# Epidemiology, risk and causation

## Conceptual and methodological issues in public health science

*Alex Broadbent*

# Background

In 2007 the PHG Foundation began funding the author of this report, Dr Alex Broadbent, to conduct research into the conceptual and methodological issues arising in connection with epidemiology. These issues include the nature of causation, methods for causal inference, the nature and communication of risk, the proper use of statistical significance testing, and the social determinants of health. The project produced a number of academic articles and included a series of workshops held at Cambridge in 2010, contributions for which form the basis of a special section of the journal *Preventive Medicine* (2011) Volume 53, issues 4-5. A book on the philosophy of epidemiology is now under contract with Palgrave Macmillan.

# Acknowledgements

This report is available from

# www.phgfoundation.org

# Contents

# Motivation

Epidemiology is the core science of public health. The biomedical sciences have attracted the attention of ethicists, political philosophers and social scientists. However, clinical medicine attracts more attention than public health, and ethical issues attract more attention than methodological and conceptual questions. These questions are the speciality of philosophers of science, epistemologists and metaphysicians. Yet to date, epistemologists and metaphysicians have not paid systematic attention to epidemiology. Nonetheless, the philosophical questions that arise in the theory and practice of epidemiology are not limited to ethics. The same is true for the translation of research findings into policy.

This project was conceived to identify and address methodological and conceptual questions arising in epidemiology, with a view to focusing attention of academics in relevant disciplines: in particular, epidemiology, philosophy and statistics.

The epidemiological literature suggests a need for work in this area. Epidemiology is a youthful scientific discipline, notwithstanding its accomplishments. Over several decades, the discussion sections of epidemiology journals have included insightful papers on topics which can only be described as philosophical: causal inference (Hill 1965), the nature of causation (Rothman 1976; Susser 1991; Rothman and Greenland 2005), the reliability of new scientific results (Ioannidis 2005), and the use of epidemiological evidence in litigation (Greenland and Robins 1988; Greenland and Robins 2000).

This tendency is not abating; indeed, it may be intensifying. Leading text books include introductory material on the philosophy of science and whole chapters on philosophical issues specific to epidemiology (notably Rothman, Greenland, and Lash 2008). As well as publishing a large number of individual papers on topics of this kind, epidemiological journals have begun setting aside space for discussion of issues such as these: In 2004 *Epidemiology* devoted space to discussion between some eminent epidemiologists and an equally eminent philosopher on the value of "risk factor epidemiology" (Greenland, Gago-Dominguez, and Castelao 2004; Haack 2004), with specific emphasis on finding a theoretical philosophical basis for causal inference in that context. The September 2008 issue of *The Journal of Public Health* contained a section of five papers whose task was "to revisit, with fresh reasoning, the whole idea of causality in the public health sciences" (Leung and Gray 2008, 217).

In January 2010 the journal *Epidemiology* again devoted a special section to a conceptual question, this time the neglected merits of absolute measures of risk, asking "why epidemiologists have come to rely almost entirely on relative measures of effect (odds ratios, risk ratios and hazard ratios), even though this practice generates considerable confusion, especially over interaction, effect modification, and the potential public health benefits associated with reported effects" (Kaufman 2010, 2).

The editors' motivations for devoting space to these matters are clearly practical, but the underlying questions are conceptual.

The fact that epidemiological journals devote space to issues of this kind demonstrates that these discussions are not considered closed by at least a substantial and important subsection of the scientific community. This project set out to identify the problems driving these discussions and provide a forum for a more sustained treatment. It also aimed to provide a forum for thoroughgoing philosophical attention to these problems.

Although philosophers have dabbled in epidemiological discussions, no philosophers to date have specialised in epidemiology. One of the aims of this project was to discover whether such a speciality made sense, both from a philosophical perspective and for the purpose of helping to clarify and make progress with the conceptual issues clouding epidemiological practice.

From the philosophical end, the timing of the project has been good. Philosophers have recently become more interested in a number of apparently disparate issues which come together in epidemiology. These include: causal inference; the use of causal modelling techniques; general causation and its relation to singular causation; the viability of "evidence hierarchies" such as those proposed under the banner "evidence-based medicine"; the nature of health and disease; and health as a property of populations (as opposed to individuals). Epidemiology provides ample material for philosophical study of these topics.

Doubt is often expressed when philosophers purport to contribute to the scientific enterprise. This project is premised on the rejection of such scepticism. Philosophers have a reputation for being better at creating problems than solving them. Again, this project is premised on the falsity of this view. Good philosophy is not always useful, but it can be. Specifically, it can be useful when it is addressed at philosophical problems arising in the course of a useful activity. Epidemiology is a useful activity, and there is ample evidence in epidemiological journals and text books that philosophical problems arise in the course of doing it.

The motivation of this project was to identify some of these problems, and to begin the process of solving them.

# Timeline

2007-8 - Postdoctoral research project, *Philosophical Issues in Epidemiology*. Postdoctoral Fellow: Dr Alex Broadbent, University of Cambridge, Department of History and Philosophy of Science. (Funding: PHG Foundation.)

2009-10 - Workshop series *Epidemiology, Risk and Genomics*, hosted at the University of Cambridge, Department of History and Philosophy of Science. (Funding: PHG Foundation.)

2011 - Report published. (Funding: PHG Foundation.)

2011 - Special section of *Preventive Medicine* published: *Epidemiology, Risk and Causation.*

2013 - Projected publication of book *The Philosophy of Epidemiology* under contract. Published by Palgrave Macmillan in the series *New Directions in the Philosophy of Science*.

# Outputs

## Publications in academic journals

Broadbent, Alex. 2008a. For Analytic Bioethics. *Clinical Ethics* **3**: 185-188.

2008b. A note on epidemiological caution.Letter to the *British Medical Journal*. www.bmj.com/cgi/eletters/335/7630/1116-b#191569.

2009. Causation and models of disease in epidemiology. *Studies in History and Philosophy of Biological and Biomedical Sciences* **40**: 302-311.

