



CHAPTER ONE

INTRODUCTION: ADVANCING METHODS IN SOCIAL EPIDEMIOLOGY

Jay S. Kaufman and J. Michael Oakes

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 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 some selected 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 determinations of states of health in populations. We define *social* epidemiology as the branch of epidemiology that considers how social interactions and purposive human activity 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 between the persons who comprise the population. Defining social epidemiology in this broad way

permits not only the analysis of how social factors serve as exposures that affect health outcomes, but also how such factors emerge and are maintained in a distinctive distribution.

Social epidemiology is thus not only concerned with the identification of new disease specific risk factors (e.g., a deficit of social capital). Social epidemiology also considers how well-established exposures (e.g., cigarette smoking, lead paint, 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 the resources that can purchase salubrious environments and competent health care. In short, social epidemiology is about social allocation mechanisms, the 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, though the same is perhaps not true for his "friend" Friday. Such interactions are central to social epidemiology, however, and it is in this way that the subfield distinguishes itself from the bulk of conventional epidemiology. Without any attention to social arrangements and institutions, epidemiologic research on humans is almost indistinguishable from an application to livestock.

It is the incorporation of purposive human interaction and agency (i.e., 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. However, social epidemiology is not a social science, at least as traditionally conceived. While the methods and models of, say, a social epidemiologist and medical sociologist might be similar, 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 explanandum is society, social forces, or the like. A social scientific study that considers and models health outcomes does so to learn about society. By contrast, the outcome variable for social epidemiology is health. While social epidemiologists may borrow theory, methods,

and constructs from social science, they do so in an effort to understand health, not social forces or related phenomena for their own sake. This means that while social epidemiology is related to the social sciences, it 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. It is simply a matter of explanatory emphasis. 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.

While 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. Further, it is too often overlooked that questions concerning the relationship between social institutions (e.g., government or societal norms) and human welfare date back to at least Hobbes, and many of the great thinkers that are more contemporary, such as Keynes, Hayek, Friedman, Sen, and Piketty, 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 cakes. In much the same way, research methods are rules and procedures that scientists 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 bakers who abide by excellent recipes tend to produce tasty snacks. There are always exceptions, but the point seems to hold.

Social epidemiological methodology is naturally the study of methods in and for social epidemiology. To reiterate a point raised in the Preface, social epidemiological 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 (Ross 1916; Eyer 1979; Hamlin 1998).

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 logically derived 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 (though we appreciate that substantive researchers may question the utility of esoteric methodological insights). 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 require additional assumptions. Rather, the solution is for methodological training that stresses the fundamental logical principles behind study design and quantitative analysis of data, and for greater enthusiasm for 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; Lieberman 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 things for preferred outcomes (Galea 2013). 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 2009). 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/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 modified through social policies and various educational or social interventions (Harper and Strumpf 2012). 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 who would have lived beneath the poverty line in the absence of this policy (Orr, Hollister, and Lefcowitz 1971; Basilevsky and Hum 1984). 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 the 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 (e.g., Holland 1986). Pearl has employed several notional conventions 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 that 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.

While the hypothesis of a causal relation between income supplementation and mortality seems plausible, it is also entirely possible that states with such programs have lower age-specific mortality risks than states 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 due 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 of interest. 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 traditional 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 (Kaufman and Harper 2013).

Even in randomized experiments, but especially in observational studies, causal inference requires a strong theoretical foundation in order to justify assumptions of causal order, of no bias due to omitted covariates, and of effect homogeneity (Naimi and Kaufman 2015). 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 (McKim and Turner 1997; Berk 2004). 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 (Kaufman and Harper 2013). Surely there are opportunity costs and risks to the public’s trust and understanding (Caplan 1988; Hogbin and Hess 1999; Greenlund *et al.* 2003).

Basic descriptive and predictive models devoid of causal import can be quite useful and, along with Berk (2004), we encourage their use. However, at some point social epidemiologists will want their results to inform social policies to improve health (Nandi and Harper 2015). Causal understanding is desirable in this case. While prediction and causality are related, they are almost always distinct. To 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 since 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. We believe 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 policymakers who will actually enact policies or interventions in order to improve the public’s health. We therefore need to promote causal explanations and to aim to build causal models. Although the yardstick is not perfection but usefulness, it does not seem that multiple-regression procedures are getting us very far (Oakes *et al.* 2015).

