REVIEW

What characterises a useful concept of causation in epidemiology?

J Olsen

It has recently been suggested that epidemiologists should avoid thinking of causes in deterministic terms. This would mean giving up the component-cause model in its original form. A model that has provided important contributions as to how we develop hypotheses, design our study, analyse data, interpret and communicate our results. The component-cause model has considerably more to offer than a simple probabilistic concept. What a causal model has to offer to the advancement of the discipline is equally important as the concept itself. It has been said that we should not hunt "the Holy Grail" (that is, determinism), if it does not exist. This line of reasoning neglects the fact that the "hunting" is more important than the "finding".

Hume defines "a cause to be an object followed by another and where all the objects similar to the first are followed by objects similar to the second."¹ In this "strong" concept, a cause is always followed by an effect. In epidemiology, effects seldom appear immediately after an exposure. When studying social, genetic, or environmental determinants of diseases, we usually have to accept an induction and latency time. Few, if any, causes are always followed by an effect—the disease.

The main concern lies in the deterministic approach to causation, which is considered to be an outdated principle.^{2,3} In two recent papers,^{3,4} the authors came to the conclusion that the causes we study in epidemiology are probabilities, and nothing more than that. Probabilities are not just euphemisms for ignorance. They are the real thing.

Determinism is completely—or almost completely—devoid of any empirical support in epidemiology. "Why hunt the Holy Grail when it does not exist,"³ wrote Karhausen believing for a while he was Socrates. Parascandola and Weed⁴ supported this point of view. Few will most probably defend a simplistic deterministic viewpoint, although this extreme was advocated by Stehbens,⁵ who resolves, "The inescapable conclusion is that the notion of cause as the role prerequisite, without which the disease cannot occur, should be retained."

Is the probability theory the official language of epidemiology or is it biology or sociology? Is it time to abandon one of the axioms of epidemiology; that diseases have causes. Should the axiom now be that diseases often, or sometimes or occa-

sionally have causes? Is chance a legitimate cause of a disease?

Hume also used a second definition, stating "...or, in other words, where the first object had not been the second would never exist." This is an entirely different definition that identifies a necessary cause in the "strong sense". Also rarely seen in epidemiology, unless we are talking about diseases where we define the disease by including the exposure in its definition.

The second definition serves as a model for counterfactual reasoning when designing our studies. We want to estimate disease occurrence among the exposed, had they not been exposed. For an exposure to be a cause, it must be true for at least one exposed, that he or she would not get the disease in question at the time he or she did, had he or she not been exposed. In our empirical estimate, this needs not be true. If sampling from the population by chance includes no susceptible, then the cause-effect relation will not manifest itself. The empirical estimates are subject to random variation that justifies our use of statistics as the mathematical language of epidemiologists. But it is not the only "language" we should use. Although the counterfactual reasoning is useful, it is not a concept of causation, rather a causality criterion. Even the logics associated with the counterfactual perform best under determinism.

It is a point of fact that diseases have more than one cause and some—perhaps many—of the causes are unknown. Therefore, the causal links we observe appear as statistical associations. Whether the association we observe is causal or not, depends on our ability to control for confounding and bias, and we may apply a set of causal criteria to make sure that we are often in the right when we claim to have identified a cause of a disease. Many will be happy with this and claim that we need no more to conduct our business. Some are reluctant in going further because we will then be trespassing on philosophers' ground. I believe this to be a mistake.

Our concept of causation should provide support for our research, which is our business. We may take a pragmatic view when deciding whether we regard a concept of causation as useful or not in our present conduct of affairs. To decide on the usefulness, the following criteria (and others) could be applied:

- 1 the concept should make sense—that is, be communicable to the users of epidemiology;
- 2 the concept should be able to explain empirical results;
- 3 the concept should open up for new avenues of research;
- 4 and finally, the concept should provide as few anomalies as possible.

Correspondence to:
Dr J Olsen, The Danish
Epidemiology Science
Centre, University of
Aarhus, Vennelyst
Boulevard 6, DK-8000
Aarhus C, Denmark;
jo@soci.au.dk

PROBABILISTIC OR DETERMINISTIC; THE TWO CONCEPTS UNDER DEBATE

Basically, only two main concepts are at present applied in epidemiology. A purely statistical concept, which states that a cause increases the probability of the disease if it is present, and the component-cause model first developed by Mackie⁶ and independently applied to epidemiology by Rothman.⁷ According to this model, component causes act together in causal fields to produce an effect. Causes are only sufficient given the other component causes are in place, and causes are only necessary for the disease caused by the causal field in question. Mackie calls these causes “weak” in opposition to Hume’s “strong” sufficient and necessary causes. His model gives meaning to the term “weak”. These causes are not globally necessary or sufficient. They are only necessary or sufficient in a certain setting, for a certain well defined subset of the population under study.