2011a. Causal inference in epidemiology: mechanisms, black boxes, and contrasts. In *Causality in the Sciences*, ed. Phyllis McKay Illari, Federica Russo, and Jon Williamson, 45-69. Oxford: Oxford University Press.

2011b. Defining neglected disease.*BioSocieties* **6** (1): 51-70.

2011c. Conceptual and methodological issues in epidemiology: An overview. *Preventive Medicine* **53**, 215-6 (special section: Epidemiology, Risk, and Causation).

2011d. What could possibly go wrong? – A heuristic for predicting population health outcomes of interventions. *Preventive Medicine* **53**, 256-9. (special section: Epidemiology, Risk, and Causation).

Epidemiological Evidence in Proof of Specific Causation, *Legal Theory*, in press.

## Special section of *Preventive Medicine*

The epidemiology and public health journal *Preventive Medicine* published a special section *Epidemiology, Risk and Causation*, guest edited by Alex Broadbent, devoted to papers arising from the 2010 workshops held in Cambridge:

Epidemiology, Risk and Causation. *Preventive Medicine* **53** (2011).

## Book

Broadbent, Alex. *The Philosophy of Epidemiology*. Under contract with Palgrave Macmillan in their series *New Directions in the Philosophy of Science*, ed. Steven French. Projected publication 2013.

# Workshops

In 2010, the PHG Foundation funded four workshops on the conceptual and methodological challenges facing epidemiology. Each workshop featured four invited speakers drawn from the fields of epidemiology, statistics and philosophy. The topics addressed at each of the workshops were:

## 1. Determinants of disease

The scope of epidemiology has been expanding in two different ways. Previously, parasite, deficiency and defect were the principle objects of study. The explosion of *genetic* knowledge since the latter part of the twentieth century has opened up a new place to look for determinants of disease, studied by *genetic epidemiologists*. At the same time, the influence of socioeconomic factors on health has attracted the attention of *social epidemiologists.*

These labels – genetic and social – are probably an oversimplification; genes and environment interact, and few epidemiologists strictly confine their attention to one or the other. Nevertheless, there is on occasion tension and even competition between genetic (or more broadly speaking biological) approaches, and social ones.

At present, genetic epidemiology tends to be relatively expensive due to genotyping costs; it makes heavier use of novel biological knowledge; and it often seeks to identify biological mechanisms. Social epidemiology is comparatively cheaper at present; it is less closely connected to developments in molecular biology or other laboratory sciences; and it is concerned with identifying "psychosocial" pathways to disease (Marmot 2006; Brunner and Marmot 2006).

With time, the relative costs of these two activities will change and may invert as the cost of genotyping comes down.

These approaches can look quite different: but are they in principle distinct? Or is there a single epidemiological ethos motivating them both? Can we generalise about which approach serves public health better? Which promises to advance understanding more?

| Speakers | Titles |
|---|---|
| **Professor Sir Michael Marmot** | Social determinants of health |
| **Professor Jonathan Wolff** | How should social policy respond to the social determinants of health? |
| **Dr Nadeem Sarwar** | Assessment of emerging strategies to predict and prevent coronary heart |
| **Professor Carlo Berzuini** | Statistical methods and causal inference |

## 2. Risk, probability and harm

There are at least two ways that risks demand our attention: by the *probability* of their materialising, and by how bad would be the *harm* if they did.

But probability and harm are different kinds of concepts, and each is a focus of controversy which working epidemiologists need to negotiate. Is probability a property of individuals (or can it be)? This view may be appealing in interpretations of quantum mechanics, where it is doubtful whether "hidden variables" can explain apparently probabilistic phenomena (Price 1996). But even if irreducible chances are the right interpretation of quantum mechanics, it is not obvious that the rationale will carry over to higher level sciences such as epidemiology, where the existence of hidden variables is beyond doubt.

Maybe, then, the appropriate understanding of probability for epidemiology is purely statistical, reducing to average frequency in a specified population. But then it is not clear what basis we have for moving from the *observed* average frequencies on which our probability estimates are based, to the *unobserved* frequencies which (in practical applications) they estimate.

The concept of harm bears more subtly but no less directly on epidemiology. For one thing, the *kind* of harm that epidemiology studies requires some thought, and is not necessarily dictated by clinical medicine.

To take the obvious example, suicide is a public health concern, but never a clinical one (even if *prevention* of suicide may be). Moreover, the range of health conditions studied by epidemiology is increasing.

The relation between harm and autonomy is also complex, and poorly articulated in the public health context. Whereas an individual smoker might be able to refuse clinical treatment for lung

cancer, the population of smokers generally cannot avoid public health initiatives on smoking. And it is not clear how, if at all, their (various) desires can or should be taken into account when counting the cost of smoking to public health.

Finally, there is a question as to whether risk is *itself* a kind of harm, so that exposing someone to an increased risk of lung cancer by passive smoking is harming them even if they do not in fact develop lung cancer. This relates in turn to legal questions about how causation is proved in toxic tort cases, where the correct presentation and interpretation of epidemiological evidence is of paramount importance.

| Speakers | Titles |
|---|---|
| **Professor Sander Greenland** | The art and (pseudo?) science of epidemiologic risk analysis |
| **Dr Mark Parascandola** | Causing harm versus causing risk |
| **Dr Stephen John** | Is a principle of saving the most at risk defensible? |
| **Professor David Spiegelhalter** | Visualising and analysing uncertainty about benefits and harms |

## 3. Explanation and intervention

The currency of contemporary epidemiology is the risk factor. But what is a risk factor? The term arises in part from the focus of epidemiology, unusually (but not uniquely) among sciences, on *public policy intervention*. This suggests interesting questions about the relation between explanation and successful intervention. How much do we need to understand before we can fruitfully intervene? Risk factors seem to offer a way to design interventions in circumstances of incomplete knowledge, either about the circumstances or the nature of the cases of illness in question. But as Jacob Henle pointed out, cataloguing the causes of disease is not sufficient for a scientific approach to medicine. It is no more scientific, Henle claims, than if a physicist were to identify as the removal of boards and beams, the cutting of ropes, the opening up of holes, and so forth as causes of falling (Henle 1844; cited in Carter 2003, 24). Identifying these "risk factors" for falling would not *by itself* bring the physicist any nearer to a theory of gravity.

