# Confounded by confounding: separating association from causation

*Huw TO Davies, Fiona LR Williams*

***Health-care professionals need to be able to distinguish causal relationships from simple associations in two main areas: when unravelling the aetiology of diseases, and when assessing the effects of therapies. In each of these the presence of confounding can seriously mislead. This short report explains the nature of confounding and outlines criteria that can be applied to help distinguish causality from mere statistical associations.***

## INTRODUCTION

The notion of 'cause' lies at the heart of health care. Without an understanding of cause and effect, diseases are unexplained, interventions and outcomes remain unconnected, and effective action is impossible. Making sense of causal relationships is thus a prerequisite for being an effective practitioner.

In addition, researchers and health professionals might be interested in more general examples of effectiveness — such as whether particular organizational configurations function well, or whether certain post-qualification educational experiences are worthwhile. Understanding causal linkages is an essential skill here also.

Health-care professionals are largely preoccupied with two specific sorts of causal relationships: those that inform the progression of health problems (disease aetiology), and those that inform their remedy (therapeutic efficacy). This article will be concerned with exploring these specific causal relationships, but the comments made are germane to the interpretation of all types of causal linkages.

## WHAT IS CAUSATION?

At its simplest, a cause is something that brings about an effect. But this simplicity hides a variety of different forms of causal relationship. Four basic types can be identified: necessary causes, sufficient causes, necessary and sufficient causes, and neither necessary nor sufficient causes (Elwood, 1992).

Necessary causes are those which must be present for a given effect to be observed. For example, infection with herpes zoster is a necessary cause of the chronic condition post-herpetic neuralgia. However, it is not a sufficient cause, as not everyone with herpes zoster infection will go on to develop post-herpetic neuralgia.

Sufficient causes are conceptually simple: a given outcome is certain to follow a sufficient cause. However, in health care, non-trivial examples of sufficient causes are unusual: even the most dangerous pathogens spare a few lucky individuals; even the most effective remedies are not infallible.

If sufficient causes are rare, then it is even more unusual in health care for causes to be both sufficient and necessary, that is, where the cause always leads to the effect and the effect never occurs without the presence of the cause. Yet, erroneously, this is how people sometimes conceive of causal relationships — as fixed, deterministic and invariant.

Far more usual is for causes to be neither sufficient nor necessary. That is, the cause may be present but the effects may be absent, or the outcomes (we can hardly call them 'effects' in this case) may be seen in the absence of the cause. How then is this a 'causal relationship' at all? The answer lies in whether or not the application of the cause increases the likelihood of the effect, i.e. whether there is a statistical relationship (usually called an 'association') between the presence of the cause and the presence of the outcome.

Herein lies the source of much confusion: the presence of a statistical relationship is neither necessary nor sufficient to impute a causal relationship. An association between two variables does not establish that these variables are causally related. Also (but more rarely) two variables may truly be causally related but a (statistical) association may not readily be apparent.

## ASSOCIATION IS NOT CAUSATION

'Association is not causation' could be tattooed over the heart of many an epidemiologist. Health professionals could also benefit, not just from knowing this mantra, but also from truly understanding its implications. Most of the important causal relationships in health (the causes of diseases and the effectiveness of treatments) are far from deterministic.

Identifying links between cause and effect thus usually consists first of identifying an association* and then trying to figure out whether this association reflects a deeper causality. But what is the difference between association and causation?

**Dr Huw TO Davies** is Lecturer in Health Services Research and Health Care Management, Department of Management, University of St Andrews, St Andrews, Fife KY16 9AL and **Dr Fiona LR Williams** is Lecturer in Epidemiology in the Department of Epidemiology and Public Health, University of Dundee, Ninewells Hospital and Medical School, Dundee

*Correspondence to: Dr HTO Davies*

*Sometimes a causal relationship may exist but the association between cause and effect may not readily be apparent. This can happen for the same reason that an association may not reflect causality: namely confounding. For example, a new effective treatment given only to sicker patients may appear no better than standard therapy simply because it is used only on sicker individuals.

An association between cause and effect (or outcome) exists if the effect (or outcome) is more common when the putative cause is present than when it is not. Thus an association is a statistical relationship between two variables. If manipulation of the application of the putative cause changes the subsequent frequency of outcome, then that association is deemed causal. An example illustrates the point.

People who carry matches are at increased risk of developing lung cancer. This is a statistical association. However, it would be perverse to say that carrying matches causes lung cancer, as any changes to match-carrying habits alone would leave the likelihood of contracting cancer unchanged. People who smoke are also at increased risk of developing lung cancer — yet going beyond this association to call the relationship 'causal' seems warranted, not least because of the evidence that people who change their smoking habits also change their subsequent health risks (Royal College of Physicians, 1971).