This second model is deterministic. It states that diseases have causes and if these causes are present in proper time windows, diseases will follow. Some of the component causes, or all of them, may, however, occur in an unpredictable manner. The model does not claim that predictive medicine is possible, not even as a principle.⁸ The model simply claims that if you have a disease, it had causes. If you analyse, for example, a traffic accident, this accident had its component causes like bad weather, a driver using a mobile phone, a vehicle that suddenly slowed down because of rain, etc. Some of these component causes are elements in other causal sequences, which may have to coexist in a certain time specific period in order for an effect (a disease) to occur. The destiny of these paths could be unpredictable. The model does, however, state that if we have identified all causal elements and if they all happen again in exactly the same order, the accident would occur once more. This regularity principle has its difficulties, in theory as well as in practice. We operate with complex biological systems, with large variation in how exposures are taken up, metabolised, distributed, and eliminated. The process is furthermore subject to differences in defence and repair mechanisms.

THE CONCEPT AND ITS OPERATIONAL CONSEQUENCES

The concept should be communicable: At present we tell a smoker that he will increase his risk of getting lung cancer 10-fold by smoking. If he gets lung cancer from smoking, it will take decades to develop, and he may even get lung cancer, should he decide not to smoke at all. Experience shows that this message is in conflict with a common sense understanding of causation, and it is apparently not very convincing. Using the component-cause model may help, as we all know that one component cause depends upon other component causes in most daily life activities. When we turn the car key, the engine will only start if it has access to fuel, the cables are intact, the battery is functioning, the engine is in order, etc. In like manner, only the susceptible get lung cancer if they smoke. Only those who have or have had all other component causes in the right sequence.

The model also provides an explanation for why a risk factor is strong or weak. The strength of the association depends upon the prevalence of the other component causes. Just like a new car is much more likely to start when we turn the key, than an old car found abandoned at the roadside.

The probabilistic model has nothing really to offer the reluctant reader of epidemiological literature. The component-cause model is in a way an attempt to explain the probabilistic model by providing reasons for its empirical behaviour. Furthermore, the component-cause model provides an explanation for the time lag between exposure and disease. Rothman⁹ calls the first part of the time period the induction

period. It is the time from the start of exposure and until the last component cause is in place. Then the disease process starts, and the time until the disease surfaces to clinical detection is called the latency period. The probability model is much less informative on the time from exposure to effect.

Linked to the ability to communicate information on causes to the users of epidemiological results is *the ability to explain empirical results*. The probabilistic model is a minimalistic concept where causes increase the probability of diseases. Preventable factors decrease the probability. What you see is what you get.

The component-cause model attempts to explain why this happens, and for some diseases we have the evidence to fit the model. In most cases, however, we have limited evidence to support a component-cause model except at least one component cause is of a probabilistic nature. So far, it is a postulate that most diseases have unknown causes that would fit into the component-cause model. The component-cause model tells us they are there—therefore we look for them.

An important criterion is which concept provides *most new ideas for research*, and the component-cause model clearly takes the lead. We have not understood the causes of diseases before we have described all causal fields and the component causes in the causal fields. A component-cause model provides inspiration for many generations to come.

The concept also plays a part in *how we design, analyse, and present data*. Multiplicative and additive models do correspond under a number of conditions to the component-cause model.¹⁰ A probabilistic model has no prior prediction to offer.

The component-cause model gave inspiration to, for example, the computerised square dance design.¹¹ In this design, two cohorts are established, one who experienced the event under study and one who did not, then the effect of changing putative determinants before the next event is examined.

A useful concept of causation should not lead to many *anomalies* that are counterproductive for achievements in research and communication. The probabilistic concept carries few anomalies because its content is limited. That is not the case for the component-cause model, which says more and therefore risks more. It is, however, not in conflict with the component-cause model that causes manifest themselves as associations with a probabilistic pattern as long as only part of the causal fields are accounted for—but there are other problems.

The component-cause model states that the strength of the association depends upon the prevalence of the other component causes in the relevant causal fields. How could we then explain dose-response relations? As the cause is neither present nor absent in the causal field, an increasing response with an increasing dose must lead to the conclusion that it now operates in a different causal field with a different set of component causes. This is not an attractive thought. The problem lies not in understanding a dose-response effect for a dose measured over time, but for a dose measured by its intensity. A cumulative dose measured over time is expected to correlate with the incidence of the disease, regardless of the prevalence of other component causes. If the other component causes fluctuate in time and must be present together with the exposure under study, long term exposure carries a higher risk than short-term exposure. This reasoning is similar to stating that the risk of traffic accidents increases with the number of miles spent on the road, as this will correlate with the risk of meeting the component causes leading to the accident, at least if driving routines do not compensate for the accumulated exposure. Speed is a measure of intensity that could increase the probability that other component causes fall into place. But in many other situations, the role of the intensity of the exposure is not developed within the component-cause model. We could suggest a new causal part leading from the external exposure to the internal biological dose modified by metabolism, distribution, and elimination. If we talk about a

dose intensity effect measure for external sources, we may have to accept a two stage component cause model or a multi-stage model, operating at different levels of causation.¹²

The component-cause model is a model of multifactorial causes, and at least four component causes are needed if none of them are necessary or sufficient in the strong sense. It takes at least four component causes to form two causal fields, each with two component causes.

