Introduction: Seeing the Forest and the Trees—From Observation to Experiments in Social Epidemiology

Lisa F. Berkman

From the Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA.

Received for publication April 1, 2004; accepted for publication April 12, 2004.

Social epidemiology, disguised in other forms and known by other names, has been with us for decades, if not centuries. During periods of rapid urbanization and industrialization in the 19th century, work by Villerme (1), Virchow (2), and Engels (3) identified the health consequences of the political economy. Durkheim (4) revealed the social patterning of suicide. For years, demographers described the social forces that shaped mortality, fertility, and population profiles. As we moved through the early and mid-20th century, a host of classically trained epidemiologists such as Frost (5), Goldberger et al. (6), and Winslow (7) and social scientists such as Sydenstricker (8), Faris and Dunham (9), and Parks (10) produced complex, nuanced understandings of how social and economic forces shape patterns of health and disease. Beginning in the 1930s, studies of mental health focused on the impact of social disorganization and social deprivation at the neighborhood level (9). In fact, the “Chicago School” of sociologists emphasized the role of neighborhoods as central to well-being. If we think of social epidemiology as focused on the central question of how social conditions give rise to patterns of health and disease in populations, we may well wonder what brings us to celebrate this collection of work at this particular moment. In the broadest sense, the field is not a new one, nor are epidemiologists the only scientists contributing to a deepening understanding of the social determinants of health.

Discussions of multilevel analysis (11), integration of group- and individual-level exposures (12), human development and life course approaches (13), and identification of mediating pathways (14) dominate this issue. These questions are central to epidemiology, not just to social epidemiology. The authors represented here address the major issues of our time—new infectious diseases such as human immunodeficiency virus/acquired immunodeficiency syndrome (15), health in developing countries (16), the growing epidemic of diabetes mellitus (17), mental disorders (18), substance abuse (19), and human growth and development (20). Again, the relevance for all epidemiologists is striking.

Building on the contributions of scientists in public health, medicine, and the social sciences, the social epidemiologists contributing to this issue are developing a new framework with which to examine the etiology of disease, patterns of population health, and modes of intervention. I would argue that the important insights of social epidemiologists are based on the training they have in the assessment of health, disease, and biologic pathways and the capacity to integrate “upstream” social dynamics into their modeling of disease causation. In this overview, my aim is to integrate some of the themes expressed by the authors in this issue and to identify some areas that, in my view, remain underemphasized. Of particular importance is the underutilization of experimental and quasi-experimental designs evaluating the potential health impacts of social and economic policies. The integration of health outcomes with experiments related to social and economic policies might well lead to more effective evidence-led policies aimed at improving population health and reducing social disparities in outcomes.

GETTING THE COUNTERFACTUAL RIGHT

Recent advances in thinking about causal inferences have strengthened work in epidemiology (21–23). Central to this endeavor is the development of counterfactual questions designed to aid epidemiologists in developing and testing specific hypotheses. Some epidemiologists have questioned...
the validity of posing counterfactuals in social epidemiology, arguing that social characteristics such as sex and race are immutable. Holland (21) initially claimed that "causes" are only those things that, in principle, can be modified. Greenland (23) and others have criticized this early work, describing the utility of identifying susceptible social groups and exploring ways to reduce social disparities. They have used these arguments to question the validity of scientific work in this area.

A counterfactual question is generally set up by asking, “What would this person’s health be like if he or she had a different exposure?” As an example, we might ask, “What would this diabetic’s risk of cardiovascular disease be if he or she didn’t have diabetes?” With regard to sex, socioeconomic position, race, or marital status, critics often say this is a useless question to pose, since we cannot change sex, race, etc. As Ana Diez Roux has framed in her paper in this issue (12), this counterfactual can, at first glance, become even more strained as we ask the multilevel question, “What would this person’s health be like if he or she had the same income but lived in a neighborhood with a lower mean income than the one he or she actually lives in?”

