AMERICAN SOCIOLOGICAL ASSOCIATION

Journal of Health and Social Behavior
51(S) S28–S40

DOI: 10.1177/0022146510383498
http://jhsb.sagepub.com
SAGE

Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications

Jo C. Phelan[1], Bruce G. Link[1,2], and Parisa Tehranifar[1]

Abstract

Link and Phelan (1995) developed the theory of fundamental causes to explain why the association between socioeconomic status (SES) and mortality has persisted despite radical changes in the diseases and risk factors that are presumed to explain it. They proposed that the enduring association results because SES embodies an array of resources, such as money, knowledge, prestige, power, and beneficial social connections that protect health no matter what mechanisms are relevant at any given time. In this article, we explicate the theory, review key findings, discuss refinements and limits to the theory, and discuss implications for health policies that might reduce health inequalities. We advocate policies that encourage medical and other health-promoting advances while at the same time breaking or weakening the link between these advances and socioeconomic resources. This can be accomplished either by reducing disparities in socioeconomic resources themselves or by developing interventions that, by their nature, are more equally distributed across SES groups.

Keywords:

health disparities, social stratification, fundamental causes, health, mortality

As we mark the fiftieth anniversary of the Medical Sociology Section of the American Sociological Association, one of the most basic and critical problems addressed by medical sociologists is a very old one: the fact that society's poorer and less privileged members live in worse health and die much younger than the rich and more privileged ones. Socioeconomic inequalities in health and mortality are very large, very robust, and very well documented. Typically, age-adjusted risk of death for those in the lowest socioeconomic level is double to triple that for the highest level (Antonovsky 1967; Sorlie, Backlund, and Keller 1995; Kunst, Feikje, and Mackenbach 1998). To illustrate, in 2005, all-cause, age-adjusted death rates for individuals between the ages of 25 and 64 were strongly related to education level for both men (at < 12 years, 821 per 100,000; at 12 years, 605; and at > 12 years, 249) and women (at < 12 years, 472; at 12 years, 352; and at >12 years, 165) (National Center for Health Statistics 2008). Similar levels of inequality are observed between income groups.

These inequalities in overall health and mortality are not only very common in modern times, but they have persisted at similar levels at least since the early nineteenth century (Antonovsky 1967). This persistence is puzzling because major diseases and risk factors that appeared to account for the

[1]Columbia University
[2]New York State Psychiatric Institute

Corresponding Author:
Jo C. Phelan, Columbia University, Department of Sociomedical Sciences, 722 W. 168th Street, 16th floor, New York, NY 10032
E-mail: jcp13@columbia.edu

inequalities seen in earlier periods (i.e., deadly infectious diseases such as diphtheria, measles, typhoid fever, and tuberculosis fueled by overcrowding and poor sanitation in low socioeconomic status homes and communities) have been virtually eradicated in the developed world. Rather than disappearing, socioeconomic status (SES) inequalities in mortality have persisted and now reflect new major causes of death including cancers and cardiovascular illness, fueled by risk factors such as poor diet, inadequate exercise, and smoking that are more common in lower SES groups. Socioeconomic inequalities in health and mortality have even survived concerted efforts to eliminate them, such as institution of the United Kingdom's National Health Service, their vast publicly-funded health care system (Black et al. 1982).

It is this persistence across time that Link and Phelan (1995) aimed to explain with their theory of fundamental causes. They reasoned that we cannot claim to understand why health inequalities exist if we cannot explain why they persist under conditions that should eliminate or reduce them, and if we can understand why they persist, this may provide clues to the more general problem of the causes of health inequalities. That is, the remarkable persistence of inequalities may provide a lever for understanding the more general fact of their existence.

In this article, we will explicate the theory as it has developed over the past 15 years, review key empirical findings, develop some refinements of the theory, address potential limits of the theory, and discuss implications for health policies that might reduce health inequalities.

THE THEORY

The theory of fundamental causes is rooted in Lieberson's (1985) concept of basic causes, which was first applied to the association between SES and mortality by House and colleagues (House et al. 1990, 1994). The theory has been developed primarily by Link and Phelan (Link and Phelan 1995; Phelan et al. 2004; Link and Phelan, forthcoming), with significant elaboration and extension by Lutfey and Freese (2005).

The primary statement of the theory appeared in 1995 in a previous special issue of the *Journal of Health and Social Behavior*. According to Link and Phelan (1995), a fundamental social cause of health inequalities has four essential features. First, it influences multiple disease outcomes, meaning that it is not limited to only one or a few diseases or health problems. Second, it affects these disease outcomes through multiple risk factors. Third, it involves access to resources that can be used to avoid risks or to minimize the consequences of disease once it occurs. Finally, the association between a fundamental cause and health is reproduced over time via the replacement of intervening mechanisms (Link and Phelan 1995). It is the persistent association of SES with overall health in the face of dramatic changes in mechanisms linking SES and health that led Link and Phelan to call SES a "fundamental" cause of health inequalities.

The Central Role of Flexible Resources for SES Inequalities in Health

According to the theory of fundamental causes, an important reason that SES is related to multiple disease outcomes through multiple pathways that change over time is that individuals and groups deploy resources to avoid risks and adopt protective strategies. Key resources such as knowledge, money, power, prestige, and beneficial social connections can be used no matter what the risk and protective factors are in a given circumstance. Consequently, fundamental causes affect health even when the profile of risk and protective factors and diseases changes radically. If the problem is cholera, for example, a person with greater resources is better able to avoid areas where the disease is rampant, and highly resourced communities are better able to prohibit entry of infected persons. If the problem is heart disease, a person with greater resources is better able to maintain a heart-healthy lifestyle and get the best medical treatment available. Because these resources can be used in different ways in different situations, we call them flexible resources.

