

ARE THERE SOCIAL DETERMINANTS OF HEALTH AND DISEASE?

RONALD A. THISTED

ABSTRACT The concept of a determinant is tied to the idea of a mechanism for action. Ideas from epidemiology, particularly the epidemiologic triad of agent, host, and environment, can help to make sense of factors that affect the absence of disease or that interfere with a mechanism that alters health. However, assembling convincing evidence for the existence of social determinants of health is a challenge, in part because of the difficulties of bridging the social and biological realms. While social contexts are measured using aggregates of individuals, disease and dysfunction occur at the individual level, leading to difficult problems of ecological inference. Although social factors have been shown to be associated with differences in mortality from specific causes, these factors account for only a small portion of the mortality from any individual cause. This suggests that the pathways through which social factors influence health are affected by their interactions with other factors.

We are seeking the principles and the causes of the things that are, and obviously of them qua being. For, while there is a cause of health and of good condition, and the objects of mathematics have first principles and elements and causes, and in general every science which is ratiocinative or at all involves reasoning deals with causes and principles, more or less precise, all these sciences mark off some particular being—some genus, and inquire into this. . . . the sciences omit the question whether the genus with which they deal exists or does not exist, because it belongs to the same kind of thinking to show what it is and that it is.

—Aristotle, *Metaphysics*, Book VI

Chair, Department of Health Studies, University of Chicago, 5841 South Maryland Avenue, MC 2007, Chicago, IL 60637.

Email: thisted@health.bsd.uchicago.edu.

This work was supported by the National Institute on Aging grant number PO1 AG18911.

Perspectives in Biology and Medicine, volume 46, number 3 supplement (summer 2003):S65–S73
© 2003 by The Johns Hopkins University Press

ROBERT SAMPSON AND JOHN CACIOPPO AND LOUISE HAWKLEY raise basic issues about the relationship of the social environment to health and well being. In thinking through these issues, it is helpful to consider first the fundamental question: “Are there social determinants of health and disease?”

The papers presented in this symposium presuppose an affirmative answer to this question. They assume that social factors can in fact determine who gets sick and who stays healthy (at least in part). They also assume that it is possible to bring evidence to bear on just how social factors bring about their effects on health. And they suggest that social factors are in some sense important modulators of health or disease. But none of these assumptions is obvious or straightforward.

The two papers discussed here shed light on what we might mean by a social “determinant” of health or disease (the conceptual issue), how we might recognize one were we to see one (the evidentiary issue), and how large their effects might be (the importance issue).

THE CONCEPT OF A SOCIAL DETERMINANT OF HEALTH

It is not clear on its face what we mean when we speak of “determinants” of health. Mere association is not enough to establish determination or causation. And even establishing an association is problematic due to the limitations of what we can directly observe.

Is race a social determinant of health? Certainly the notion of race is a social construct, and there are notable mortality differences between whites and blacks in the United States. Yet we should be reluctant to say that race per se—the socially constructed label—causes poor health and early death. Rather, we are inclined to attribute these differences to multiple correlates of race, both biological and social. The problem with calling race itself a (social) determinant of health is precisely that we have only the correlation: we have no knowledge of how or even whether race by itself predisposes to poor health outcomes.

Mitchell Wong and colleagues (2002) argue that the difference in mortality between blacks and whites in the United States is due in large measure to differences in deaths from four causes: hypertension, HIV, diabetes, and trauma (principally homicide). This conference provides accounts for how particular social factors could contribute to each of these diseases. Passive coping strategies coupled with the cumulative need for mounting a stress response can lay the groundwork for hypertension (Cacioppo and Hawkley 2003), and the constant vigilance required in communities with high crime and low income is a perfect setting for this to occur. The social circumstances that lead to obesity could account for increased diabetes (Rubenstein 2002). Interestingly, however, the social factors that might account for differences in hypertension or diabetes are different from the neighborhood collective efficacy characteristics that affect homicide (Sampson 2003). And these are different still from the structure of social

networks that can account for differences in rates of transmission of sexually transmitted infections such as HIV between blacks and whites (Laumann and Youm 1999). The differences between these social factors suggest that race is coincidental rather than causative, which justifies our reluctance to attribute mortality differences to unspecified effects of race per se. Indeed, having a plausible mechanism by which action is accomplished is a key element of our notion of determination. Thus, Cacioppo and Hawkley note that social isolation in the elderly is a risk factor for major causes of death, but until there is a plausible mechanism by which lack of social connectedness leads to increased risk of hypertension, for instance, we are left only with a puzzling (but reproducible) coincidence.