This objection appears to be borne out by the subsequent history of epidemiology: the most dramatically successful interventions have tended

to be closely linked to improvements in our understanding. But is that historical claim correct? If so, it would seem that our ability to *explain why* illness occurs is linked to our ability to *cure and prevent* it. But how? Do risk factors encourage epidemiologists, and their colleagues studying functional genomics, knock-out models, and similar, to seek explanations?

Or do they allow potentially dangerous misunderstandings about the scientific and practical import of the results they represent? Is it a naïve mistake to suggest that public health interventions might be best served by seeking general explanations on the model of the physical sciences? The answers to these questions bear directly on the future direction of epidemiology and how it can appropriately employ its conceptual tools.

| Speakers | Titles |
|---|---|
| **Dr Michael Joffe** | Difference-making, discovery, mechanism and effectiveness |
| **Professor Philip Dawid** | How much do we need to understand in order to decide what to do? |
| **Professor Alfredo Morabia** | Until the lab takes it away from epidemiology |
| **Professor Alexander Bird** | Intervention in medicine and inference to the best explanation |

## 4. Causation in epidemiology

There has been a rash of epidemiological literature on the traditionally philosophical topic of the nature of causation (a small selection: Rothman 1976; Robins and Greenland 1989; Susser 1991; M. Parascandola and Weed 2001; Rothman and Greenland 2005). Why is this? Epidemiologists often seek to *infer* causation, but then so do many other scientists. There are at least three (and probably more) more likely reasons that causation is an issue in epidemiology.

First, it is a science relying heavily on observational studies; this means that the difference between cause and coincidence cannot be easily reduced to a difference between repeatability and coincidence.

Second, epidemiology deals with general causal claims, such as 'Smoking causes lung cancer'. These claims are hard to interpret. Do they express a relation between two properties, smoking and lung cancer? Or are they (exception-ridden) generalisations over individual cases of smoking causing lung cancer? These questions relate to the interpretation of probability, discussed in the second workshop, since causal generalisations in epidemiology are usually probabilistic. There is a related question here about whether epidemiology deals with laws of nature. Are there epidemiological laws, or are the relevant laws all biological – or even physical? And what is the relation between each of the aforementioned concepts of general causation and ceteris paribus laws?

Finally, the identification of causes in many sciences is tied up with the process of explanation. In epidemiology, however, causes may not be explanatory. An epidemiological study might establish that *smoking causes lung cancer*, to the satisfaction of governments and the general public, without *explaining why people get lung cancer*. Intuitively, this is because the link between smoking and lung cancer is not illuminated by establishing that smoking causes lung cancer.

This last point links back to the topic of the third workshop, since it is unclear how the identification of causes for the purpose of devising interventions relates to causal explanation (which has traditionally received a great deal more attention from philosophers of science). Moreover, the *kind* of cause that we focus our attention on may bear both on the chances of a good explanation and of devising a good intervention. For example, identifying readily manipulable socioeconomic causes of illness may offer promising interventions but little understanding; with genetic causes, the reverse may be true. These are, of course, disputable claims, relating to the topic of the first workshop.

This final workshop therefore tied together some of themes of the previous three.

| Speakers | Titles |
|---|---|
| **Dr Olaf Dekkers** | Causation in epidemiology |
| **Professor Nancy Cartwright** | The long road from it-works-somewhere to it-will-work-for-us |
| **Professor Dan Hausman** | How can practical causal generalisations guide practice |
| **Professor John Worrall** | Causality in medicine: getting back to the Hill top |

# Analysis

## Summary

The most important point to arise out of this project is the most obvious one: that there is considerable scope for academic research into the conceptual and methodological foundations of epidemiology, and that this research is of practical importance to the development of the discipline and to the interpretation and deployment of its results in clinical and public health contexts.

In particular, there is scope for further research in the following areas:

- Causation and causal inference in epidemiology

- Risk: interpretations, communication, and relation between population and individual

- The treatment of uncertainty in both academic and policy contexts

None of these topics is new, either to philosophers or to epidemiologists, but the need to focus specifically on epidemiology in studying each of these topics has not been clearly recognised, and is elaborated further below.

In addition to topics, it has been possible to identify a number of interfaces between different domains across which epidemiological research travels, and which stand in need of attention:

- The epistemic/normative interface

- The conceptual/applied interface

- The research/policy interface

Again, the contribution of this project lies not in any novelty claimed for these distinctions, but for the recognition of a need to focus specifically on epidemiology at each of these interfaces. This need is elaborated further below.

## Topics

### *Causation and causal inference*

The nature of causation is an old and difficult philosophical problem. One way to formulate it is by contrasting causal and non-causal sequences of events. You see the footballer swing his foot towards the ball. When it makes contact, the ball flies into the back of the goal. At the same time, you also see the referee scratch his nose. When his hand touches his nose, the ball flies into the back of the goal. The first sequence of events is causal: the footballer's kick *caused* the ball to fly into the back of the net. The second it not: the referee's scratch did not cause the ball to fly into the back of the net.

There are, of course, many visible differences between these two sequences of events. The problem is that none of these visible differences is causation: so philosophers try to say what the difference is. For example, we might note that if the footballer had not kicked then the goal would not have been scored; but that if the referee had not scratched his nose, the goal would still have been scored. Or we might note that kicks of this sort are always or often followed by goals, whereas nose-scratches are not. But whatever we say about the difference, we must do more than simply point to our experience of those events. This means that both the nature of causation, and our knowledge of it, are topics of philosophical interest.

Both the metaphysics and epistemology of causation are also topics of epidemiological interest. There are a number of differences, however, between philosophical and epidemiological approaches. Most importantly, philosophers have focussed primarily on singular causation, that is, causation between single events (such as the footballer's kick and the ball entering the goal). Epidemiologists, however, focus primarily on general causation, that is, the relation (though technically it may not be a relation) between two variables, factors, or classes of events that holds when it is true to say that one causes the other (such as holds between the inhalation of asbestos and the disease mesothelioma, for example).