Measurement of Social Phenomena

It was the poet Yeats (1938) who grasped the essential idea with the words “measurement began our might.” Yet while there can be no doubt that measurement of biological phenomena is quite advanced and the field of psychometrics has aided progress on individual-level measures, such as IQ and depression (Nunnally and Bernstein 1994), measures of social phenomena and other aggregate constructs remain remarkably crude and lacking (Lazarsfeld and Menzel 1961; Duncan 1984). 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; O'Brien, Sampson, and Winship 2015).

It is unclear why so little progress has been made on the measurement of constructs fundamental to social epidemiologic inquiry, especially because we presume consensus agreement on the basic consequences of measurement error: it has been well known, since at least 1877, that measurement error complicates (attenuates or accentuates) effects (Yatchew and Griliches 1984; Nunnally and Bernstein 1994; Gustafson 2004; Jurek *et al.* 2005). Surely one reason for the slow pace of progress is that such work 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. Thus, the constituent "cells" in social epidemiology think, choose, grow, and emote—sometimes with brutish self-interest, sometimes irrationally, and sometimes by habit or whim. This greatly complicates things. Other reasons for the slow progress probably include the fact that work on social measurement within epidemiology rarely offers high reward. Under current scientific norms and practice, there seems to be little incentive to conduct such basic, if not mundane, research. For better or worse, it is clear that conventional epidemiology has not devoted much attention to the social sciences (Oakes 2005), which means social epidemiologists interested in measurement "stick their neck out" when conducting such research. On a more skeptical note, we also see merit in Berk's (2004, p. 238) lament: "Many investigators appear to proceed as if fancy statistical procedures can compensate for failures to invest in proper data collection." Progress in measurement is key to our advancement, and more attention should be devoted to it.

Multilevel Methods

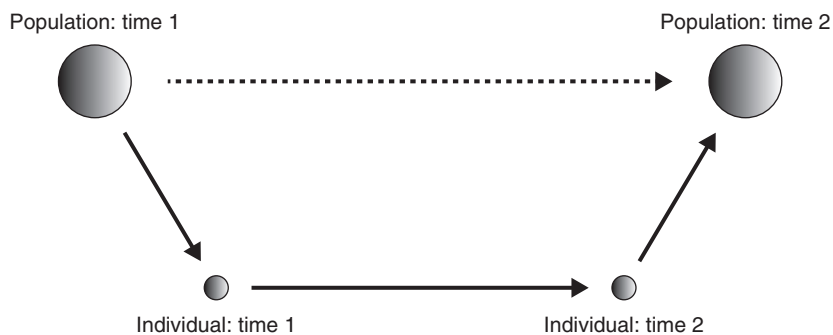
Much has been learnt over the last two decades about multilevel theory and multilevel models (Diez-Roux 2000; Kaplan 2004). This seems like a good thing because the point of such discussions—that context matters—is on target and important. Yet while various authors have ably considered some of the statistical issues of the multilevel regression model (see Chapter 15), 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 (Suzuki *et al.* 2012). Asked differently, is there a group without the specific individuals who constitute it? Another question concerns 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 easy recipe or simple conclusion. Obviously, a full treatment is far beyond our scope here.

So as to understand multilevel theory better 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. According to Coleman, Weber was trying to explain how society changed 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 between the individuals who made up the societies. According to Weber (2002), it was the adoption and 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 formalize the issues better by drawing a trapezoidal figure (which we affectionately call the "Coleman bathtub"). We adopt this pedagogical device and present a similar figure below (Figure 1.1). While simple on its face, this figure contains a great deal of information useful for advancing multilevel methods in social epidemiology.

FIGURE 1.1. CONCEPTUAL FRAMEWORK FOR MULTILEVEL THINKING



Adapted from Coleman 1990, Figure 1.3.

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. His or her goal might be to try and 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 him- 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 and/or implies change at the societal or population level.

What matters most 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 mythological 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. While difficult to study, the former idea appears to be 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? Simply put, the term endogenous preferences implies that what we like and dislike is at least partly learned from others and the constraints faced (Bowles 1998). This is to say, our circumstances affect our preferences if not our entire world view. It follows that ideas of socialization and endogenous preferences imply that our own likes and dislikes are social constructed, which is a slippery and controversial conclusion since it implies that free-will is also a suspect notion (Sunstein 1986, 2014).

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 antismoking demonstrations. In order to keep this discussion accessible and brief, we greatly oversimplify and say 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). To see this, 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 over another. However, a paradox arises because, since the probability that anyone's vote will be decisive approaches zero, an individual has no incentive to waste even a moment to vote. So why do so many 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 an 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 no individual would have wanted: 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 (i.e., population-level) change. For the most part, social epidemiology has not addressed these fundamental issues in theory, measurement, or analysis. There is much work to be done here.