It is their capacity to be used flexibly by individuals and groups that places resources of knowledge, money, power, prestige, and beneficial social connections at the center of fundamental cause theory. Their flexible use tells us why SES gradients tend to reproduce themselves over time. This focus on resources and their deployment does not deny the importance of antecedent causes of the resources themselves that lie in the social, economic, and political structures of society. In fact, fundamental cause theory is deeply connected to the sociological study of stratification in this way—the resources highlighted in fundamental

cause theory must come from somewhere, and theories of the origins of inequalities are the best source for understanding these processes. To understand how flexible resources might facilitate the creation of new mechanisms linking SES and health, consider the following example. Screening for several cancers has become possible over the past few decades, making it feasible to detect cancer or its precursors earlier, thereby helping to prevent mortality from these cancers. Since the screening procedures represent relatively recent technological advances, one can imagine a time before the procedures existed, when resources had no bearing on access to cancer screening because the procedures did not exist. There was no mechanism linking SES to screening access to health. But after the screening procedures were developed, people with more resources could use those resources to gain access to the life-saving screens. Link et al. (1998) presented evidence from the Behavioral Risk Factor Survey showing that screening rates for cervical and breast cancer are indeed associated with education and income.[1] A new mechanism had emerged to link social conditions to health outcomes. The idea is that this process extends beyond this example to many, many others.

The flexible resources that are central to fundamental cause theory operate at both individual and contextual levels. At the individual level, flexible resources can be conceptualized as the "cause of causes" or "risk of risks" that shape individual health behaviors by influencing whether people know about, have access to, can afford, and receive social support for their efforts to engage in health-enhancing or health-protective behaviors. In addition, resources shape access to broad contexts that vary dramatically in associated risk profiles and protective factors. For example, a person with many resources can afford to live in a high SES neighborhood where neighbors are also of high status and where, collectively, enormous clout is exerted to ensure that crime, noise, violence, pollution, traffic, and vermin are minimized, and that the best health-care facilities, parks, playgrounds, and food stores are located nearby. Once a person has used SES-related resources to locate in an advantaged neighborhood, a host of health-enhancing circumstances comes along as a package deal. Similarly, a person who uses educational credentials to procure a high-status occupation inherits a package deal that is more likely to include excellent health benefits and less likely to involve dangerous conditions and toxic exposures. In these circumstances, the person benefits in numerous ways that do not depend on his or her own initiative or ability to personally construct a healthy situation; it is an "add on" benefit operative at the contextual level. These contexts may be meso (families) or macro levels (a congressional block that opposes changes in health care policy that would shift the distribution of health care away from high SES groups to the uninsured), formal (employer or trade union) or informal (social networks). The clearest example of fundamental cause theory occurs when groups explicitly push for better health conditions for their members. But the health-enhancing use of group resources can operate at a less explicit level. Consider Cockerham's (2005) ideas about the influence of status groups on health lifestyles. According to Cockerham, social norms and other social supports, such as the health-product industry, reinforce distinctive health lifestyles in different status groups, and the lifestyles of high SES groups are particularly healthy ones. In these instances status groups do not explicitly advocate for health-enhancing conditions, but rather members form cultural practices around food, exercise, and other health-related circumstances that influence the behavior of status-group members. These lifestyles are shaped by the extant stock of health knowledge and pecuniary resources generally available in particular status groups—a circumstance that generally leads to healthier lifestyles in higher status groups. For example, it is almost unheard of for snacks offered at meetings held at the Mailman School of Public Health at Columbia University to not include multiple varieties of fruits; Dunkin Donuts, in contrast, are rare indeed. It is not as if the people who order these snacks explicitly consider the health impact of their choices each time a decision is made. Instead, cultural practices shaped over time lead them to order the conventional, and the conventional in this context is generally healthy fare.

KEY EMPIRICAL FINDINGS

Empirical tests of the theory are not obvious or straightforward. A demonstration of socioeconomic inequalities in health or mortality, even ones that persists over time, does not in itself constitute support for the theory. It is precisely the nearly ubiquitous inverse association between SES and mortality that the theory attempts to explain.

Demonstrating this association in any particular circumstances cannot adjudicate between fundamental causes and other possible explanations of those facts.

Empirical support for the theory relies on evaluating the four essential features of a fundamental cause of health inequalities (Link and Phelan 1995). In the following sections, we present key findings bearing on each of these components: (1) evidence that SES influences multiple disease outcomes; (2) evidence that SES is related to multiple risk factors for disease and death; (3) evidence that the deployment of resources plays a critical role in the association between SES and health/mortality; and (4) evidence that the association between SES and health/mortality is reproduced over time via the replacement of intervening mechanisms.

Evidence That SES Is Related to Multiple Disease Outcomes via Multiple Risk Factors

The first two propositions are strongly supported by empirical data. Low SES is related to a multiplicity of diseases and other causes of death. The broad generality of this association can be summarized with two sets of facts: (1) Low SES is related to mortality from each of the broad categories of chronic diseases, communicable diseases, and injuries (Pamuk et al. 1998; National Center for Health Statistics 2008), and (2) low SES is related to mortality from each of the 14 major causes of death in the International Classification of Diseases (Illsley and Mullen 1985).

There is also clear evidence that SES is associated with numerous risk and protective factors for disease and other causes of death, both currently and in the past. These include smoking, sedentariness, and being overweight (Lantz et al. 1998; Link 2008); stressful life conditions (Turner, Wheaton, and Lloyd 1995; House and Williams 2000); social isolation (House and Williams 2000; Ruberman et al. 1984); preventive health care (Dutton 1978; Link et al. 1998); and crowded and unsanitary living conditions, unsanitary water supplies, and malnutrition (Rosen 1979).

Lutfey and Freese (2005) describe this component of the theory as involving a "massive multiplicity of mechanisms." They suggest that, because fundamental cause processes are "holographic," such a multiplicity of mechanisms should be found in all or most particular instances in which SES and health outcomes are connected. Using an ethnographic analysis, they use the example of routine diabetes care in two socioeconomically contrasting clinics to articulate several concrete ways in which differential health outcomes emerge in the two clinics. For example, the clinic serving higher SES patients provided better continuity of care, and the higher SES patients encountered fewer costs of complying with treatment regimens and had more knowledge about diabetes. Similar analyses conducted in a variety of contexts relating to treatment or prevention of a variety of diseases would enrich our understanding of the pathways through which SES influences health and longevity.