It is well known to both philosophers and epidemiologists that causation is a topic of interest in each of their fields. What this project has shown, however, is just how different the topic looks in each context. It is very tempting for philosophers in particular to doubt this (as I did at the beginning of this project). We are apt to suppose that the problems epidemiologists address are either rephrasings of familiar philosophical concerns, or else not philosophical concerns at all. But often they are neither, implying that they are new, or at least unstudied, philosophical problems.

The point can be illustrated with two examples, one concerning the metaphysics (or nature) of causation and one concerning its epistemology (causal inference). As I have already indicated, philosophers interested in the metaphysics of causation have focussed largely, though not exclusively, on singular causation, while epidemiologists focus on general causation. Philosophers have a lot to learn about general causation in this context, and in particular about its complexity.

The most widely-used philosophical example of a general causal claim is "Smoking causes cancer", but this is at best a distant descendent of the output of actual epidemiological work on smoking. In particular, epidemiologists seek to measure the strength of the effects they identify, and to compare the strength of these effects. This is simply not something on the radar of most philosophers working on causation (an exception is Sober 1988). This oversimplification is illustrated by the fact that many, though not all, philosophers use the term "general causation" interchangeably with "causal generalisation". Epidemiology shows that this is an error, since it deals neither in singular causal claims nor in generalisations about singular causal relations, and especially not exceptionless or universal generalisations (as Dan Hausman emphasised in his presentation at Workshop 4).

This suggests that an instance of general causation and a causal generalisation are not the same thing, any more than a general interest is the same as an interesting generalisation. Other philosophers have suggested that general causation is a relation between "types" of event (Eells 1991). But this view does not help make sense of the quantitative aspect of causal claims in epidemiology. Moreover it is hard to reconcile with the fact that many such claims are relative to a particular population, in a way that claims about abstracted types are not.

The second example of an unstudied philosophical problem in this area concerns causal inference. Philosophers are accustomed to the idea that causation is not directly perceptible and thus that we often, perhaps always, infer that it is present in a given case.

Again, the study of epidemiology suggests that there is more complexity to causal inference, especially at the general level, than might be suspected. In particular the distinction between internal and external validity, familiar to epidemiologists, has philosophical ramifications.

A study is internally valid when its conclusions hold for the studied group. It is externally valid when its conclusions hold for some target group that is partly or wholly outside the studied group. When a study is internally valid and has a causal conclusion, it may nevertheless fail to be externally valid. Thus a study may warrant a causal inference for the group studied, but not for some target or wider group.

Philosophers have not generally paid much attention to these two components of causal inference. Yet they are clearly important in a public health context. Just because something works somewhere does not mean it will work "for us", to use Nancy Cartwright's phrase.

The insensitivity of extant philosophical theory to these two steps is related to the oversimplified treatment of causation itself: there is simply no ready philosophical toolkit for framing a distinction between the assertion that X causes Y in a studied group, and the assertion that X causes Y in some target group. The underlying assumption is that either X causes Y or it does not. Tautological though this may sound, a study of epidemiology suggests that it is false. X may cause Y in the studied group but the study may fail to be externally valid, and it may turn out that X does not cause Y in some context that is either more general or more pertinent to the aims of the study.

These two examples are quite specific, and, along with similar examples, they receive their full expression in academic activities and outputs of the project, especially discussion at the workshops, articles in the special issue of *Preventive Medicine*, and the author's forthcoming book.

In the course of this project, a number of more general points have emerged, which also deserve emphasis. Simply confirming that causation and causal inference are still live issues in philosophy and epidemiology is important, especially in epidemiology. It is important that epidemiologists, and epidemiology students, appreciate that there are methodological problems in their discipline which have not been solved, and where conceptual work is still necessary to reach even the most pragmatic goals. The importance of this appreciation can be seen in the common occurrence of various misunderstandings, which were the subject of much discussion at the workshops.

For example, Hill's famous "viewpoints" for deciding whether an association is causal are frequently elevated to the status of "criteria", despite his explicit (and correct) insistence that they must not be treated as such. He was clear that they are only aids for answering "the fundamental question - is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?" (Hill 1965, 299). This was a point emphasised by both Alexander Bird and John Worrall at the workshops.

Relatedly, the attempt to substitute statistical significance testing for causal inference, criticised in Hill's paper, was a focus of considerable discussion at the workshops (especially by Sander Greenland and Mark Parascandola). It is important to continue to stress that there is no formal or algorithmic method of causal inference, and that causation itself is not well-conceptualised or understood. Any claims to the contrary are false, and, in the context of epidemiology, dangerous.

## Risk

Risk is a topic of growing interest in political philosophy and the philosophy of science. The word "risk" also occurs frequently in epidemiology. Possibly the simplest point to make here is that the epidemiological use of the word "risk" often does not correspond to risk as it is commonly understood, or as it is understood by philosophers studying risk.

In epidemiology, a risk is a kind of measure: it is the number of new cases of a disease or other health event occurring within a given time period. Philosophers, on the other hand, typically conceptualise risk as the product or other function of the probability of a harm occurring and how harmful it is if it does occur. The epidemiologic measure takes no account of the harmfulness of the outcome and is not related to any probabilities without further assumptions; it is simply a fraction. This is a terminological matter, but it has clear potential to cause confusion, especially in the public reporting of epidemiological results and in feeding epidemiological results into policy decisions.

There are also conceptual difficulties in expressing risk, and especially in communicating it. David Spiegelhalter discussed several of these in his presentation at Workshop 2.

It is not at all clear how to measure risk in a meaningful way. One option is to pick a risky exposure, such as the eating of bacon sandwiches, and quantify other risks in terms of that risk. Thus the riskiness of paragliding every weekend might be expressed in terms of the number of bacon sandwiches one would have to eat to reduce one's life expectancy to that of a person who paraglides every weekend. However, as this example illustrates, it is not always easy to find suitable measures. Moreover the method relies on an ability to conceptualise the riskiness of eating a bacon sandwich in the first place, an ability which, arguably, many of us do not possess.

