PHILOSOPHY OF EPIDEMIOLOGY

Alex Broadbent

The Nature of Epidemiology

Epidemiology is the study of the distribution and determinants of disease and other health states in human populations by means of group comparisons for the purpose of improving population health (Broadbent 2013, 1). But this definition does not make obvious the distinctive character of epidemiology, nor its philosophically interesting aspects.

Epidemiology is a detective-like science, seeking to identify exposures or traits that cause ill health, or trying to find out what the effects of certain exposures or traits might be. It relies on a mix of evidence types, depending on the topic under study. In clinical epidemiology, where one is testing the effect of a medication, it is possible to construct large and methodologically strong studies, randomized controlled trials (RCTs), where individuals are randomly assigned between treatment and control group. In theory, this allows the probability of various kinds of bias being responsible for the finding to be calculated, although in practice, RCTs often fail to live up to their on-paper promise. For example, they remain open to financial bias (where the funder’s financial interests influence the finding) and to publication bias (where findings of no effect or negative effect are less likely to be published than positive findings).

However, often, RCTs are impossible, and studies are “observational,” meaning that the investigator does not assign subjects to exposure/treatment or control group, but passively observes the groups as s/he finds them. For example, the effects of smoking are studied entirely through observational studies, because of the ethical and practical difficulty of forcing people to take up or stop smoking. This makes epidemiologists methodologically creative in coming up with ways to get the most out of apparently unpromising investigative conditions, and it also forces them to take different kinds of evidence from a wide range of sources.

The unifying principle of epidemiology is normative. Epidemiologists seek to find out things that will improve population health. Not all epidemiologists are happy with this value-laden goal, but it is hard to make sense of epidemiology if that is not the goal. If improving population health were not the goal, then the discipline would not cohere. Because the unifying principle comes from a goal and not from the field of study itself, the scope of epidemiology is unlimited: anything that might give rise to differences in health status between identifiable populations is a potential topic of study, including environment, occupation, nutrition, lifestyle, genes, and medication. Likewise, the outcomes that epidemiologists measure are not necessarily lined up with existing medical disease classifications. The combined effect is an expansive pressure on the scope of medicine. This is most evident in calls of some social epidemiologists for...
physicians to take an active interest in reducing socioeconomic inequality, on the basis that socioeconomic inequality is detrimental to population health (Marmot 2006; Venkatapuram and Marmot 2009).

Epidemiology is quite unlike a typical philosopher's picture of science. Besides often proceeding without experiment, as already discussed, it also lacks theory, in the sense that philosophers use the term. Epidemiologists do not seek to fit their findings into an overarching mathematical framework, like physicists or economists (surprisingly, since on paper economics and epidemiology might look similar to a visiting alien). Nor do they work within the auspices of a grand non-mathematical theoretical framework, as biologists do with the theory of evolution. Epidemiologists do not aim to offer a comprehensive description or understanding of their domain—indeed, their domain is not clearly delimited, as already mentioned, but is rather defined with reference to the goal of improving population health. Thus a truth-apt epidemiological claim about the empirical world, such as “Smoking causes lung cancer,” is not helpfully seen as a piece of epidemiological theory. It is a piece of causal knowledge, but it is not a theory in the sense that physics or even biology has theories—grand, overarching, often mathematical structures explaining or describing large chunks of the observable world. Epidemiologists do not devise theories in this grand sense. Asked to identify a piece of epidemiological theory, an epidemiologist would be more likely to come up with something like the fact that you can estimate a relative risk from an odds ratio given the assumption that the disease is rare. Thus epidemiological theory consists, not in what philosophers would call theoretical claims about the nature of the world, but of a bundle of methods for finding things out.

There are many different philosophical approaches to science, and it is perhaps dangerous to generalize. However, even those philosophers who place method in the foreground tend to see science as a matter of developing overarching theoretical frameworks having, ideally, universal application. Indeed, those philosophers who have emphasized method (such as Carl Hempel and Karl Popper) have tended to be the most enamored of the “physics model” of science. The idea that science is distinguished by its methods is certainly familiar to philosophers, but the idea that these methods might not be directed at developing theory, rather than solving piecemeal problems like a detective, is not widespread.

This disconnect between the philosopher's received image of science and the actual nature of epidemiology gives familiar philosophical topics new aspects as they arise in epidemiology. This is what makes the study of those topics within epidemiology fruitful for the discipline of epidemiology, since there really is room for conceptual and methodological advances in this young and developing science. The detective-like nature of epidemiology gives rise to questions about the inferences it seeks to make. These inferences concern causation, prediction, and explanation—all central philosophical topics. These topics will be prominent themes in the remainder of this chapter, which is divided into three sections, as follows. First, we will explore the significance of epidemiological thinking for disease classification. Second, we will consider recent thinking about causal inference in epidemiology. Third, we will look at prediction in epidemiology, and the philosophical questions it raises.