Overall, we agree with Coleman and others who believe 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 sociostructural 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,” which transmit the consequences of an individual’s action to other individuals and yield macro-level phenomena (Coleman 1990). Cutting-edge work in multilevel theory includes that by Cetina and Cicourel (2014), Durlauf’s (2002) paper on social capital, Durlauf and Young’s (2001) edited volume on dynamic social interactions, and Bowles’ (2009) 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. We did not foresee many of the developments between the first and second editions of this textbook, and it is difficult to know how our subfield will further 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. We made this suggestion a decade ago in the first edition of this book, yet in the subsequent decade it was primarily economists, not social epidemiologists, who embraced this strategy (Duflo, Glennerster, and Kremer 2007). While Hannan and Glymour discuss many aspects of community trials and natural experiments in, respectively, Chapters 11 and 19 of this text, 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 their 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 a design strategy to avoid confusion. 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 that of Nock and Rossi (1979) who used the method to understand the *independent* effects of factors considered when judging a household's socioeconomic status. Likewise, Schwappach (Schwappach and Koeck 2004) employed the method to understand judgments about medical errors better. 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 continues to be work by evolutionary economists and their like-minded kin who use simple experiments to understand social interactions and outcomes better (Sunstein 2000; Henrich *et al.* 2005; Dechenaux, Kovenock, and Sheremeta 2014; Heller *et al.* 2015). 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, say, a regression model are compared to 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 (i.e., model building) without prosecution and adjudication (i.e., validation) is worthless, if not irresponsible.

This is an especially acute concern when investigators have unlimited access to the data in order to try an unlimited range of model specifications without accounting for the multiplicity of tests or the overfitting of chance fluctuations in the dataset. When there is no accounting for “fishing” and authors report models as though they were specified in advance of seeing that data, 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

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 (e.g., Goran and Khaled 1995; Finan,

Larson, and Goran 1997; Thomsen *et al.* 2002; Craig *et al.* 2003; Beekley *et al.* 2004; Vander Weg *et al.* 2004). Indeed, diagnostic medicine has not tolerated poorly validated instruments since the publication of Ransohoff and Feinstein's landmark paper (1978) (see also Zhou, Obuchowski, and McClish 2002). Yet analogous practices in social epidemiology are still rare. The historical reason was that researchers rarely had access to a second independent sample from their target population, largely due to cost constraints in data collection or the uniqueness of some data sources such as national health examination surveys.

Fortunately, it has long been possible to validate a model with the same data used to estimate it (Hastie, Tibshirani, and Friedman 2001). Advances in internal cross-validation methods since the publication of the first edition of this book have transformed the landscape of model selection in applied research, and provided software tools that can largely avoid "fishing" by leaving model specification to machine-learning algorithms that provide optimal fit through cross-validation using random partitions of the original data set (van der Laan *et al.* 2007). We worry that overfitting is still rampant in social epidemiology, and that such practices are impeding scientific progress and improvements to public health. However, given the new computer-intensive technologies available for model selection, there is no longer any excuse for this state of affairs. Social epidemiology applications are starting to appear (Mirelman *et al.* 2016), and by the third edition of this book we expect this to be the norm in our field.

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 utilized in most social epidemiological 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 here is to tell the story of the exposures and outcomes in the specific sociohistorical 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, since 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 ideographic 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 which, 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 US economic embargo (Orduñez-García *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. However, 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. 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* (Klinenberg 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. By contrast, several social sciences recognize that the narrative historical approach is as an essential tool for investigating and characterizing the complex relations between social arrangements and their consequences (King, Keohane, and Verba 1994).

The bulk of the current volume is organized around the paradigm of the experimental trial as the standard for scientific inference. However, 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, subtly, uniqueness and peculiarity.

