



PART ONE

BACKGROUND

COPYRIGHTED MATERIAL



CHAPTER ONE

INTRODUCTION: ADVANCING METHODS IN SOCIAL EPIDEMIOLOGY

J. Michael Oakes and Jay S. Kaufman

The aim of this brief introductory chapter is to highlight some of the fundamental methodological issues facing social epidemiology. In many cases, these are the background issues that this volume's contributing authors have weaved into each of the chapters that follow.

It is necessary to first define social epidemiology and social epidemiologic methodology, as these definitions underlie all of the discussion that follows. Subsequently, we discuss three fundamental issues that typically arise in the application of social epidemiologic methodology. We conclude by offering a short and speculative discussion on methods not included in this text that may help advance the field beyond its present limitations.

What Is Social Epidemiology?

Epidemiology is the study of the distribution and determinants of states of health in populations. We define *social* epidemiology as the branch of epidemiology that considers how social interactions and collective human activities affect health. In other words, social epidemiology is about how a society's innumerable social arrangements, past and present, yield differential exposures and thus differences in health outcomes among the persons who comprise the population. Defining social epidemiology in this broad way permits the analysis of not only how social

factors serve as exposures that affect health outcomes but also how such factors/exposures emerge and are maintained in a distinctive distribution.

Social epidemiology is thus concerned with more than the identification of new disease-specific risk factors (for example, a deficit of social capital); it also considers how well-established exposures, such as cigarette smoking, lead paint, and lack of health insurance, emerge and are distributed by the social system. With such a focus, social epidemiology must consider the dynamic social relationships and human activities that ultimately locate toxic dumps in one neighborhood instead of another, make fresh produce available to some and not others, and permit some to enjoy resources such that they can purchase salubrious environments and competent health care. In short, social epidemiology is about social allocation mechanisms (that is, economic and social forces) that produce differential exposures that often yield health disparities, whether deemed good or bad.

Social epidemiology is different from the bulk of traditional epidemiologic practice, which tends to operate with a model based on the fictitious Robinson Crusoe. Recall that this character is someone in an environment devoid of social context, whose health depends only on biological relationships and the vicissitudes of island weather. Social interaction and thus political and economic power play no role in Robinson's health, although the same is perhaps not so true for his "friend" Friday. Such interactions are central to social epidemiology, however. Without any attention to social arrangements and institutions, epidemiologic research on humans is almost indistinguishable from an application to, say, livestock.

It is the incorporation of purposive human interaction and agency (that is, social coordination and conflict) that links social epidemiology to the social sciences and raises enormous methodological obstacles to inference—obstacles that leading social scientists have long sought to overcome. But social epidemiology is not a social science, at least as traditionally conceived. Although the methods and models of, say, a social epidemiologist and medical sociologist might be identical, the distinction between social epidemiology and social science lies in the focus, outcome variable, or more formally the "explanandum" of each discipline. The goal of social science—including sociology, economics, political science, and anthropology—is to understand and explain the social system. In other words, social science's outcome variable (that is, explanandum) is society, social forces, or the like. A social scientific study that considers/models health outcomes does so to learn about society. By contrast, the outcome variable for social epidemiology is health. Whereas social epidemiologists may borrow theory, methods, and constructs from social science, they do so in an effort to understand health rather than social forces or related phenomena. This means that social epidemiology, although related to the social sciences, firmly remains a branch of epidemiology. Accordingly, social epidemiology should not discount the potential impact of genes,

microbes, or other factors frequently found in other subfields within epidemiology. The inevitable decline in the importance of (sub)disciplinary boundaries is a necessary step for the integration of these diverse considerations, as it frequently requires multidisciplinary teams to properly address the important research questions in their true complexity.

Although each day seems to bring more interest and activity in social epidemiology, it is important to appreciate that the questions we consider are anything but new. Not only did the ancient Greeks wonder about the relationship between social conditions and health but John Snow's famous cholera investigations, which many say mark the dawn of epidemiology and germ theory more generally, were infused with the same paradigm. Furthermore, what is too often overlooked is that questions concerning the relationship between social institutions (for example, government or societal norms) and human welfare date back to at least Hobbes and many great, more contemporary, political thinkers such as Keynes, Hayek, Freedman, and Sen, who continue to contribute to insights into the fundamental normative question: *How must we organize . . . to improve health?*

What Is Social Epidemiologic Methodology?

Methods are rules or procedures employed by those trying to accomplish a task. Sometimes such rules or procedures are written down. For example, cookbooks provide methods for baking better cookies or cakes. In much the same way, research methods are rules and procedures that researchers working within a disciplinary framework employ to improve the validity of their inferences. At risk of taking the analogy too far, researchers who abide by good research methods may more reliably produce valid inferences in much the same way that bakers who abide by excellent recipes tend to produce tasty cookies and cakes. There are always exceptions, but the point seems to hold generally.