Epidemiology and Disease Classification

One striking effect, or perhaps cause, of the growth of epidemiology in the latter part of the 20th century was the rise of the view that many or perhaps all diseases are multifactorial. Prior to the rise of multifactorial thinking, it seemed natural to classify infectious diseases by their causes, since their causes come ready-classified as distinct species of bacteria or virus. It is much harder to classify chronic, non-communicable diseases (CNCDs), such as cancer or coronary heart disease, by their causes, since the various cases of cancer or coronary heart disease do not
display any obvious pattern of causes. Some factors are more commonly found—perhaps much more commonly—among those with these diseases than among those without, but there will be cases of these diseases among those without these exposures and cases where the disease fails to occur even with them.

Lung cancer provides a case in point. Before the widespread uptake of smoking, it was virtually unknown and was not identified as a distinct form of cancer. Nonetheless, it does occur in non-smokers. And among smokers, it is not an especially common disease—the majority of smokers do not get lung cancer. Thus there is little appetite for defining lung cancer as “smoking-itis” or “smoking disease,” even though lung cancer is very rare in the absence of smoking. In particular, we do not know what makes the difference between those smokers who succumb to lung cancer and those who do not. So even though smoking is far and away the most significant cause of lung cancer, we do not go so far as to define lung cancer with reference to this cause. Contrast cholera, where the definition of the disease does mention a particular cause—namely, *vibrio cholerae* in the small intestine. This is part of the definition of cholera: it distinguishes cholera from non-choleric fever and diarrhea.

Much of the classic epidemiological writing on the multifactorial nature of disease focuses on the multiple causes of various diseases (e.g., Rothman 1976). However, this is misleading: everything is “multifactorial” in the sense of arising from many causal factors and not one alone. The real question—the question that would make sense of the literature on multifactorialism in epidemiology—is not “How many causes do diseases have?” but “Can diseases be defined or classified causally?”

This issue arose because of a prevailing view in the first part of the 20th century that diseases should be classified by their causes, and by just one cause. This view was pushed hard by the originators of germ theory, notably Robert Koch: “. . . each disease is caused by one particular microbe—and by one alone. Only an anthrax microbe causes anthrax; only a typhoid microbe can cause typhoid fever” (Koch 1876; quoted in Evans 1993, 20).

This principle may be appealing for infectious diseases (though even there, it does not make room for diseases arising from symbiosis between two or more infectious agents, such as swine fever, which arises from the symbiotic action of a bacterium and a virus), but it is clearly not appealing for something like cancer, for which we simply cannot identify anything like a universal general cause that is present when cancer is present and absent when cancer is absent. As industrialized countries passed through the epidemiologic transition (where the main causes of mortality shift from infectious diseases to CNCDs), epidemiologists found this view of disease unhelpfully restrictive. Hence the trend toward multifactorialism.

This current trend toward multifactorialism amounts to a view that diseases need not be classified by reference to causes in any strict way. Interestingly, this is in many ways an echo of the way diseases were thought about before the advent of germ theory in the late 19th century. Jacob Henle (who gave his name to a loop in the kidneys) issued this wonderful complaint:

Only in medicine are there causes that have hundreds of consequences or that can, on arbitrary occasions, remain entirely without effect. Only in medicine can the same effect flow from the most varied possible sources. One need only glance at the chapters on etiology in handbooks or monographs. For almost every disease, after a specific cause or the admission that such a cause is not yet known, one finds the same horde of harmful influences—poor housing and clothing, liquor and sex, hunger and anxiety. This is just as scientific as if physicists were to teach that bodies fall because boards or beams are removed, because ropes or cables break, or because of openings, and so forth.

(Henle 1844; quoted in Carter 2003, 24)
This sort of complaint might equally be directed at some areas of contemporary epidemiology, which identify risk factors for disease that, though clearly part of the causal story in some cases of disease, are absent in other cases, and often present where the disease is not.

The fact that Henle was complaining about it in 1844 shows that multifactorialism is not new, and this ought to raise a question mark over the modern trend toward it. To be specific, it raises the question: what is at stake in classificatory decisions about disease entities? Clearly, Henle and Koch felt that it mattered whether diseases were classified by their causes or not. Note that Henle is not criticizing the medical science of his day for poor causal inference. In the analogy, he is not denying that breaking ropes and beams, openings, and so forth cause falling. He is rather pointing out that merely cataloguing the causes of falling is not a good explanatory theory (he says not "scientific") of the phenomenon.

