Social Conditions as Fundamental Causes of Disease*

BRUCE G. LINK
Columbia University and New York State Psychiatric Institute

JO PHELAN
University of California, Los Angeles


Over the last several decades, epidemiological studies have been enormously successful in identifying risk factors for major diseases. However, most of this research has focused attention on risk factors that are relatively proximal causes of disease such as diet, cholesterol level, exercise and the like. We question the emphasis on such individually-based risk factors and argue that greater attention must be paid to basic social conditions if health reform is to have its maximum effect in the time ahead. There are two reasons for this claim. First we argue that individually-based risk factors must be contextualized, by examining what puts people at risk of risks, if we are to craft effective interventions and improve the nation’s health. Second, we argue that social factors such as socioeconomic status and social support are likely “fundamental causes” of disease that, because they embody access to important resources, affect multiple disease outcomes through multiple mechanisms, and consequently maintain an association with disease even when intervening mechanisms change. Without careful attention to these possibilities, we run the risk of imposing individually-based intervention strategies that are ineffective and of missing opportunities to adopt broad-based societal interventions that could produce substantial health benefits for our citizens.

Epidemiology has been enormously successful in heightening public awareness of risk factors for disease. Research findings are frequently and prominently publicized in the mass media and in rapidly proliferating university-based health newsletters. Moreover, there is evidence that the message has been received and that many people have at least attempted to quit smoking, include more exercise in their daily routine, and implement a healthier diet. With few exceptions, however, the new findings generated within the field of epidemiology have focused on risk factors that are relatively proximate “causes” of disease, such as diet, cholesterol, hypertension, electromagnetic fields, lack of exercise, and so on. Social factors, which tend to be more distal causes of disease, have received far less attention.1 This focus on more proximate links in the causal chain may be viewed by many, not as a limitation or bias, but as the rightful progression of science from identifying correlations to understanding causal relationships (e.g., Potter 1992). In fact, some in the so-called “modern” school of epidemiology (e.g., Rothman 1986) have explicitly argued that social conditions such as socioeconomic status are mere proxies for true causes lying closer to disease in the causal chain.

This focus on proximate risk factors, potentially controllable at the individual level, resonates with the value and belief systems of Western culture that emphasize both the ability of the individual to control his or her personal fate and the importance of doing so (Becker

* We thank Bernice Pescosolido, Sharon Schwartz, and Sarah Rosenfield for helpful comments. This work was supported in part by NIMH grants MH46101 and MH13043. Address communications to Bruce G. Link, Epidemiology of Mental Disorders, 100 Haven Avenue, Apartment 31D, New York, NY 10032.
This affinity between cultural values and the focus of contemporary epidemiology undoubtedly contributes to the level of public interest in epidemiological findings, and probably influences funding priorities as well. Thus modern epidemiology and cultural values conspire to focus attention on proximate, individually-based risk factors and away from social conditions as causes of disease.

This is not to say that the role of social factors in disease causation has been neglected in all quarters. Medical sociologists and social epidemiologists have kept alive classical epidemiology’s (e.g., Susser, Watson, and Hopper 1985) concern with social conditions and have made major strides toward documenting and understanding the connections between social factors and disease. However, we believe there are conceptual pitfalls that sometimes lead medical sociologists and social epidemiologists themselves to unwittingly reinforce the emphasis on proximate, individual-level risk factors. One of these pitfalls is that, in the process of elucidating the mechanisms connecting social conditions to health and illness—an important and desirable activity—we may, over time, lose interest in and come to neglect the importance of the social condition whose effect on health we originally sought to explain. Also, our tendency to focus on the connection of social conditions to single diseases via single mechanisms at single points in time neglects the multifaceted and dynamic processes through which social factors may affect health and, consequently, may result in an incomplete understanding and an underestimation of the influence of social factors on health.

Our purposes here are to highlight the accomplishments of medical sociologists and social epidemiologists in advancing our understanding of social conditions as causes of disease, to underscore the critical importance of continued work in this direction, and to offer two conceptual frameworks that we hope will facilitate and enhance this research. First, we discuss the importance of “contextualizing” risk factors—that is, attempting to understand how people come to be exposed to individually-based risk factors such as poor diet, cholesterol, lack of exercise, or high blood pressure—so that we can design more effective interventions. Second, we introduce the notion that some social conditions may be “fundamental causes” of disease. A fundamental cause involves access to resources, resources that help individuals avoid diseases and their negative consequences through a variety of mechanisms. Thus, even if one effectively modifies intervening mechanisms or eradicates some diseases, an association between a fundamental cause and disease will reemerge. As such, fundamental causes can defy efforts to eliminate their effects when attempts to do so focus solely on the mechanisms that happen to link them to disease in a particular situation. We conclude by discussing the implications of these ideas for research and social policy.