Social epidemiologic methodology is naturally the study of methods in and for social epidemiology. To reiterate a point raised in the Preface, social epidemiologic methodology includes not only the broad collection of study design, measurement, and analytic considerations that has evolved over the previous century in mainstream epidemiology but also methods needed to address social epidemiology's special or unique questions and data. This latter group of methods arises more clearly from the social sciences, although a long tradition of considering these points in relation to communicable disease is also discernable in the history of epidemiology (Eyler 1979; Hamlin 1998; Ross 1916).

Methodological research is largely concerned with studying the logic of and improving techniques for scientific inference. The broad objective is to learn what

conclusions can and cannot be drawn given specified combinations of assumptions and data (Manski 1993). Because methodologists strive to determine what conclusions may be legitimately drawn given a set of assumptions, it is natural that this group of researchers often views existing practice more skeptically. Many methodologists might readily propose that a fundamental problem in applied research is that substantive investigators frequently fail to face up to the difficulty of their enterprise. We would venture to guess that many of the contributors to this volume would themselves articulate a similar position, that much published research is naïve with respect to assumptions being relied upon and to the many alternate explanations being ignored. The solution to this problem is rarely the use of more elaborate statistical methodology, however, as such solutions tend to be more assumption-laden rather than less so. Rather, the solution is for methodological training that stresses the fundamental logical principles behind study design and quantitative analysis of data and for greater rigor in the criticism of such models. Disciplines that become overly fascinated with the technique of analysis can easily become distracted from more elemental issues in the logic of inference, a nagging concern in economics, sociology, and other social sciences (Leamer 1983; Lieberson and Lynn 2002).

Three Fundamental Issues

In this section we briefly comment on three issues fundamental to social epidemiologic methodology: causal inference, measurement, and multilevel methodology.

Causal Inference

Perhaps the most fundamental and yet intractable problem of all research, especially observational research, is that of causal inference. The centrality of this concern rests with the need to have science be successfully predictive of the future and thus serve as a guide for how human activity may manipulate conditions for preferred outcomes. Because social epidemiology seeks to identify the effects of social variables, we must necessarily adopt a model of human agency that posits various actions taken or not taken and their consequences (Pearl 2000). Because a causal effect is *defined* on the basis of contrasts between various of these (potentially counterfactual) actions, many authors argue that we must immediately exclude non-manipulable factors, such as individual race or ethnicity and gender, from consideration as causes in this sense (Kaufman and Cooper 1999). The modifiable exposures that are typically of interest to social epidemiologists include

factors such as income, education, and occupation, which are potentially influenced through social policies or by various specific educational or social interventions. For example, the existence of a governmental income supplementation program changes income distributions in the population, allowing some families to live above the poverty line that would have lived beneath it in the absence of this policy (Basilevsky and Hum 1984; Orr et al. 1971). The contrast of these two policy regimes or between many specific variations of this intervention is the basis for the definition of a causal effect of interest in etiologic observational research.

For simplicity of exposition, consider a binary outcome ($Y = 1$ if disease occurs during the period of observation, $Y = 0$ otherwise), although extension to other outcome distributions is straightforward. For example, suppose that $Y = 1$ represents a subject in the defined population dying before the end of follow-up, whereas $Y = 0$ indicates that the subject is alive at end of follow-up. Consider social exposure $X = 1$ as the policy that provides income supplementation up to the poverty line and $X = 0$ as the absence of such a policy. As a notational convention to represent intervention, many sources in the statistical and epidemiologic literature make use of a subscript on the outcome variable ($Y_{X=x}$) to indicate the variable conditioned on forcing the target population to exposure level x (for example, Holland 1986). Pearl has employed several notational conventions (Pearl 2000, p. 70), including the “SET” notation, which expresses intervention as $\text{SET}[X = x]$. Using this notation, the outcome distribution under the various interventions is readily expressed as $\Pr(Y = y | \text{SET}[X = x])$, which may be translated as the probability of an outcome Y being the value y given the value of intervention X is set at x . These distributions of Y enable computation of outcome contrasts between all possible values of x taken by X . For example, for the causal effect of income supplementation on mortality, common contrasts would include the difference or ratio between the risk of death in the target population during the specified time period if the income supplementation policy were in effect versus if it were not in effect.

Although the hypothesis of a causal relation between income supplementation and mortality seems plausible, it is also entirely possible that states or counties with such programs have lower age-specific mortality risks than states or counties without such programs for extraneous reasons. If this were true, it would suggest that some part of the empirical association observed between income supplementation and mortality may arise not from the causal link between them but rather owing to their mutual response to other conditions, such as the level of the state cigarette tax, which affects both revenues available for income supplementation and the death rate through its effects on smoking behavior.