Just as one might hope that a theory of falling will unify all the various cases of falling and not merely provide a catalogue of causes (e.g., boards breaking, ropes snapping, etc.), one might hope that a disease classification system expresses or is underwritten by an explanatory theory. In this context, an explanation would not be a law like the theory of gravity, but would be a causal difference between cases with the disease and cases without. Hence the feeling of dissatisfaction that some have expressed about contemporary risk factor epidemiology (Vandenbroucke 1988), which catalogues causes fairly enough, but does not attempt to offer explanations of the difference between cases of disease and cases of health. Multifactorialism about disease licenses the development of a catalogue of causes, but does not put pressure on medical science to systematize this catalogue so as to arrive at something more explanatory (Broadbent 2009). It also licenses the development of medications that are not "magic bullets" but that reduce the risk at the population level, without necessarily helping all or even most patients who take them (Stegenga forthcoming). (Statins and selective serotonin reuptake inhibitors are two examples of such medications.) If we settle for an understanding of etiology consisting of risk factors only, then we must also settle for interventions on risk factors only, and the best these will do is reduce the probability of a disease. Yet we know there have been "magic bullets" in the history of medicine (such as penicillin). Whether it is reasonable to insist upon magic bullets or not, the present point is that the effects of our conceptual framework for thinking about disease affect medical practice, can encourage or discourage medical breakthroughs, and can shape the development of medications and the assessment of their effectiveness.

Causal Inference

The paradigm case of causal inference in modern epidemiology is the discovery that smoking causes lung cancer. This episode involved significant methodological developments (in study design, e.g., in relation to case-control and cohort study methodologies) and conceptual developments (insights into how to use given data to draw inferences, e.g., the argument that a potential confounder could not be entirely responsible for an association if its relative prevalence is lower than the observed risk ratio [Cornfield et al. 1959]). It also saw the authoring of some famous commentaries on causal inference by epidemiologists and statisticians. The two most famous are those of Austin Bradford Hill (Hill 1965) and the principles of causal inference included in the first Surgeon General’s report on the effects of smoking (Advisory Committee to the Surgeon General of the Public Health Service 1964). Thus, for example, Hill lists strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy as properties of an association that one can look for when seeking to make a causal inference.
However, it is important to note that Hill insists that the presence or absence of none of these properties is decisive on its own. Hill’s nine viewpoints have often been misunderstood as “criteria,” but this is not how Hill intended them:

None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on 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, 11)

Hill is effectively implementing *inference to the best explanation* (IBE), some years before that topic was paid comprehensive attention by philosophers of science. The core idea of IBE is that, in some circumstances, we can infer the truth of an explanation on the basis that it is good. For example, if you hear a roaring noise outside the window, you can reasonably infer that a bus or car is revving its engine. The noise *could* be coming from a recording, or from a helicopter, but these are probably less good explanations of the sound because they are more complex or raise further questions (e.g., why would someone play a recording of an engine outside the window?). In Peter Lipton’s terminology (Lipton 2004), Hill was in effect listing a number of criteria that would need to be considered in order to decide whether an explanation was “lovely”—whether it was a good explanation. If so, then one can make a causal inference, albeit tentatively. But there is not a formal balancing of factors, no strict checklist; sometimes very strong evidence of one kind (say, an overwhelmingly strong association) might outweigh the lack of evidence of another kind (say, the lack of biological plausibility, relative to our current biological knowledge). Instead, there is a focus on qualitative judgment taking into account as much evidence as possible, from as many sources as possible—and, crucially, considering the question of whether there is a better possible explanation.

This reflects the character of the evidence that was available concerning smoking and lung cancer. The direct epidemiological evidence was entirely based on observational studies. Many of these were case-control studies, and it is difficult to base causal inferences on case-control studies alone. (In a case-control study, cases of the disease are identified, and the prevalence of the exposure among cases is compared to the prevalence of the exposure among controls, who do not have the outcome.) Large-scale cohort studies were started, but since lung cancer takes a long time to develop, the results of these studies were not decisive at that time. (In a cohort study, a “cohort” of people are followed, and their exposures and outcomes are recorded. In a famous cohort study on smoking running from 1951 to 2001, Richard Doll and Austin Bradford Hill wrote to registered British doctors requesting information on their smoking habits.) Laboratory evidence from animal experiments showed carcinogenic effects of tar painted onto the ears of rats, but it is a long way from there to the conclusion that tar carried in smoke inhaled into the lungs of a human will cause cancer in the lung. Time-trend data showed that lung cancer incidence, previously virtually unknown, had increased about 20 years after the widespread uptake of smoking, but time-trend data invites the slogan “correlation is not causation”: the fact that two factors trend in the same way does not mean that one is causing the other; it could equally be explained by some other factor causing both trends.