Another option is to pick an outcome, such as death, which is objective and whose harmfulness can be assumed to be held reasonably constant across evaluations. We can then seek to express the risk of an activity in terms of the number of deaths per unit of activity. However it is clear that policy makers, and perhaps many members of the public, do not hold the value of death equal across such evaluations. Professor Spiegelhalter noted government initiatives to encourage young people to countenance a reasonable and controlled degree of risk in the context of sporting and outdoor activities. He noted, however, that on some measures, horse-riding and taking ecstasy are about equally risky. The government does not

encourage taking ecstasy, and in fact forbids it. This suggests either a governmental view that it is worse to die from taking ecstasy than falling off a horse, or ignorance of the relative riskiness of these activities.

Whichever it is, there is a need for conceptual clarification in the deployment of statistics in public decisions concerning how dangerous a given exposure is and how its riskiness relates to attempts to regulate it. Epidemiology is implicated because it supplies many of the relevant statistics, and it has a scientific and civic duty to render them as clear and as fit for purpose (whatever that may be) as it can.

A particularly important area in which risk poses conceptual difficulties concerns the bearing of population risks on individual risks. At a population level, "risk" is (in epidemiology) a clearly defined term. It tells us how many new cases of the disease or other harm occur within a specified time period, as a proportion of the population size at the start of the time period.

But what does this mean for an individual patient, or a consumer, or a doctor seeking to treat an individual patient? An individual either develops the disease or does not. So the risk for the "population" consisting of that individual alone is either 100% or 0%. What the individual will want to know, arguably, is her *probability* of developing the disease.

The interpretation of probability, especially of the occurrence of an individual event, is a philosophical topic, and one which has direct bearing on the use of data derived from population studies in clinical settings, consumer decisions, court rulings, and other occasions where a decision must be taken about an individual in light of population-level data. To ignore such data is surely wrong; but the correct way to accommodate it is far from obvious.

## Uncertainty

Uncertainty is a feature of most, if not all, of our knowledge. There is little, if anything, that we know without room for doubt, nor with certainty so absolute that greater certainty is inconceivable.

Epidemiology is a particularly uncertain activity, because epidemiologists are often studying associations about which little is known. The underlying process giving rise to an association may be entirely or partially mysterious, and may remain so even after epidemiological evidence suggests that the association is robust. For example, we know that smoking causes cancer even though work to identify and understand the action of the carcinogens in tobacco smoke is ongoing. As Alfredo Morabia argued in his presentation to Workshop 4, epidemiology actively investigates a phenomenon only until it is sufficiently well-understood for the laboratory sciences to take over. After that, epidemiologists may be called upon to make predictions about the spread of a disease, but work on identifying and understanding the causes (analytic epidemiology) ceases. This means that, by its nature, (analytic) epidemiology is at the forefront of our knowledge, where it inevitably confronts uncertainty.

Considerable discussion at the workshops focused on the proper treatment of uncertainty. In particular, the use of "statistical significance" was discussed. Sander Greenland and Mark Parascandola both discussed limitations on the use of "p-values". Two such limitations are particularly important.

First, p-values test at most one kind of error: the probability that an observed association arose "by chance", where the latter phrase indicates that the variables in question occur randomly with respect to each other, and that the association observed on this occasion is the result of that random occurrence. (Errors of this kind are often referred to as *false positives* or *type 1 errors*). This means that p-values are, at best, a guard against wrongly inferring a false hypothesis. They are no guard at all against failing to infer a true hypothesis (often referred to as *false negatives* or *type 2 errors*). If p-values are not low enough then, on standard methodologies, no inference will be made - but the hypothesis under test may nonetheless be true. We lack a widely accepted method for guarding against this sort of error, even though it is potentially just as harmful. Moreover, an excessive p-value does not even mean that a causal inference is unwarranted. That will depend on methodological factors that cannot be meaningfully reduced to that statistical measure.

The other important limitation of the use of p-values is that the "null hypothesis" is rarely subjected to the same scrutiny as the hypothesis under test. Mark Parascandola pointed out that if the null hypothesis is applied to studies or meta-analyses of large numbers of variables, it amounts to the hypothesis that everything is due to chance, a presupposition which, if seriously entertained, would undercut the scientific enterprise altogether.

Sander Greenland pointed out that the use of the p-values in many contexts amounts to an unwarranted bias towards the null, by encouraging the view that the null hypothesis is to be held until disproved. In fact, we cannot regard the null hypothesis as proven by a failure to find a statistically significant association. We cannot conclude that there is no association. All we can immediately conclude, in the absence of further evidence or repeated studies, is that we *still don't know* whether there is an association.

The misuse of statistical tests in the context of quantifying uncertainty by estimating the probability of error is related to the incorrect substitution of significance testing for causal inference, mentioned previously. In both cases, the emerging picture is the same: the existence of mathematical tools sometimes misleads epidemiologists into thinking that they have more (or less) certainty than they really do; and sometimes it leads them into substantive errors. The lesson is likewise the same: that appreciation of the conceptual foundations of the discipline, and of the persistently informal, unalgorithmic nature of key components of its methodology, is essential to its success.

## Interfaces

### *The epistemic / normative interface*

Epidemiology, or its outputs, move across a number of interfaces, in ways that are conceptually interesting and sometimes challenging. Most obvious among these is the interface between epistemic activities, concerning the gathering of knowledge, and normative activities, concerning evaluation and recommendation. Epidemiology employs the methods of natural sciences, but it is so closely tied with human concerns that in some forms it might also be seen as a social science.

In his presentation at Workshop 1, Michael Marmot argued that medical professionals are duty-bound to identifying and eliminating causes of ill-health, even when these include poverty or inequality of social status – things normally considered the preserve of politics or perhaps economics, not of medicine, and certainly not of the natural sciences. Jonathan Wolff, in his presentation, sought to work out what, if anything, should be done if Professor Marmot is correct that socioeconomic inequality causes ill health.

One thing that Professor Marmot's work shows is how epidemiology can exert an expansive pressure on medicine. Epidemiological methods can be applied without much sensitivity to subject matter. It is possible to assess the effect of social inequality on health just as long as it is possible to

measure poverty and health. This paves the way for arguments such as Marmot's, explicitly advocating an expansion of the scope of medical attention.

