Commentary: What can epidemiology accomplish?

Sharon Schwartz*1 and Ezra Susser1,2

Over the past 20 years epidemiological studies of risk factors have led the way to a great many discoveries useful to public health. Some particularly striking examples are found in child health. We can point to the prevention of Sudden Infant Death Syndrome by sleeping position of the infant, the prevention of neural tube defects by periconceptional folate supplementation, and the improvement of vision and other health outcomes by giving children vitamin A supplements.

Yet over the same time period, the discipline was shaken by events that made the limitations of studying risk factors all too apparent. Notably, epidemiologists were caught unprepared by the AIDS pandemic. The identification of sexual and drug use behaviours as risk factors for HIV transmission was a timely and extremely important, but nowhere near sufficient, response to what became a public health catastrophe. Such events have provoked many of us to rethink our designs and methods. This is reflected in a surge of experimentation with other approaches, including mathematical modelling of infectious disease epidemics and other dependent happenings, analysis of historical trends, multilevel studies that take account of contextual variables, life course studies that examine the developmental origins and trajectory of health and disease, and designs, such as Mendelian randomization, that make use of genomics to answer epidemiological questions.

The limitations of risk factor studies are taken up in the accompanying article from Buchanan et al.2 Their critique indicates many of the central weaknesses and underscores that these pertain to genetic as well as non-genetic factors. Thus, current practice fosters a tendency towards larger studies of smaller effects. Contributing to the distortion is an incongruity between various formal procedures of causal inference and the actual process by which it occurs. Furthermore, the premises of our research designs, and the constraints imposed by these premises, are not often made explicit. Consequently, the designs used to detect risk factors are sometimes applied to questions for which they are not well suited, for example, questions about historical trends. With these difficulties in mind, the authors liken the search for causes of complex diseases to the alchemist’s quest for the Philosopher’s Stone. They ask whether there is any realistic prospect for major advances in our understanding and prevention of chronic diseases.

Our answer is a cautious yes. First, the identification of risk factors is important to public health, and epidemiology’s track record in this area is quite good. Attesting to their continuing relevance, our staple cohort and case–control designs played a major part in the breakthroughs in child health noted earlier.

Second, we are optimistic that the current round of experimentation will generate new methods to answer many of the questions that fall outside the purview of risk factor studies. Many of the efforts to develop new approaches share common themes. For example, there is a growing recognition of the need to differentiate levels of causation. To understand a historical trend in disease for a whole population, it is not sufficient to study only the factors that account for variation in the disease risk of individuals within the population.3 Changes in the characteristics of the population as a whole, over time, need to be examined. A new approach is unlikely to be widely accepted until there is a landmark achievement demonstrating its utility. As yet, there is no equivalent to the discovery of microorganisms in the 19th century (which ushered in infectious disease epidemiology) or smoking as a cause of lung cancer in the 20th century (which ushered in chronic disease epidemiology). But we may be optimistic that as we move from recognition of the problems to the development of new ways of thinking and new designs and methods, signal achievements will follow.

Third, we see considerable continuity as well as some discontinuity between the problems faced by epidemiologists of yesteryear and of today. Currently, a disease is generally considered ‘simple’ if it is known to be closely related to an infectious vector or a specific gene. However, the causes of these diseases are not really simple. Before the identification of such causes these diseases often appeared very complex. Afterwards, they were simplified, by redefining the disease according to the presence of an infectious agent or genetic abnormality, and/or by focusing on features critical to prevention. The discovery of causes often happened despite the complexity rather than because of the simplicity of the diseases being investigated. The authors themselves draw attention to this point. They note that only rarely is a cause truly necessary and sufficient. The effect of exposure to an infectious agent on disease in individuals and populations tends to be contingent and somewhat unpredictable. We usually cannot tell which individuals will fall prey to a particular infectious disease nor accurately predict the rise and fall of infectious epidemics across time and place. Genetic

---

1 Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA.
2 Department of Psychiatry, Epidemiology of Brain Disorders at the New York State Psychiatric Institute, New York, USA.
* Corresponding author. E-mail: sbs5@columbia.edu
heterogeneity imposes a layer of complexity on even ‘Mendelian’ diseases; hundreds of different mutations may be involved in the transmission of such a disease, albeit within different families. It is, therefore, unclear why they draw such a sharp distinction between the prospects for epidemiology in the past (when it was undeniably successful) and in the present. We suggest that over successive eras of epidemiology, the complexity of the diseases confronted has not changed as much as the methods of investigating them. Each era has been characterized by both success in identifying some kinds of causes and myopia towards others. The present does not compare unfavourably with the past in this respect.