In this situation, the only way to make a causal inference is to triangulate from the various kinds of evidence available. Each piece of evidence alone is weak, but the combination, handled correctly, amounts to a strong case. An example of such “triangulation” concerns the use of time-trend data, which, as mentioned, is not impressive evidence for causality on its own. The main alternative explanation for the association between smoking and lung cancer was
the “constitutional hypothesis,” which suggested that smoking and lung cancer might share a genetic cause of some kind. This hypothesis does not, however, accommodate time-trend data, since it is not plausible that a genetic variant would have spread fast enough to explain the increase in lung cancer incidence. Thus a piece of evidence that is not compelling on its own becomes an important part of the larger case, by refuting a competitor hypothesis.

Against this background, it is not surprising that a qualitative, informal approach to causal inference should have become dominant. However, contemporary epidemiology is witnessing a trend in the opposite direction, toward formal methods for causal inference. This trend emphasizes the importance of study design. Because of this emphasis, it implicitly locates causal inference within studies, or with data that can be amalgamated into one data set. The implication is that a causal inference is something that you do with a data set, using formal tools. This is not how causal inference was done in all historical cases: for example, there is no formal way of amalgamating the results of experiments involving painting tar on the ears of rats with observational data about smoking and lung cancer.

The core of the contemporary approach is the idea that, in order to make a meaningful causal claim, one must clearly specify the intervention one has in mind to bring about the difference between the exposed group and the unexposed group. It is not enough, according to this approach, merely to seek to investigate the effects of obesity on mortality, for example. There are many ways to intervene so as to reduce (or increase) body mass index (BMI), and these may have different effects on mortality (Hernán and Taubman 2008). Exercise and diet may have different effects; smoking or amputation will reduce BMI (the measure we are supposing is used in this case to detect obesity), but are likely to have quite different effects on mortality from either exercise or diet. This means that if one conducts a study of the association between obesity and mortality, one cannot make reliable claims about the potential outcome of reducing obesity in the study population. It would depend on how obesity was reduced. Proponents of this line of reasoning argue that this inability to identify the potential outcome—the inability to assert a counterfactual—means that such a study is not well-placed to provide evidence for a causal inference.

This “potential outcomes approach” (POA) insists that investigators clearly specify the counterfactuals whose truth they are investigating. The POA is also often thought to imply a kind of ranking among study designs, with experimental studies coming out well, because in such studies the investigator actually makes an intervention to create the exposed and control groups. A well-specified intervention is causally uniform: it has the same effect on the outcome of interest in different cases. The intervention “reduce obesity” does not satisfy this criterion with respect to the outcome “mortality” because the different ways that you might reduce obesity have different effects on mortality.

One of the confusions that arises in connection with the POA is the idea that experimental studies, in which you “do” the intervention rather than just observe an exposure of interest, guarantee that interventions are well-specified. That is not correct, because one may under-specify an intervention, even if one actually “does” it. For example, “one hour of strenuous exercise a day” could include boxing, powerlifting, running, and so forth, and these exercises might have different effects on mortality. So well-specified interventions are not guaranteed in experimental trials.

Nonetheless, the emphasis on clearly specified alternatives lends itself better to implementation in experimental studies than in observational ones. And among observational studies, it lends itself much better to cohort studies than case-control studies, which—despite their enormous influence in the history of epidemiology—are barely discussed at all in the POA literature, receiving just one page in the authoritative POA textbook (Hernán and Robins 2015, 98).
Thus, the nature of causal inference in epidemiology is undergoing active and energetic rethinking. There are real philosophical issues at stake in this rethinking. These can be divided into two main kinds. One concerns the nature of causation. Do causal claims only make sense if they are relativized to a clearly specified intervention? Must the intervention be humanly possible or only physically possible—either to be admissible in a causal question or to make the answer useful? Must we be able to actually move tectonic plates in order to understand or use the claim “Movements of tectonic plates cause earthquakes”? The other set of issues concerns the methodological implications of the POA. Is the focus on precise formulation of causal questions incompatible with the kind of triangulation that is so central to the history of epidemiology? Does it make space for study designs that are admittedly weak on their own but may provide important pieces of a larger puzzle, as with time-trend data on smoking and lung cancer? Is causal inference an informal, qualitative judgment, located “between” studies, or is it a formal, quantitative finding, located “within” studies?

