PERGAMON

Social Science & Medicine 58 (2004) 1929–1952

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology

J. Michael Oakes*

Division of Epidemiology and Population Research Center, University of Minnesota, 1300 South 2nd Street, Minneapolis, MN 55454, USA

Abstract

The resurgence of interest in the effect of neighborhood contexts on health outcomes, motivated by advances in social epidemiology, multilevel theories and sophisticated statistical models, too often fails to confront the enormous methodological problems associated with causal inference. This paper employs the counterfactual causal framework to illuminate fundamental obstacles in the identification, explanation, and usefulness of multilevel neighborhood effect studies. We show that identifying useful *independent* neighborhood effect parameters, as currently conceptualized with observational data, to be impossible. Along with the development of a dependency-based methodology and theories of social interaction, randomized community trials are advocated as a superior research strategy, one that may help social epidemiology answer the causal questions necessary for remediating disparities and otherwise improving the public's health.

Keywords: HLM; Mixed model; Cluster trial; Community trial; Counterfactual; Assignment mechanism; Propensity score

Introduction

A "neighborhood effect" is the independent causal effect of a neighborhood (i.e., residential community) on any number of health and/or social outcomes (Jenks & Mayer, 1990; Mayer & Jenks, 1999). Of interest have been (a) so-called *contextual* effects, which presumably emerge from within-neighborhood social interactions, and (b) so-called *integral* effects that emerge from toxic dumps, parks, sidewalks, etc. (Ozonoff et al., 1987; Geschwind et al., 1992; Susser & Susser, 1996; see Diez-Roux, 1998), though one need not distinguish between them.[1]

Epidemiologists have long recognized that people residing in different areas have differing health outcomes (cf. Macintyre, Maciver, & Sooman, 1993; McMichael, 1999; Catalono & Pickett, 2000; Lawson, 2001). Most would agree that spatial variation in morbidity and mortality is somehow associated with the clustering of genetic predispositions, cultural norms, opportunity structures, and/or environmental conditions. By definition, advantaged neighborhoods offer cleaner, safer, and less stressful environments as compared to, say, ghetto areas. It would be shocking to learn that such contexts did not somehow impact health. The question is about magnitude, mechanism, and mutability: How big are the effects, how do they emerge, and how might such information be exploited to improve the public's health?

*Tel.: +1-612-924-6844; fax: +1-612-624-0315.

E-mail address: oakes@epi.umn.edu (J.M. Oakes).

[1] We postpone a precise definition of a neighborhood effect, which appears to have been confused in most every empirical study to date (Manski, 1995). But whereas social scientists typically focus entirely on context effects due solely to social interactions, social epidemiologists are equally interested in the effects of socially mediated *integral* variables (such as toxic dumps) on health. We therefore include them into our working definition of neighborhood effect, and thereby distinguish ourselves from the primary objectives of social science. Results are independent of this decision.

0277-9536/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.socscimed.2003.08.004

Social scientists have long suffered an interest in neighborhood effects, which they view as a special case of context effects—the *raison d'etre* of social science. Most know Durkheim (1951 (1897)) aimed to show that social forces (e.g., norms and values) external to the individual influenced suicide and that Weber (1958 (1905)) aimed to show how religious ideology shaped economic behavior. But many distinguished contemporary social scientists—notably Merton (1949), Lazarsfeld (Lazarsfeld & Menzel, 1961), Blau (Blau, 1960; Blau & Duncan, 1967), Coleman (Coleman, 1958, 1990; Coleman et al., 1966), Sewell (Sewell, 1964; Sewell & Armer, 1966), Blalock (1984), Hauser (1970, 1974), Jenks and Mayer (Jenks & Mayer, 1990; Mayer & Jenks, 1999), Bandura (1986), Sampson (Sampson, 1991; Sampson, Raudenbush, & Earls, 1997), Massey (Massey and Denton, 1993; Massey, Gross, & Shibuya, 1994); Wilson (1987), Manski (Manski, 1993b, a, 1995, 2000), Arrow (1971, 1994), Akerlof (1970, 1976, 1980, 1997); Bowles and Gintis (Bowles & Gintis, 1977, 2000; Gintis, 2000; Bowles & Gintis, 2002), Shively (Shively, 1974, 1987; Achen & Shively, 1995), and King (1997)—have ably addressed related questions from an analytic/statistical perspective.