EVIDENCE LINKING SOCIAL CONDITIONS TO DISEASE

We begin with a brief review of the evidence concerning the connection between social conditions and illness. For the purposes of this paper, we define social conditions as factors that involve a person’s relationships to other people. These include everything from relationships with intimates to positions occupied within the social and economic structures of society. Thus, in addition to factors like race, socioeconomic status, and gender, we include stressful life events of a social nature (e.g., the death of a loved one, loss of a job, or crime victimization), as well as stress-process variables such as social support.

Forty years of medical sociology have uncovered numerous examples of the social patterning of disease. Most obvious is the ubiquitous and often strong association between health and socioeconomic status. Lower SES is associated with lower life expectancy, higher overall mortality rates and higher rates of infant and perinatal mortality (Buck 1981; Dutton 1986; Illsley and Mullen 1985; Adler et al. 1994; Pappas et al. 1993). Moreover, low SES is associated with each of the 14 major cause-of-death categories in the International Classification of Diseases (Illsley and Mullen 1985), as well as many other health outcomes, including major mental disorders (Dohrenwend et al. 1980; Kessler et al. 1994). Other examples of the social patterning of disease are plentiful. Males have higher mortality rates at all ages (Walsh and Feldman 1981), as well as higher rates of coronary heart disease (Syme and Guralnik 1987), chronic respiratory diseases (Colley 1985) and ulcers (Gazzard and Lance 1993).
There are pronounced gender differences in rates of various forms of cancer (Prout, Colton, and Smith 1987) and mental disorder (Dohrenwend et al. 1980; Kessler et al. 1994). African Americans have higher rates of overall mortality and infant mortality (Dutton 1986; Miller 1987), renal failure (Challah and Wing 1983), and stroke (Pedoe 1982a) than do Whites, but lower rates of coronary heart disease (Pedoe 1982b); cancer rates also differ by race and ethnicity (Prout et al. 1987). Both physical and mental disorders vary with marital status and population density (Kelsey 1993; Benenson 1987; Robins et al. 1984), and certain religious groups such as Mormons and Seventh Day Adventists have lower risks of some types of cancer (Saracci 1985).

In addition, the tremendous growth and success of the stress paradigm have added considerably to the evidence for an association between social conditions and disease (Dohrenwend and Dohrenwend 1981; Turner and Marino 1994; Turner, Wheaton, and Lloyd 1995). Stressful life events have been linked to heart disease, diabetes, cancers, stroke, fetal death, major depression, and low birth weight in offspring (Miller 1987; Brown and Harris 1989; Shrout et al. 1989). Research has also extended to the domains of social support (Berkman and Syme 1979; House, Landis, and Umberson 1988; Thoits 1982; Turner 1981; Turner, Grindstaff, and Phillips 1990; Turner and Marino 1994) and coping (Pearlin and Schooler 1978), which have been shown to be associated with health and well-being in their own right.

The evidence reviewed to this point clearly establishes a strong and pervasive association between social conditions and disease. But medical sociologists and social epidemiologists have taken the field considerably beyond a description of the social patterning of disease. Important advances in establishing a causal role for social factors have focused on two major issues—the direction of causation between social conditions and health and the mechanisms that explain observed associations. In what follows we present prominent examples of work on these two issues.

Concerning the issue of causal direction, important controversies surround some of the relationships between social conditions and health. For example, does low SES cause poor health, or does poor health cause downward mobility? Does social support reduce morbidity and mortality, or does illness restrict social interaction and thereby lead to social-support deficits? Social epidemiologists have used three general strategies to address these questions.

One approach uses quasi-experimental strategies which involve locating conditions under which alternative explanations make different predictions about observable facts. This approach is exemplified by Dohrenwend’s (1966) quasi-experiment designed to test social selection and social causation explanations for the association between SES and specific mental disorders. The two explanations make different predictions about rates of disorder in advantaged and disadvantaged ethnic groups, when socioeconomic status is held constant. The recent culmination of Dohrenwend’s work on this problem, based on a large-scale epidemiological study in Israel (Dohrenwend et al. 1992), concluded that social causation was stronger than social selection in producing the inverse association of SES to major depression in women, and substance abuse and antisocial personality in men. For schizophrenia, however, the evidence was more supportive of the social-selection explanation.