## CONFOUNDED CONFOUNDING

Assessing the relationship between two variables is complicated. The variables can appear related not because of any causal link but because each is separately related to a third factor. This is called confounding. Consider the example above — people who carry matches are at increased risk of developing lung cancer. Yet carrying matches and lung cancer only appear to be closely related (a statistical association) because each is closely linked to a third factor: smoking itself. In this case smoking is called a confounding variable.

A second example is the paradoxical finding that people who can swim are more likely to drown than those who cannot. Such an association between swimming ability and subsequent drowning seems to throw in doubt the value of swimming lessons. However, a moment's reflection suggests that perhaps those who cannot swim steer clear of water in general and water sports in particular. Thus the apparent relationship (a real statistical association) may be less important than at first sight — a possible explanation is variation in a third variable 'propensity to go near deep water'. Many apparent relationships (associations) may diminish in interest and importance when the full range of confounding factors is considered.

The examples given were chosen precisely because any causal relationships are implausible and the confounders are obvious. They thus provide easy illustrations of confounding. In practice, confounders are rarely so conspicuous. Sometimes the postulated relationship may be superficially plausible or even attractive — and an uncritical or inexperienced reader may fail to consider alternative explanations.

A further difficulty is that descriptions of empirical associations are rarely provided in a neutral manner. For example, reports of associations between diet and health may assert that 'consumption of produce X reduces disease Y by 20%'; reports of treatment success might similarly claim that 'patients having the new treatment have fewer side effects and better outcomes'. Such phrasing (for example, using the present tense) may encourage assumptions of causality when the data themselves confirm only past associations. Whether causality can be inferred is an altogether trickier question.

## CRITERIA FOR CAUSALITY

Deciding what is causal and what is not has a long history in epidemiology. Over 100 years ago Koch described a set of 'postulates' for determining whether an infectious agent was the cause of a malady (Koch, 1880, 1886). These helped greatly in unravelling the aetiology of infectious disease.

Of wider applicability are the criteria identified in the 1960s by the celebrated epidemiologist Sir Austin Bradford Hill (Hill, 1965). These outline a number of considerations for those attempting to identify relationships as causal. Although the criteria have been expanded and at times criticized (Elwood, 1992; Weed, 1997), they nonetheless form a useful framework for thinking about the possible interpretation of empirical associations.

### Temporal relationship

That causes should precede effects is axiomatic. When the evidence for any association comes from prospective studies (prospective cohorts or trials for example), such a temporal relationship may be obvious. However, when the data come from retrospective cohorts, case-control studies, cross-sectional studies or case reports then this relationship may not be so apparent. A lack of the correct time sequence (causes preceding effects) may be enough to reject outright any notion of causality.

### Strength

The strength of any association between putative cause and effect (or outcome) is sometimes seen as an indication of the likelihood of causality. Large effect sizes (say three or more for odds ratios or relative risks; Davies, 1998a; Davies et al, 1998) do not of themselves guarantee causality but they do mean that any confounding relationship or bias has to be at least as big. Such large biases would, one would hope, be relatively easy to uncover.

### Dose–response relationship

A more sophisticated judgment of the strength of any association is whether or not the relationship becomes stronger as more of the putative cause is applied. A classic example is the relationship between carboxyhaemoglobin in the blood and subsequent health outcomes (Beaglehole et al, 1993). Low levels of carboxyhaemoglobin (say 20%) are associated with mild headache. However, the outcomes become progressively more severe as the percentage increases. At levels of 50% nausea and blackouts are experienced, at 70% unconsciousness ensues, over 80% and death is the most likely result.

Although the presence of a dose–response relationship is often

taken as supportive of a causal relationship there are definite reasons to be cautious. For example, the presence of a powerful confounder that has a strong dose–response relationship with the putative cause may produce an apparent dose–response relationship between that cause and the outcome of interest.

Crucially, the absence of a dose–response relationship does not prove the absence of any causal relationship. For example, the cause may only operate above a certain threshold, having a step-wise relationship with the outcome rather than a gradually increasing effect. Alternatively, the association may be U-shaped or J-shaped, such as that between alcohol consumption and cardiovascular disease (Shaper et al, 1988; Kemm, 1993). Here there is a dose–response relationship, but only within certain ranges of the cause (alcohol consumption).