Still, the authors’ critique is sharp and serves the useful purpose of challenging us to improve risk factor studies. Most of us recognize that our staple designs will probably play an important role for the foreseeable future and will still be needed after new approaches are developed. So it was particularly jarring for epidemiologists in recent years when these tried and true approaches proved surprisingly fallible, as several major findings from large epidemiological studies were overturned in clinical trials. Acknowledging that these designs have failed us in some instances, we need to consider ways to build in safeguards against inappropriate causal inferences. As discussed briefly below, and in more detail elsewhere we propose that current approaches could be improved by more clearly specifying what we hope to accomplish in these studies and how we go about it.

What can risk factor studies accomplish?

One way to prevent inappropriate inferences is to better specify what we can and cannot accomplish within the constraints of our research designs. Adhering to the analogy of the alchemists’ search for the Philosopher’s Stone, the goal itself was misguided. No methods could have been devised to achieve it.

A legitimate goal within the rubric of risk factor studies is to isolate some (though not all) of the factors that are in the totality of causes of a disease. In other words, what we can do is identify factors that play a role in causing the disease under certain circumstances. As we proceed to establish a given cause, the specification of antecedents, mediators, and co-factors helps to make our tests more stringent and, thereby, solidify our inferences.

An illusory goal would be to produce the one ‘true’ effect of such a cause. This would require the relative risk to be the same across populations and over time. A constant effect seems illusory even for infectious and Mendelian diseases with identified agents. Because most causes are insufficient, they require the presence of other risk factors to cause disease. Since most phenomena are influenced by social, economic, and material conditions, it is likely that the prevalence of these factors will vary significantly across time and place yielding different effect estimates. Therefore, the relationships found in a particular study will always be to some degree population specific.

This does not mean that our findings cannot have general significance. Rather, it means that the generalizability of our findings will depend on learning how and under what circumstances a cause can act, and on understanding the disease process related to the uncovered cause. Understanding how a particular genetic mutation leads to disease will help us to understand the role of other mutations that have similar actions even in societies where the original mutation does not exist.

When we understand how environmental factors cause disease, we may be able to prevent disease in the future in groups that do not yet have this factor but may later acquire it. The designs best suited to explore the social and biological contingencies under which a risk factor will lead to the disease may or may not be risk factor studies, depending on the biology of the disease, the feasibility of various approaches, and our hypotheses.

Another illusory goal would be to accurately predict disease for individuals. Epidemiology is a population-based science that uses individuals as the units of analysis but can only make statements about the average effects for these individuals. The more we understand about causes and the contingencies required for causes to lead to disease, the more closely we can tailor advice to particular individuals. But epidemiologically derived profiles of individual risk are inherently imperfect. Disease risk is an ever changing probability depending on past experience and current and future behaviours and context.

This limitation need not prevent us from using epidemiological findings as a basis for public health policy. Disease rates can be reduced through decreasing causes without identifying the individuals in whom disease was prevented. This is true even for infectious diseases where the agent, the necessary cause, has been identified. For example, we make vaccination policies without knowing which particular individuals would have had the disease without the vaccine.

How do we go about it?

The goal of detecting a cause is realistic but never easy. As emphasized by Buchanan et al., there is no formula that can be used to determine if we have found one. In drawing conclusions from our data, statistical inference, causal criteria, Popperian falsification, and Bayesian approaches each have something to offer, but all are fallible and none can proclaim to have a rule. Our conclusions necessarily depend on consensus derived from principled arguments. Progress will be halting and we will continue to make mistakes. This is especially true for causes with small effects.