References

- Basilevsky, A. and Hum, D. (1984) *Experimental Social Programs and Analytic Methods*, Academic Press, New York.
- Beeckley, M.D., Brechue, W.F., deHoyos, D.V., *et al.* (2004) Cross-validation of the YMCA sub-maximal cycle ergometer test to predict VO₂max. *Research Quarterly Exercise and Sport*, 75, 337–342.
- Berk, R. (2004) *Regression Analysis. A Constructive Critique*, Sage, Thousand Oaks, CA.
- Bowles, S. (1998) Endogenous preferences: the cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36, 75–111.
- Bowles, S. (2009) *Microeconomics: Behavior, Institutions, and Evolution*, Princeton University Press, Princeton, NJ.
- Box, G. (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.
- Cetina, K.K. and Cicourel, A.V. (eds) (2014) *Advances in Social Theory and Methodology (RLE Social Theory): Toward an Integration of Micro-and Macro-Sociologies*, Routledge, London, England.
- Coleman, J.S. (1990) *The Foundations of Social Theory*, Belknap Press, Cambridge, England.
- Craig, C.L., Marshall, A.L., Sjostrom, M., *et al.* (2003) International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, 35, 1381–1395.
- Dechenaux, E., Kovenock, D., and Sheremeta, R.M. (2015) A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18 (4), 609–669.
- Diez-Roux, A.V. (2000) Multilevel analysis in public health research. *Annual Review of Public Health*, 21, 171–192.
- Duflo, E., Glennerster, R., and Kremer, M. (2007) Using randomization in development economics research: a toolkit. *Handbook of Development Economics*, 4, 3895–3962.

- Duncan, O.D. (1984) *Notes on Social Measurement: Historical and Critical*, Russell Sage, New York.
- Durlauf, S.N. (2002) On the empirics of social capital. *Economics Journal*, 112, 459–479.
- Durlauf, S.N. and Young, H.P. (2001) *Social Dynamics*, Brookings Institution Press, Washington, DC.
- Eyler, J.M. (1979) *Victorian Social Medicine: The Ideas and Methods of William Farr*, Johns Hopkins University Press, Baltimore, MD.
- Feldman, H.A., McKinlay, J.B., and Potter, D.A., *et al.* (1997) Nonmedical influences on medical decision making: an experimental technique using videotapes, factorial design, and survey sampling. *Health Service Research*, 32, 343–366.
- Finan, K., Larson, D.E., and 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.
- Galea, S. (2013) An argument for a consequentialist epidemiology. *American Journal of Epidemiology*, 178 (8), 1185–1191.
- Goran, M.I. and 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. and Morgenstern, H. (2001) Confounding in health research. *Annual Review of Public Health*, 22, 189–212.
- Greenlund, K.J., Neff, L.J., Zheng, Z.J., *et al.* (2003) Low public recognition of major stroke symptoms. *American Journal of Preventative Medicine*, 25, 315–319.
- Gustafson, P. (2004) *Measurement Error and Misclassification in Statistics and Epidemiology*, Chapman & Hall/CRC, Boca Raton, FL.
- Hamlin, C. (1998) *Public Health and Social Justice in the Age of Chadwick: Britain, 1800–1854*. Cambridge University Press, New York.
- Hardin, G. (1968) Tragedy of the commons. *Science*, 162, 1243–1248.
- Harper, S. and Strumpf, E.C. (2012) Social epidemiology: questionable answers and answerable questions. *Epidemiology*, 23 (6), 795–798.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer-Verlag New York, Inc., New York.
- Henrich, J., Boyd, R., Bowles, S., *et al.* (2005) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small Societies*, Oxford University Press, New York.
- Heller, S.B., *et al.* (2015) Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago, no. w21178. National Bureau of Economic Research working paper.
- Hogbin, M.B. and Hess, M.A. (1999) Public confusion over food portions and servings. *Journal of American Dietetic Association*, 99, 1209–1211.
- Holland, P.W. (1986) Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Jurek, A.M., Greenland, S., Maldonado, G., and 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.S. and Cooper, R.S. (1999) Seeking causal explanations in social epidemiology. *American Journal of Epidemiology*, 150, 113–119.