Of special import are the similarities between the epidemiologist's neighborhood and the educational scientist's school for they lead us to methodological work on "school effects" by Raudenbush and colleagues (Raudenbush & Bryk, 1986; Raudenbush & Whillms, 1995; Raudenbush & Sampson, 1999a), among others (e.g., Coleman et al., 1966; Aitkin & Longford, 1986; Goldstein, 1995). The problem for educational science is how to estimate the independent effect of good teachers or school administrators on student achievement. The analogous problem for social epidemiologists is to estimate the independent effect of toxic dumps, locally promulgated smoking policies, or inducible increases in social networking, on a neighborhood's health. Both problems share two fundamental characteristics. First, they are typically analyzed with non-experimental (i.e., observational) data. Second, people/students are nested within neighborhoods/schools, which yields a hierarchical data structure where measurements are taken on both individuals and the groups in which they act. How to address these problems is the central concern of this paper.

From an epidemiological perspective, it is difficult to understate the importance of studying contexts such as neighborhoods (cf. Cassel, 1976; McMichael, 1999; Susser, 1999; Berkman, Glass, Brissette, & Seeman, 2000; Krieger, 2001). Social forces, above and beyond any individual, have been repeatedly shown to play an important role in how we perceive, measure, and address health and illness (cf. Parsons, 1951; Starr, 1982; Rose, 1985; Clark, Potter, & McKinlay, 1991; Barr, 1995; McKinlay, 1996; Feldman et al., 1997). Even if we could measure more person-level characteristics, including them all in a model for health violates the principle of parsimony, to say nothing of the theoretical arguments against the atomistic fallacy, biophysical reductionism, or other such "Robinson Crusoe" assumptions (Link & Phelan, 1995; Kaplan, 1996; Susser, 1998). It just may be that "upstream" causes, i.e., those that systematically affect people through their neighborhoods or social groups, are more amenable to interventions designed to improve the public's health.

Focused epidemiologic interest in neighborhood effects dates back to Chadwick's sanitation efforts, circa 1842. Contemporary efforts begin with Cochran, who in the early 1960s developed multivariable regression to help the city of Baltimore estimate the effect of public housing on health-related outcomes (Salsburg, 2001a, b). Motivated by advances in social epidemiology, multilevel theories and sophisticated statistical models, interest in such questions has recently surged. Pickett and Pearl (2001) reviewed 25 "neighborhood effect" studies published in epidemiology journals since the mid-1980s, the thrust of which was that investigators detected small but consistent "context" effects associated with group-level socioeconomic status (SES) on health outcomes. Diez-Roux et al. (2001) recently published "Neighborhood of Residence and Incidence of Coronary Heart Disease" in the *New England Journal of Medicine*, the thrust of which was that people living in lower SES neighborhoods had higher incidence of cardiovascular disease (CVD), independent of their individual-level SES. What is more, the broad notions are now central to our greater discipline: a naive Medline search for "multilevel" or "contextual" in TITLE revealed over 1200 citations; an unrestricted search of the same key words yielded almost 5000 citations. Enthusiasm for such studies is understandable for they exemplify the effect of social forces, emergent contexts, and social relationships on health—the *raison d'etre* of social epidemiology.