Evidence that the Deployment of Resources Plays a Critical Role in the Association between SES and Health

Central to fundamental cause theory is the idea that resources of money, knowledge, power, prestige, and beneficial social connections are critical to maintaining a health advantage. Empirically testing the importance of resources per se is difficult, because it requires the identification of situations in which the ability to use socioeconomic resources can be analytically separated from SES itself (e.g., situations in which high SES persons are prevented from using their resources to gain a health advantage). If the utilization of resources is critical in maintaining health or prolonging life, then in situations in which the resources associated with higher status are of no use, high SES should confer no advantage, and the usually robust association between SES and health or mortality should be greatly reduced.

One such situation occurs when the causes and cures of fatal diseases are unknown. In these circumstances, socioeconomic resources cannot be used to avoid death due to these diseases, because it is not known how the resources should be deployed. Thus, to the extent that the ability to use socioeconomic resources is critical in maintaining SES inequalities in mortality, there should be strong SES gradients in mortality for causes of death that are highly preventable —for which we have good knowledge and effective measures for prevention or treatment. However, for causes of death about which we know little regarding prevention or treatment, SES gradients in mortality should be much weaker. Consistent with this prediction, Phelan et al. (2004) found that socioeconomic inequalities in mortality were significantly

more pronounced for causes of death that were reliably rated by two physician-epidemiologists as being highly preventable (such as lung cancer and ischemic heart disease), and thus more amenable to the application of flexible resources than for causes that were rated as not very preventable (such as brain cancer and arrhythmias). Although they do not address or explicitly test fundamental cause theory, three other studies that reported evidence on this issue also found that the SES-mortality association was stronger for preventable causes of death (Dahl, Hofoss, and Elstad 2007; Marshall et al. 1993; Song and Byeon 2000).

Evidence for the validity and generality of these findings is strengthened by another study that employed a similar research strategy but (1) examined a different set of causes of death, (2) confined attention to treatment rather than including prevention, (3) used a different and more objective measure of amenability to treatment, and (4) examined racial and ethnic differences as opposed to socioeconomic ones.[2] Tehranifar et al. (2009) identified, prior to hypothesis testing, cancers that are more or less amenable to treatment and examined whether racial-ethnic differences in disease-specific mortality varied according to the degree to which that disease is amenable to available medical intervention. This study used five-year survival rates for 53 different cancer sites as a measure of effectiveness of treatment and/or early detection methods. Consistent with fundamental cause theory, survival disparities comparing disadvantaged minority groups (African Americans, American Indians, and Hispanics) to whites were substantially greater for cancers that were more amenable to treatment (e.g., cancers with five-year relative survival rates $\geq$ 70%, such as bladder, breast, and prostate cancers) than they were for cancers that were less so (e.g., cancers with five-year relative survival rates $<$ 40%, such as liver, pancreatic, and esophageal cancers).

These studies show that, somewhat ironically, one way in which fundamental cause theory can be tested is by looking for exceptions to the strong SES gradient in health or mortality that is almost always observed—exceptions in which the ability to use resources to gain a health advantage is blocked. In these examples, the use of socioeconomic resources to improve health is blocked because risk factors are unknown and treatments do not exist (Phelan et al. 2004; Tehranifar et al. 2009). Other situations in which resources may be unhelpful or even harmful may be exploitable for testing of the theory. Examples are situations in which prevailing medical recommendations are subsequently discovered to be harmful (Carpiano and Kelly 2007) and the case of old age, when the growing frailty of the body may place limits on the effectiveness of interventions (Phelan et al. 2004).

Evidence That the Association between SES and Health/Mortality Is Reproduced over Time via the Replacement of Intervening Mechanisms

The fourth essential feature of SES as a fundamental cause of health inequalities is that the association between SES and health/mortality is reproduced over time via the replacement of intervening mechanisms. This key element of the theory arose from two sets of observations: (1) The SES-mortality association persisted over time despite the decline of mechanisms (e.g., poor sanitation and widespread death from infectious disease) that formerly provided important links between SES and mortality; and (2) new, previously weak or absent mechanisms currently link SES and mortality (e.g., smoking, exercise, diet, and cardiovascular disease). These observations are consistent with the idea that socioeconomic inequalities in health are reproduced via the replacement of intervening mechanisms. To more fully evaluate this component of the theory, however, more direct evidence was needed showing the emergence of new mechanisms. In particular, the theory predicts that new mechanisms arise following the development of new knowledge or medical intervention related to some disease, because higher SES individuals and groups are better equipped to take advantage of the new knowledge. Therefore, a key empirical question is whether the SES-health gradient shifts in favor of high SES individuals following the development of new knowledge. This evidence is particularly persuasive if the health outcome for which a shifted gradient is observed is directly related to the emergent knowledge, for example, if an advance in heart disease treatment furthers the advantage of high SES individuals in terms of heart disease mortality. Just as important is evidence that, in the absence of advances in knowledge, the SES gradient in relevant health outcomes remains fairly steady.

Several such analyses have now been conducted. Phelan and Link (2005) examined selected causes of death for which great strides in prevention or treatment were made over the last half of the twentieth century (heart disease, lung cancer,

and colon cancer), and for which much less progress had been made over the same period (brain cancer, ovarian cancer, and pancreatic cancer). Looking at age-adjusted death rates by race and by county-level SES, they reported that, for the causes of death where little had been learned about treatment or prevention, mortality rates stayed fairly steady, and the degree of inequality based on race and SES stayed fairly steady as well. By contrast, for the causes of death where gains in treatment and prevention had been significant, overall mortality rates declined while race and SES gradients shifted in the direction of relatively higher mortality for the less advantaged group.