In these examples, the utility of the counterfactual depends completely on getting it “right.” Getting it right means thinking through and identifying what we really think is the causal component of a risk factor. In fact, that is precisely why counterfactual arguments have come to be so compelling. For social epidemiologists, the identification of a social group at risk (e.g., a particular racial/ethnic group or sex, the poor or isolated, recent immigrants) is usually based on the premise that something about the group’s social or economic experience is risky, not something about their innate biologic or genetic background (24). Thus, in most cases, we hypothesize that it is the experience of discrimination that places African Americans at risk, not their skin color or racial background per se. The correct counterfactual is not “What would it be like for an African American to be Caucasian?” but rather “What would it be like for an African American to experience less discrimination?” The correct counterfactual related to hypotheses about why middle-aged US men have higher rates of coronary heart disease than women is not “What would this man’s risk be if he were a woman?” Rather, it may be “What would the American man’s risk of coronary heart disease be if gender roles did not encourage specific high-risk behaviors?” Similarly, we may decide that the explicit comparison regarding US children who live in poor families headed by a single parent is a comparison of their households with single-parent households in other countries, because the related risk we are considering is the social policies we have in place for single parents in the United States. Our implicit comparison might otherwise have been between such households and economically better-off two-parent households.

In all of these cases, working on expressing the right counterfactual is extremely useful in social epidemiology, as it is in other areas. Getting the counterfactual correct pushes us to identify and then subsequently test what we think of, often implicitly, as the etiologic agent. It helps us make the right comparisons using observational data. As we will see, it can help guide us in the development of studies using experimental or quasi-experimental designs and approaches. Being thoughtful about a counterfactual can lead us to identify what is modifiable about a social experience.

**EXPERIMENTS: THE ACHILLES’ HEEL OF SOCIAL EPIDEMIOLOGY**

In the epidemiologic, biologic, and social sciences, there is a hierarchy of study designs in which experimental evidence carries more weight than do results from observational approaches. As someone who has struggled with ways to study social networks, poverty, status, and discrimination, all of which are conditions that do not lend themselves to easy or ethical manipulations in the health arena, I have embraced the strengths of observational studies. Observational studies that can sort through temporal ordering and confounding offer powerful approaches to our understanding of disease causation. They do not, however, tell us everything, and even strong studies can sometimes generate misleading results. First and foremost, observational studies do not tell us whether we can reduce the risk of poor health by changing the exposure. Only experiments can do that. I am not suggesting that we stop conducting observational studies. I am, however, suggesting that we learn to exploit the power of experiments to help us identify causal “agents” in the social environment, to clarify critical time periods in the development of disease risk, and to implement new interventions for improving health. Observational studies take us only part of the way.

**The randomized clinical trial: the queen or the pawn?**

In medicine and public health, the experimental paradigm has taken the form of randomized clinical trials. Classically, in such trials, individuals are randomly assigned to a treatment or control group, both participants and experimenters are blind with regard to treatment assignment, and all subjects are included in the final intention-to-treat analysis. We know the rules; they are given in most introductory epidemiology textbooks. Trials testing the efficacy of drugs or specific surgical procedures are often, though not always, well suited to this approach. For the most part, implementers of behavioral interventions have followed this biomedical paradigm, with modest results (25–27). Studies of social and psychological interventions have started to follow randomized clinical trial designs, most often showing small or null effects of the intervention (28, 29).

While randomized clinical trials may have been used to test primary prevention efforts, one might argue that they are best suited to testing secondary or tertiary prevention efforts among persons who are at high risk or clinically ill. They also work well in populations in which we can expect that our interventions will have an impact on the mediating mechanism relatively quickly (e.g., antihypertensive agents influence blood pressure), as well as when we expect outcome events to accrue during the relatively short time frame (a few years) of a study. They are less well suited to studies of primary prevention, though there are some notable exceptions to this, especially with regard to vaccine trials. Randomized clinical trials are difficult to implement when it
is impossible to blind the intervention or when individuals must alter their own behavior in substantial ways. To the degree that the intervention is difficult for participants to implement, the effect of the intervention will be diluted by the number of people not getting the “dose” we hypothesize is needed to obtain a benefit. Intention-to-treat analyses that do not capture this specific treatment effect have sometimes been referred to as tests of a policy or as attempts to offer the “treatment” rather than tests of the effect of the treatment per se (30). However, intention-to-treat analyses do offer important controls for selection, and from a policy perspective they offer more realistic assessments of what effects are likely to result from policies once they are enacted. Randomized clinical trials have been used to evaluate social interventions in a number of domains (30–32), though they have been used more often in medicine and public health.