As we think about mechanisms, it is helpful to consider how we conceptualize nonsocial determinants of health and disease, and how they might relate to factors properly labeled as social. The epidemiologic triad is a key concept of epidemiology that is helpful here. Epidemiologists note that the spread of infectious disease requires three factors to be present simultaneously: an agent, a suitable environment, and a susceptible host. Removing any one factor makes transmission impossible. The presence or absence of any one of these factors can be due to a combination of biological and social phenomena.

To take a classic epidemiologic example, cholera is caused by the *Vibrio cholerae* bacteria; we might call this a biological determinant of disease. But *V. cholerae* is only the agent. To contract the disease, one must ingest the bacteria when it is infectious, which is at its height within five hours of passing through the human GI tract (Merrell et al. 2002). Poor sanitation is the most common way in which this is made possible, and poor sanitation is in turn often a consequence of social conditions—a lack of infrastructure and poverty as, for example, in Bangladesh. However, different social forces among the more wealthy in industrialized nations can also lead to transmission, as, for instance, in the case of a 1974 outbreak in Portugal due to contamination of bottled water (Blake et al. 1977). Thus the “environment” arm of the epidemiologic triad could be a result of social, primarily economic, factors. Cultural factors, too, can play a role. For example, John Snow’s evidence in the 1854 London cholera epidemic that cholera was waterborne was ignored, even suppressed, because of the universally held belief that the disease was caused by bad air. Furthermore, susceptibility of the host can be substantially reduced by use of modern vaccines, but these are too expensive to use widely in underdeveloped countries and are prohibited in some areas due to fear of adverse side effects. In short, except for the biologic agent, much of the precondition for infectious disease may result from social rather than biological factors.

It can be also be said that decidedly noninfectious maladies, such as hypertension and homicide, can be understood through the model of hosts and agents interacting in a suitable environment, and that it is this interaction that describes the mechanism of action by which the agent operates. Both Sampson and

Cacioppo and Hawkley focus on social environments, and account for differences in health and well being by the social environment's effects on individuals.

Cacioppo and Hawkley focus on the social network (or lack thereof) in which an individual is embedded. More accurately, they consider the individual's perception of isolation or embeddedness, and they argue that those who feel isolated use passive strategies to cope with stress, rather than using the active strategies employed more often by the socially embedded. These passive coping strategies lead directly both to different physiological responses and to increased cumulative stress over time, which in turn reinforces the detrimental physiological response. Both the cumulative load and the hemodynamic response to stress lay the physiological groundwork for hypertension later in life. Thus, to return to our race example, there is a compelling story linking the psychosocial state to poor long-term health. To the extent that African Americans perceive themselves to be socially isolated, Cacioppo and Hawkley's work suggests a mechanism that could provide a partial account for the black-white mortality differences associated with hypertension.

Sampson, too, focuses on a social environment, namely, the neighborhood. He shows, for instance, that a neighborhood's collective efficacy can account in part for its homicide rate or crime victimization rate. Because we can credit the many ways in which neighbors watching out for one another could result in fewer criminal acts, we are prepared to say that collective efficacy is one determinant of neighborhood crime. Indeed, to the extent that neighborhoods in which African Americans find themselves have lower collective efficacy, this theory can provide a partial mechanism for the differences in mortality between whites and blacks due to homicide. Thus, the notion of a mechanism of action is a key factor underlying our concept of social effects on health.

PROBLEMS OF EVIDENCE

Even if we were to take association as a preliminary indicator for a causative role, it is not entirely clear what should count as evidence even for association. The reasons for this difficulty are important, because they raise challenging methodological issues that must be addressed in order to marshal convincing evidence for the nature, strength, and importance of social influences on health. Three key issues—ecological inference, measurement and longitudinal inference, and feedback—can be highlighted by three familiar sayings: (1) “The whole is greater than the sum of its parts” (Aristotle), (2) “One never steps into the same stream twice” (Heraclitus), and (3) “No man is an island” (Donne). I shall address only the first of these in depth.

Ecological Inference

Disease (if not health) is played out in the individual often through loss of biological function, while the social causes are, by definition, social: they involve

interactions among individuals and not between those individuals' biological organ systems. Consequently, the measurable features of social context are on different entities than the measurable features of disease. Thus, there are two levels at work here, and understanding their interplay requires some form of multilevel analysis.

Indeed, this is a theme in Sampson's work. Neighborhood features are collective phenomena, neither aggregations of individual properties nor mere perceptions of individuals. Sampson would argue that the degree to which a street corner is dangerous is a very different thing than a resident's perception of how dangerous the street corner is. The two must be assessed in different ways, and may have different effects on health and behavior.