Subsequent studies have gone much further in drawing specific connections between gains in knowledge and subsequent changes in relevant disease outcomes. Carpiano and Kelly (2007) analyzed changes in breast cancer incidence following the widely publicized findings from the Women's Health Initiative (WHI) that linked hormone replacement therapy to increased breast cancer risk (Haas et al. 2004). In the following two years, consistent with the racial pattern in the use of hormone replacement therapy (Haas 2004; Hulley et al. 1998), breast cancer incidence among white women age 50 and older, the age group most likely to have been using hormone therapy before the WHI study results were publicized, dropped precipitously, while incidence among black women in that age group stayed fairly steady (Carpiano and Kelly 2007). These findings were confirmed by another study that also considered county-level median household income and breast tumor estrogen (ER) receptor status (Krieger, Chen and Waterman 2010). That study found the decline in breast cancer incidence after the WHI study publication to be limited to white women, aged 50 and older, who were residents of high income counties and had estrogen-positive breast tumors (the type of tumor most likely to be affected by hormone replacement).

Chang and Lauderdale (2009) studied the impact of statins (an effective and expensive medication to lower cholesterol) on socioeconomic gradients in cholesterol levels. Using nationally representative data from 1976 to 2004, they found that those with higher income initially had higher cholesterol levels, but that the SES-cholesterol association then reversed and became negative in the era of widespread statin use.

Link (2008) traced changes in knowledge, beliefs, and behavior that followed the discovery of a causal link between cigarette smoking and lung cancer, and that eventually led to strong socioeconomic gradients in smoking. Scientific evidence strongly linking smoking to lung cancer emerged in the early 1950s. To assess changes that may have occurred in the decades following the production of this new knowledge, Link (2008) analyzed multiple public opinion polls assessing smoking beliefs and behaviors. Evidence from the first surveys conducted just as the scientific evidence was emerging in 1954 showed that, while most people had heard about the findings, only a minority believed that smoking was a cause of lung cancer, and no educational gradient in this belief was evident. Nor was smoking behavior strongly linked to educational attainment in 1954. Over the subsequent 45 years, as people began to adopt the belief that smoking is a cause of lung cancer, sharp educational gradients opened up in this belief. Additionally, people of higher education were less likely to start smoking and more likely to quit, thereby generating a strong SES gradient in smoking behavior (Link 2008). A new and powerful mechanism linking SES to an important health behavior had emerged.

The studies just described are particularly valuable for their ability to pinpoint temporal connections between particular developments in knowledge and technology surrounding specific diseases, on the one hand, and changes in SES-related health gradients predicted by the theory, on the other. Moreover, these studies address major diseases that are important causes of death. However, there is always the possibility that these cases are not representative of the situation that holds more generally when new health knowledge or technology develops. For this reason, the more systematic and comprehensive analysis of Glied and Lleras-Muney (2008) is particularly valuable. This study provides evidence that the results of the case studies reported above are indeed generalizable. Like Phelan et al. (2004) and Tehranifar et al. (2009), Glied and Lleras-Muney conducted a systematic test based on a comprehensive set of diseases. In fact, Glied and Lleras-Muney repeated their analysis with two separate data sets: the Mortality Detail Files from the National Center for Health Statistics, and the Surveillance Epidemiology and End Results cancer registry. They operationalized the development of life-saving knowledge and technology, or "innovation," in two ways. In the first they used the rate of change in mortality over time to indicate progress in addressing mortality due to particular diseases, the assumption being that the greater the decline in mortality, the greater the progress that has been made. In the second,

they used the number of active drugs approved to treat particular diseases, with the assumption that more progress has been made where more new drugs have been developed to treat disease. They found, consistent with the theory of fundamental causes, that educational gradients became larger for diseases where greater innovation had occurred.

In summary, evidence has accumulated that is consistent with each of the four components of fundamental cause theory. Empirical testing of the theory is accelerating, and studies are now being conducted by researchers other than the theory's originators. This is a desirable development, as it raises confidence that the theory is being subjected to scientific scrutiny.

RETURNING TO THE THEORY: REFINEMENTS AND LIMITATIONS

Refinements to Fundamental Cause Theory

The theory has two sets of implications for continuity and change in health inequalities over time. The theory's basic principle—that a superior collection of flexible resources held by higher SES individuals and the collectivities to which they belong allow those of higher SES to avoid disease and death in widely divergent circumstances—leads to the prediction that, at any given time, greater resources will produce better health, and consequently inequalities in health and mortality will persist as long as resource inequalities do.

At the same time, this long-term stability in the association between SES and health/mortality results from the amalgamation of effects across many specific processes and conditions. New knowledge and technology relating to innumerable diseases emerges constantly. The nature of the new knowledge varies, and the social conditions in which this knowledge emerges also vary. As a result, while in general new knowledge and medical development about a disease will lead to a shift in the disease gradient in favor of higher SES individuals and groups, they will not all have an identical impact on this gradient. Another reason for the long-term stability in the SES-mortality association is that old mechanisms wane to be replaced by new ones. Again, the demise of mechanisms is not a uniform process: Some mechanisms have long lives, others short ones. In this section, we take steps toward understanding some of the conditions that lead to variations in the processes of mechanism generation and demise. Our aim is not only to strengthen the theory but to understand how it may be possible to weaken new mechanisms connecting SES and disease/mortality, and how old ones may be undermined.

Specifying Conditions that Modify the Impact of New Knowledge on Health Inequalities

The situation that most clearly exemplifies fundamental cause processes is one in which we initially know nothing about how to prevent or cure a disease, and there is no association between SES and morbidity or mortality due to that disease. Then, upon discovery of modifiable risk or protective factors, an inverse association between SES and the disease in question emerges. But other situations that differ from this prototype are not only possible but to be expected.

One factor that should modify the impact of emergent knowledge is the pre-existing SES distribution of the disease at the time of a new advance in prevention or treatment.[3] The pre-existing association between the disease and SES is unlikely to be null for two reasons. First, when new knowledge and technology emerge, it is often the case that prior knowledge and technology have already shaped the association between SES and disease; the new knowledge will further shape this association. Second, even in the absence of previous knowledge about its risk and protective factors, a disease may be influenced by factors that are associated with SES, either directly or inversely. For example, before cholesterol was identified as a risk factor for cardiovascular disease its levels were likely higher in higher SES populations because such populations had greater access to relatively expensive fatty foods.