Yet due largely to persistent and complex methodological obstacles, along with a lack of attention to them, the *causal effect* of neighborhood contexts on health continues to confuse and elude us (see Hook, 2001). There appear to be no multilevel neighborhood effect studies with observational data, including those cited above, that directly confront causal inference.[2] What is

[2] This paper makes the important distinction between observational and experimental studies. The community trials work by David Murray and others is not included here because it is the sole focus of the later part of this paper. What is more, we ignore the vast literature aiming to estimate the effect of contexts on persons with aggregate (i.e., group-level) data. Although differences between ecological (aggregate) regression and individual-level regression has been of interest since Quetelet (circa 1831) first posited the notion of a *L'homme*

more, despite over three decades of methodological discussion (cf. Hauser, 1970, 1974; Stipak & Hensler, 1982; Blalock, 1984; Gray, 1989; Swaminathan, 1989; Von Korff, Koepsell, Curry, & Diehr, 1992; Manski 1993a; DiPrete & Forristal, 1994; Draper, 1995; Raudenbush & Whillms, 1995; Duncan, Connell, & Klebanov, 1997; Blakely & Woodward, 2000; Greenland, 2001, 2002), it is evident that many social epidemiologists are not clear on exactly what multilevel models are or how they may be used to estimate and interpret *causal* neighborhood effects.

This paper aims to advance our understanding and assist social epidemiologists designing, conducting and/or reviewing multilevel neighborhood effect studies. The first section motivates a model for estimating neighborhood effects. The second section develops, with causal inference in mind, the now common multilevel model for observational data. The third section adopts a critical methodological perspective to show the impossibility of estimating useful neighborhood effects with a regression model, of which the multilevel model is a special case. In order to avoid criticizing without an alternative, the fourth section shows the relationship between multilevel neighborhood effect models and randomized community trial designs, and argues the latter appears to be the best bet for estimating useful neighborhood effects. We conclude by summarizing findings and suggesting that in concert with the development of both a dependency-based methodology and a rigorous theory of social interaction, the community trial be viewed as the canonical experimental design for a social epidemiology seeking to actually improve the public's health.

Methodologists will appreciate the enormous and subtle complexities to be addressed. Such issues are typically handled in isolation and with great precision. But this creates a disparate and technical literature often inaccessible to social epidemiologists. We presume the dearth of tailored and richly annotated translations explains much of the current confusion. Pursuant to remedy, we build our case slowly, take liberties with nomenclature (equations, hats, Greek letters and such), employ a conversational style, and include abundant footnotes and citations. Reaching a broader audience is consistent with our ultimate goal of encouraging a more thoughtful social epidemiological methodology, more likely to provide valid inferences, and thereby the basis for scientific understanding and sound policy recommendations for improving the public's health.

Motivating a causal model for neighborhood effects

This section aims to motivate a causal model for estimating neighborhood effects. The word "cause" is central for it illuminates practical obstacles and potential solutions, and because it links this paper to the important and growing interdisciplinary interest in causal inference (cf. Heckman & Smith, 1995; Manski, 1995; Sobel, 1995; McKim & Turner, 1997; Greenland, Pearl, & Robins, 1999; Kaufman & Poole, 2000; Pearl, 2000; Robins, 2001; Greenland, 2002; Rosenbaum, 2002; Shadish, Cook, & Campbell, 2002). It is always best if causality is established by relating specific outcomes to biological mechanisms and pathways (Beyea & Greenland, 1999; Macintyre & Ellaway, 2000). But it may suffice to understand how such mechanisms are mediated, modified, and otherwise shaped by social interaction.

We begin with first principles: let Y be a normally distributed random variable with realized values y_{ig}, where i indexes an individual nested within group g, the neighborhood unit in which she resides. Imagine that y_{ig} is some measure of CVD risk, for example. Assume two kinds of neighborhoods, bad and good, assign to them labels g_0 and g_1, respectively. The fundamental question is: What would person i's health be under alternative neighborhood conditions, g_0 and g_1? In other words, how would moving to a new neighborhood affect i's risk for CVD?