**Plausibility**

That a causal relationship should be plausible is intuitively sensible, but plausibility is limited by the contemporaneous knowledge base. For instance, in the 19th century, when Oliver Wendell Holmes first proposed that contagion was responsible for puerperal sepsis many sceptics dismissed this as simply implausible (Holmes, 1842–3). That Holmes was right should serve as a stark warning of the limitations of this requirement. Nonetheless, when coming to some judgment as to whether an observed statistical association reflects an underlying causality, there is no doubt that plausibility is highly influential.

Judgments of plausibility reflect an assessment about how reasonable the findings of an association are in the light of what else is known, especially about biological mechanisms. Thus postulating a causal relationship between smoking and lung cancer gains credence because of known carcinogens in tobacco smoke that are in contact with lung tissue.

Equally important, however, is the fact that plausibility may provide erroneous support for specious relationships. This is particularly true when it comes to assessing treatments. There are many examples of therapies once thought efficacious but now recognized as of doubtful value (Crombie and Davies, 1996): the surface plausibility of treatments such as radical mastectomy for breast cancer, and neuroablation for chronic pain wrongly convinced many of their efficacy. Of course, as critics of Oliver Wendell Holmes discovered, a lack of plausibility may also be shaky grounds for rejecting causality.

**Consistency**

Any substantial causal relationship (factors causing disease, or treatments providing relief) might be expected to have consistency across different individuals and places. When associations are seen only in specific subgroups (defined by age, gender, ethnicity, disease severity or whatever), and there is no substantive biological rationale for the different effects in the subgroups, then we would quite reasonably suspect that the relationship was artefactual.

Again, this criterion provides no hard-and-fast guidance, but a consistent relationship between a cause and an effect seen in different studies conducted on different study populations at different times is reasonably seen as supportive of the belief that the relationship is indeed causal.

**Reversibility**

A crucial aspect of any relationship between a putative cause and a supposed effect is that manipulation of the cause is followed by changes in the frequency of the effect. Therefore evidence that removing a cause is associated with a diminution of the effect can be taken as supportive that the relationship is indeed causal.

This criterion is again not without problems, once more because of confounding. It is feasible that, unless there is true control over which individuals avoid the putative cause, that those choosing to avoid it would in any case be less susceptible to the outcome.

**Specificity**

A final criterion is that of specificity: that is, the cause should produce specific rather than generalized effects. This is relevant to infectious diseases where, for example, the cholera vibrio produces only cholera. However, the concept is not particularly helpful when considering non-infectious diseases where disease causation is mostly multi-factorial and when causes often produce an array of different health outcomes. Smoking, for example can cause cancers of the lung, larynx and nose, as well as contributing to cardiovascular disease (Wynder, 1988).

**Guidance not guarantees**

Given the number of caveats associated with each of the criteria it is clear that although these issues are frequently headlined 'causal criteria' they serve much more usefully as a framework for thinking about the extent of the evidence. None of the above provide incontrovertible support for causality, although the first (the temporal relationship) may allow its rejection. Their importance lies in their usefulness as a structure for both research enquiry and study interpretation.

## CHANCE VARIABILITY

The findings from all studies are subject to the play of chance. That is, any observed association may be a spurious finding due simply to random variability. A substantial real relationship may appear weak or even non-existent for the same reason.

Assessment of either of these two possibilities can only occur once spurious systematic reasons for the relationship (biases) have been discounted (Brennan and Croft, 1994). At the very least, reported measures of association (usually relative risk, odds ratio or number needed to treat; Davies, 1998a) should be surrounded by a confidence interval to aid interpretation (Davies, 1998b).

## SOURCES OF EVIDENCE

Evidence for causal relationships come from a range of different study designs — each with its own strengths

and weaknesses. The first crucial difference between designs is whether the investigator had experimental control over which individuals were exposed to the putative cause. Exposing people to harmful causes is unethical so this type of study is limited to assessing the benefits of interventions. Randomized controlled trials provide the strongest evidence in support of causal relationships, primarily because the randomization controls for both known and unknown confounding variables (Cassens, 1992).

When the investigator is unable to control who gets exposed to the putative cause and who does not, the study is said to be observational in design. Observational studies provide weaker evidence than experimental studies but are, nonetheless, the only ethical approach when assessing the causative agents of disease. In all observational studies the presence of confounding is the biggest bugbear. Even within the observational framework different study designs are available and again they have different strengths and weaknesses.

Prospective cohort studies tend to provide more robust evidence than either retrospective cohort studies or case-control designs (Abramson, 1984). Simple cross-sectional studies (surveys) are particularly weak at assessing causality, as they may be unable to clarify the temporal relationship between different variables.