The task is to contrast the proportion of the target population who would die if subjected to a policy of income supplementation to the proportion who would die

if there was no policy in place for income supplementation: $\Pr(Y = 1 | \text{SET}[X = 1])$ versus $\Pr(Y = 1 | \text{SET}[X = 0])$. The problem in observational data is that nothing is actually SET, and so we must manipulate the observed quantities in some way to more validly estimate the causal effect. Clearly the crude contrast of observed mortality proportions, $\Pr(Y = 1 | X = 1)$ versus $\Pr(Y = 1 | X = 0)$, is not adequate, as these conditional probabilities may differ not only because of the causal effect of X but also because of the correlated perturbation in X and Y by their common cause.

The usual epidemiologic solution is to condition in some way on measured covariates that represent the common causes of X and Y . The logic behind this strategy is that within the categorizations of the covariates, there can be no confounding by these quantities (Greenland and Morgenstern 2001). Formally, this adjustment provides a statistically unbiased estimate of the true causal effect for X on Y when, within each stratum of covariate Z , observed exposure X is statistically independent of the potential response ($Y | \text{SET}[X = x]$) for each imposed value x (Rosenbaum and Rubin 1983). To the extent that one can enumerate and accurately measure all of the important common ancestors of exposure and outcome, this conventional epidemiologic solution is entirely adequate for the specification of the desired causal effect from observational data in point-exposure studies with no interference between units. For exposures related to human behavior, however, the task of identifying and measuring these common antecedents is often daunting.

Even in randomized experiments, but especially in observational studies, causal inference requires a strong theoretical foundation to justify assumptions of causal order, of no bias due to omitted covariates, and of effect homogeneity. This level of theoretical justification is often lacking in epidemiology, and is especially uncommon in social epidemiology (Oakes 2004). Regression modeling is particularly insidious in this regard, as the method has become so routine as to seem facile, when, in fact, the statistical and the extra-statistical assumptions required are often heroic (see Berk 2004; McKim and Turner 1997). Some authors are assiduously cautious with their language, yet many others imply causal relationships when they employ euphemisms such as “effect,” “impact,” “influence,” “dependent variable,” or “outcome” (Oakes 2004). The motivations are laudable, but in the end such “findings” may do more harm than good. Surely there are opportunity costs and risks to the public’s trust and understanding (Caplan 1988; Greenlund et al. 2003; Hogbin and Hess 1999).

Basic descriptive and predictive models devoid of causal import can be quite useful (Berk 2004). But at some point policymakers will want to use the results of social epidemiologic investigation to improve health, and causal understanding is desirable in this case. While prediction and causality are related, they are almost

always distinct because the latter is tied to action rather than observation. Too see this, recall that a rooster's crow does not raise the sun, but it predicts it with regularity. Such an alarm clock may be quite helpful to the sleepy farmer. But this model is merely predictive, because no matter how many times the sleepy farmer might get his rooster to crow later, the sun will rise in accordance with a completely different causal mechanism.

The subfield of social epidemiology is now suitably mature and sophisticated that we must state our analytic goals more clearly: does an author seek a causal, predictive, or perhaps "merely" descriptive model? Unlike fields such as climatology, social epidemiologists are often interested in actually enacting policies or interventions in order to improve the public's health. We therefore need to privilege causal explanations and to *aim* to build causal models. The yardstick is not perfection but usefulness, but it does not seem that multiple-regression procedures are getting us very far in this regard (Berk 2004).

Measurement of Social Phenomena

It was the poet Yeats (1938) who grasped the essential idea with the words "measurement began our might." Yet, although there can be no doubt that measurement of biological phenomena is quite advanced and that the field of psychometrics has aided progress on individual-level measures, such as IQ and depression (Nunally and Bernstein 1994), measures of social phenomena and other aggregate constructs remain remarkably primitive (Duncan 1984; Lazarsfeld and Menzel 1961). For example, several authors have revealed a striking lack of attention to the measurement of the central construct of socioeconomic status (SES) in health research (Oakes and Rossi 2003). The situation appears even worse when it comes to measures of ecological settings such as neighborhoods, schools, and workplaces. The fact is that the methodology needed to evaluate these measure remains in its infancy (Sampson 2003).

It is unclear why so little progress has been made on the measurement of constructs fundamental to social epidemiologic inquiry, especially in light of a consensus agreement on the basic consequences of measurement error: it has been known for over 100 years that measurement error generally biases (attenuates or accentuates) effects (Gustafson 2004; Jurek et al. 2005; Nunally and Bernstein 1994; Yatchew and Griliches 1984). Surely one reason for the slow pace of progress is that the task is difficult. Unlike counting red blood cells or calculating a subject's body mass index, relevant constructs in social epidemiology are always between persons and are often group-level phenomena. This means that such measures reflect complex functions of individual action, interactions, and largely unknown feedback systems: this greatly complicates things. Other reasons for the