In the second strategy, medical sociologists and social epidemiologists identify social risk factors that cannot reasonably be conceived as having been caused by an individual’s illness condition. Exemplifying this strategy is a study by Hamilton and colleagues (1990) concerning the effects of plant closings on auto-workers’ mental health. The investigators compared workers who were laid off because of a plant closing, workers who anticipated being laid off, and workers whose plant was not closing, and found that those laid off were more likely to experience negative mental-health consequences—particularly if they were minorities and of low SES. Since the illness condition of the workers cannot be thought of as having caused the plant closing, the differences between the groups studied are more clearly interpretable as the effects of social conditions on health. This strategy was also employed by Fenwick and Tausig (1994) in a study that used the Census-based unemployment rate for an individual’s occupation in a longitudinal design to show that when unemployment is higher, workers’ job satisfaction, decision-making latitude, and well-being are lower. Again, since workers’ health cannot cause the aggregate unemployment rate, the results are more readily interpretable as demonstrating
the influence of social conditions on health-related outcomes. Finally, studies of stressful circumstance have used this strategy by identifying “fateful” life events that are unlikely to have been caused by an individual’s behavior (e.g., death of spouse, plant closing). Thus when Shrout et al. (1989) found the odds of developing major depression to be more than three times as high among people experiencing a recent fateful life event, the association was more clearly interpretable as an effect of social conditions.

The third strategy adopted by social epidemiologists to clarify causal direction involves the use of longitudinal designs. When such designs can clearly place the emergence of an illness or an illness exacerbation before or after the social condition under study, a great deal can be learned about the direction of cause between the two. Unfortunately, some longitudinal studies do not allow clear inferences about time order and therefore do not provide the definitive evidence about causality that is sometimes attributed to them (see Link and Shrout 1992). Still, some notable studies have identified social conditions that clearly predate health outcomes and show that the social conditions predict morbidity and mortality even when competing risk factors are held constant. For example, Berkman and Syme (1979) used baseline data on social networks, collected in 1965, to predict mortality during the subsequent nine years. They found a near doubling of risk for mortality among those low on a social-network index as compared to those high on the index. Although this study controlled for many competing risks (smoking, obesity, physical activity, etc.), it did not include measures derived from a physical exam. A subsequent study by House, Landis, and Umberson (1988) did include a baseline physical exam and controlled for blood pressure, cholesterol levels, and other biomedical variables. These investigators found associations between social relationships and mortality that were similar to those reported by Berkman and Syme. This line of work has continued to become more and more refined. For example, in a recent longitudinal study, Berkman and colleagues (1992) have shown that a measure of perceived support collected before the occurrence of a heart attack predicts survival following the heart attack net of an impressive array of biomedical and psychosocial control variables. Other social variables have also been effectively studied with longitudinal designs. For example, Catalano and colleagues (1993) related job layoffs to the emergence or reemergence of alcohol abuse, and Lin and Ensel (1989) and Ensel and Lin (1991) showed that stressful circumstances predicted subsequent health and mental-health outcomes.

Thus, while medical sociologists and social epidemiologists have not denied the possibility that illness affects social conditions (Johnson 1991), they have, at the same time, demonstrated a substantial causal role for social conditions as causes of illness.

Research identifying the mechanisms linking social conditions to disease has also done much to move social epidemiology beyond the description of social patterns of disease. Consider, for example, the job-stress model of Karasek and colleagues that provides evidence for one mechanism linking SES to coronary heart disease among men. These investigators have shown that “job strain,” characterized by a combination of high job demands and low decision latitude, is more common in lower status jobs and is associated with coronary heart disease (Karasek et al. 1988; Schnall et al. 1990) and elevated levels of ambulatory blood pressure both on and off the job (Schnall et al. 1992). Another example is the work of Mirowsky and Ross (1989), who elucidate the mechanisms that might account for social patterns of distress. They present evidence showing that alienation and perceived control over life circumstances underlie many social conditions that put people at risk for elevated levels of psychological distress. Consider as a final example a study by Rosenfield (1989) that sought to understand mechanisms producing gender differences in symptoms of depression and anxiety. Rosenfield shows that women have higher symptoms of depression and anxiety. This work shows that women have higher symptom levels than men when they are overloaded by work and family demands or when they experience low power as a consequence of being out of the labor market. Moreover, the common mechanism underlying low power and role overload is a decreased sense of personal control, which is in turn related to symptoms of anxiety and depression.

Link and Dohrenwend (1989) explicitly advocate the approach of elucidating mechanisms because of its value in clarifying the relative merit of competing explanations for social patterns of disease. The rationale is that alternative explanations for these patterns, such as social
causation and social selection, frequently imply different intervening mechanisms. Thus, evidence about which intervening mechanisms account for the association can help answer questions concerning causal direction and other competing explanations. Moreover, if causal links between distal factors (e.g., SES) and more proximal factors (e.g., occupational stress, diet) can be drawn, as Karasek et al. (1988), Mirowsky and Ross (1989), Rosenfield (1989), and others (Lenon 1987; Pearlvin et al. 1981; Link, Lennon, and Dohrenwend 1993; Umberson, Wortman, and Kessler 1992) have done, it becomes increasingly clear that social conditions are causes exerting indirect effects on disease outcomes, rather than mere proxies as Rothman (1986) and others might claim.

But are there unintended and undesirable consequences of an approach that focuses on intervening mechanisms? We believe there are. Despite the obvious benefits of such an approach, it is possible that in its enactment, one may inadvertently contribute to the focus on factors that are closer to disease in the causal chain. The intervening mechanism becomes the new and exciting “next step,” while the social conditions become the old, passé “starting point.”

