A Historical Framework for Social Epidemiology

LISA F. BERKMAN AND ICHIRO KAWACHI

Epidemiology is the study of the distribution and determinants of states of health in populations (Susser 1973). Ever since John Graunt (1662) counted deaths in county parishes in England in the seventeenth century, social variations in morbidity and mortality have been observed. Early studies often centered on the ill effects of poverty, poor housing conditions, and work environments. By the nineteenth century, physicians such as Villerme (1830) and Virchow (1848) refined observations identifying social class and work conditions as crucial determinants of health and disease (Rosen 1963). Durkheim wrote eloquently about another profound social experience, that of social integration and how it was related to patterns of mortality, especially suicide (1897). So, in many ways, the idea that social conditions influence health is not new. Social epidemiology, however, is.

As the public health movement developed in the United States and Great Britain in the nineteenth and early twentieth centuries, attention was drawn to the increased risk of disease among the poor (Rosen 1975; Duffy 1990). Efforts to improve their physical environments (e.g., housing, noxious work environments and water supply), sanitation, nutrition, and access to immunization were the primary focus of public health professionals. With broad improvements in the physical environment in the United States, Great Britain, and much of northern Europe, countrywide increases in life expectancy occurred. Based on this observation, many scientists forecast large-scale reductions of social disparities in health (Kadushin 1964). Perhaps no other phenomenon has augured the need for the perspective of social epidemiology as clearly, however, as the continued maintenance and recent growth of social inequalities in health in many countries. Thus, while diseases have come and gone, some infectious diseases have been eradicated, others have emerged, and a host of noninfectious diseases have dominated the profile of causes of death and disability, social inequalities in health remain. These persistent patterns call for an epidemiologic approach to understanding disease etiology that incorporates social experiences as
more direct causes of disease and disability than is the customary view.

Fortunately, many forces have converged to permit the development of this field. Among the most critical has been the development of work on stress and physiologic responses to stressful experiences. Building on the fundamental work by Cannon (1935) and Selyé and Wolff (1973), health psychologists, neuroendocrinologists, and physiologists have made it clear that stressful conditions may exact a direct toll on the body, offering powerful biological models that link external stressors to physiologic responses capable of influencing disease development and prognosis. Work on psychophysiology, psychoneuroimmunology, and most recently on allostatic load has helped trace biologic pathways as well as specific behaviors and exposures to noxious agents that link social conditions to important health outcomes. (Cohen 1988; Kiecolt-Glaser et al. 1996, 1997; McEwen 1998).

The second factor has been a progressive increased blurring of the distinction between "psychosomatic" illness and other physical illnesses. Whereas it was formerly believed that some diseases were caused by psychological states with little biological basis and others were purely "physical," we now understand that in almost all cases this distinction is false. Most psychosomatic diseases involve varied genetic and environmental determinants, and all states of health and disease are influenced to some extent by psychosocial conditions. Rarely for any disorders is there a single necessary and sufficient cause of disease. The breakdown of this artificial dichotomy is critical to advancing knowledge in the coming decades: Diseases are no longer classified as psychosomatic or not.

A third theoretical development in understanding the distribution of risk in populations further enhances our ability to launch a solid investigation of social factors and health. In 1992, Geoffrey Rose (1992), an eminent epidemiologist, wrote a small book on the strategy of preventive medicine. In this landmark work, small only in size, Rose pointed out that rarely are either risk factors or disease binary in nature. In most cases, risks are distributed along a continuum and small shifts in the distribution of risk throughout a population can make large differences in the health status of that population. Furthermore, understanding the dynamics of why some populations have certain distributions leads to very different etiologic questions than asking why some individuals are in the tails of the distribution. Pursuing this population-based strategy, rather than a high-risk strategy, leads to framing very different questions and utilizing very different preventive approaches. The population strategy is of central importance to social epidemiology and it has been traditionally the mainstay of public health.