Any assessment of causality needs to take account of the inherent weaknesses of the design of the empirical studies that suggest such a relationship. In addition, the actual execution of any study may be flawed, introducing further possibilities of bias. Critical appraisal of study quality should be added to the criteria outlined above. Established tools for carrying out such an appraisal, specific to the different study designs, can ease this task (Sackett et al, 1991, 1997; Crombie, 1996).

## CONCLUSIONS

Statistical associations abound but may not reflect meaningful causal relationships. The presence of confounding should always be suspected and creative reflection on any reported association may suggest many plausible confounding variables. Bradford Hill's criteria for assessing causality offer a useful framework for thinking about the presence of causality but they provide no firm guarantees. Also useful are a number of critical appraisal checklists (Sackett et al, 1991, 1997; Crombie, 1996) for assessing the various study types that may lead to claims of causal linkages. HM


Abramson JH (1984) *Survey Methods in Community Medicine: An Introduction to Epidemiological and Evaluative Studies.* Churchill Livingstone, Edinburgh

Beaglehole R, Bonita R, Kjellstrom T (1993) *Basic Epidemiology.* World Health Organization, Geneva: 126

Brennan P, Croft P (1994) Interpreting the results of observational research: chance is not such a fine thing. *Br Med J* **309:** 727–30

Cassens BJ (1992) *Preventive Medicine and Public Health.* 2nd edn. National Medical Series. Williams and Wilkins, Baltimore

Crombie IK (1996) *The Pocket Guide to Critical Appraisal.* BMJ Publishing, London

Crombie IK, Davies HTO (1996) *Research in Health Care: Design Conduct and Interpretation of Health Services Research.* John Wiley and Sons, Chichester

Davies HTO (1998a) Interpreting measures of treatment effect. *Hosp Med* **59:** 499–501

Davies HTO (1998b) Assessing chance variability in treatment trials. *Hosp Med* **59:** 650–2

Davies HTO, Crombie IK, Tavakoli M (1998) When can odds ratios mislead? *Br Med J* **316:** 689–91

Elwood JM (1992) *Causal Relationships in Medicine. A Practical System for Critical Appraisal.* Oxford Medical Publications, Oxford

Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine: Section of Occupational Medicine. Royal Society of Medicine, London

Holmes OW (1842–3) The contagiousness of puerperal fever. *N Engl J Med Surg* **1:** 503–30

Kemm J (1993) Alcohol and heart disease: the implications of the U-shaped curve. *Br Med J* **307:** 1373–4

Koch R (1880) *The Aetiology of Traumatic Infective Diseases.* Translated by W Watson Cheyne. New Sydenham Society, London: 21

Koch R (1886) *The Aetiology of Tuberculosis.* Translated by S Boyd. In: Watson Cheyne W, ed. *Microparasites in Disease.* New Sydenham Society, London: 65

Royal College of Physicians (1971) *Smoking and Health Now.* Pitman Medical and Scientific Publishing Company Limited, London

Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Little, Brown and Company, Boston, Massachusetts

Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1997) *Evidence Based Medicine: How to Practice and Teach EBM.* Churchill Livingstone, London

Shaper AG, Wannamethee G, Walker M (1988) Alcohol and mortality in British men: explaining the U-shaped curve. *Lancet* **ii:** 1267–73

Weed DL (1997) On the use of causal criteria. *Int J Epidemiol* **26:** 1137–41

Wynder EL (1988) Tobacco and health: a review of the history and suggestions for public health policy. *Public Health Rep* **103:** 8–18


### KEY POINTS

- Association is not causation: a statistical association should not be thought of as a causal relationship without considering whether or not manipulation of the putative cause will lead to changes in the frequency of the specified effect.
- Confounding occurs when any two variables appear to be related because each is separately related to a third (the confounding variable).
- Confounding can hide a true causal relationship or can suggest one when none is present.
- Consideration of a number of criteria can help establish the presence of causality. Foremost among these are that the putative cause should precede the effect, the supposed mechanism should have some external biological plausibility, the relationship between the putative cause and supposed effect should exhibit a 'dose–response' relationship (more of the cause leading to more of the effect), and the relationship should be consistent across study groups.
- Causality criteria such as those of Bradford Hill offer a framework for thinking about causality. They do not provide guarantees about its presence or absence.
- Assessment of any biases such as confounding needs to take place before the assessment of the play of chance.
- Different study designs have different strengths in assessing causality. Experimental control can remove problems of confounding which are always present in observational studies. Even so, all studies may be flawed in design, execution and analysis and thus should be subjected to rigorous critical appraisal.