London, during the 1995 heatwave. J Epidemiol Community Health 1998;52:482-6.
18. Lucas RM, Ponsonby AL. Ultraviolet radiation and health: friend and foe. Med J Aust 2002;177:594-8.
19. Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modelling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol 2003;159:1077-86.
20. Bach J. The effect of infections on susceptibility to autoimmune and allergic diseases. New Engl J Med 2002;347:911-20.
21. The world health report 1998–Life in the 21st century: a vision for all. Geneva: World Health Organization; 1998.
22. Rothman KJ, Adami HO, Trichopoulos D. Should the mission of epidemiology include the eradication of poverty? Lancet 1998;352:810-3.
23. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. Am J Epidemiol 1999;149:887-97.