The fields of physiology and psychosomatic, social, and preventive medicine as well as medical sociology and health psychology have all made important contributions to the development of social epidemiology (See Rosen 1975 for an excellent history of preventive medicine in the United States). But the seeds of social epidemiology have also grown from within epidemiology itself. In the late 1960s and 1970s, epidemiologists such as John Cassel, Mervyn Susser, S. Leonard Syme, Saxon Graham, Lawrence Hinkle, Al Tyroler, Sherman James, and Leo Reeder started to develop a distinct area of investigation in epidemiology centered on the health impact of social conditions, particularly cultural change, social status and status inconsistency, and life transitions. Their work drew heavily on that of epidemiologists who worked earlier in the century such as Goldberger and Sydenstricker (Goldberger et al. 1929), who investigated the etiology of pellagra, and Wade Hampton Frost, whose work on tuberculosis was seminal (Maxcy 1941). They also drew deeply from medical sociology (Freeman et al. 1963) and the work of psychiatric epidemiologists (Faris and Dunham 1939; Hollingshead and Redlich 1958; Leighton 1959; Srole et al. 1962). Syme (1965) explained that investigations of the
A HISTORICAL FRAMEWORK FOR SOCIAL EPIDEMIOLOGY

"social etiology of disease attempted to systematically examine variations in the incidence of particular diseases among people differentially located in the social structure and attempt[ed] to explore the ways in which their position in the social structure tended to make them more vulnerable, or less, to particular disease."

In a seminal article, Saxon Graham (1963) discussed the social epidemiology of selected chronic illnesses. While never giving an explicit definition of social epidemiology, he suggested that a union of sociology with the medical sciences would produce a new and more successful epidemiology. Graham went on to say that achieving a coherent and complete theory of disease causation would require obtaining social and biological data that are consistent with each other with regard to a specific disease (Graham 1963, p. 72). Thus, he argued, one must understand how membership in a social group relates to behavior patterns, to exposure to "vehicles" for transmitting agents, to direct tissue changes, and finally to disease. Graham aimed to identify specific social circumstances that led to a chain of events in which specific behaviors were linked to specific diseases. His classic example involved Percival Pott's analysis of scrotal cancer in chimney sweeps. Parallel to his analysis of Pott's studies, much of his early work dealt with smoking and dietary and sexual behaviors that were associated with different social groups and thus more proximally linked to specific diseases. His classic example involved Percival Pott's analysis of scrotal cancer in chimney sweeps. Parallel to his analysis of Pott's studies, much of his early work dealt with smoking and dietary and sexual behaviors that were associated with different social groups and thus more proximally linked to specific diseases. In seeking to understand the large-scale social patterning of disease in terms of individual behaviors of group members, Graham's great contribution to epidemiology was his ability to incorporate this multilevel thinking into the field.

Almost a decade later, in the mid-1970s, two epidemiologists, John Cassel and Mervyn Susser, more explicitly tackled the methodologic controversies and paradigm shifts inherent in incorporating a deeper understanding of the social influences of disease into epidemiologic thinking. Armed with evidence from the previous decade, John Cassel (1976) in the fourth Wade Hampton Frost Lecture to the American Public Health Association stated that "the question facing epidemiologic inquiry is, are there categories or classes of environmental factors that are capable of changing human resistance in important ways and making subsets of people more or less susceptible to ubiquitous agents in our environment." In this classic paper "The Contribution of the Social Environment to Host Resistance," he argued that environmental conditions capable of "producing profound effects on host susceptibility" involve the presence of other members of the same species, or more generally, certain aspects of the social environment (Cassel 1976, p. 108).

Building on the work of Hinkle (1973) and stress researchers such as Cannon (1935), Dubos (1965), and Selyé and Wolff (1973), Cassel posited that at least one of the properties of stressful situations might be that the actor is not receiving adequate evidence that his actions are leading to anticipated consequences. Today we might cite situations of powerlessness brought on by social disorganization, migration, discrimination, poverty, and low support at work as prime examples of this situation. Cassel also outlined a series of protective factors that might buffer the individual from the deleterious consequences of stressful situations. The property common to these processes is "the strength of the social supports provided by the primary groups of most importance to the individual" (Cassel 1976, p. 113). Thus, consolidating the findings gathered by epidemiologists doing empirical work on status and status incongruity (Syme et al. 1965; Hinkle 1973), rapid social change and disorganization (Cassel et al. 1961; James and Kleinbaum 1976), acculturation and migration (Marmot and Syme 1976), and social support and family ties (Nuckolls et al. 1972; Pless and Satterwaite 1972), Cassel laid out an intellectual agenda for social epidemiology that provided the groundwork for decades to come.