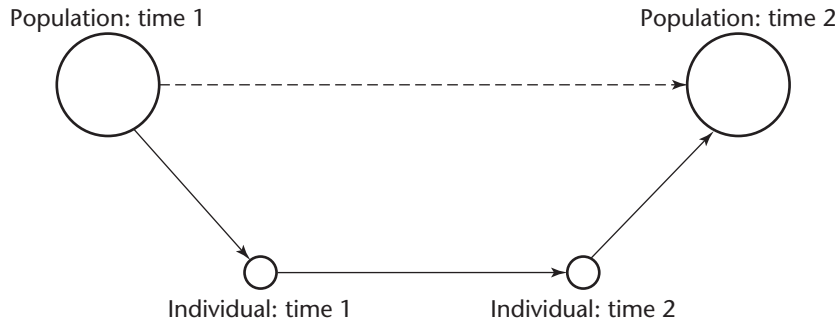
slow progress probably include the fact that there is little practical incentive for work on social measurement within epidemiology. For better or worse, it is clear that conventional epidemiology has not devoted much attention to the social sciences (Oakes 2005), which means that whereas health outcome assessments may be thoroughly scrutinized by reviewers and editors, social exposure assessment may be accomplished crudely or reflexively without drawing much negative attention (Jones and Cameron 1984). As Berk (2004, p. 238) laments, “Many investigators appear to proceed as if fancy statistical procedures can compensate for failures to invest in proper data collection.” Progress in conceptualization and measurement is key to advancement of social epidemiology, and more attention should be devoted to it.

Multilevel Methods

Much has been published in recent years within social epidemiology about multilevel theory and multilevel models (Diez-Roux 2000; Kaplan 2004). This is clearly a salutary development for the field because the point of such discussions—that context matters—is timely and important. Yet, whereas several scholars have ably considered some of the statistical issues of the multilevel regression model (see Chapter Thirteen), few have fully discussed the fundamental methodological issues inherent in a true multilevel methodology, namely, an approach that incorporates the critical and dynamic tension between individuals and groups. At some point several slippery questions must be considered, including whether a group is an entity independent of its constituents. Asked differently, is there a group without the specific individuals who comprise it? Another question is how groups or aggregate phenomena change over time; what are the mechanisms? These issues rest at the core of multilevel theory and models, and more attention needs to be devoted to them. To be sure, such issues are difficult, and we can offer no facile recipe or simple conclusion. Obviously, a full treatment is far beyond our scope here.

To better understand multilevel theory, we turn to the work of Coleman, who in 1990 tried to present the key issues by discussing Weber’s 1905 classic explanation of the rise of capitalism in the Protestant West (Coleman 1990; Weber [1905] 1958). According to Coleman, Weber was trying to explain how society evolved from pre-capitalistic to capitalistic by describing changes that occurred among individuals within the societies under investigation. Weber’s research question was how and why some societies changed so dramatically over a relatively brief period of time. For purposes here, the important point is that Weber’s explanation of social change rested on the changes within and among the individuals who made up the societies. According to Weber, it was the adoption and

FIGURE 1.1. CONCEPTUAL FRAMEWORK FOR MULTILEVEL THINKING



(Adapted from Coleman 1990, fig. 1.3)

internalization of the Calvinist religious ethic by individuals that eventually led to the growth and dominance of capitalism at the societal level.

Coleman tried to better formalize the issues by drawing a trapezoidal figure (which we affectionately call the “Coleman bathtub”). We adopt this pedagogical device and present a similar figure (Figure 1.1). Although simple on its face, this figure contains a great deal of useful information for advancing multilevel methods in social epidemiology.

The larger circle to the left represents the population or society at time one or *before* any change. The larger circle to the right represents the population at time two or *after* some change. Alone, these two larger circles represent a change in a population/society over time. That is, the two larger circles and the dotted-line arrow linking them represent our central question: how did society change? In concrete terms, one might observe a change in the rate of cigarette smoking over time. A social epidemiologist observing this might ask how and why this change occurred. Her goal might be to try to explain this change so that better interventions to reduce the smoking rate could be developed and tested. How then might we understand the social or epidemiologic change?

A methodological individualist, Coleman insists that such social change comes about only through changes in individual people and their interactions. Societal or group-level change does not just happen mysteriously without the involvement of actual persons; social change must be grounded in the activity of constituent individuals. It follows that the change in smoking rates can only be explained by understanding what happened to the smokers and non-smokers, and their relationships, under investigation.

Individual change is diagrammed in Figure 1.1. The smaller circle to the left represents a given person living in the society at time one. The smaller circle to the right represents the same person at time two, after some change. The arrow linking this person at time one to himself or herself at time two represents personal growth or change, a psychological (or perhaps medical) phenomena. Note well that, however interesting, this change is not our focus here; indeed, for social epidemiology, personal change is only important to the extent it reflects or implies change at the societal or population level.

Most important here are the (near) vertical arrows to the left and to the right. The downward pointing arrow to the left, from the larger circle to the smaller, represents the impact or influence of society on an individual. This is the *macro-to-micro* transition. The arrow to the right, from the smaller to the larger circle, represents the impact of the individual on society. This is the *micro-to-macro* transition. Together, these “micro-macro” transitions represent the most important but most difficult methodological challenge for a multilevel social epidemiology. The fundamental questions are how and why society “gets into” individuals and how and why do individuals interact to produce complex social organizations and related outcomes.