Still, by being much more explicit about how we go about the task, and making some modest emendations to our practice, we could reach useful conclusions with greater frequency. We suggest one strategy along these lines, which might be termed ‘quasi-Popperian informal Bayesian’. It interweaves several approaches, while proceeding systematically and explicitly from theory to conclusions.

In this construction of how epidemiology should proceed, before we embark on a study we have theories about how a disease develops. By a theory, we mean a set of connected testable propositions that are more abstract than simple descriptions of observed regularities, but less inclusive than laws of nature. Our theories derive from clinical observations, past research, articulated debates, serendipitous findings, and informed hunches. How elaborated the theories are depends on the state of knowledge about the disease under investigation.

In formulating our theories, the causal criteria offer valuable guidance. The criteria of strength, consistency, and specificity can be used as probes for understanding and integrating extant research. To do so requires recognition of the premises behind each criterion and consideration of its applicability to the specific disease under investigation. For example, we may
consider the strength and consistency of the current evidence and postulate explanations for inconsistencies between previous studies. How well tested are the hypothesized causal effects? How consistent are past results? Are the inconsistencies between studies due to differences in methods? How large are the effects found in previous studies? Do the effect sizes differ and, if so, are they within the range of sampling variations, or do they represent real effect modification due to true differences in the contexts under which the exposure has an effect? Postulated answers to these questions form the factual underpinnings of our theories and the basis for the development of competing theories that warrant testing. In weaving these ideas into reasonable theories about disease causation the criteria of coherence, biological plausibility, and analogy are useful.

The theories can then be elaborated by specifying the mediators and causal partners implied by these theories and identifying patterns of results that would discriminate among competing theories. Reference to these guidelines helps to systematize our thinking throughout this process.

The evaluation of past evidence is inevitably tainted by subjective elements and uncertainty about the data. Theory development is limited by the state of current knowledge, normative constraints, and the creativity required to consider the full array of competing hypotheses. Bayesian approaches provide guidance in weighing these uncertainties and gauging the level of confidence we have in our theories before we begin testing them. We can also make explicit the types of data that would significantly change our belief in the competing theories and design studies to collect these data.

Studies are designed to provide strong tests of theories.7 Strong tests, developed along Popperian lines, 8 are those that confront hypotheses derived from theories. They can shift the balance of evidence towards or away from theories about disease causation and are, therefore, capable of changing our confidence in them. Ideally, all reasonable extant theories for a phenomenon are tested until only the most viable theories are left standing. Studies that pit one theory against another can be especially useful in diminishing the number of viable theories. The corollaries of competing theories are spelt out to develop tests that would produce different results under these theories.

Not only our theories but also our tests are all fallible.9 Recognition of this fallibility is a first step to mitigating it. Faulty logic plagues the development of corollaries to be compared with the data and limits the extent to which the data speak to the theory. In addition, our data provide only approximate and fallible representations of the facts in the population studied.

Statistical tests and confidence intervals help us to assess the precision and reliability of the data. Sensitivity analyses can be used to probe our data for validity. Sensitivity analyses examine the robustness of our data patterns to errors related to bias and chance.10,11 We examine our data making different assumptions about the amount of measurement error, unmeasured confounding, and other methodological artefacts. Trustworthy data will show the same basic pattern of results over a broad range of assumptions. For example, if we suspect that part of the association between our exposure and disease is due to recall bias, we can reanalyse the data to see how large this bias would have to be to appreciably change the association. If the results only change under unreasonable scenarios, we have more confidence in our data patterns.

The combination of the original confidence in our theories and the confidence in our data should move us to change our view of the original theories. Whether or not our new belief is sufficient to declare a risk factor a cause depends in part on how well the data confronted the theory, whether new problems arose that make our conclusion suspect, and the number of plausible alternative explanations that remain. But our decision will also depend in part on what is at stake if we are wrong in either direction. Something is lost and something gained in both action and inaction. Laying out the processes involved will not ensure agreement or the validity of conclusions drawn but will help to reveal the issues that need to be resolved to reach consensus.