In a provocative series of articles, Mervyn
Susser has written that epidemiology must broaden its base and move beyond its focus on individual-level risk factors and “black box epidemiology” to a new “multilevel ecoepidemiology” (Susser 1994a,b, 1998; Susser et al. 1996a,b). The foundations for much of this framework can be seen in his 1973 book, Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology. In the introduction to that volume, Susser stated that epidemiology shares the study of populations, in a general way, with other population sciences such as sociology, human biology, and population genetics. In affirming common methodologic and conceptual ground with other sciences involved in the study of society, he explained that “states of health do not exist in a vacuum apart from people. People form societies and any study of the attributes of people is also a study of the manifestations of the form, the structure and the processes of social forces” (Susser 1973, p. 6). In other chapters, Susser discussed how agent, host, and environment models, the most basic organizing principles of epidemiology, could be framed as an ecological system with different levels of organization.

Susser’s recent work again emphasizes that epidemiology is, in essence, ecological since the biology of organisms is determined in a multilevel, interactive environment. Identifying risks at the individual level, even multiple risks, does not sufficiently explain interactions and pathways at that level, nor does it incorporate the social forces that influence risks to individuals.

GUIDING CONCEPTS IN SOCIAL EPIDEMIOLOGY

We define social epidemiology as the branch of epidemiology that studies the social distribution and social determinants of states of health. Defining the field in this way implies that we aim to identify socioenvironmental exposures that may be related to a broad range of physical and mental health outcomes. Our orientation is similar to other subdisciplines of epidemiology focused on exposures (e.g., environmental or nutritional epidemiology) rather than those areas devoted to the investigation of specific diseases (e.g., cardiovascular, cancer, or psychiatric epidemiology). We focus on specific social phenomena such as socioeconomic stratification, social networks and support, discrimination, work demands, and control rather than on specific disease outcomes. While future studies may reveal that some diseases are more heavily influenced by social experiences than others, we suspect that the vast majority of diseases and other health outcomes such as functional status, disability, and well-being are affected by the social world surrounding us all.

Like environmental and nutritional epidemiology, social epidemiology must integrate phenomena at the margins of what is defined as its domain. For instance, psychological states, behaviors, and aspects of the physical or built environment are influenced by social environments and vice versa. Borders at the periphery of any field, and social epidemiology is no exception, are bound to be fuzzy. We make no attempt to draw clean lines encircling the field. Because it is important for social epidemiologists to consider related areas, we have included sections in this volume on psychological states and behaviors that are closely related to the social experiences which are our primary concern. If we err on the side of blurring boundaries, we must balance that with precision in defining explicit testable hypotheses in our work. Without hypotheses that can be clearly supported or refuted, without having a clear understanding of temporal sequencing or biological plausibility, and without articulated theories and specific concepts to guide empirical investigation, we will not be able to make progress.

The rest of this chapter outlines several concepts that are important to the field of social epidemiology. These concepts are not offered as universals to be uncritically accepted but rather as useful and sometimes challenging guides that transcend the study of any single exposure.
A POPULATION PERSPECTIVE

Individuals are embedded in societies and populations. The crucial insight provided by Rose’s (1992) population perspective is that an individual’s risk of illness cannot be considered in isolation from the disease risk of the population to which she belongs. Thus, a person living in Finland is more likely to die prematurely of a heart attack compared to someone living in Japan, not just because any particular Finnish individual happens to have a high level of cholesterol, but because the population distribution of cholesterol levels in Finnish society as a whole is shifted to the right of the Japanese distribution. The level of cholesterol that might be considered “normal” in Finnish society would be grossly abnormal and a cause for alarm in Japan. Moreover, we know from detailed studies of migrants that the basis for these population differences are not genetic (Marmot and Syme 1976). For instance, Japanese immigrants to America take on the coronary risk profiles of their adopted country.

Although Rose’s initial examples involved the examination of risk factors for heart disease, we now recognize that his insight has broad applicability to a swath of public health problems, ranging from aggression and violence, mental health, to the effects of poverty and material deprivation on health. Fundamentally, Rose’s insight harks back to Durkheim’s discovery about suicide: that the rate of suicide in a society is linked to collective social forces. There are a myriad reasons why any individual commits suicide, yet such individuals come and go while the social rate of suicide remains predictable.

The crucial implication of Rose’s theory for social epidemiology is that we must incorporate the social context into explanations about why some people stay healthy while others get sick. Applying the population perspective into epidemiological research means asking “Why does this population have this particular distribution of risk?”, in addition to asking “Why did this particular individual get sick?” Furthermore, as Rose pointed out, the greatest improvements in population health are likely to derive from answering the first question, because the majority of cases of illness arise within the bulk of the population who are outside the tail of high risk.