The evolution of the stress paradigm is a good example of such an inadvertent downgrading of the issue which provided the initial impetus for research. The social causation/social selection controversy concerning the association between socioeconomic status and mental disorder spawned an interest in stressful life events as a direct operationalization of the adversity that might be experienced in lower SES contexts (Dohrenwend and Dohrenwend 1969, 1981). When a consistent but modest association between stressful events and illness was identified (see Rabkin and Struening 1976), investigators elaborated the model to consider social support and coping as potential modifiers. Now researchers are invested in understanding the mechanisms linking these factors with disease. Also, research on the biological consequences of stress (e.g., immune status and elevated catecholamines) is seen as an exciting new development (e.g., Cohen, Tyrrell, and Smith 1991). In general, interest has followed the most recent step in the progression toward disease outcomes, while concern with the earlier foci has dissipated to a point where some express disinterest in factors such as the causation/selection issue and the role of stressful life events in causing illness (but see Pearlvin 1989; Dohrenwend 1990; Angermeyer and Klusman [1987] for dissenting views). Indeed Angermeyer and Klusman (1987) documented a sharp decline in the number of publications focused on social class and mental disorder in the period from 1966 to 1985, while the number of articles on stress and psychiatric disorders increased rapidly during the same period. To the extent that interest in mechanisms increases at the expense of more fundamental social conditions, medical sociologists may unwittingly contribute to the emphasis on individual factors and play into the hands of those who argue that social factors have only a modest role in disease causation.

To this point, we have described two characterizations of social conditions as causes of disease that either adventently or inadvertently downplay their importance. One of these is the outright declaration that social factors are only proxies for true causes. This position is demonstrably unwarranted given the achievements of medical sociology and social epidemiology over the past few decades. The other characterization, which may be partially constructed by medical sociologists and social epidemiologists themselves, is the view that social factors serve as starting points whose main function is to point the direction to more proximal risk factors. We take sharp issue with both of these characterizations. In the next two sections, we develop two concepts that illustrate the critical importance of social factors in disease causation, provide conceptual frameworks for future research in this area, and point to the problems that may ensue if the role of social conditions is neglected by researchers and policymakers. These are the ideas of “contextualizing risk factors” and “fundamental causes.”

CONTEXTUALIZING RISK FACTORS

We suggest that medical sociologists and social epidemiologists need to counter the trajectory of modern epidemiology toward identifying risk factors that are increasingly proximate to disease—ones for which “biological plausibility” can be argued. One way they can do this is by “contextualizing” individually-based risk factors. By this we mean that investigators must (1) use an interpretive framework to understand why people come to be
exposed to risk or protective factors and (2) determine the social conditions under which individual risk factors are related to disease. We present examples that illustrate both these principles.

First, an important strategy for reducing the threat of AIDS is to educate the public concerning the steps they must take as individuals to reduce their risk of contracting or infecting others with the HIV virus. Clearly, however, some people are better able to take advantage of this information than others. By contextualizing risk factors for AIDS, we may be able to understand why some people cannot avoid the risk. For example, homeless or other extremely poor women who turn to prostitution as a survival strategy may not have the options or resources that would enable them to refuse to engage in risky sexual behaviors, no matter how well informed they may be about the risks they face. This example suggests that medical sociologists and social epidemiologists need to contextualize risk factors by asking what it is about people’s life circumstances that shapes their exposure to such risk factors as unprotected sexual intercourse, poor diet, a sedentary lifestyle, or a stressful home life.

Our second example concerns the increasing attention being paid to the public health problem posed by contamination of meat, poultry, and eggs with E. coli and salmonella bacteria. The public has been warned to rinse and cook meat and poultry thoroughly and to carefully wash hands, knives, cutting boards, and so on. Because some follow these safety guidelines more assiduously than others, one can imagine a risk profile of individual behaviors that might predict bacterial infection. These precautions are only necessary, however, when the food that reaches the marketplace is contaminated. Government actions in the 1980s that reduced the number of government inspectors and deregulated the meat-processing industry have created the need for vigilance on the part of individuals. While the current approach to the problem focuses on the individual, it can readily be seen that economic and political forces shape individuals’ exposure to this risk. This example suggests that medical sociologists and social epidemiologists need to contextualize by asking under what social conditions individual risk factors lead to disease and whether there are any social conditions under which the individual-level risk factors would have no effect at all on disease outcome.

While the importance of contextualizing risk factors may seem obvious, if we take a hard look at even some of the most influential areas of research in medical sociology, we will find that much more of this kind of contextualizing is needed. Consider again the stress paradigm. While there are hundreds if not thousands of studies relating stressful circumstances to health outcomes, until the recent efforts of Turner and colleagues (Turner and Marino 1994; Turner, Wheaton, and Lloyd 1995), there was very little even descriptive data about the social origins of stressful circumstances (but see Smith 1992; Goldberg and Comstock 1980).