Macro-to-micro transitions may come as resource constraints, social norms, laws, and all other such forces that affect individual behavior. Especially important are the concepts of socialization and endogenous preferences. Although difficult to study, the former idea appears easily understood: socialization is the process of learning and internalizing the rules of proper behavior and the consequences of behaving improperly. Parents and teachers socialize offspring and students. What then of the related notion of endogenous preferences? The term *endogenous preferences* implies that what we like and dislike is at least partly learned from others and the constraints faced (Bowles 1998). Simply put, our circumstances affect our preferences if not our entire world view. These notions of socialization and endogenous preferences may be interpreted as implying that our own likes and dislikes are all social constructs, which is a slippery and controversial conclusion because it seems to cast suspicion on the very existence of free will and individual volition (Sunstein 1986).

Moving in the other direction, micro-to-macro transitions may come as efforts of individuals to change laws, lower-prices, or promote collective actions, such as anti-smoking demonstrations. To keep this discussion accessible and brief, we shall greatly oversimplify and assert that all micro-to-macro transitions may be viewed as collective actions where individuals somehow act together for seemingly common goals. Collective action problems are ubiquitous in society and well studied in the social sciences. The key point is that there are fundamental interdependencies and interactions among persons engaged in a social goal, which means

that simple aggregations of presumed individual behavior fails to explain or predict outcomes (Olson 1971). Consider two notable examples of collective action problems: voting, and protection of a field for grazing sheep. First, the issue of voting in an election is at once simple and complex. Simply understood, persons vote to express their preference for one candidate or object to another. But a paradox arises, because as the probability that anyone's vote will be decisive approaches zero, an individual has no incentive to waste even a moment in order to vote. So why do so many people do it? More generally, why does any voluntary group effort occur when individuals typically have no incentive to participate? The second example of collective-action phenomena may be found in the so-called "commons problem." In short, the classic commons problem occurs when individual sheep farmers have incentive to graze more sheep (Hardin 1968). The trouble is that when all shepherds do so the common land is overgrazed, the sheep starve, and each farmer loses his fortune. This is a collective action problem that illustrates how individuals seeking their own self-interest can yield collective outcomes that no individual would want; in other words, private rationality can lead to collective irrationality.

Both micro-macro transitions highlight the crucial role of interdependencies in social phenomena that affect social (that is, population-level) change. For the most part, social epidemiology has not addressed these fundamental issues in theory, measurement, or analysis, leaving much work still to be done. Coleman and others have suggested that the best way through this thicket is to conceptualize the micro-macro transitions not with respect to particular persons or even any persons but rather as a system of socio-structural positions that tend to emerge from the characteristics of the micro-macro transitions. Accordingly, the transitions can be conceived of as the "rules of the game" that transmit the consequences of an individual's action to other individuals and yield macro-level phenomena (Coleman 1990). New and insightful work in multilevel theory includes Durlauf's 2002 article on social capital, Durlauf and Young's 2001 edited volume on dynamic social interactions, and Bowles' 2004 novel microeconomics text.

Advancing Further Still

Although it seems appropriate to briefly comment on some potential steps beyond this volume that would appear to enhance the practice and import of social epidemiology, we do so with some trepidation. It is simply difficult to know how our subfield will evolve or co-evolve with more mainstream epidemiology. Nevertheless, some speculation on three approaches may be useful for discussion, debate, and further study.

First, success might be enhanced if social epidemiologists considered and conducted more randomized experimental studies. While Hannan and Glymour discuss many aspects of community trials and natural experiments in Chapters Fourteen and Seventeen, respectively, it is worth pointing out that there have been other applications of experimental methods that seem potentially useful to social epidemiology. The first type includes efforts to manipulate constructs important to social epidemiology through laboratory-like factorial experiments. For example, McKinlay et al. (2002) used videotape vignettes in an experiment aimed to determine: (1) whether patient attributes (specifically a patient's age, gender, race, and socioeconomic status) independently influence clinical decision-making; and (2) whether physician characteristics alone (such as gender, age, race, and medical specialty) or in combination with patient attributes influence medical decision-making (see also Feldman et al. 1997). If nothing else, such efforts are useful because they clearly require sharply formed *a priori* hypotheses and offer some control over confounding.

Somewhat relatedly, there may be benefit in resurrecting the seemingly overlooked method of factorial surveys, which aim to experimentally examine judgments and preferences by combing factorial experiments with survey methods (Rossi and Nock 1982). Classic examples include the work of Nock and Rossi (1979), who used the method to understand the *independent* effects of factors considered when judging a household's socioeconomic status. More recently, Schwappach and Koeck (2004) employed the method to better understand judgments about medical errors. Furthermore, though rarely used in this fashion, the method would seem to hold some promise for understanding variation in social norms (Rossi and Berk 1987). Finally, there is the growing and recent work of evolutionary economists and their like-minded kin who use simple experiments to better understand social interactions and outcomes (Henrich et al. 2005; Sunstein 2000). Paying greater attention to such work and extending it would seem to hold great promise for social epidemiology.