THE SOCIAL CONTEXT OF BEHAVIOR

Over the last several decades, a huge number of clinical trials have been launched to modify individual behavioral risk factors such as alcohol and tobacco consumption, diet, and physical activity. By and large, the most successful have been those which incorporated elements of social organizational changes into interventions. We now understand that most behaviors are not randomly distributed in the population. Rather, they are socially patterned and often cluster with one another. Thus, many people who drink also smoke cigarettes, and those who follow health-promoting dietary practices also tend to be physically active. People who are poor, have low levels of education, or are socially isolated are more likely to engage in a wide range of risk-related behaviors and less likely to engage in health-promoting ones (Matthews et al. 1989; Adler et al. 1994). This patterned behavioral response has led Link and Phelan (1995) to speak of situations that place individuals “at risk of risks.”

Understanding why “poor people behave poorly” (Lynch et al. 1997) requires a shift in understanding—specific behaviors once thought of as falling exclusively within the realm of individual choice occur in a social context. The social environment influences behavior by (1) shaping norms, (2) enforcing patterns of social control (which may be health-promoting or health-damaging), (3) providing or not providing environmental opportunities to engage in certain behaviors, and (4) reducing or producing stress for which certain behaviors may be an effective coping strategy, at least in the short term. Environments place constraints on in-
individual choice. Incorporating the social context into behavioral interventions has led to a whole new range of clinical trials that take advantage of communities, schools, and work sites to achieve behavioral change (see Sorensen et al. 1998 and Chapter 11).

CONTEXTUAL MULTILEVEL ANALYSIS

The understanding that behavior is conditioned by society yields a more general appreciation of the need for contextual analysis in epidemiology. As Susser (1998) noted, “risk factor epidemiology in its pure form exploits neither the depth and precision of micro-levels nor the breadth and compass of macro-levels.” Conceptions of how culture, policy, or the environment influences health remain fuzzy and speculative if one analyzes only the independent effects of individual-level risk factors. Ecological analysis, a central part of both epidemiology and sociology early in this century, offered an approach to the study of environments, but it lost a great deal of respectability because of problems related to the ecological fallacy (e.g., drawing individual inferences from grouped data; see Chapter 14). It was difficult, if not impossible, to rule out reverse causation (that the illness influenced residential relocation) in many studies. In fact, it was this latter problem that plagued many of the early studies on psychiatric disorder and community disorganization.

In the past few years, however, it has become apparent that just as there are ecologic-level exposures in environmental and infectious disease epidemiology, so are there valid ecologic-level exposures related to the social environment that are not adequately captured by investigation at an individual level (Macintyre et al. 1993; Kaplan 1996; Kawachi and Kennedy 1997; Kawachi et al. 1997). For example, the number of grocery stores, parks, the condition of housing stock, and voter participation may be critical determinants of behaviors, access to care, or illness. These ecologic-level exposures call for innovative methods (Jones and Moon 1993; DiezRoux et al. 1997). The assessment of exposures at an environmental or community level may lead to an understanding of social determinants of health that is more than the sum of individual-level measures. Although important questions remain about the appropriate level of environmental assessment (e.g., neighborhood, city, state, country), the disentangling of compositional versus contextual effects, and the pathways linking such environmental exposures to individual health outcomes, ecological analyses offer a valuable research tool to epidemiologists. When coupled with individual-level data, they offer the critical advantages available in the form of multilevel analyses.

A DEVELOPMENTAL AND LIFE-COURSE PERSPECTIVE

In general, epidemiologists have only crude tools with which to explore developmental and life-course issues. Cumulative risk and latent periods are familiar terms but we often lack methods to deal with them adequately. Yet there is intriguing evidence that such perspectives may yield valuable insights. In fact, social epidemiologists working in the 1960s and 1970s implicitly adopted a life-course perspective in testing theories about status incongruity in which the stressful experiences being studied resulted from having grown up in one situation or as a member of one status group and then having shifted to either a higher or lower status. (See Syme et al. 1965 for an excellent discussion of this.)