Why is it so important that we strive to contextualize risk factors? One reason is that efforts to reduce risk by changing behavior may be hopelessly ineffective if there is no clear understanding of the process that leads to exposure. For example, there are powerful social, cultural, and economic factors shaping the diet of poor people in the United States. Consequently, providing information about healthy diet to poor people and exhorting them to follow nutritional guidelines is unlikely to have much impact. Without an understanding of the context that leads to risk, the responsibility for reducing the risk is left with the individual, and nothing is done to alter the more fundamental factors that put people at risk of risks.

This line of thinking suggests that medical sociologists and social epidemiologists should turn on its head the now-popular tendency to examine risk factors that are ever closer to disease in a causal chain. Rather, it suggests that it is just as important to face the other direction and search for the factors that put people at risk of risks. It exhorts researchers both to explore the social origins of risks and to ask whether individually-based risk factors are context-dependent in the sense of influencing health outcomes only within the context of a specific set of social conditions.

FUNDAMENTAL CAUSES

In addition to the obvious need to contextualize risk factors, medical sociologists and social epidemiologists need to take as their task the identification and thorough consideration of social conditions that are what we term “fundamental causes” of disease. We call them
“fundamental” causes because, as we shall see, the health effects of causes of this sort cannot be eliminated by addressing the mechanisms that appear to link them to disease. The possibility that some social conditions have this fundamental quality with regard to health was first presented by House and colleagues in a discussion of potential reasons for the persistent association between SES and disease (House et al. 1990, 1994). We elaborate upon these ideas to build our concept of fundamental social causes of disease.

The Case of SES and Disease. The idea that social conditions might influence health was forcefully asserted by nineteenth-century physicians who founded the field of social medicine. Virchow (1848), for example, declared that “medicine is a social science.” And, of course, it was in part the strong association between indicators of poverty and health that supported this claim. The reasons for the powerful association were also thought to be apparent, residing in the dire housing, sanitation, and work conditions of poor people at the time (Rosen 1979). With tremendous medical advances and extensive public health initiatives, the incidence of such diseases as diphtheria, measles, typhoid fever, tuberculosis, and syphilis declined dramatically. In addition, in modern welfare states, poor people’s access to care increased substantially. By the 1960s, many of the factors that had been identified as linking SES to disease had been addressed, and one might have expected the association to wane and perhaps disappear altogether. Indeed, this is exactly the conclusion that Charles Kadushin reached in a 1964 article in Sociological Inquiry (Kadushin 1964). Startled that social scientists had not recognized the demise of the SES gradient in health, Kadushin reminded his readers that most of the mechanisms thought to produce SES differences in health in the United States had been addressed and that “as countries advance in their standard of living, as public sanitation improves, as mass immunization proceeds, and as Dr. Spock becomes even more widely read, the gross factors which intervene between social class and exposure to disease will become more and more equal for all social classes” (1964:75). As a result, Kadushin declared, Americans from the lower classes are no more likely to develop disease than those from the middle or upper classes.

Of course, Kadushin’s prediction turned out to be dramatically incorrect as indicated by studies (cited above) documenting an enduring or even an increasing (Pappas et al. 1993) association between SES and many disease outcomes. But what was wrong with Kadushin’s reasoning? Hadn’t he engaged in logic that most of us not only accept but take for granted? Having implicitly drawn the path model with SES as the distal factor that is linked to disease by more proximal risk factors, and having observed that the proximal risk factors in the model had been largely eliminated as causal agents, he concluded that the SES-disease association should have disappeared. But it didn’t.

On the face of it, the reason Kadushin’s 1964 prediction turned out to be wrong is readily apparent when one compares the intervening risk factors he considered to the intervening risk factors identified by Adler and colleagues in their 1994 review of socioeconomic status and health. The “gross” risk factors of sanitation and immunization that Kadushin mentioned are replaced in the Adler and colleagues’ review by risk factors that include smoking, exercise, and diet, among others. Further, the evidence suggests that several of the risk factors mentioned by Adler and colleagues were not important intervening mechanisms when Kadushin wrote. Before the 1960s, for example, there was no evidence that rates of smoking were higher among lower SES individuals. Rather, the association emerged during the 1960s because people of higher socioeconomic status were likely to start smoking and more likely to quit if they had started (Ernster 1988; Novotny et al. 1988). Similar changes have occurred in other risk-related behaviors. For example, in considering the strong evidence that declines in coronary heart disease have been greatest among people of higher socioeconomic status, Beaglehole (1990) pointed to the fact that higher SES individuals have been better informed about and more able to implement changes in health behaviors like smoking, exercise, and diet. The result has been a widening of the gap in rates of heart disease between the rich and the poor (Beaglehole 1990). Thus studies of the association between SES and disease over the past several decades reveal an important fact—the risk factors mediating the association have changed. As some risk factors were eradicated, others emerged or were newly discovered. As new risk factors became apparent, people of higher SES were more favorably situated to know about the risks and to have the resources that allowed them to engage in protective efforts to avoid them.
From one vantage point, this account of the association between SES and disease might be seen as a curious story in the history of social epidemiology—an instance in which unique historical events pulled the rug out from under an otherwise reasonable hypothesis put forward by Kadushin in 1964. Far more likely, however, is the possibility that the effect of SES on disease has endured—despite radical changes in intervening risk factors—because a deeper sociological process is at work. If so, what happened over the past several decades will continue to happen and if, at this particular point in time, we presume that an understanding of the SES-disease association lies in tracing the mechanisms that currently appear to link the two, time will prove to be as wrong as Kadushin was. This will occur, we argue, because SES is a fundamental cause of disease.