Much of the work looking at social processes from this point of view has focused on outcomes such as crime rather than health. In many ways, making inferences about crime seems more straightforward than drawing inferences about health effects. Individual crimes, more than individual diseases, appear to result from a small number of discrete choices of behavior over a relatively short period of time, and the mechanisms of carrying out specific acts of criminal violence are well understood. There is a counterpart of the epidemiologic triad: armed robbery or homicide requires motive, means, and opportunity—not so different from environment, agent, and susceptible host. And Sampson's work is compelling because it helps us understand directly how neighborhood social structure can have an effect on each of these components.

Cacioppo and Hawkley take a different tack that is multilevel in a quite different way. They focus on perceived social isolation, which they argue can trigger biological mechanisms that lead to disease. While both inputs and outputs are at the individual level in Sampson's sense, the levels crossed here are between the psychosocial and the biological.

For Cacioppo and Hawkley, health is maintained by the organism's internal regulatory mechanisms. Social factors can affect health through the demands placed on these regulatory processes as mediated by psychological perceptions of them. In exactly the same way that walking into a warm room causes me to perspire, walking into a room full of lawyers prepared to take my deposition will cause my heart rate to increase, my stomach to knot, and will initiate palpitations. In this example, my perceived social isolation, while profound, is of short duration. Were I to face similar stressors every day—as many people must face threatening environments on a daily basis—the cumulative effect on my regulatory processes could have side effects such as hypertension and poorer sleep.

It is the interaction across levels, then, that is common to the two approaches: the interaction of psychological responses to external social contexts with biological regulatory processes for Cacioppo and Hawkley, and the interaction of the collective with the individual for Sampson.

Multilevel measurement settings involving collectives and individuals have long been recognized to be subject to problems of ecological inference and the

so-called “ecological fallacy.” Expressed in its simplest form, the association between two variables measured at aggregated levels may not reflect—or even be in the same direction as—the association between the same two variables when viewed at the level of the individuals who make up those aggregates. Thus, for instance, an investigation of discrimination in admissions to graduate programs at Berkeley in 1973 found that women applicants were much more likely than men to be denied admission to the departments studied. However, in almost all of these departments fewer men were admitted than women! Approximately 45 percent of men were admitted, while only 35 percent of women were. However, women applied to those departments that were very difficult to get into (such as English) at very high rates compared to men, while departments that were relatively easy to obtain admission in (such as Computer Science) attracted relatively large numbers of male applicants. At the same time, “hard” departments attracted many more applicants of both sexes than did “easy” departments (Bickel, Hammel, and O’Connell 1975).

Sampson addresses the issue of ecological interference by focusing on the effects of neighborhood characteristics *per se*. The key social mechanisms he investigates are not simple aggregates of individual characteristics (percent black, percent below poverty level), but much richer properties that are built up from individuals’ mutual behaviors and expectations (collective efficacy, neighborhood physical disorder). This requires direct measurement of the neighborhood itself—a nascent field of endeavor that Sampson terms “ecometrics.” In this arena, Sampson and his colleagues are at the forefront, developing methods for measuring qualities and processes of social environments, and putting them on a solid foundation comparable to that underlying psychological measurement.

Longitudinal versus Cross-Sectional Inference

Even if there were no difficulties in connecting static features of the aggregate (such as neighborhood or household) to static features of individuals’ health, a second methodological problem arises from the fact that both social processes and states of health play out over time. Purely cross-sectional analyses can only hint at possible relationships. Moreover, even relatively short-term longitudinal studies may not be adequate to uncover fundamental relationships. Although the time frames may overlap, the critical periods that determine the relationship between a particular social context and particular health outcomes may not. Robert Fogel (2003) has catalogued some convincing examples. Cacioppo and Hawkey, too, recognize that it may well be the cumulative effects of social interactions, embodied in the concept of allostatic load, that are responsible for initiating the cascade leading to adverse health effects. Sampson conceptualizes neighborhood as a dynamic process, and he recognizes that existing tools are weak for measuring such processes. This recognition underlies his call for longitudinal studies designed from the outset to measure both individual and ecological characteristics.

Sampson also argues convincingly that spatially proximate neighborhoods interact with one another to produce effects on health and well-being. Processes that play out in both time and space require the new methods that econometrics will one day provide.

Feedback

A third issue that raises difficult methodological challenges is the likelihood that one's social environment not only affects health and disease but is affected by them as well. For instance, infirmities may lead an individual to move (or be moved) to a nursing home, with consequent changes involving the individual's family dynamic, network of social contacts, psychological state (including perceived social isolation), geographic neighborhood, and economic status. Sampson recognizes this difficulty in dealing with the selective migration account of the poverty and poor health correlation. Exploring in detail the idea that social forces and health (which itself is a construct dependent in part on social consensus) are mutually determinative over a life course, will be a fruitful area for future study.

