Preventive Medicine 53 (2011) 246–249

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

The gap between evidence discovery and actual causal relationships

Michael Joffe

Imperial College London, UK

ARTICLE INFO

Available online 3 August 2011

Keywords:
Causal relationships
Causation
Epidemiological methods
Mechanism
Difference-making
Stochastic causation
Multiple causation
Regularity
Variation

ABSTRACT

The concept of causation in epidemiology can be illuminated by situating the discussion within a more general concept of causation in biology: "a causal relationship is one that has a *mechanism* that by its operation *makes a difference*". Mechanism and difference-making are complementary, and discovery can proceed from either direction; each type of evidence can be qualitative or quantitative. An explanation becomes fully convincing only when supported by both. In biology, causation is typically stochastic and/or multiple. Multiple causation can be analysed statistically/epidemiologically, even though it is not truly (ontologically) stochastic. This requires some degree of *regularity in the outcome variable*, plus *sufficient variation in the exposure(s)*. The analysis then demonstrates co-variations between exposure(s) and outcome that *regularly* occur. Rose's important distinction of "causes of incidence" and "causes of cases" should be reconceptualised in terms of epidemiological *visibility*, raising the possibility of epidemiological "dark matter".

The concept of causation in biology

The literature on causation in epidemiology tends to focus on the epidemiological method itself, and what can be learnt by using it (Bhopal, 1999; Parascandola and Weed, 2001; Rothman and Greenland, 2005; Susser, 1973). It may, however, be instructive to situate the discussion within a more general concept of causation in biology. In particular, it is important to separate how knowledge is obtained (epistemology) from how we think about what actually exists (ontology).

The study of how organisms function is composed of instances in which an entity such as an enzyme, gene or neurotransmitter brings about some change in the body. Thus, the enzyme pepsin begins the conversion of dietary protein into a form in which it can be absorbed. The *SRY* gene turns a mammalian embryo from the default female sex into a male, through a cascade of effects ending with hormonal changes such as a high production of testosterone. A neurotransmitter enables the firing of one neuron (nerve cell) to be transmitted to another neuron or to a muscle. Most laboratory biological research is concerned with uncovering and confirming the existence of such causal mechanisms. Since the late nineteenth century, it has achieved considerable success in so doing.

Thus, the concept of "mechanism" is central to this major branch of biology. Philosophers of science have found it difficult to pin this concept down precisely, possibly because they have traditionally focused mainly on the study of physics, which is a particularly difficult case: classical (non-quantum) physics is largely characterised by deterministic causation, and can be expressed precisely in mathematical terms (and quantum physics raises very difficult specific issues in relation to causality). Biology, in contrast, deals successfully with a more messy type of reality — and may therefore be useful as a source of a more general model of causation, that can then be applied (*mutatis mutandis*) in other circumstances.

At first glance, epidemiology is in sharp contrast to this mechanistic focus. It studies the effect of a causal factor such as a toxin on disease rates: uranium miners are found to have a higher rate of lung cancer than non-miners from the same background. But the strength of epidemiology is that it provides evidence on differences in outcome between groups defined by different exposures, evidence that corresponds to some biological mechanism that can be elucidated by other methods — in this instance, that the miners inhale radioactive particles that act on the DNA of the bronchial cells in such a way that cancer is likely to result at a later date. In fact, epidemiologists are increasingly adapting their methods so as to investigate mechanisms. For example, the study of the likely effects of physical activity on certain cancers involves hypotheses that involve several distinct pathways, and measurements are made to assess their roles (Neilson et al., 2009).

In each type of case, the basic explanatory account depends on a mechanism or capacity (Cartwright, 1989) that brings about an alteration. In general, then, a causal relationship is one that has a *mechanism* that by its operation *makes a difference* (Joffe, 2011). "Makes a difference" here can indicate a change in the probability or timing of occurrence of an event, or in its magnitude or severity. This concept of causation is distinct from agency (human decision making), which has additional features, and is beyond the scope of this paper.

An implication is that the alteration brought about by the mechanism occurs *over time*. This is distinct from non-causal differences that *exist* between categories of background variables, e.g. the rates of a disease may differ between males and females, but this

E-mail address: m.joffe@imperial.ac.uk.

0091-7435/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ypmed.2011.07.011

does not amount to difference-*making*. For example, the higher rate of breast cancer in women than men can be traced to metabolic differences between the two sexes (for example high endogenous estrogens in females), which do play a causal role over time. The observed sex difference is due to differences between processes in the two sexes that are themselves causal. In addition, a series of additional background conditions is typically required, e.g. acidity for pepsin, functioning receptors for testosterone to bring about masculinisation, and a receptive neuron or muscle fibre for a neurotransmitter.

Mechanistic accounts generally find it hard to deal with "negative" cases, i.e. where an effect is caused by the absence of something. But when applied to biology, the concept "a causal relationship is one that has a *mechanism* that by its operation *makes a difference*", can deal with this. For example, the pigment melanin is normally produced in the skin and eyes as the result of a group of genes, but if one of these is inherited in an altered form so that the gene no longer functions properly, the offspring is an albino. The causation of pigmentation in a normal offspring is that these genes initiate a cascade of processes (the mechanism) that end with the production of melanin (the difference); if the mechanism is absent or non-functioning, the difference is not brought about, so there is no melanin.