The reason that prior associations between risk factors or diseases and SES are important for fundamental cause theory is that the new knowledge has greater utility for those who have the disease or risk factor. Notably, if the initial association between SES and the disease is *inverse* such that people of lower SES are at greater risk, an effective intervention can reduce inequalities in that disease. This is because more people of low SES are likely to benefit from the intervention, because more of them have the disease initially. This can be true even if persons of higher SES who have the disease are more likely to gain access to and benefit from the intervention than lower SES persons who have the disease. We call this a "give back

effect" (Link and Phelan, forthcoming), because the initial inverse SES-to-disease association provides a starting point that allows the new knowledge about the disease to "give back" some equality even though it may also exemplify a fundamental cause process in which the knowledge is not distributed equally across socioeconomic groups. For example, smoking is a risk factor that has been influenced by knowledge of its harmful effects such that what was once a direct SES-to-smoking association has become a sharply graded inverse association, and one reason that SES is related to smoking-related diseases.

In this context a "give back" effect would arise if a new intervention blocked the effect of smoking on heart disease or lung cancer mortality. Even if this new intervention was itself maldistributed by SES, a "give back" effect might arise because smoking is so much more common in low SES populations; in other words, there are more people at the low end who can benefit from the new intervention. Importantly, from a fundamental cause perspective, if the intervention had been discovered earlier, before an SES-to-disease association in smoking emerged, and if the intervention had been maldistributed by SES at that time, the intervention would have created an inverse association between SES and lung cancer or heart disease.

Mechanism Demise and Death

Whereas it is understandable that empirical tests have focused on the creation of mechanisms that produce health inequalities, fundamental cause theory is predicated on the idea that mechanisms are *replaced*. Replacement requires that old mechanisms wane in importance over time. In fact, the theory emerged in part *because* prominent risk-factor mechanisms associated with vicious infectious diseases declined in significance as germ theory, improved sanitation, and vaccination came into existence. Thus, understanding the demise and death of mechanisms linking flexible resources to disease is an important area that needs more development and testing. We offer two examples that may help others develop this area of inquiry more fully.

Salk's discovery of the polio vaccine is an example of a mechanism that was very short-lived. Before his discovery, people of all resource levels could be afflicted, including, for example, President Franklin Roosevelt. After the discovery, resource-rich individuals were more likely to receive the vaccine and be protected. A mechanism linking resources to health existed, but only for a short time. The vaccine was quickly approved for widespread distribution to the U.S. population, and polio was virtually eradicated here. Other mechanisms remain potent for a very long time. For, example the discovery of the pap test for the early detection and prevention of cervical cancer has existed since the 1940s. Early on, access to the test was shaped by flexible resources creating an inequality in the use of this life-saving screen that remains prominent today. As these examples suggest, some mechanisms become long-lasting while others have short lives. If we can understand what leads to the demise of mechanisms, and especially how that decline is related to flexible resources, we may open avenues to speed such a demise and reduce health inequalities. Indeed, much of the public health significance of fundamental cause theory may reside in understanding how the link between flexible resources and health-relevant risk and protective factors has been broken.

Limits on Fundamental Cause Theory: Countervailing Mechanisms

Whereas the previous sections elaborated fundamental cause theory, here we consider conditions that place limits on the theory.

We believe readers will agree that health and longevity are desirable, but they are not all that a person may want. Other things being equal, those with more resources can be expected to deploy those resources to increase health. But there are undoubtedly situations in which the goals of health and long life compete with and may cede dominance to other important life goals. Perhaps desiderata such as power, manliness, or beauty are sometimes more powerful motivators than health, and are pursued to the detriment of health. Lutfey and Freese (2005) refer to these competing goals as "countervailing mechanisms." The potential for countervailing mechanisms does not threaten the truth-value of fundamental cause theory, because "fundamental relationships do not require that all of the pathways between X and Y support the relationship. The only requirement is that the effects of [countervailing] mechanisms are cumulatively smaller than the effects of mechanisms producing the fundamental relationship" (Lutfey and Freese 2005:1365). However, to the extent that countervailing mechanisms are called upon *post hoc* to explain results that do not support the theory,

countervailing mechanisms pose a challenge to the falsifiability of the theory. For this reason, as well as for the fuller understanding of health inequalities, it is desirable to attend to countervailing mechanisms systematically, as Lutfey and Freese argue, and attempt to move the consideration of countervailing mechanisms from *post hoc* to *a priori*.

We first note that the connection between SES and health is an extremely powerful one and that goals that successfully compete with those of health and long life must surely be quite potent. For example, the goal of health attainment has been powerful enough to override or socially reconstruct many aspects of pleasure and pain—which would seem to be basic and powerful forces in their own right—among the socioeconomically privileged. Erstwhile pleasures such as well marbled steaks are eschewed by higher SES groups in favor of sushi-grade tuna. Similarly, in the past, exhausting physical activity was considered something that high SES people were fortunate enough to be able to avoid. Now, "no pain no gain" prevails in the most expensive health clubs. Cigarette smoking, although highly addictive, as well as sexual practices that increase the risk of HIV/AIDS, have also been significantly altered by high SES groups in the name of health attainment. We also note that the goals of health and longevity are strongly supported by social norms and other forms of social support among high status groups as part of the beneficial health lifestyle associated with high SES (Cockerham 2005). We suggest that the power of health attainment to shape the behavior of high SES individuals is largely due to these social forces, and we propose that successful countervailing mechanisms are also likely to be embedded in strong social norms and support.