As the component-cause model states that the strength of association depends upon the prevalence of other component causes, we have little reason to expect many exposures to have the same quantitative effects in different populations. Meta-analysis then becomes an exercise to examine heterogeneity in effect measures, rather than to make combined risk estimates using the entire sample size to estimate its confidence limits. Combining results into one general estimate is a rather naive idea under the component-cause model.

CAUSES AT DIFFERENT LEVELS

The component-cause model provides a measure of aetiological fractions that leads to the fact that aetiological fractions do sum up to more than 100%.¹³ This should not be seen as an anomaly, but a point of fact that is easily explained by the model.

The terms distal and proximal determinants have no obvious significance in the component-cause model unless component causes need a given time sequence to operate and the terms apply to time from exposure to disease. Most causal fields could, however, be perceived as layers of causal fields where one causal field leads to an exposure that is part of a different causal field. Proximal determinants are close to biological causal fields in this model. Distal determinants operate on the causation to exposures rather than the direct effects. Proximal determinants are in causal fields that lead to the disease, and distal determinants in causal fields that lead to exposures of relevance for the pathogenesis. A probabilistic model may similarly be developed to cover several steps in the causal process.

Accepting the component-cause model also leads directly to a measure of the proportion of susceptibility to the exposure in the population.¹⁴

A probability concept leads to “black box” epidemiology. The component-cause model attempts to open this black box.

CONCLUSION

The ontology of disease causation is a matter for philosophers, but the consequences of using different models in studying the aetiology of diseases should be subject to discussion—also among epidemiologists.

How we choose to view causation in epidemiology plays a part in how we teach, do research, and evaluate and comment our research findings. Rothman describes, for example, that he uses the concept of causation throughout his teaching of epidemiology.¹⁵ We should welcome a debate of which concept of causation will serve us best and take an active part in it.

The deterministic component-cause model has served us well in risk communication, in the interpretation of empirical results, and in setting up new hypotheses. It has consequences that we perhaps are not willing to accept at present, the deterministic aspect being one of them. Going back to a minimalistic “black box” model may, however, be too conservative. Statistics may be the mathematical language of parts of epidemiology, but epidemiology is more than statistics. It is a health science with a touch of demography, sociology, and biostatistics.

The best concept of causation is the concept that provides the most interesting and useful results. The concept should of course make sense—also for the users of epidemiological findings.

ACKNOWLEDGEMENT

The activities of the Danish Epidemiology Science Centre are funded by a grant from the Danish National Research Foundation.

REFERENCES

- 1 Hume D. *An enquiry concerning human understanding*. Oxford: Oxford University Press, 1974.
- 2 Karhausen LR. Causation: The elusive grail of epidemiology. *Med Health Care Philos* 2000;**3**:59–67.
- 3 Karhausen LR. Commentary: coda—a Socratic dialogue: Plato. *Int J Epidemiol* 2001;**30**:710–11.
- 4 Parascandola M, Weed DL. Causation in epidemiology. *J Epidemiol Community Health* 2001;**55**:905–12.
- 5 Stehens WE. The concept of cause in disease. *J Chron Dis* 1985;**38**:947–50.
- 6 Mackie JL. *The cement of the universe. A study of causation*. Oxford: University Press, 1974:340.
- 7 Rothman KJ. Causes. *Am J Epidemiol* 1976;**104**:587–92.
- 8 Olsen J. Some consequences of adopting a conditional deterministic causal model in epidemiology. *Eur J Public Health* 1993;**3**:204–9.
- 9 Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986.
- 10 Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol* 1980;**112**:467–70.
- 11 Olsen J, Schmidt MM, Christensen K. Evaluation of nature-nurture impact on reproductive health using half-siblings. *Epidemiology* 1997;**8**:6–11.
- 12 Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology* 2001;**12**:114–22.
- 13 Olsen J. Causes and Prevention. *Scand J Soc Med* 1991;**19**:1–6.
- 14 Khoury MJ, Flanders WD, Greenland S, et al. On the measurement of susceptibility in epidemiologic studies. *Am J Epidemiol* 1989;**129**:183–90.
- 15 Rothman KJ. A first course in epidemiologic principles and methods. In: Olsen J, Saracci R, Trichopoulos D, eds. *Teaching epidemiology*. Oxford: Oxford University Press, 2001:63–75.