Deficiency diseases provide another type of example. The human body has evolved in such a way that it requires vitamin D, which can be obtained from sunlight or from the diet. It is involved in pathways (the mechanism) that promote the healthy mineralisation, growth and remodelling of bone (the difference). If the quantity of vitamin D is insufficient, then rickets or osteomalacia results, respectively in children and adults. Thus, if the pathways are inadequately supplied with the vitamin (deficient mechanism) then the bones become abnormal (the difference is now insufficient, and falls short of healthy development).

Within this overall concept, there are some special cases: some instances of causal relationships involve necessary and/or sufficient causes, whilst others involve multiple causation or stochastic causation — this depends on the system being studied, and has to be judged on a case-by-case basis. They have different methodological implications. Multiple and/or stochastic causation is typical of biology, and is probably the most frequent situation found in other disciples apart from classical physics. Biology therefore provides a rich set of examples for methodological investigation that has wider implications.

This concept has been derived from consideration of the structure of explanations in two different branches of biology, and provides a characterisation of the real world (ontology) from a biological viewpoint. It is compatible with the classic discussion of causal inference by Hill (1965); see also Joffe (2011). However, it only specifies a general structure; it has little or no content. Knowledge of *specific* causal relationships is obscure without evidence of different kinds. This is the epistemological question: how scientific method succeeds in obtaining reliable knowledge.

The two basic types of evidence on causal relationships

If a causal relationship is one that has a *mechanism* that by its operation *makes a difference*, then there are in principle two ways of acquiring evidence about it: understanding the mechanism, and seeing what difference it makes — "difference-making". They are widely acknowledged to be distinct (Cartwright, 1989; Pearl, 2000, 2002).

The process of evidence gathering can start from either direction. If mechanistic understanding comes first, the question arises whether the mechanism is actually operating in practice. For example, animal experiments have shown that anti-androgens that interfere with testosterone synthesis can impair male development *in utero* (Fisher et al., 2003). This has been suggested as having caused the large rise in testicular cancer in some populations during the twentieth century. However, the health effects in the animals are small unless exposures are used that are far higher than real human exposures; furthermore, exposures to these substances began too late to explain the start of the rise in cancer (Joffe, 2010). The mechanism is plausible but the difference-making evidence shows that something else must be causing (most of) the observed trend.

Conversely, difference-making evidence on its own is a "black box", leaving the unanswered question of how the difference is brought about. A classic case is cigarette smoking and ischemic heart disease: the observation of an approximately two-fold increased risk was at first regarded as biologically implausible. However, further research succeeded in establishing the biochemical causal pathways.

Mechanistic and difference-making types of evidence have complementary roles. To be fully convincing, a causal explanation needs to be supported by both aspects, and this affects even how initial findings are interpreted. A stable association needs to have some *prima facie* plausibility: nicotine-stained fingers would be unlikely as a cause of lung cancer! Similarly, a suggested mechanism has to correspond to some difference observable in the real world. Mechanistic evidence on the tendency of hot tarry smoke to cause lung tumours when inhaled does not necessarily mean that in practice cigarette smoking will have this effect. For example there could be insufficient quantities of the chemicals involved, so complementary evidence of the difference-making type is indispensable. Furthermore, difference-making evidence is also necessary to indicate the size of the resulting effect, both to calibrate difference-making with mechanism, and to check whether that particular mechanism is a sufficient explanation of the effect or whether it only makes a minor contribution.

Mathematics, determinism and causation

In biology, difference-making evidence is often mathematical (see examples below). But it may also be qualitative: if the motor nerve to a muscle is cut or damaged, the muscle fibre can no longer contract in response; and if it is intact then stimulating the nerve causes muscular contraction. Similarly, the evidence for a mechanism can be quantitative or qualitative. The chemical nature of neuromuscular transmission was established by measuring the delay between the firing of the nerve and muscle fibres, a quantitative step. But the subsequent steps in this story were established by non-quantitative means: visualisation of vesicles in nerve endings by electron microscopy, analysing them chemically, finding that this substance specifically stimulates muscle fibres, and discovery of specific receptors for it on the muscle fibre surface. Thus, it is not generally true that "science is measurement" (Cartwright, 1989) — it may also be qualitative observation.

In addition, there is a reason to query the role of mathematics as the language of causal relationships *in principle*. Mathematics is expressed by equations, and these are symmetrical around the equals sign. In contrast, causation is inherently asymmetrical, as is clear from the example of the neuromuscular junction (although it is true that a subset of causal processes is reversible, e.g. some chemical reactions). There is thus a potential mismatch of mathematics with causality. In epidemiology, it is sensible to talk about lung cancer incidence as a function of smoking, both causally and mathematically. But whereas it could be possible to write an equation of smoking as a function of lung cancer, this would have no causal meaning. Equations can represent non-causal as well as causal relationships.

The view put forward in the previous paragraph is not generalisable across all of science. There is a major exception: deterministic causation, which characterises much of non-quantum physics. In this special case, the system is accurately predictable both backwards and forwards. For example, Halley described the periodicity of "his" comet in 1705, and by extrapolating back it can easily be calculated that it would have been visible also in 1066, thus explaining