One such motivation that may meet these conditions is status attainment. In Lutfey and Freese's (2005) ethnographic analysis of diabetes treatment, the pursuit of status, for example, occupational success or staying thin, sometimes led higher SES diabetic patients to behave in ways detrimental to the management of their disease. Similarly, Courtenay (2000) suggests that signifiers of masculinity such as the denial of weakness and engagement in risky or aggressive behavior often undermine men's health. Thus, the pursuit of masculine status may help explain the fact that women, who are generally lower resourced than men, live longer than men, a fact that would not be predicted by fundamental cause theory.[4] It seems, then, that status pursuit is one potential countervailing mechanism to the SES-health association. In the context of particular empirical studies, researchers may be able to consider *a priori* whether the situation under study is one in which the goals of health and social status are likely to collide. Additional motivations that might potentially be powerful enough to operate as countervailing mechanisms to SES include power, affiliation, self-esteem, identity, freedom, creation, and leisure (Maslow 1943; Max-Neef, Elizalde, and Hopenhayn 1989).

Note that, in most circumstances, we would expect the goal of good health to be compatible with goals of power, self-esteem and so on, and we would expect higher SES individuals to use their resources to achieve more of all these desiderata than lower SES persons would be able to. Still, the example of status pursuit as a countervailing mechanism suggests that there will be instances when other powerful motivations that are more readily attained by high SES persons work to the detriment of health. In those situations, the usual association between resources and health should be attenuated. Also note that these countervailing mechanisms may create conditions when SES will not operate as a fundamental cause of health and mortality, but they do not negate the power of SES as a fundamental cause of unequal life chances more generally.

IMPLICATIONS FOR HEALTH POLICY

The fundamental cause approach leads to very different policies for addressing health inequalities than does an individually oriented risk-factor approach. The latter asks us to locate modifiable risk factors that lie between distal cause (such as SES) and disease, and to intervene in those risk factors. By addressing intervening factors, the logic goes, we will eliminate health disparities.

Our approach points to the pitfalls of this logic and suggests that developing new interventions, even when beneficial to health, is very likely to increase social inequalities in health outcomes. The idea that medical progress often leads to increased health inequality leads to an obvious conundrum: Must we choose between improving overall levels of health and reducing inequalities in health? Some argue that continued inequalities in health outcomes are acceptable as long as overall health improves or that some improvement is achieved for

most social groups. We, on the other hand, are committed to reducing health inequalities, but it seems wrong-headed to oppose advances in health knowledge and technology because those may increase inequalities. We see no reason not to make both outcomes important goals, simultaneously pursuing better overall health and reduced inequalities.

We suggest some general strategies that we believe will lead to improved overall population health without further widening social inequalities in health. Our approach points to policies that encourage advances while breaking or weakening the link between these advances and socioeconomic resources, either by reducing disparities in socioeconomic resources themselves, or by developing interventions that, by their nature, are more equally distributed across SES groups.

Reduce Resource Inequalities

The first recommendation falls outside the explicit domain of health policy, but according to fundamental cause theory is intimately tied to it. The theory stipulates that people and collectivities use their knowledge, money, power, prestige, and social connections to gain a health advantage, and thereby reproduce the SES gradient in health. The most direct policy implication of the theory is that, if we redistribute resources in the population so as to reduce the degree of resource inequality, inequalities in health should also decrease. Policies relevant to fundamental causes of disease form a major part of the national agenda, whether this involves the minimum wage, housing for homeless and low-income people, capital-gains and estate taxes, parenting leave, social security, head-start programs and college-admission policies, regulation of lending practices, or other initiatives of this type. We argue that all these policies are health-relevant policies and that understanding how they are relevant should be claimed as an essential part of the domain of medical sociology.

Contextualize Risk Factors

Potential interventions that seek to change individual risk profiles should first identify factors that put people at risk of risks, for example, power disadvantages that prevent some people from adopting safe sex strategies or neighborhood environments that make healthful foods unavailable. This will avoid the enactment of interventions aimed at changing behaviors that are powerfully influenced by factors left untouched by the intervention.

Prioritize the Development of Interventions that Do Not Entail the Use of Resources or that Minimize the Relevance of Resources

As we seek to create interventions to improve health, we need to ask if an intervention is something that anyone can potentially adopt, or whether the benefit will only be available to people with the necessary resources. Fundamental cause theory suggests that health inequalities based on SES can be reduced by instituting health interventions that automatically benefit individuals irrespective of their own resources or behaviors. Examples are the manufacture of automobiles with air bags as opposed to relying on the use of seatbelts; providing health screenings in schools, workplaces, and other community settings rather than only through private physicians; providing health care to all citizens rather than only to those with the requisite resources; requiring window guards in all high-rise apartments rather than advising parents to watch their children carefully; thoroughly inspecting meat rather than advising consumers to wash cutting boards and cook meat thoroughly; adding folic acid to grains rather than recommending that supplements be taken by pregnant women to prevent neural tube defects in developing embryos; requiring landlords to keep homes free of lead paint hazards rather than warning parents to protect their toddlers from chipped paint. In some cases, such as this last example, existing risks will be greater in low-income neighborhoods and contexts, and special enforcement of these policies may be required in those contexts. In each example, the former solution does not give an advantage to those with greater resources, because individual resources are unrelated and irrelevant to benefiting from the intervention.

However, even if we become far more creative in developing contextually based interventions that blanket an entire population with health benefit, addressing many health problems will still require individual resources and action. In these cases, resource-rich persons are likely to fare better. Even in these cases, however, we can influence the trajectory of inequalities by attending to the type of interventions we adopt. When we create interventions that are expensive, complicated and time-consuming to carry out, and difficult to distribute broadly, we are likely to create health disparities (Chang and Lauderdale 2009). Conversely, to the extent that we develop interventions that are relatively affordable and easy to disseminate and use, we should be able to reduce the

degree to which new interventions give advantage to high SES persons. Goldman and Lakdawalla (2005) analyzed two case studies supporting the idea that the introduction of difficult-to-implement treatments (in their analysis, HAART treatment for HIV/AIDS) lead to increased SES inequalities in health outcomes, whereas treatments that are simpler and require less effort (in their analysis, beta-blockers to reduce hypertension) reduce such inequalities. As Chang and Lauderdale (2009) suggest, this principle should also apply to cost: New interventions that are less expensive should result in smaller SES-based health inequalities than those that are more expensive. Chang and Lauderdale also point out, importantly, that, "technologies that have the potential to contract disparities will not do so unless they also diffuse broadly" (Chang and Lauderdale 2009:257). We add that a necessary ingredient of successful diffusion will be broadly disseminated and clearly stated information about how an intervention can help one's health, where that intervention is available, whether and how much of it is covered by health insurance plans, and, if not, how much it will cost individuals.