This setup conforms to the counterfactual causal framework, sometimes called Rubin's (1976, 1991) model, favored in many scientific disciplines (Sobel, 1995; Kaufman & Poole, 2000; Pearl, 2000; Maldonado & Greenland, 2002; Shadish et al., 2002). By exploiting a few such developments, we can better see the obstacles between us and our answer. Although we simplify greatly here, the framework permits us to formally ask whether *hypothetically* changing neighborhoods from g_0 to g_1 affects person i's health endpoint, y_{ig}, by writing the treatment effect due to such a change for *each* person as $y_{i1} - y_{i0} = \Delta_i$. Since we are interested in populations, we write the average causal effect (ACE) of neighborhoods *across* persons as $\sum_{i=1}^{n} \Delta_i n^{-1} = \bar{\Delta} = \text{ACE}$.[3] Our goal now is to estimate this *population* parameter, but many obstacles prevent us from doing so directly.

An obvious problem with counterfactual estimation is that we cannot observe the counterfactual. If i is in g_0 she cannot simultaneously be in g_1. Controlled

(footnote continued)
moyen (average man) (Stigler, 2002), the efforts of more contemporary scholars to infer the effect of contexts on individuals from aggregated data produced nothing of substantive import (Achen & Shively, 1995); the recent efforts of King (1997) perhaps notwithstanding. In any case, this paper assumes hierarchically nested individual-level data. For contemporary insights and cautions about purely ecological studies, see Morgenstern (1995) and Greenland (2001).

[3] ACE is but one parameter of potential interest. Others might include proportion of positive gains or loss minimization. Such extensions are beyond the scope of this paper.

experiments are conducted to approximate the unobservable counterfactual conditions with a *substituted* comparison group (Maldonado & Greenland, 2002). Most believe randomization asymptotically ensures our observed comparison group is substitutable with the unobservable counterfactual comparison group, and thus identical to the treatment group in every way except for the treatment itself. Randomization thus permits actual estimation (a real-world approximation) of our desired causal contrast $\bar{\Delta}$.

Imagine now a real-world experiment to *estimate* $\bar{\Delta}$, the ACE of neighborhood context on health endpoint y. Assume g_0 is an observed "control" neighborhood and g_1 an observed "treatment" neighborhood—perhaps g_1 has less crime, more parks is more politically active, or is replete with affordable vegetarian restaurants. If we let some variable T indicate assignment of persons to g_1 and we randomize a whole bunch of people, we could write a simple OLS regression model (e.g., t-test) for estimating our "treatment effect" as

$$y_{ig} = a + b_1 T + e_{ig}. \quad (1)$$

Conditional on established statistical assumptions,[4] model 1 would yield a consistent and unbiased estimate our treatment effect, $b_1 = \bar{\Delta}$. In other words, if we could randomize people to neighborhoods, we could easily estimate the true causal health effect of moving from a bad neighborhood to a good one.

Of course, we cannot easily randomize people to neighborhoods.[5] Even if it were ethical and constitutional, the "Moving to Opportunity" (MTO) study (see Goering & Kraft, 1999; Katz, Kling, & Liebman, 2001) and other such experiments show that it is very expensive, difficult, and time-consuming to do so (Heckman & Smith, 1995; Rossi, 1997). Since counterfactuals cannot be observed and randomization of people to neighborhoods is, for the most part, impractical, we *appear* to be left with observational designs for estimating neighborhood effects. This is troublesome since in the absence of control over assignment mechanisms we can never be certain our observed control group(s) can be substituted for our desired but unobservable counterfactual group(s) (Campbell & Stanley, 1963; Cook & Campbell, 1979; Heckman, 1979, 1992; Rosenbaum, 2002). The upshot is that enormous difficulties for causal inference obtain, so many in fact that overcoming them has been the driving force behind almost all statistical/methodological research to date. Nevertheless, we are heartened by Leamer (1983), who argues persuasively that the real enemy is not observational data but "whimsical inference" with it.