Fundamental Social Causes of Disease. Our discussion of SES to this point has focused on its persistent association with disease despite changes in intervening mechanisms. However, we have not yet explicitly indicated why SES, or any fundamental cause, might maintain this kind of enduring relationship with disease.

The reason for such persistent associations, and the essential feature of fundamental social causes, is that they involve access to resources that can be used to avoid risks or to minimize the consequences of disease once it occurs. We define resources broadly to include money, knowledge, power, prestige, and the kinds of interpersonal resources embodied in the concepts of social support and social network. Variables like SES, social networks, and stigmatization are used by medical sociologists and social epidemiologists to directly assess these resources and are therefore especially obvious as potential fundamental causes. However, other variables examined by medical sociologists and social epidemiologists, such as race/ethnicity and gender, are so closely tied to resources like money, power, prestige, and/or social connectedness that they should be considered as potential fundamental causes of disease as well.

An additional condition that must obtain for fundamental causes to emerge is change over time in the diseases afflicting humans, the risks for those diseases, knowledge about risks, or the effectiveness of treatments for diseases. If no new diseases emerged (such as AIDS), no new risks developed (such as pollutants), no new knowledge about risks emerged (as about cigarette smoking in the 1950s and 1960s), and no new treatments were developed (such as heart transplants), the concept of fundamental social causes would not apply. In such a static system, as risk factors known to intervene between a social cause and disease are blocked, the association between the social cause and disease would decline in lockstep. But, of course, this is nothing like the situation humans have ever confronted with regard to health. In the context of a dynamic system with changes in diseases, risks, knowledge of risks, and treatments, fundamental causes are likely to emerge. The reason is that resources like knowledge, money, power, prestige, and social connectedness are transportable from one situation to another, and as health-related situations change, those who command the most resources are best able to avoid risks, diseases, and the consequences of disease. Thus, no matter what the current profile of diseases and known risks happens to be, those who are best positioned with regard to important social and economic resources will be less afflicted by disease.

The foregoing reasoning suggests two further attributes of fundamental causes. Because a fundamental cause involves access to broadly serviceable resources, it influences (1) multiple risk factors and (2) multiple disease outcomes. This is an important observation, because it alerts us to the possibility that the association between a fundamental cause and disease can be preserved through changes either in the mechanisms or in the outcomes. The idea that multiple mechanisms may contribute to a persistent association between a cause and an effect comes from sociologist Stanley Lieberson. Lieberson (1985) proposed that some causes, which he called “basic causes,” have enduring effects on a dependent variable because, when the effect of one mechanism declines, the effect of another emerges or becomes more prominent. We have already described the example of the changing role of mechanisms like smoking, exercise, and diet in relation to the association between socioeconomic status and disease. While these variables were no doubt always linked to disease, their connection to socioeconomic status changed when knowledge about their importance in health became available. We take the idea that a cause can affect multiple health outcomes from social epidemiologist John Cassel. Cassel (1976) points out that some social factors make individuals
vulnerable, not to a specific disease, but to a wide array of diseases. As a result, investigations of the relationship of such social factors to specific manifestations of disease are of limited utility. Since only one manifestation of the social cause is measured in such studies, the full impact of the social cause goes unrecorded (also see Aneshensel 1992; Aneshensal, Rutter, and Lachenbruch 1991; Cullen 1984). However, in addition to underestimating the full impact of social causes at any given time, a narrow focus on one disease at a time misses the possibility that changes in particular disease outcomes can lead to enduring associations between fundamental causes and disease overall. When health surveillance or immunization systems fail and old diseases begin to reemerge (TB, measles) or when new diseases enter a population (AIDS), they do so in the context of existing social conditions that are ripe environments for producing mechanisms that link fundamental social causes to new or reemerging diseases. Thus, for example, before 1980, SES was linked to the intravenous use of drugs, which in turn had negative health consequences. But with the emergence of AIDS, this SES-linked risk factor came to have an even more potent effect on health. Indeed, AIDS will likely become a significant contributor to SES differentials in mortality in the time ahead due to the rapid spread of infection in low-income areas (Brunswick et al. 1993). Similarly, the reemergence of drug-resistant tuberculosis is striking poor inner-city populations to a far greater extent than it is higher-status suburban areas.