CONCLUSION

The theory of fundamental causes attempts to explain why the association of SES to health and mortality has persisted despite the demise of risk factors and diseases that appeared to explain the association. Mounting evidence in support of the theory of fundamental causes begins to suggest that the theory is not just an interesting idea but very possibly a valid explanation of persistent SES inequalities in health and mortality. We believe this empirical support warrants the investment of medical sociologists in (1) further empirical analyses using a variety of methodologies to give greater weight to the body of research, to specify and elaborate the processes at work, and to find conditions that may block these processes and (2) developing elaborations, extensions, and modifications of the theory itself. We also believe the accumulated evidence warrants serious attention to the implications of the theory for health policy. Those implications are that, to achieve greater equality in matters of life, death, and health, the connection between socioeconomic resources and health-beneficial preventive measures and treatments must be broken or diminished, by reducing the magnitude of inequalities in socioeconomic resources themselves and/or by minimizing the extent to which socioeconomic resources buy a health advantage. By attending to these principles, we believe we can move toward the important dual goals of continuing to improve overall population health while distributing that health more equally.

FUNDING

This work was supported by a Young Investigator Award granted to Professor Tehranifar by the Lance Armstrong Foundation.

NOTES

1. We acknowledge recent debate and changes in guidelines with regard to screening interval and age at initiation of screening mammography and pap tests. However, convincing evidence supports the effectiveness of these screens in reducing cancer mortality and morbidity (U.S. Preventive Services Task Force 2009; ACOG Committee on Practice Bulletins—Gynecology 2009).
2. Fundamental cause theory was developed to explain the enduring effects of SES on health and mortality. It is possible that other social statuses, such as race, ethnicity, or gender, also have enduring associations with resources of money, knowledge, power, prestige, and beneficial social connections, and with health and mortality, and that they may also operate as fundamental causes. Even if not, however, race and ethnicity are currently strongly related to resources and consequently would be expected to behave similarly to SES in analyses such as Tehranifar's (Tehranifar et al. 2009), which focus on the current health context.
3. We thank David Mechanic for this insight.
4. Recent research suggests that, when health behaviors of women come to resemble those of men more closely, the female mortality advantage declines (Preston and Wang 2006).

REFERENCES

ACOG Committee on Practice Bulletins—Gynecology. 2009. "ACOG Practice Bulletin No. 109: Cervical Cytology Screening." *Obstetrics and Gynecology* 114:1409–20.

Antonovsky, A. 1967. "Social Class, Life Expectancy, and Overall Mortality." *Milbank Memorial Quarterly* 45:31–73.

Black, Douglas, J. N. Morris, Cyril Smith, and Peter Townsend. 1982. *Inequalities in Health: The Black Report*. Middlesex, England: Penguin.

Carpiano, Richard M. and Brian C. Kelly. 2007. "Scientific Knowledge as Resource and Risk: What Does Hormone Replacement Tell Us About Health Disparities." Presented at the annual meeting of the American Sociological Association, August 2007, New York.

Chang, Virginia W. and Diane S. Lauderdale. 2009. "Fundamental Cause Theory, Technological Innovation, and Health Disparities: The Case of Cholesterol in the Era of Statins." *Journal of Health and Social Behavior* 50:245–60.

Cockerham, William. 2005. "Health Lifestyle Theory and the Convergence of Agency and Structure." *Journal of Health and Social Behavior* 46:51–67.

Courtenay, Will H. 2000. "Constructions of Masculinity and their Influence on Men's Well-Being: A Theory of Gender and Health." *Social Science and Medicine* 50:1385–1401.

Dahl, Espen, Dag Hofoss, and Jon I. Elstad. 2007. "Educational Inequalities in Avoidable Deaths in Norway: A Population Based Study." *Health Sociology Review* 16:146–59.

Dutton, Diana H. 1978. "Explaining the Low Use of Health Services by the Poor: Costs, Attitude, or Delivery Systems." *American Sociological Review* 43:348–68.

Glied, Sherry and Adriana Lleras-Muney. 2008. "Technological Innovation and Inequality in Health." *Demography* 45:741–61.

Goldman, D. P. and D. N. Lakdawalla. 2005. "A Theory of Health Disparities and Medical Technology." *Contributions to Economic Analysis and Policy* 4:1–30.

Haas, J. S., C. P. Kaplan, E. P. Gerstenberger, and K. Kerlikowske. 2004. "Changes in the Use of Postmenopausal Hormone Therapy after the Publication of Clinical Trial Results." *Annals of Internal Medicine* 140:184–88.

House, James S., Ronald C. Kessler, A. R. Herzog, R. P. Mero, A. M. Kinney, and M. J. Breslow. 1990. "Age, Socioeconomic Status, and Health." *The Milbank Quarterly* 68:383–411.

House, James S., James M. Lepkowski, Ann M. Kinney, Richard P. Mero, Ronald C. Kessler, and A. Regula Herzog. 1994. "The Social Stratification of Aging and Health." *Journal of Health and Social Behavior* 35:213–34.

House, James S. and David R. Williams. 2000. "Understanding and Reducing Socioeconomic and Racial/Ethnic Disparities in Health." Pp. 81–124 in *Promoting Health: Intervention Strategies from Social and Behavioral Research*, edited by B. D. Smedley and S. L. Syme. Washington, DC: National Academy Press.

Hulley, S., D. Grady, T. Bush, C. Furberg, D. Herrington, B. Riggs, and E. Vittinghoff. 1998. "Randomized Trial of Estrogen Plus Progestin for Secondary Prevention of Coronary Heart Disease in Postmenopausal Women: Heart and Estrogen/Progestin Replacement Study (HERS) Research Group." *Journal of the American Medical Association* 280: 605–13.