There has been a great deal of attention devoted recently to the discussion of whether we should believe the results of randomized clinical trials or observational studies when the results from different studies differ in important ways. Many of these discussions have occurred over topics such as hormone replacement therapy, diet, and other interventions in which behavioral action on the part of participants is central. These discussions have great relevance for social interventions as well.

Over the last 10 years, I have been involved in the development and implementation of a clinical trial among postmyocardial infarction patients who were either socially isolated or depressed (33). The intervention was a psychosocial one aimed at the reduction of subsequent cardiac events and mortality through the reduction of the two main psychosocial conditions, depression and social isolation. The trial showed no difference between the intervention group and the usual-care group in terms of the primary risk endpoint of subsequent cardiac events or death. This has left the behavioral medicine community discouraged, even though in many ways the trial was a huge success. For instance, the trial recruited a large number of women and members of racial/ethnic minority groups. We kept participants involved in the trial and succeeded in modifying the mediating mechanisms we hypothesized were central to the intervention. Over the past several years, a number of large behavioral and psychological interventions have been disappointing in terms of altering health outcomes, even when prospective observational studies have shown strong and consistent associations (28, 29, 34). The reasons for the lack of effect may be the same as those for many other types of trials. That is, perhaps 1) our intervention was not powerful enough; 2) it could not affect the outcomes soon enough, given the late stage of disease progression; 3) the effort to reverse the intermediate outcomes (depression and social isolation) was made too late; 4) the cumulative, potentially lifelong exposures involved had already taken their toll; or 5) there was too much improvement among persons in the usual-care arm of the trial.

On the basis of results like this, many social epidemiologists are reluctant to subject their findings to an experimental design that they believe is not well suited to the factors being studied. Many are skeptical about the value of experimentation based on our past efforts with randomized clinical trials, believing that experimental designs may not fit well with social epidemiologic paradigms. They say that we are bound to fail with such designs. In response to these viewpoints, I would suggest that there is a great deal that can be done about these conditions. In fact, if we go back to the above discussion about counterfactuals, we can develop some sensitive designs that will test our hypotheses well and simultaneously guide our intervention efforts to improve population health. Our fears of using experimental designs are based on the narrow biomedical paradigm of the randomized clinical trial. This paradigm often leads us to focus on individual-level interventions, behavioral changes that are decontextualized, and short-term follow-up. Thus, the biomedical paradigm frequently translates into selecting high-risk populations and/or to studying persons who already have clinical disease.

Unintended consequences: the health effects of social and economic policies and experiments

If we shift the biomedical paradigm and look at large-scale social or economic policies or relevant experiments, we maintain the power that experimental or strong quasi-experimental designs provide and test interventions that are designed at an ecologic level appropriate to the level at which we, as social epidemiologists, suspect that action and change may most effectively occur. For instance, with regard to tobacco and alcohol consumption, evidence points to the effectiveness of national- and state-level taxation policies in reducing consumption. The use of large-scale social experiments may not solve all of the problems inherent in experimental designs (for instance, the acceptance of the intervention in the comparison group), but they will deal effectively with many. Furthermore, Cook (35), a social scientist who has been involved in evaluating the effects of social interventions, has called for an end to “black box” experiments and has recommended that investigators put more thought into the design of experiments leading to an identification of mediating and moderating mechanisms.