Another point at which epidemiology potentially crosses the border between epistemic and normative is in the handling of uncertainty, discussed previously. The choice of a value for statistical significance is arbitrary in relation to the statistical significance tests themselves. It depends on a judgement as to how serious the consequences of error will be, and also, as discussed, on weighing the relative perils of falling into different kinds of error.

## The conceptual / applied interface

Epidemiology deals in both practical and theoretical matters. But the distinction between the practical and theoretical is not very precise. We can split it into two more precise distinctions. The first is between conceptual and applied; the second is between research and policy. Epidemiology works across both these interfaces.

The interface between conceptual and applied work in epidemiology is evident in good epidemiological text-books (e.g. Szklo and Nieto 2007; Rothman, Greenland, and Lash 2008). These focus on teaching epidemiological methods with a view to their application to the sorts of problems that epidemiologists need to solve. Yet they often also devote space, sometimes considerable space, to conceptual issues – more specifically, to the way in which conceptual issues can arise during the attempted application of epidemiological methods. Thus the conceptual sections of epidemiology text books – the chapters on causation, for example – do not simply set down a conceptual framework for the student to absorb. They describe the lack of such a framework (if they are honest) and indicate how this lack can make an epidemiologist's work difficult.

What this means is that epidemiologists cannot be simply technicians. They must also have some appreciation of the conceptual foundations of the discipline, and in particular of the issues arising in the topic areas described above. If they do not then they are likely to commit errors, such as exhibiting an unwarranted bias towards the null hypothesis, failing to properly balance the risk of different kinds of errors, making bad causal inferences, or failing to make good ones.

## The research / policy interface

The other dimension of the more general interface between practical and theoretical work is the distinction between scientific research and health policy making. As noted above, epidemiology has elements in common with both natural and social sciences. Its methods may be scientific, but its objectives are often thoroughly human.

One particularly interesting example of the interface between research and policy is in Geoffrey Rose's "paradox of prevention" (Rose 1992). The paradox arises from the fact that a large number of people at low risk can generate a larger number of adverse outcomes than a small number at high risk. This means that, in cases where individuals are distributed along a risk profile (for example, across a range of cholesterol levels as a risk factor for heart disease), it may be more effective to intervene on the majority at a moderate to low risk than to intervene on those at high risk. The paradox is that most of these individuals will receive no benefit from the intervention, but will presumably suffer the attendant harms, if only the harm of inconvenience. John Worrall emphasises the fact that many (perhaps 95%) of those who take statins receive no benefit, but do suffer the attendant risks of taking the drug.

However, effective prevention strategies often require individuals to make exactly this sort of sacrifice. Vaccination is the best-known example. At Workshop 4, Stephen John used this feature of preventive strategies to bring out philosophical questions about the direct and indirect benefits that an individual can derive from a preventive strategy.

Dr John also sought a clear formulation of the paradox of prevention. Arriving at a clear formulation is of practical as well as academic interest, since individuals choosing whether to participate in preventive programs such as vaccination will inevitably ask themselves what are the likely risks and benefits for they themselves.

# Practical significance

The following points of practical significance are salient in the foregoing analysis.

### The need for conceptual clarity in the use of health statistics

A number of workshop contributions highlighted ways in which both statistical analysis and reliance on statistics in policy may rest on implicit and unwarranted assumptions. Greater conceptual clarity on the part of those performing and using statistical analyses, and a critical approach to statistical warrant for policy decisions, are necessary to prevent this.

### The need for clarity in the use of statistical significance testing in particular

A particularly prominent case where conceptual clarity is often lacking is in the use of statistical significance testing. The underlying message is that statistical significance must not be confused with significance simpliciter. An association might fail to be statistically significant due to the size of the study, yet the nature of the study might nevertheless provide convincing evidence for an inference. Moreover statistical significance tests do not estimate the probability of wrongly failing to make an inference, and do not estimate the practical costs of wrongly failing to make an inference. A result might be significant because it suggests a serious hazard even though it is not statistically significant. Morever, statistical significance is no guarantee of truthfulness. For a p-value of 0.05 we would expect 1 in 20 independent studies of null associations to wrongly pronounce that there is an association. Statistical significance is thus neither necessary nor sufficient for significance more generally and must always be supplemented by methodological reasoning and, where appropriate, consideration of the practical implications of both Type 1 and Type 2 errors.

### The difficulty of causal inference and its continued resistance to formal methods

Causal inference is difficult. Statistical tools can help. However, there is no algorithm or mathematical tool which can answer the most important question: is there a causal relationship underlying the observed association, and if so what are its properties? Moreover the application of formal tools often requires some prior causal interpretation. This does not prevent them from being useful, but it does show that causal inference remains prior to any statistical or other mathematical analysis. Contrary to what some have asserted, causation has not yet been mathematized (cf. Pearl 2000, xiii).

### The importance of distinguishing between internal and external validity, and the difficulties attendant on applying or generalising the results of a study to a target or wider group

Even if the foregoing cautions are all taken into account, further thought is required before using the results of a study, or even several studies, to make claims about other populations. There are many reasons why a drug or a policy might work on one occasion but not another; and many reasons why an exposure might cause ill health on one occasion but not another. As Nancy Cartwright emphasizes in her contribution to the special issue of *Preventive Medicine*, ensuring that evidence meets strict criteria is only one part of establishing a sound basis for policy.

Another, equally important, part is finding some reason to think that the study provides a guide for what will happen in the circumstances with which the policy is concerned. This point has sometimes been neglected due to an overriding emphasis on quality of evidence, as opposed to relevance or applicability.

### The importance of continued methodological development in epidemiology, and that epidemiologists and policy makers have some understanding of the live methodological and conceptual debates within the discipline

Epidemiology is a young science, with a number of central conceptual questions unresolved or only partially resolved. It is also an unusual science, in that the body of epidemiological knowledge is not primarily factual, but methodological. Epidemiology is in large part a collection of methods for finding things out on the basis of scant evidence, and this by its nature is difficult. If working epidemiologists and policy-makers possess some understanding of the live conceptual debates in epidemiology then they will be better able to avoid erroneous certainty.