THE PROBLEM OF IMPORTANCE

Cacioppo and Hawkey and Sampson suggest very different mechanisms by which the social milieu can affect health, and they investigate very different multi-level models. Moreover, the effects of each mechanism do not seem to account for much of the variation in health. Should this bother us?

The way we think about genetic determinants of disease may provide a helpful metaphor. With a few notable exceptions such as Huntington's disease and cystic fibrosis, most diseases are not completely determined by genetic makeup. For instance, the strongest genetic determinant of breast cancer is a mutation of the BRCA1 gene in Ashkenazi Jewish women. Yet having this mutation is neither necessary nor sufficient for developing breast cancer. Indeed, it is estimated that only 12 percent of breast cancers in these women are due to the BRCA1 mutation. Of those women with the mutation (about 2 percent of this population) approximately 60 percent will develop breast cancer by age 70 (Warner et al. 1999). Thus, while it is clearly an important factor in women who carry the BRCA1 mutation, other factors must also be at play that either promote or inhibit the gene's effect, since two out of five carriers will not develop the disease. Geneticists refer to this phenomenon as "incomplete penetrance," and it is an important idea to keep in mind when considering social factors such as neighborhood structure or quality of social interactions: they may influence, but they will rarely determine health outcomes by themselves.

The four principal causes of death accounting for the mortality difference between blacks and whites—hypertension, HIV, diabetes, and homicide (Wong et al. 2002)—account for 43 percent of the difference in life-years lost. But being

black is not a very good predictor of dying from any one of these causes, even though whites die from each at a lower rate. For instance, of all deaths among blacks, 4.2 percent are attributed to diabetes, and 2.8 percent to homicide, compared to 2.7 percent and 0.4 percent, respectively, for whites (Anderson 2002). Social factors to which blacks may have higher exposure than whites may be responsible for some of these differences, but the vast majority of individuals of both races exposed to those factors will not die of diabetes or homicide. In short, social determinants of health or disease are likely to have very low penetrance. As with low-penetrance genetic causes of poor health, we must seek explanations in the interactions between the individual, the individual's environment, and possible initiating factors—the interplay of host, environment, and agent.

CONCLUSION

The Sampson and Cacioppo and Hawkey studies both report that objective measures of social connectedness are much less associated with outcomes than are perceptual measures either of neighbors' willingness to act or of one's own social embeddedness. Both suggest that (objective) social connections set the stage for, but do not determine, positive effects on health or well being. Sampson notes that "social ties create the capacity for informal social control, but it is the act of exercising control rather than the existence of social networks per se that is related to crime." In a sense, these authors describe patterns of active coping to be associated with better health-related outcomes. The relationship, if any, between the individual's coping style and the perceived social-control aspects of collective efficacy would be a rich area to investigate.

It is essential to understand how social forces act by themselves and in concert with other partial determinants—social, behavioral, psychological, political, economic, as well as biological—to effect changes in health. Viewing social factors from the standpoint of the epidemiologic triad, it is important to learn when they are acting to provide a suitable environment for poor health or disease, when they are acting as the agent by which poor health is initiated or maintained, and when they act by increasing susceptibility to other environmental, social, or biological risks.

REFERENCES

- Anderson, R. N. 2002. Deaths: Leading causes for 2000. *Natl. Vital Stat. Rep.* 50(16):1–88.
- Bickel, P. J., E. A. Hammel, and J. W. O'Connell. 1975. Sex bias in graduate admissions: Data from Berkeley. *Science* 187(4175):398–404.
- Blake, P. A., et al. 1977. Cholera in Portugal, 1974. II. Transmission by bottled mineral water. *Am. J. Epidemiol.* 105(4):344–48.
- Cacioppo, J. T., and L. C. Hawkey. 2003. Social isolation and health, with an emphasis on underlying mechanisms. *Perspect. Biol. Med.* 46(Suppl.):S39–S52.

- Fogel, R. W. 2003. Changes in the inequality in health outcomes in the 20th Century. *Perspect. Biol. Med.* 46(Suppl.):S24–S38.
- Laumann, E. O., and Y. Youm. 1999. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. *Sex. Transm. Dis.* 26(5):250–61.
- Merrell, D. S., et al. 2002. Host-induced epidemic spread of the cholera bacterium. *Nature* 417(6889):642–45.
- Rubenstein, A. H. 2002. Health implications of obesity. Paper presented at symposium, Social Determinants of Health and Disease, 15–16 November 2002, University of Chicago.
- Sampson, R. J. 2003. The neighborhood context of well-being. *Perspect. Biol. Med.* 46(Suppl.):S53–S64.
- Warner, E., et al. 1999. Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. *J. Natl. Cancer Inst.* 91(14):1241–47.
- Wong, M. D., et al. 2002. Contribution of major diseases to disparities in mortality. *N. Engl. J. Med.* 347(20):1585–92.