There is not space here to enter into these debates (but see Further Reading). However, it is worth noting that the POA is partly motivated by the idea that unless one clearly specifies one’s intervention, one is unable to make good predictions about the effects of a proposed intervention. This connection is both welcome and misguided. It is welcome because, as the POA has noticed, there is a lack of clear thinking about prediction and about the way causal knowledge can be used to predict, as we will see in the next section—both within epidemiology and much more broadly. It is misguided because the POA sets the generation of useful predictions as a necessary condition for a causal claim to be well-formulated. This is too strong, as we will see in the next section as well.

Prediction

Prediction is clearly central in epidemiology, and epidemiologists often use predictive language and concepts. For example, they will describe a risk factor as predicting a certain outcome. However, the exact nature of epidemiological prediction has received less attention than one might expect within the theoretical branches of epidemiology. This may be because the paradigmatic problem of epidemiology, the question of whether smoking causes lung cancer, was about causal inference. Once that causal inference was made, good predictions appeared to follow fairly automatically. Before smoking, lung cancer was rare; smoking was responsible for the increase in the incidence of lung cancer; thus it seems barely to require discussion that reducing smoking prevalence will result in a decrease in lung cancer incidence.

However, the apparent ease of moving from a piece of causal knowledge to a prediction in this instance is misleading. There is no denying that causal knowledge is useful for predicting and, in particular, for predicting what will happen if one intervenes to bring a certain situation about. However, it is still a further step to make a good prediction using causal knowledge. This can even be seen in the context of smoking and lung cancer, where some false predictions were made. Reducing the quantity of tar in cigarettes appeared not to help much, since smokers adapted their habits to get the tar anyway (e.g., puffing more, covering ventilation holes with their fingers). And advising smokers to take shallower puffs appeared to make matters worse, since—it turns out—the tissue at the top of the lungs is more susceptible to carcinogens in tobacco smoke. (Both cases are discussed in Parascandola 2011.)

The dangers of assuming that causal knowledge is sufficient, or almost sufficient, for prediction are part of what drives the POA, as already noted. Merely knowing that obesity causes a certain excess mortality does not allow you to predict how your obesity-reduction program will affect mortality. To take another example, establishing that acetaminophen can cause asthma (for which there is inconclusive evidence) does not license a recommendation not to give
acetaminophen to children who are at risk of asthma, as one eminent pediatrician suggested (McBride 2011). It may be that fever and alternative fever-reducing medication would be just as likely to cause asthma, or more so (Broadbent 2012). This difficulty also connects with the difficulties previously identified for risk factor epidemiology: a causal risk factor is typically difficult to use for prediction in a public health context, precisely because the knowledge that a certain risk factor is causal does not tell you what the effect will be of an intervention on that risk factor.

Even if the causal knowledge has been obtained through asking a causal question that corresponds to a well-specified intervention, the most that licenses you to predict is what would have happened had the unexposed group been exposed or vice versa. It does not allow you to predict with confidence beyond your study group. This is sometimes called the problem of external validity (where internal validity is the soundness of a causal inference within a study, and external validity is the transportability of that conclusion to other populations). This is not an entirely happy formulation because it implies that a study can be rendered externally valid, in some general way. A better way to put the question is Nancy Cartwright’s question: “Will this work for me?” A prediction in relation to a particular population must always take account of the properties of that population, and must take an all-things-considered, total-evidence view as to whether the same outcome will be found as was found in the studies that are used as an evidence base. Otherwise, the possibility remains open that some difference between the study and target populations matters for the prediction (Cartwright 2011; Broadbent 2011a).

It does not help matters that the philosophy of science is very poor in literature on prediction as a topic in its own right. Prediction receives far less attention from philosophers than do explanation, causation, and laws of nature (Douglas 2009; Broadbent 2013, 84–89). If there is one area ripe for new work in the applied philosophy of science, it is prediction, and epidemiology is a prime example of a science where philosophical attention to prediction could be really useful.

References


hernan/causal-inference-book/.
the Royal Society of Medicine 58: 259–300.
Koch, R. 1876. Verfahren Zur Untersuchung Zur Conserviren Und Photographie Der Bakterien, Beitrag Der
Pflanzen. Breslow: Cohn's Bier.
versity Press.
Determinants of Health Research.” Bioethics 23 (2): 79–89.

Further Reading

Broadbent, Alex. 2013. Philosophy of Epidemiology. New Directions in the Philosophy of Science. London
and New York: Palgrave Macmillan.
Broadbent, Alex. 2015. “Causation and Prediction in Epidemiology: A Guide to the Methodological
Revolution.” Studies in History and Philosophy of Biological and Biomedical Sciences, 54, 72–80.
the Royal Society of Medicine 58: 259–300.