Second, it seems prudent to devote greater attention to cross-validation—a procedure where predicted values from, for example, a regression model are compared with actual observations. Cross-validation is one of the true tests of a (statistical) model because until tested, parameter estimates are shielded from scrutiny and perhaps public view because true values are not known—sampling variability offers enormous protection (Kennedy 1988). Box (1994) draws an analogy to a criminal investigation: no matter how good it might be, detective work (that is, model building) without prosecution and adjudication (validation) is worthless if not irresponsible.

Ironically, the medical and public health literatures, especially as related to obesity, are replete with cross-validation studies focused on validating instruments or biological relationships (for example, Beekley et al. 2004; Craig et al. 2003;

Finan et al. 1997; Goran and Khaled 1995; Thomsen et al. 2002; Vander Weg et al. 2004). Indeed, diagnostic medicine has not tolerated non- or poorly validated instruments since the publication of Ransohoff and Feinstein's landmark paper (see also Zhou et al. 2002). Yet as far as we know, no social epidemiologic models have been formally validated or tested in this way. The reason, it would seem, is that researchers rarely have access to a second independent sample from their target population. Presuming awareness of validation methods, the fact is that second samples are expensive. It is possible to validate a model with the same data used to estimate it (Hastie et al. 2001). But building and validating a model with the same data, even subsets of it, can be very misleading; the model is likely to appear better than it really is. This is because it is too easy to capitalize on chance or a particular realization of the stochastic process (Browne 2000; Zucchini 2000). Once again, Berk (2004, p. 130) captures the point:

model selection can lead to the problem of "overfitting." If a goal of data analysis is to make inferences from a sample to a population or to natural processes that generated the data (or to forecast), testing lots of different regression models can lead to a final model that reflects far too many idiosyncrasies in the sample . . . the final fit is then an overfit. . . .

Although a general problem for all observational science, overfitting may be particularly rampant in social epidemiology, and such mistakes may serve to impede scientific progress and improvements to the public health.

Finally, we note that the most widely applied method for evaluating the impact of social exposures on health is one that is not covered to any extent in this volume or used in most social epidemiologic research, despite its importance and arguable advantages in relation to other methods. This is the qualitative or narrative historical approach. In broad outline, the basic idea is to tell the story of the exposures and outcomes in the specific socio-historical context in which they actually occurred rather than in an abstract and idealized context defined by statistical models. The strength of this approach is clearly that it does not presume to state some set of universal rules that exist for all vaguely similar situations at all times but rather is the explanatory narrative of one unique configuration of events. The weakness of this approach is exactly the same: if we only know how exposure and outcome were related in one particular instance in the past, of what practical use is this information to us for the future? Furthermore, if no generalization to other settings is formally justified, then the explanatory mechanism proposed by the author is not prospectively testable and therefore not refutable, because those exact circumstances will never be replicated. Instead, critique can only come in the form of counterarguments and alternative explanations, and therefore the

evaluation of competing explanations remains necessarily subjective. This is the fundamental tension between the idiographic and nomothetic scientific paradigms.

Narrative historical depictions can certainly be highly quantitative, in the sense that they involve numerical summaries of the events that occurred. These depictions may also be characterized by specific causal explanations in the form of counterfactuals (that is, arguing that events are the results of specific precipitating conditions that, had these conditions not pertained earlier, would have come out differently). For example, from 1991 through 1994, there was an epidemic of neuropathy in Cuba in which more than 50,000 people experienced vision loss. The causal explanation appears to be an acute nutritional deficiency subsequent to the collapse of the Soviet Union (which had subsidized the Cuban economy) and concomitant tightening of the U.S. economic embargo (Ordunez-Garcia et al. 1996). This explanation is causal because it implies that, had the Soviet subsidies continued, the epidemic would have been reduced or avoided entirely. But it differs in numerous ways from the inferences gleaned from statistical models. For example, although the factual conditions may be represented with great precision, the outcome distribution under the counterfactual condition is not generally identified quantitatively in the narrative approach. Indeed, an important strength of this analytic approach is that it successfully avoids the seductive generality of statistical models, the results of which are described in universal terms, without reference to the specific circumstances in which the data-generating mechanism operated. And by representing the counterfactual outcome distribution qualitatively as opposed to quantitatively, this also avoids the illusion of numerical precision for contrasts that fall outside the realm of the observed data (King and Zeng 2003).