In sum, a fundamental social cause of disease involves resources that determine the extent to which people are able to avoid risks for morbidity and mortality. Because resources are important determinants of risk factors, fundamental causes are linked to multiple disease outcomes through multiple risk-factor mechanisms. Moreover, because social and economic resources can be used in different ways in different situations, fundamental social causes have effects on disease even when the profile of risk factors changes radically. It follows that the effect of a fundamental cause cannot be explained by the risk factors that happen to link it to disease at any given time.

Research Implications. All too frequently, even those of us who believe that social conditions are important for health are lulled into thinking that the best way to understand and ultimately address the effects of social conditions is to identify the intervening links. Indeed, it is precisely this reasoning that Adler and colleagues use to assert that psychologists have an important role to play in addressing the SES-disease association—the risk factors they identified were individually-based behaviors that psychologists are well-equipped to address. But the concept of a fundamental cause sensitizes us to the possibility that fundamental social causes cannot be fully understood by tracing the mechanisms that appear to link them to disease. To be sure, a focus on mechanisms can help identify variables more proximal to health, and if such risks are addressed, the health of the public can be improved. However, in the context of a dynamic system in which risk factors, knowledge of risk factors, treatments, and patterns of disease are changing, the association between a fundamental social cause and disease will endure because the resources it entails are transportable to new situations. If one genuinely wants to alter the effects of a fundamental cause, one must address the fundamental cause itself.

There are two implications of this reasoning. First, medical sociologists and social epidemiologists need to be careful in interpreting and communicating the meaning of research involving social factors, intervening mechanisms, and disease. Specifically, if the social factor is a fundamental cause, one cannot claim to have accounted for its effects by having “explained” its association with the inclusion of intervening variables in a path or regression model. Second, to understand associations between fundamental causes and disease, medical sociologists need to examine the broader determinants of the resources that fundamental causes entail. This distinctly sociological enterprise will link medical sociologists to the broader discipline in a productive way as we seek to understand how general resources like knowledge, money, power, prestige, and social connections are transformed into the health-related resources that generate patterns of morbidity and mortality.
POLICY IMPLICATIONS

Mechanic and Aiken (1986) argue that, rather than leading to specific policies, the main contribution of social science research to social change is through its influence on the way policymakers and the general public think about social and health problems. This paper aims to make such a contribution by drawing attention to the progress made by medical sociologists and social epidemiologists in recent years and by offering the concepts of “contextualizing risk factors” and “fundamental causes.”

On its own, the focus on individually-based risk factors that has dominated much population-based medical research in recent years is inadequate. To be sure, this focus is a compelling attention-getter, since the findings it generates are highly relevant to any given individual. They can lead to “personal policy” changes such as a reduction in fat intake, a little exercise, or an aspirin a day—actions that individual people can control personally. But those who craft policy for populations can be led astray if their purview is narrowly limited to a focus on individually-based risk factors. This paper reveals two reasons why this is so. First, without understanding the social conditions that expose people to individually-based risk factors, interventions will fail more often than they should. This will occur because interventions will be targeted to behaviors that are resistant to change for unrecognized reasons. The consequence will be that lives and money are wasted, and the American public will lose confidence in our ability to implement changes that really improve health. Second, some social conditions are fundamental causes of disease and as such cannot be effectively addressed by readjusting the individually-based mechanisms that appear to link them to disease in a given context. If we wish to alter the effects of these potent determinants of disease, we must do so by directly intervening in ways that change the social conditions themselves.

The issues addressed in this paper suggest three general criteria that policymakers should use in evaluating whether to commit funds to a proposed health intervention:

(1) Consistent with the idea of contextualizing risk factors, policymakers should require that all interventions seeking to change individual risk profiles contain an analysis of factors that put people at risk of risks. This will avoid the enactment of interventions aimed at changing behaviors that are powerfully influenced by factors left untouched by the intervention. If the evidence is to come from studies that involve the experimental manipulation of a risk factor, policymakers should require confirmation that the intervention works outside of the experimental context. The reason for this is that, by experimentally manipulating the risk factor (e.g., diet or exercise), the researchers have removed from consideration the social factors that determine exposure to the risk factor in the natural environment. Outside the experimental context, the factors that put people at risk of risks may dominate, resulting in the intervention’s ultimate failure.

(2) Consistent with the concept of fundamental causes, health policymakers should consider whether a proposed intervention will have an impact on just one disease or whether, because of its influence on a fundamental cause, it will affect many diseases. An intervention that has even a modest impact on many diseases may be far more important than one that has a relatively strong impact on just one.