With respect to neighborhood effects research, the trouble with observational designs is that people, to use a shorthand expression, are "selected" into neighborhoods; they are not randomly distributed.[6] If neighborhoods are imagined to be treatments that affect health, then the "selection" of people to them confounds effect estimates. To see this, appreciate that social stratification (i.e., social selection) means that people of lower SES systematically live in neighborhoods g_0 whilst people of high SES live in other neighborhoods g_1, there is very little overlap or mixing. It follows that the effects of some neighborhood g on any person i's health are clouded by i's personal SES background, which we know affects health (Oakes & Rossi, 2003). More technically, if background differences are collected in a vector of measured variables, $\mathbf{X}$, and/or a vector of unmeasured variables $\mathbf{Z}$, then any between-person differences in the expected values of $\mathbf{X}$ and $\mathbf{Z}$ might confound the effect of the neighborhood on health. Confounding muddles analyses, at times so much so that it renders the estimates useless (Leamer, 1983; Kaufman & Cooper, 1999; Winship & Morgan, 1999; Greenland & Morgenstern, 2001).

In his aforementioned effort to help Baltimore estimate the effect of public housing on health, Cochran realized randomization of people to neighborhoods was impossible and that confounding threatened his results. He developed multivariable regression in order to statistically equate observed groups (Salsburg, 2001a, b). The idea was to include (one or more) conditioning variables $\mathbf{x}$ in model 1. In our case, $\mathbf{x}$, might be measures associated with a person's "selection" of one neighborhood over another. Whatever these measures turn out to be, the resultant *quasi*-experimental (see Cook and Campbell, 1979; Shadish et al., 2002) model for estimating the treatment effect, T, could now be written as the familiar analysis of covariance (ANCOVA),

$$y_{ig} = a + b_1 T + \mathbf{b_2 x} + e_{ig}. \quad (2)$$

[4] The six assumptions for classic regression are (a) model linearity, $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$; (b) no perfect collinearity, $\mathbf{X}$ is $n \times K$ with rank K; (c) no confounding, $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$; (d) no clustering, $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\mathbf{I}$; (e) $\mathbf{X}$ is fixed; and (f) normally distributed errors, $\boldsymbol{\varepsilon}|\mathbf{X} \sim N[\mathbf{0}, \sigma\mathbf{I}]$. See Greene (1997, p. 235).

[5] It is probably more appropriate to say "households" rather than "people." We use the latter for simplicity.

[6] We put quotes around 'select' and call it a shorthand expression since an often subtle and certainly complex sorting and selection process is apparent. Many people are left little choice in residential location; economic inequalities, racism, and other structural phenomena appear to be the dominant forces at work (cf. Rossi, 1955; Logan, 1978; Shlay & Rossi, 1981; Rabin, 1989; Farley, Steeh, & Krysan, 1994; Massey et al., 1994; Schill & Wachter, 1995). But using the 'selection' expression is not just a useful fiction, it relates the "problem" to a larger statistical literature (see Heckman, 1979; Berk & Ray, 1982; Winship & Mare, 1992; D'Agostino, 1998; Rosenbaum, 2002) which we illustrate below.

It appears as if we can once again identify our causal effect; this time, however, we say it is conditional on $\mathbf{x}$ and write $b_1|\mathbf{x} = \bar{\Delta}$. But things are more complicated now, and we state two reasons why.

First, in order to infer a causal effect for T, we must be sure the expected value of our model's error term is conditionally independent of our "selection" variables and combinations thereof (see Heckman, 1979; Clogg & Haritou, 1997; Greene, 1997; Robins, Hernan, & Brumback, 2000; Wooldridge, 2000). Somewhat technically, we can write this as[7]