- Kaufman, J.S. and Harper, S. (2013) Health equity: utopian and scientific. *Preventative Medicine*, 57 (6), 739–740.
- Kennedy, P. (1988) *A Guide to Econometrics*, 4th edn, MIT Press, Cambridge, MA.
- King, G. and Zeng, L. (2007) When can history be our guide? The pitfalls of counterfactual inference. *International Studies Quarterly*, 51, 183–210.
- King, G., Keohane, R.O., and Verba, S. (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton University Press, Princeton, NJ.
- Klinenberg, E. (2002) *A Social Autopsy of Disaster in Chicago*, University of Chicago Press, Chicago, IL.
- Lazarsfeld, P.F. and Menzel, H. (1961) On the relation between individual and collective properties, in *Complex Organizations: A Sociological Reader* (ed. A. Etzioni), Holt, Rinehart, and Wintson, New York.
- Leamer, E. (1983) Let's take the con out of econometrics. *American Economic Review*, 73, 32–43.
- Lieberson, S. and 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 *Sociological Methodology 1993*, vol. 23 (ed. P.V. Marsden), Blackwell Publishers, for the American Sociological Association, Washington, DC, pp. 1–56.
- McKim, V.R. and Turner, S.P. (1997) *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, Notre Dame University Press, Notre Dame, IN.
- McKinlay, J.B., Lin, T., Freund, K., and 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.
- Mirelman A.J., Rose S., Khan J.A., Ahmed S., Peters D.H., Niessen L.W., Trujillo A.J. (2016) The relationship between non-communicable disease occurrence and poverty-evidence from demographic surveillance in Matlab, Bangladesh. *Health Policy Plan*, 31(6): 785–792.
- Naimi, A.I. and Kaufman, J.S. (2015) Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. *Current Epidemiology Reports*, 2 (1), 52–60.
- Nandi, A. and Harper, S. How consequential is social epidemiology? A review of recent evidence. *Current Epidemiology Reports*, 2 (1), 61–70.
- Nock, S.L. and Rossi, P.H. (1979) Household types and social standing. *Social Forces*, 57, 1325–1345.
- Nunnally, J.C. and Bernstein, I.H. (1994) *Psychometric Theory*, McGraw-Hill, New York.
- Oakes, J.M. (2004) The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social Science and 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., Andrade, K.E., Biyoow, I.M., and Cowan, L.T. (2015) Twenty years of neighborhood effect research: an assessment. *Current Epidemiology Reports*, 2 (1), 80–87.
- Oakes, J.M. and Rossi, P.H. (2003) The measurement of SES in health research: current practice and steps toward a new approach. *Social Science and Medicine*, 56, 769–784.
- O'Brien, D.T., Sampson, R.J., and Winship, C. (2015) Econometrics in the age of big data: measuring and assessing “broken windows” using large-scale administrative records. *Sociological Methodology*, 45, 101–147.

- Olson, M. (1971) *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, Cambridge, MA.
- Orduñez-García, P.O., Nieto, F.J., Espinosa-Brito, A.D., and 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., and Lefcowitz, M.J. (1971) Income Maintenance: interdisciplinary approaches to research. *Institute for Research on Poverty*, Madison, WI.
- Packard, R.M. (1989) *White Plague, Black Labor: The Political Economy of Health and Diseases in South Africa*, University of California Press, Berkeley.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference*, 2nd edn, Cambridge University Press, New York.
- Ransohoff, D.F. and 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. and 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. and Berk, R.A. (1987) Varieties of normative consensus. *American Sociological Review*, 50, 333–347.
- Rossi, P.H. and Nock, S.L. (1982) *Measuring Social Judgements: The Factorial Survey Approach*, Sage Publications, Beverly Hills, CA.
- Sampson, R.J. (2003) Neighborhood-level context and health: lessons from sociology, in *Neighborhoods and Health* (eds I. Kawachi and L.F. Berkman), Oxford University Press, New York, pp. 132–146.
- Schwappach, D.L. and Koeck, C.M. (2004) What makes an error unacceptable? A factorial survey on the disclosure of medical errors. *International Journal of Quality 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 and Economics*, Cambridge University Press, New York.
- Sunstein, C.R. (2014) *Why Nudge?: The Politics of Libertarian Paternalism*, Yale University Press, New Haven, CN.
- Suzuki, E., Yamamoto, E., Takao, S., *et al.* (2012) Clarifying the use of aggregated exposures in multilevel models: self-included vs. self-excluded measures. *PLoS One*, 7 (12), e51717.
- Thomsen, T.F., McGee, D., Davidsen, M., and 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.
- van der Laan, M.J., Polley, E.C., and Hubbard, A.E. (2007). Super learner. U.C. Berkeley Division of Biostatistics Working Paper 222 posted at <http://biostat.bepress.com/ucbbiostat/paper222/>
- Vander Weg, M.W., Watson, J.M., Klesges, R.C., *et al.* (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. (2002). *The Protestant Ethic and the Spirit of Capitalism: And Other Writings*, Penguin, New York.

- Yatchew, A. and Zvi, G. (1984) Specification error in probit models. *Review of Economics and Statistics*, 67, 134–139.
- Yeats, W.B. (1938) “*Under Ben Bulbin*” in *Last Poems and Two Plays*, Irish University Press, Shannon.
- Zhou, X.-H., Obuchowski, N.A., and McClish, D.K. (2002) *Statistical Methods in Diagnostic Medicine*, John Wiley & Sons, Inc., Hoboken, NJ.
- Zucchini, W. (2000) An introduction to model selection. *Journal of Mathematical Psychology*, 44, 41–61.