(3) Health policymakers concerned with broad social conditions as causes of disease should regard with skepticism interventions that focus only on intervening variables but claim to address the broader social condition. Even an “effective” intervention that addresses the identified risk factor will, in the long run, fail to eliminate the effect of a fundamental social condition. In a changing state of affairs, the resources that accrue to the more advantaged allow them to regain the health advantage that may have been dented temporarily by the intervention.

If one wishes to address fundamental social causes, the intervention must address inequality in the resources that fundamental causes entail. Many people and some medical sociologists believe that this is impractical—even to think about—because, for them, inequality is so firmly entrenched that nothing much can be done about it. Believing this, the only reasonable thing to do is to focus on more proximal individually-based risk factors, even if doing so has little long-run bearing on the association between fundamental social causes and disease.

But this reasoning is shortsighted. There are many policies that have a direct bearing on the extent of inequality in our society and thus on the extent to which people from different social
circumstances have access to health-related resources. To be sure, these policies are rarely discussed with reference to their health implications. Still, policies relevant to fundamental causes of disease form a major part of the national agenda, whether this involves the minimum wage, housing for homeless people, capital-gains taxes, parenting leave, head-start programs, or other initiatives of this type. Such policy initiatives often lie outside the realm of influence and expertise of health policy experts. Yet if fundamental causes are potent determinants of disease, the potential health impact of these broad policies needs to be thoroughly understood—a task that medical sociologists and social epidemiologists should take up more thoroughly than they have. Ideally, a research-based “health impact statement” should accompany such plans, and health experts should be trained in the skills needed to produce such a statement.

CONCLUSION

The dominant focus in epidemiology and perhaps the American culture in general is on individually-based risk factors that lie relatively close to disease in a causal chain. But this focus overlooks important sociological processes and, as a result, could lead us to actions that limit our ability to improve the nation’s health. We have focused on two concepts—contextualizing risk factors and fundamental causes—that direct our attention to precisely those factors that are left unexamined in the currently dominant orientation to research on risk factors for disease. If future research by medical sociologists and social epidemiologists increases our understanding of the processes implied by these concepts, we will be better positioned as a society to further improve the nation’s health.

NOTES

1. Using the American Journal of Epidemiology as an indication of the current emphasis of epidemiological research, we reviewed the 240 articles published between November of 1992 and 1993. Excluding methodological reports (N = 44) and studies focused exclusively on descriptive epidemiology (N = 15), we found that only 13.3 percent (24/181) of the articles focused on risk factors that could be construed as social in nature. Moreover, because many of these articles examined race, ethnicity or gender, without explicit reference to the social aspects of these characteristics, our figure of 13.3% should be considered an upper-bound estimate of the journal’s focus on social factors.

2. We include stigmatization because it is so closely tied to the prestige system (Goode 1978). Prestige, or the general standing that an individual holds in the eyes of others, is an important resource that is likely to have many implications for health—either indirectly through resources like money, power, or social connections, or more directly through what a person and/or those around him/her believe he/she deserves from the social environment. Stigmatization is important because it involves the denial of the benefits of prestige.

3. We focus here on fundamental social causes of disease. It is possible to conceive of fundamental psychological or biological causes as well. For example, at the psychological level, one might consider a mastery orientation to be a resource that would be linked to many mechanisms and thus to many diseases. Similarly, at the biological level, the immune system might be conceptualized as a resource that would influence many specific mechanisms and thus many disease outcomes. In either of these cases, the association between the fundamental cause (mastery or immune system) and disease outcomes would likely endure even if the specific mechanisms were to change. Our main point regarding social factors as fundamental causes is not that fundamental causes should be taken seriously because they are often social, but rather that social conditions need to be taken seriously because they are often fundamental causes.

4. In addition to the factors we consider here, it is important to recall that this individually-based risk factor approach can also have deleterious effects by shifting an excessive portion of the blame to the individual. When research focuses attention on individually-based causes of disease, the onus is often taken off broader-based conditions. Morbidity and mortality due to tobacco is attributed to an individually-based bad habit rather than to a heavily advertised, government-subsidized, highly profitable killer industry.
FUNDAMENTAL CAUSES OF DISEASE

REFERENCES


Bruce G. Link is associate professor of public health at Columbia University and research scientist at New York State Psychiatric Institute. His interests lie in understanding the sources of particular types of inequality, its legitimation, and its consequences as these bear on the social patterning of health and illness. This interest is reflected in his work on the association between socioeconomic status and major mental disorders and the possible role that occupational conditions may play in this association, research on the health and well-being of homeless people, and research on the social and economic adversities engendered by the stigma of mental illness.

Jo Phelan is assistant professor of sociology at the University of California, Los Angeles. Her research interests include homelessness, social stigma, the impact of social conditions on health and illness, and attitudes concerning inequality and its legitimacy.