We know relatively little about the health impact of most large-scale social and economic policies implemented over the last few decades in both the United States and throughout the world. For instance, there has been a minimal amount of research on the health effects of income supplementation programs, ranging from the minimum wage to the earned income tax credit to Social Security to labor policies (such as flex-time and parental/family leave). While it is true that some large-scale programs do not easily lend themselves to experimental or quasi-experimental evaluation, many social interventions have had periods in which they were tested using an experimental design (e.g., the High/Scope Perry Preschool Study (36) and the Moving to Opportunity Project (37)). The Moving to Opportunity Project compared groups of young people initially living in high-poverty public housing (37). An experimental group of participants was offered federal Section 8 housing-choice vouchers valid only for living in a low-poverty area; another group was offered Section 8 vouchers without geographic restriction; and the third group served as a control group. Females in both treatment groups benefitted across a number of domains, from
education to reductions in risky behavior and better mental and physical health. Males in both treatment groups were more likely to engage in risky behavior and to experience worse physical health. The Moving to Opportunity Project is an excellent example of evaluating a policy using a strong experimental design and incorporating a number of health measures as outcomes.

Other policies (welfare-to-work policies, immigration policies, civil rights laws) have been implemented in such a way that we have been able to make comparisons with people who were very similar but did not have the benefits of the policy. For instance, some research on the implementation of Social Security for Americans over the age of 65 years has compared men and women aged 64 years with those aged 65 years (38). The effects of welfare-to-work policies have been tested by examining the behavioral changes that occurred among children whose mothers received benefits during a rapid change in policies (39). In the latter study, Chase-Lansdale et al. (39) reported that as women leave welfare for work, there are few poor health consequences for their children. Adolescent children in particular may do better. Chase-Lansdale et al. speculated that some of these neutral or positive outcomes may be related to the high rates of employment among the women who left the welfare rolls. These high rates were possible in strong economic times, but rates might be much worse during economic downturns for such a vulnerable segment of poor women with young children. While it is sometimes hard to know whether the effects of these policies are specific to the actual policy or to other policies or economic trends occurring simultaneously, these evaluations have the advantage of being unrelated to individual selection issues, a major point of contention in observational epidemiologic studies. The most problematic part of these efforts is not finding a suitable experimental setting in which to evaluate the impact of social and economic policies but finding valid indicators of health that can be linked to evaluation efforts.

Attention to the evaluation of public and private social, economic, and environmental policies has been discussed in terms of health impact assessments in a number of settings over the past several years (40, 41). As those authors have noted, such impact assessments have both strengths and weaknesses. Health impact assessments often have forecasting as well as evaluation components. Forecasting future health impacts based on previous observational data seems a much riskier venture at this time than integrating health assessments into ongoing policy evaluations. The latter would enable us to develop evidence-led policies that might well result in reductions in health disparities and improvements in population health. Such linkages would enable us to identify some of the specific actions that the public and private sectors might take to reduce social disparities in health. Coupled with an iterative research agenda based on observational studies and a clear set of hypotheses as to mediating and moderating mechanisms, social epidemiologists would have the opportunity to identify the social conditions across the life course that lead to poor health. Building on the results of social and economic experiments would also provide us with valuable guidance regarding what we can do to improve population health and reduce social disparities in morbidity and mortality.

Over the last 30 years, epidemiologists have focused predominantly on proximate health risks conceptualized at the level of the individual and have subsequently promoted interventions that rely on change emerging from individual-level behavioral choices. The modest results of many of the public health interventions based on this model have led to a new generation of studies in which individuals are seen in a social and environmental context. Ecosocial models of health and disease position individuals in the social and economic context in which they live in ways that permit us to consider new sets of health risks and protective factors. The articles in this issue of Epidemiologic Reviews attest to the value of this perspective. Determinants of population health include an array of macro-level social, economic, and environmental conditions that are shaped by numerous public- and private-sector social, economic, and environmental policies. Our next challenge is to incorporate rigorous evaluations of the large-scale social and economic changes and related policies that have influenced the enormous increases in life expectancy and both reductions and increases in health inequality that have occurred over the last century. Capitalizing on experimental and quasi-experimental designs will allow us to investigate these difficult issues, which many epidemiologists have been reluctant to address.

ACKNOWLEDGMENTS

This work was supported by the Russell Sage Foundation Program on Inequality and by the Robert Wood Johnson Foundation Scholars in Health and Society Program. The author thanks Drs. Maria Glymour, Nancy Krieger, and Leonard Syme for their helpful comments on this paper.

REFERENCES

9. Faris RE, Dunham HW. Mental disorders in urban areas. Chi-
cago, IL: University of Chicago Press, 1939.