Important social epidemiologic works that adopt this analytic strategy include Randall Packard's *White Plague, Black Labor* (Packard 1989) and Eric Klinenberg's *Heat Wave: A Social Autopsy of Disaster in Chicago* (2002). Unfortunately, however, this approach lends itself more naturally to book-length treatment or, at very least, to the longer article lengths typical of the humanities and social sciences. The restrictive length and structuring requirements of many biomedical journals make it almost impossible to engage in these kinds of arguments in our mainstream epidemiology journals. One notable exception is the "Public Health Then and Now" column in *The American Journal of Public Health*. By contrast, several social sciences recognize that the narrative historical approach is an essential tool for investigating and characterizing the complex relations between social arrangements and their consequences (King et al. 1994).

The bulk of the current volume is organized around the paradigm of the experimental trial as the standard for scientific inference. But for social epidemiology to thrive in the decades to come, we must also become comfortable with the

realization that some scientific questions will not be answered best by treating observational data as though they arose from controlled experiments. For some highly complex systems, such as human social structures, the costs of generality in terms of oversimplification and unjustified assumptions may easily be too great to warrant the fantasies of regression equations and exogenous errors and the like. If the statistical models must become so baroque that they obscure rather than facilitate understanding and insight, then it is time to consider alternate approaches that more readily acknowledge subtlety, uniqueness, and peculiarity.

References

- Basilevsky, A., & Hum, D. (1984). *Experimental social programs and analytic methods*. New York: Academic Press.
- Beekley M. D., Brechue W. F., deHoyos D. V., Garzarella L., Werber-Zion G., Pollock M. L. (2004). Cross-validation of the YMCA Submaximal Cycle Ergometer Test to Predict VO₂max. *Research Quarterly for Exercise and Sport*, 75, 337–342.
- Berk, Richard. (2004). *Regression analysis. A constructive critique*. Thousand Oaks, CA: Sage Publications.
- Bowles, Samuel. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36, 75–111.
- Bowles, Samuel. (2004). *Microeconomics: Behavior, institutions, and evolution*. Princeton, NJ: Princeton University Press.
- Box, George. (1994). Statistics and quality improvement. *Journal of the Royal Statistical Society, Series A* 157, 209–229.
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108–132.
- Caplan, A. L. (1988). Professional arrogance and public misunderstanding. *Hastings Center Report*, 18, 34–37.
- Coleman, James S. (1990). *The foundations of social theory*. Cambridge: Belknap.
- Craig, C. L., Marshall, A. L., Sjoström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, 35, 1381–1395.
- Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, 21, 171–192.
- Duncan, O. D. (1984). *Notes on social measurement: Historical & critical*. New York: Russell Sage.
- Durlauf, S. N. (2002). On the empirics of social capital. *Economics Journal*, 112, 459–479.
- Durlauf, S. N., & Young, H. P. (2001). *Social dynamics*. Washington, DC: Brookings Institution Press.
- Eyler, J. M. (1979). *Victorian social medicine: The ideas and methods of William Farr*. Baltimore: Johns Hopkins University Press.
- Feldman, H. A., McKinlay, J. B., Potter, D. A., Freund, K. M., Burns, R. B., Moskowitz, M. A., et al. (1997). Nonmedical influences on medical decision making: An experimental technique using videotapes, factorial design, and survey sampling. *Health Serv Res*, 32, 343–366.
- Finan, K., Larson, D. E., & Goran, M. I. (1997). Cross-validation of prediction equations for resting energy expenditure in young, healthy children. *Journal of the American Dietetic Association*, 97, 140–145.

- Goran, M. I., & Khaled, M. A. (1995). Cross-validation of fat-free mass estimated from body density against bioelectrical resistance: Effects of obesity and gender. *Obesity Research*, 3, 531–539.
- Greenland, S., & Morgenstern, H. (2001). Confounding in health research. *Annual Review of Public Health*, 22, 189–212.
- Greenlund, K. J., Neff, L. J., Zheng, Z. J., Keenan, N. L., Giles, W. H., Ayala, C. A., et al. (2003). Low public recognition of major stroke symptoms. *American Journal of Preventive Medicine*, 25, 315–319.
- Gustafson, P. (2004). *Measurement error and misclassification in statistics and epidemiology*. Boca Raton, FL: Chapman & Hall/CRC.
- Hamlin, C. (1998). *Public health and social justice in the age of Chadwick: Britain, 1800–1854*. New York: Cambridge.
- Hardin, G. (1968). Tragedy of the commons. *Science*, 162, 1243–1248.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. New York: Springer-Verlag New York, Inc.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2005). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small societies*. New York: Oxford University Press.
- Hogbin, M. B., & Hess, M. A. (1999). Public confusion over food portions and servings. *Journal of the American Dietetic Association*, 99, 1209–1211.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Jones, I. G., & Cameron, D. (1984). Social class analysis—An embarrassment to epidemiology. *Community Medicine*, 6, 37–46.
- Jurek, A. M., Greenland, S., Maldonado, G., & Church, T. R. (2005). Proper interpretation of non-differential misclassification effects: Expectations vs. observations. *American Journal of Epidemiology*, 34, 680.
- Kaplan, G. A. (2004). What's wrong with social epidemiology, and how can we make it better? *Epidemiologic Reviews*, 26, 124–135.
- Kaufman, J., & Cooper, R. (1999). Seeking causal explanations in social epidemiology. *American Journal of Epidemiology*, 150, 113–119.
- Kennedy, P. (1988). *A guide to econometrics* (4th ed.). Cambridge, MA: MIT Press.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- King, G., & Zeng, L. (2003). *When can history be our guide? The pitfalls of counterfactual inference*. (vol.) Retrieved December 12, 2005, from <http://gking.harvard.edu/preprints.shtml>: Harvard University.
- Klinenberg, E. (2002). *A social autopsy of disaster in Chicago*. Chicago: University of Chicago Press.
- Lazarsfeld, P. F., & Menzel, H. (1961). On the relation between individual and collective properties. In A. Etzioni (Ed.), *Complex organizations: A sociological reader*. New York: Holt, Rinehart, and Winston.
- Leamer, E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73, 32–43.
- Lieberman, S., & Lynn, F. B. (2002). Barking up the wrong branch: Scientific alternatives to the current model of sociological science. *Annual Review of Sociology*, 28, 1–19.