# Philosophical signficance

In the foregoing analysis, the following points of philosophical significance are salient.

### The complexity of the sorts of claims about general causation that epidemiology makes, as a science primarily interested in general causation

Philosophers have tended to work with very simple claims, such as "smoking causes lung cancer", when thinking about general causation. Epidemiologists tend to make much more complex claims. This project has suggested that there are important philosophical issues lurking in the complexities; but it is impossible to say what they are in advance of a thorough philosophical treatment.

### The complexity of causal inference, especially to general causation, as exemplified by the distinction between internal and external validity

Philosophers have given a considerable thought to the way in which we make causal inferences. They have focused, naturally enough, on simple examples and idealised cases. They have also focused on experimental scenarios, in which an intervention is made (Semmelweis's work is a classic in philosophy of science, for instance). Philosophers have not yet caught up with the advent of sophisticated observational methods. For example there has been no thorough and extended philosophical analysis of the cohort study, the case-control study, or indeed many other central epidemiological methods. Nor have philosophers given much explicit thought to the question of how the results of a study may be generalised: they have tended to focus on the problem of inferring causation from an experimental result, and not on the question of how – if causation can be inferred – that conclusion might be applied elsewhere. These are clear gaps in the philosophical literature and it is to be hoped that philosophers of science will attend to them soon.

### The "paradox of prevention" as a genuine philosophical problem

The "paradox of prevention" (a term coined by Geoffrey Rose (1992)) arises when reducing the risk of persons in medium- to low-risk groups has a larger impact on the overall risk in the population than reducing the risk of persons in high-risk groups only. The paradox arises from the fact that interventions on persons on these groups will typically offer little or no benefit to those individuals (or even incur costs), despite the effect on the health of the population. Vaccination and the reduction of cholesterol levels are two well-known examples. In his contribution to the special issue of *Preventive Medicine*, Stephen John argues that this is not a mere pragmatic awkwardness but a real philosophical paradox (at least on some contractualist views of the source of political

obligation). As such it deserves to be studied by philosophers.

### The relation between risks applying to populations and individual risks

As epidemiologists use the term, "risk" has no application to the individual. Yet ignoring the risks derived from studying populations in making decisions about individuals would be foolish. In the clinical setting, the evidence-based medicine movement has advocated attaching very great weight to evidence derived from studies of large populations. One criticism they have faced is that the individual case may differ in important ways which render the evidence irrelevant. More generally there is an open philosophical question about what the exact rational bearing of population-level data on an individual is. The problem has also been explored in legal contexts, when plaintiffs have sought to rely on epidemiological evidence. This problem is a general theme of many efforts to use the output of epidemiological work and it raises questions which a philosopher can help to answer.

### The apparent fact that epidemiology does not focus either on theory or experiment, meaning that most philosophical characterisations of science are of limited or no applicability to epidemiology

It is evident from even a cursory acquaintance with epidemiology that experiment does not play the central role that it does in some other sciences. Less immediately evident is the fact that epidemiologists do not develop theory in the way that other scientists do; or rather, the "theory" they develop is of a different sort, and not what a philosopher would call "theory". Epidemiologists develop methods for finding out about things. They do not, however, develop an overarching view of the way that things are in the domain they study. Epidemiologists employ theoretical knowledge of this kind from other disciplines – biology, medicine, and others. But epidemiological findings, for example that smoking causes lung cancer, are not assimilated to a body of "epidemiological facts", in the way that medical or biological facts might be. This is because epidemiology focuses on finding things out, and developing better methods for doing so. This makes it very suitable for philosophical study. Moreover, the lack of theory (as philosophers use the term) and experiment makes it hard to see how some of the central philosophical views of science apply to epidemiology. This suggests that they need reconsidering.

**The need for and interest of a specific focus on the philosophy of epidemiology within the philosophy of science**

If so many promising topics of philosophical interest can arise in the course of this project, it seems very likely that a systematic philosophical study of epidemiology will prove worthwhile. No such study has yet been undertaken, and there is not as yet a widely-recognised sub-area within the philosophy of science devoted to epidemiology. This project suggests that there should be philosophy of epidemiology.

# Directions for further research

The project has thrown up a large number of research questions, and they differ depending on the interests of the parties involved and their contact with the project. Rather than seek to list them, this report identifies three important general areas for further research.

## Philosophy of epidemiology

This project suggests that there is scope for a distinct focus on epidemiology within philosophy of science. There is no established "philosophy of epidemiology", as there is of physics, biology, economics, medicine, and so forth. This project has demonstrated the richness and diversity of intellectual problems that epidemiology brings together. It has also demonstrated the practical need for academic study of these problems. The simplest and most direct way to satisfy this need is to promote the philosophy of epidemiology as a distinct focus within the philosophy of science.

The author is writing a book on this topic and has obtained funding for further academic meetings.

## Epidemiology and law

A second direction for more research concerns the role of epidemiological evidence in litigation. This is a topic that Mark Parascandola and Sander Greenland discussed in Workshop 3 as well as in a variety of prior publications (Greenland and Robins 1988; Greenland and Robins 2000; Robins and Greenland 1989; Greenland 2004; Mark Parascandola 1998). It is also something that legal academics have discussed (see especially Wright 1988; Wright 2008).

There are a number of difficulties in using epidemiological evidence in litigation. Two are particularly salient. The first is the application of statistical tests of significance to either bolster or refute causal inferences which are not, in fact, bolstered or refuted by the data in question (a topic treated thoroughly by Sander Greenland in his contribution to the special issue of *Preventive Medicine*). The second is the question of whether and if so how epidemiological findings can be applied to individual litigants, especially in proof of causation. Both of these are pressing concerns, because epidemiological evidence is sometimes the only evidence available for proof of the causal element of liability. As Professor Morabia emphasized at Workshop 3, epidemiology is actively engaged in researching a disease only when that disease is not tractable to laboratory scientists. Legal controversy is also much more likely to arise when the grounds of a claim are novel. This is a systemic reason for the fact that epidemiology sometimes becomes embroiled in legal controversy, and provides a reason for further academic work on the proper use of epidemiological evidence in litigation.