Illsley, Raymond and Ken Mullen. 1985. "The Health Needs of Disadvantaged Client Groups." Pp. 389–402 in *Oxford Textbook of Public Health*, edited by W. W. Holland, R. Detels, and G. Know. Oxford, England: Oxford University Press.

Krieger, N., J. T. Chen, and P. D. Waterman. 2010. "Decline in U.S. Breast Cancer Rates after the Women's Health Initiative: Socioeconomic and Racial/Ethnic Differentials." *American Journal of Public Health* 100:S132–S139.

Kunst, A. E., G. Feikje, and J. P. Mackenbach [EU Working Group on Socioeconomic Inequalities in Health]. 1998. "Occupational Class and Cause Specific Mortality in Middle Aged Men in 11 European Countries: Comparison of Population Based Studies." *British Medical Journal* 316:1636–42.

Lantz, P. M., J. S. House, J. M. Lepkowski, D. R. Williams, R. P. Mero, and J. Chen. 1998. "Results from a Nationally Representative Prospective Study of U.S. Adults." *Journal of The American Medical Association* 279:1703–08.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Link, Bruce G. 2008. "Epidemiological Sociology and the Social Shaping of Population Health." *Journal of Health and Social Behavior* 49:367–84.

Link, Bruce G., M. Northridge, J. Phelan, and M. Ganz. 1998. "Social Epidemiology and the Fundamental Cause Concept: On the Structuring of Effective Cancer Screens by Socioeconomic Status." *The Milbank Quarterly* 76:375–402.

Link, Bruce G. and Jo C. Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior* 35(Extra Issue):80–94.

———. Forthcoming. "Social Conditions as Fundamental Causes of Health Inequalities." In *Handbook of Medical Sociology*, 6th ed., edited by C. Bird, P. Conrad, and A. Fremont. Nashville, TN: Vanderbilt University Press.

Lutfey, Karen and Jeremy Freese. 2005. "Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic

Visit for Diabetes." *American Journal of Sociology* 110:1326–72.

Marshall, Stephen W., Ichiro Kawachi, Neil Pearce, and Barry Borman. 1993. "Social Class Differences in Mortality from Diseases Amenable to Medical Intervention in New Zealand." *International Journal of Epidemiology* 22:255–61.

Maslow, A. H. 1943. "A Theory of Human Motivation." *Psychological Review* 50:370–96.

Max-Neef, Manfred, Antonio Elizalde, and Martín Hopenhayn. 1989. "Human Scale Development: An Option for the Future." *Development Dialogue: A Journal of International Development Cooperation* 1:7–80.

National Center for Health Statistics. 2008. *Health United States 2008: With Chartbook on Trends in the Health of Americans*. Hyattsville, MD.

Pamuk, E. R., D. Makuc, K. Heck, C. Reuben, and K. Lochner. 1998. *Socioeconomic Status and Health Chartbook: Health, United States, 1998*. Hyattsville, MD: National Center for Health Statistics.

Phelan, Jo C. and Bruce G. Link. 2005. "Controlling Disease and Creating Disparities: A Fundamental Cause Perspective." *The Journals of Gerontology* 60B(special issue II):27–33.

Phelan, Jo C., Bruce G. Link, Ana Diez-Roux, Ichiro Kawachi, and Bruce Levin. 2004. "Fundamental Causes of Social Inequalities in Mortality: A Test of the Theory." *Journal of Health and Social Behavior* 45:265–85.

Preston, S. H. and H. Wang. 2006. "Sex Mortality Differences in the United States: The Role of Cohort Smoking Patterns." *Demography* 43:631–46.

Rosen, G. 1979. "The Evolution of Social Medicine." Pp. 23–50 in *The Handbook of Medical Sociology*, 3rd ed., edited by H. Freeman, S. Levine, and L. Reeder. Englewood Cliffs, NJ: Prentice Hall.

Ruberman, William, Eve Weinblatt, Judith D. Goldberg, and Banvir S. Chaudhary. 1984. "Psychological Influences on Mortality after Myocardial Infarction." *New England Journal of Medicine* 311:552–59.

Song, Yun-Mi and Jai Jun Byeon. 2000. "Excess Mortality from Avoidable and Non-Avoidable Causes in Men of Low Socioeconomic Status: A Prospective Study in Korea." *Journal of Epidemiology and Community Health* 54:166–72.

Sorlie, P., M. Backlund, and J. Keller. 1995. "U.S. Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study." *American Journal of Public Health* 85:949–56.

Tehranifar, P., A. I. Neugut, J. C. Phelan, B. G. Link, Y. Liao, M. Desai, M. B. Terry. 2009. "Medical Advances and Racial/Ethnic Disparities in Cancer Survival." *Cancer Epidemiology, Biomarkers and Prevention* 18:2701–08.

Turner, R. Jay., Blair Wheaton, and Donald A. Lloyd. 1995. "The Epidemiology of Social Stress." *American Sociological Review* 60:104–25.

U.S. Preventive Services Task Force. 2009. "Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 151:716–26.

Bios

Jo C. Phelan is professor of sociomedical sciences at the Mailman School of Public Health of Columbia University. Her research interests include social stigma, conceptions of mental illness, the impact of the "genetics revolution" on the stigma of mental illness, attitudes and beliefs relating to social inequality and its legitimation, and social inequalities in health and mortality. In collaboration with Bruce Link, she developed the argument that frames social conditions as fundamental causes of disease.

Bruce G. Link is professor of epidemiology and sociomedical sciences at the Mailman School of Public Health of Columbia University, and a Research Scientist at New York State Psychiatric Institute. His interests include the nature and consequences of stigma for people with mental illnesses, the connection between mental illnesses and violent behaviors, and explanations for associations between social conditions and morbidity and mortality.

Parisa Tehranifar is assistant professor of epidemiology at the Mailman School of Public Health of Columbia University. Her research combines social science and epidemiologic perspectives and methods to studies of social inequalities in cancer and other chronic disease risk.