- Manski, C. F. (1993). Identification problems in the social sciences. In P. V. Marsden (Ed.), *Sociological Methodology 1993* (Vol. 23, pp. 1–56). Washington, DC: Blackwell Publishers, for the American Sociological Association.
- McKim, V. R., & Turner, S. P. (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. Notre Dame, IN: Notre Dame University Press.
- McKinlay, J. B., Lin, T., Freund, K., & Moskowitz, M. (2002). The unexpected influence of physician attributes on clinical decisions: Results of an experiment. *Journal of Health and Social Behavior*, *43*, 92–106.
- Nock, S. L., & Rossi, P. H. (1979). Household types and social standing. *Social Forces*, *57*, 1325–1345.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw Hill.
- Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science & Medicine*, *58*, 1929–1952.
- Oakes, J. M. (2005). An analysis of AJE citations with special reference to statistics and social science. *American Journal of Epidemiology*, *161*, 494–500.
- Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science & Medicine*, *56*, 769–784.
- Olson, M. (1971). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Ordunez-Garcia, P. O., Nieto, F. J., Espinosa-Brito, A. D., & Caballero, B. (1996). Cuban epidemic neuropathy, 1991 to 1994: History repeats itself a century after the “amblyopia of the blockade.” *American Journal of Public Health*, *86*, 738–743.
- Orr, L. L., Hollister, R. G., & Lefcowitz, M. J. (1971). *Income maintenance: Interdisciplinary approaches to research*. Madison, WI: Institute for Research on Poverty.
- Packard, R. M. (1989). *White plague, black labor: The political economy of health and diseases in South Africa*. Berkeley: University of California Press.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
- Ransohoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine*, *299*, 926–930.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*, 41–55.
- Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry. Part I. *Proceedings of the Royal Society of Medicine, Series A*, *92*, 204–240.
- Rossi, P. H., & Berk, R. A. (1987). Varieties of normative consensus. *American Sociological Review*, *50*, 333–347.
- Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgements: The factorial survey approach*. Beverly Hills, CA: Sage.
- Sampson, R. J. (2003). Neighborhood-level context and health: Lessons from sociology. In I. Kawachi & L. F. Berkman (Eds.), *Neighborhoods and health* (pp. 132–146). New York: Oxford.
- Schwappach, D. L., & Koeck, C. M. (2004). What makes an error unacceptable? A factorial survey on the disclosure of medical errors. *International Journal for Quality in Health Care*, *16*, 317–326.
- Sunstein, C. R. (1986). Legal interference with private preferences. *The University of Chicago Law Review*, *53*, 1129–1174.
- Sunstein, C. R. (2000). *Behavioral law & economics*. New York: Cambridge University Press.

- Thomsen, T. F., McGee, D., Davidsen, M., & Jorgensen, T. (2002). A cross-validation of risk-scores for coronary heart disease mortality based on data from the Glostrup Population Studies and Framingham Heart Study. *International Journal of Epidemiology*, *31*, 817–822.
- Vander Weg, M. W., Watson, J. M., Klesges, R. C., Eck Clemens, L. H., Slawson, D. L., & McClanahan, B. S. (2004). Development and cross-validation of a prediction equation for estimating resting energy expenditure in healthy African-American and European-American women. *European Journal of Clinical Nutrition*, *58*, 474–480.
- Weber, M. (1958). *The Protestant ethic and the spirit of capitalism*. New York: Scribners. (Original work published 1905).
- Yatchew, A., & Griliches, Z. (1984). Specification error in probit models. *Review of Economics and Statistics*, *67*, 134–139.
- Yeats, W. B. (1938). Under Ben Bulben. In *Last Poems and Two Plays*. Shannon: Irish University Press.
- Zhou, X-H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. New York: Wiley.
- Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*, 41–61.