## Statistical significance

A third important direction for further research concerns statistical significance testing. This was a recurrent theme at the workshops. There is considerable discontent among epidemiologists and statisticians at the way in which statistical significance testing is taught and employed in many contexts, including by eminent statisticians. This suggests that there is scope for further research in at least two directions. First, the use of statistical methods for testing for "false negatives" as well as false positives might be more widely taught. Second, procedural rules might be sought for deciding whether a statistical significance test is being correctly applied. Whether these are achievable goals is not something that the author is in a position to judge. But it is clear that much of the confusion surrounding the use of statistics comes from the difficulty of understanding it. Since it is not feasible for everyone to become a statistician, the development of simple models and heuristics for understanding statistics – such as those being developed by Professor Spiegelhalter – must be a priority.

# References

Broadbent, Alex. 2008a. For Analytic Bioethics. *Clinical Ethics* 3: 185-188.

Broadbent, Alex. 2008b. "A note on epidemiological caution." Letter to the *British Medical Journal*. http://www.bmj.com/cgi/eletters/335/7630/1116-b#191569.

Broadbent, Alex. 2009. Causation and models of disease in epidemiology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 40: 302-311.

Broadbent, Alex. 2011a. Causal inference in epidemiology: mechanisms, black boxes, and contrasts. In *Causality in the Sciences*, ed. Phyllis McKay Illari, Federica Russo, and Jon Williamson, 45-69. Oxford: Oxford University Press.

Broadbent, Alex. 2011b. Defining neglected disease. *BioSocieties* 6 (1): 51-70.

Brunner, Eric, and Michael Marmot. 2006. Social organization, stress, and health. In *Social Determinants of Health,* ed. Michael Marmot and Richard Wilkinson, 6-30. Second. Oxford: Oxford University Press.

Carter, K. Codell. 2003. *The Rise of Causal Concepts of Disease*. Aldershot: Ashgate.

Eells, Ellery. 1991. *Probabilistic Causality*. Cambridge: Cambridge University Press.

Greenland, Sander. 2004. The Need for Critical Appraisal of Expert Witnesses in Epidemiology and Statistics. *Wake Forest Law Review* 39 (2): 291-310.

Greenland, Sander, and James Robins. 1988. Conceptual Problems in the Definition and Interpretation of Attributable Fractions.*American Journal of Epidemiology* 128 (6): 1185-1197.

Greenland, Sander, and James Robins. 2000. "Epidemiology, Justice, and the Probability of Causation." *Jurimetrics* 40: 321.

Greenland, Sander, Manuela Gago-Dominguez, and Jose Esteban Castelao. 2004. The Value of Risk-Factor ('Black Box') Epidemiology.*Epidemiology* 15 (5): 529-535.

Haack, Susan. 2004. An Epistemologist Among the Epidemiologists. *Epidemiology* 15 (5): 521-522.

Henle, Jacob. 1844. Medicinische Wissenschaft und Empirie. *Zeitschrift fur rationelle Medizin* 1: 1-35.

Hill, Austin Bradford. 1965. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 58: 259-300.

Ioannidis, John P.A. 2005. Why Most Published Research Findings Are Wrong. *PLoS Medicine* 2 (8): e124.

Kaufman, Jay S. 2010. Towards a More Disproportionate Epidemiology. *Epidemiology* 21 (1): 1-2.

Leung, Gabriel M., and Selena Gray. 2008. Translational public health: rehearsing the evidence until the task is done. *The Journal of Public Health* 30 (3): 217-218.

Marmot, Michael. 2006. Health in an unequal world: social circumstances, biology, and disease. *Clinical Medicine* 6 (6): 559-572.

Parascandola, M., and D. L Weed. 2001. Causation in epidemiology. *Journal of Epidemiology and Community Health* 55: 905—912.

Parascandola, Mark. 1998. What is Wrong with the Probability of Causation? *Jurimetrics* 39: 29-44.

Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Price, Huw. 1996. *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.

Robins, James, and Sander Greenland. 1989. The Probability of Causation under a Stochastic Model for Individual Risk. *Biometrics* 45: 1125-1138.

Rose, Geoffrey. 1992. *The Strategy of Preventive Medicine*. Oxford: Oxford University Press.

Rothman, Kenneth J. 1976. Causes. *American Journal of Epidemiology* 104 (6): 587—592.

Rothman, Kenneth J, and Sander Greenland. 2005. Causation and Causal Inference in Epidemiology. *American Journal of Public Health (Supplement 1)* 95 (S1): S144—S150.

Rothman, Kenneth J, Sander Greenland, andTimothy L Lash. 2008. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins.

Sober, Elliot. 1988. Apportioning Causal Responsibility. *The Journal of Philosophy* 85 (6): 303-318.

Susser, Mervyn. 1991. What is a cause and how do we know one? *American Journal of Epidemiology* 133 (7): 635--647.

Szklo, Moyses, and F. Javier Nieto. 2007. *Epidemiology: Beyond the Basics*. 2nd ed. Boston, Toronto, London, Singapore: Jones and Bartlett Publishers.

Wright, Richard. 1988. Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts. *Iowa Law Review* 73: 1001-1077.

Wright, Richard. 2008. Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof. *Loyola of Los Angeles Law Review* 41: 1295-1344.

# About the PHG Foundation

The PHG Foundation is an independent, non-profit organisation based in Cambridge, UK, with the mission making science work for health. We identify the best opportunities for 21st century genomic and biomedical science to improve health and tackle disease in ways that are rapid and effective, equitable and responsible. This entails work to promote the prompt translation of scientific innovation into medical and public health policy and practice.

We provide knowledge, evidence and ideas to stimulate and direct well-informed debate on the potential and pitfalls of key genomic and biomedical developments, and to inform and educate stakeholders – policy makers, health professionals, patients and the public. We also provide expert research, analysis, health services planning and consultancy services for governments, health systems, and other non-profit organisations.

# phg
foundation
*making science
work for health*