Example

This process is exemplified in the investigations of social causation of schizophrenia. Social causation posits that social disadvantage can play a role in the development of this disease. Based on this causal idea, it was hypothesized that low social class is a risk factor for schizophrenia. This hypothesis predicts that low social class should—like any risk factor—be associated with schizophrenia and be temporally antecedent to it. An association between low social class and schizophrenia was readily established. In fact, early evidence for this association helped stimulate the idea of social causation.

It was far more difficult, however, to discern the temporal order. Reverse causation was a potent possibility. On the one hand, as predicted by the theory of social causation, low social class might increase the risk for the disease. On the other hand, as predicted by a competing theory of social selection, early manifestations of schizophrenia might cause individuals to attain a lower than expected social class. In a classic study, Goldberg and Morrison devised a means to discriminate between these two theories.12 They reasoned that under social causation but not social selection, the low socioeconomic status of the family of origin—as well as that of the afflicted individual—would be associated with schizophrenia. They found that low socioeconomic status of the family of origin was not associated with schizophrenia. They concluded that social causation did not match the data. Later studies, using various approaches directed towards addressing remaining questions came to similar conclusions.13

Thus, investigators over a period of time generated a theory, formulated hypotheses, elaborated a set of predictions from these hypotheses, designed studies to test the predictions in empirical data, and ultimately rejected the idea, at least in its original form. Foreseeing and specifying the competing explanations for the exposure disease association was a crucial part of this process. This enabled investigators to design studies that afforded a rigorous test of the causal theory.

Conclusion

We are confident that epidemiologists can develop alternative approaches to risk factor studies. There is already a discernible shift away from a narrow focus on static proximal risk factors towards consideration of context, contingency, and non-linearity. It is increasingly clear that we cannot understand the whole by studying only its constituent parts and vice versa. Thus, we see epidemiologists edging towards a paradigm that includes complexity.
At the same time, we think that we can improve our track record in these studies through careful consideration and testing of explicit causal theories. For we agree with Hill\textsuperscript{14} (p. 300) that ‘All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.’

References
\begin{enumerate}
\end{enumerate}

Commentary: Understanding sources of complexity in chronic diseases—the importance of integration of genetics and epidemiology

Kathleen Ries Merikangas,\textsuperscript{1*} Nancy C.P. Low\textsuperscript{1} and John Hardy\textsuperscript{2}

Progress in identifying genes for nearly all known Mendelian diseases spurred by mapping of the human genome generated a nearly universal expectation that the same research strategies would eventually be successful in identifying genes for complex diseases, such as heart disease, obesity, cancer, diabetes, and many common psychiatric conditions. As the failure to identify and replicate genes for complex disorders has become increasingly apparent, enthusiasm for rapid success has waned. In ‘Dissecting Complex Disease: The Quest for the Philosopher’s Stone?’ Buchanan et al.\textsuperscript{1} admonish geneticists and epidemiologists who are engaged in searching for either genetic or environmental risk factors for complex diseases that there may be no gold at the end of the rainbow (or that the philosopher’s stone will generate base metals rather than gold; or that there is no ‘fountain of youth.’)\textsuperscript{2}

Although the target readership of this essay is not obvious, chronic disease epidemiologists alone are clearly the wrong audience. The concepts that the authors discuss are well recognized by most serious epidemiologists engaged in research that includes genetic risk factors. The major points raised in this essay on the characteristics of complex human diseases

\begin{enumerate}
\item Section on Developmental Genetic Epidemiology, National Institutes of Health, National Institute of Mental Health, 35 Convent Drive, MSC#3720, Bethesda, MD 20892, USA.
\item Laboratory of Neurogenetics, National Institute of Aging, National Institute of Mental Health, 35 Convent Drive, Bethesda, MD 20892, USA.
\item Corresponding author. Section on Developmental Genetic Epidemiology, National Institutes of Health, National Institute of Mental Health, 35 Convent Drive, MSC#3720, Bethesda, MD 20892, USA.
\item E-mail: merikank@mail.nih.gov
\end{enumerate}