Three hypotheses have been proposed (Power and Hertzman 1997) to explain early life influences the onset of disease in middle and late life. The first is that some exposure in early childhood could influence developmental processes—particularly brain development during periods of great plasticity. By molding patterns of response during these “critical stages,” early life experiences would then make the individual vulnerable or resistant to various
diseases in adulthood (Barker 1992). This model is similar to that of latency models. The second hypothesis is one of cumulative disadvantage and is outlined by several medical sociologists (Ross and Wu 1995). Disadvantage in early life sets in motion a series of subsequent experiences that accumulate over time to produce disease after 30, 40, 50, or 60 years of disadvantage. The third hypothesis is that while early experiences set the stage for adult experiences, it is really only the adult experiences that are directly related to health outcomes. For instance, low educational attainment in earlier life might matter only in so far as it constrains the range of job opportunities and job experiences. These three models lay out a framework within which to examine life-course issues. Our aim here is not to conclude that there is strong evidence to support one or another of them, nor in fact to advocate an overly deterministic, developmental model of disease causation at all, but rather to suggest that this perspective provides a lens through which to examine how social factors may influence adult health.

GENERAL SUSCEPTIBILITY TO DISEASE

Wade Hampton Frost (1937) noted that at the turn of the 20th century there was nothing that changed "nonspecific resistance to disease" as much as poverty and poor living conditions. In referring to this altered resistance, Frost suggested that it was not just increased risk of exposure among the poor that produced high prevalence rates of tuberculosis: It was something about their inability to fight off the disease—their increased susceptibility to disease once exposed—that contributed to high rates of disease in poor populations.

Cassel, Syme, and Berkman (Cassel 1976; Syme and Berkman 1976; Berkman and Syme 1979) built on this idea when they observed that many social conditions were linked to a very broad array of diseases and disabilities. They speculated that social factors influence disease processes by creating a vulnerability or susceptibility to disease in general rather than to any specific disorder. According to the general susceptibility hypothesis, whether individuals developed one disease or another depended on their behavioral or environmental exposures as well as their biological or genetic makeup. But whether they became ill or died at earlier ages or whether specific socially defined groups had greater rates of disease depended on socially stressful conditions.

As originally proposed, the concept of general susceptibility or psychosocial "host resistance" was a powerful and intuitively appealing metaphor but not well grounded biologically. It was not until research in social epidemiology became more integrated with research in neuroscience and psychoneuroimmunology that clear biological mechanisms were defined, at least as potential pathways leading from stressful social experiences to poor health. Neuroendocrinologists had identified classic stress mediators such as cortisol and catecholamines as well as less well understood mediators such as dehydroepiandrosterone (DHEA), prolactin, and growth hormone, and they knew that these affected multiple physiologic systems. By linking evidence from both fields, researchers showed that some stressful experiences activate multiple hormones and thus might not only affect multiple systems but could also produce wide-ranging end-organ damage. Furthermore, recent advances in understanding variable patterns of neuroendocrine response with age suggest that the cumulative effects of stress, or even stressful experiences that have taken place during development, may alter neuroendocrine-mediated biological pathways and lead to a variety of disorders from cardiovascular disease to cancer and infectious disease (Meany et al. 1988; Sapolsky 1996; McEwen 1998).

These developments in aging research suggest new ways in which stressful experiences may be conceptualized as accelerating the rate at which we age or changing the aging process itself (Berkman 1988).
ceptual shift relates well to earlier notions of general susceptibility.

CONCLUSION

In recent decades, the discipline of epidemiology has witnessed the birth of multiple subspecialties such as environmental, nutritional, clinical, reproductive, and most recently, genetic epidemiology (Rothman and Greenland 1998). The central question of social epidemiology—how social conditions give rise to patterns of health and disease in individuals and populations—has been around since the dawn of public health. But the rediscovery of this question through the lens of epidemiology is a relatively recent phenomenon. As demonstrated in the contributions to this volume, social epidemiologists are now applying concepts and methods imported from a variety of disciplines ranging from sociology, psychology, political science, economics, demography, and biology. The multidisciplinary nature of the venture makes the research both new and suited to tackle the problems at hand. Social epidemiology has already yielded many important findings during the relatively brief period of its existence, yet important discoveries remain to be made. By sharpening the tools we have to capture the powerful social forces experienced by individuals and communities, as well as by strengthening our methods of inquiry, we may look forward to further decades of insight into how society shapes the health of people. With rigorous attention to issues related to the social context, biological mechanisms, and the timing and accumulation of risk, we can hope to identify the ways in which the structure of society influences the public’s health.

REFERENCES


Social class and mental illness. New York: John Wiley.