$$E(e_{ig}|\mathbf{x}) = 0. \quad (3)$$

Though not computationally necessary to calculate regression estimates, this requirement ensures that our neighborhood effect estimates are not confounded with background characteristics.[8] The trouble is that if the vector $\mathbf{x}$ contains many variables, which it often will in social epidemiologic investigations, we will likely find sparse data in any given (cross-tabulated) cell. Sparse data mean more assumptions (e.g., linearity, homoskedasticity, etc) are necessary. Since many such assumptions cannot be tested in the data, our confidence in making useful inferences is undermined (Robins & Greenland, 1986; Robins et al., 2000; Woolridge, 2000; Freedman, 2002). We must rely on a *causal theory* to help us specify which variables to measure and include in a model (Leamer, 1978). Unlike the relatively simple theories tested in randomized experiments, causal theories for observational designs must be elaborate so as to rule out competing explanations.[9] Moreover, we must also consider unmeasured variables, $\mathbf{z}$, which by definition can never be included, for their effects might cause *hidden* bias (Rosenbaum, 2002). A causal theory is once again needed, but this time to assess the sensitivity of our results to assumptions (Seltzer, 1993; Greenland, 1996; Greene, 1997; Manski & Nagin, 1998; Rosenbaum, 2002). Fortunately, statisticians have developed methods to address this conditional independence issue. Below we will show how such methods are implicitly incorporated in multilevel models.

The second problem, altogether ignored by model 2, is that people are clustered within neighborhoods; in fact, it is clustering that defines neighborhoods. As a result of "selection," social interaction and common exposures, it is more than likely that people within neighborhoods are more alike than people between neighborhoods. Consequently, any two people *within* an identifiable neighborhood yield less *independent* information than if people were randomly distributed across the geographic landscape. This fact should be intuitively obvious, since we often can guess things about (i.e., stereotype) an individual when we know what "part of town" he lives in. Within-group dependency is troublesome because an inviolable assumption of OLS regression (i.e., model 2) is that observations are independent (i.e., non-correlated errors) of one another; technically the assumption is written as $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\mathbf{I}$ (Greene, 1997). In a famous quote, Cornfield (1978) warned that treating clustered data as if it were independent was an exercise in self-deception, to be strongly discouraged.[10]

Fortunately, statisticians have also developed useful tools for modeling hierarchically clustered (i.e., error correlated) data. One class of models has been popular in neighborhood effect and other such contextual analysis. Though essentially equivalent, these are referred to as multilevel, hierarchical linear, random coefficient, random effect, mixed, empirical Bayes, Laird-Ware, or variance component models.[11] Diez-Roux (2000, 2002), Greenland (2000), Sullivan et al. (1999), and Murray (1998) offer accessible introductions of such models for epidemiologists. Raudenbush and Bryk (2002) offer an exemplary presentation from a

[7] We should also condition on the treatment, T, but for purposes here we omit this symbolism.

[8] We are indebted to Dr. Jay Kaufman, University of North Carolina, for clarifying this crucial point. Although details are beyond the scope of this paper, constraint 3 (note 4) is meant to imply not only homoskedacity but also a zero covariance between errors and predictors, $E(X\varepsilon) = \sigma_{X\varepsilon} = 0$. Note, however, that this relationship is not observable. For precise terminology, we recommend the remarkably lucid discussion of Clogg and Haritou (1997).

[9] In his important paper on causal inference with observational data, Cochran (1955) lamented not knowing any social scientists who wrote out lists of possible alternative causes *before* they began to analyze data. With respect to neighborhood effects research, at least, same goes for us: we do not know of many researchers who consider alternative hypotheses before they begin their research.

[10] Many others deftly address the problem of within-group dependencies and efforts (e.g., means-as-outcomes) to overcome them. See Aitkin and Longford (1986).

[11] The theoretical foundation of multilevel models lies in variance component methodology, which in its modern form dates back to Fisher's work circa 1925 (Draper, 1995). A ground-breaking advance came when Lindley and Smith (1972) formulated their empirical Bayes regression model, but it was not until the introduction of the EM algorithm (Dempster, Laird, & Rubin, 1977) that computational feasibility was obtained. Laird and Ware (1982) popularized the model for biostatisticians, Bryk, Raudenbush, Goldstein, and Mason for social scientists (Mason et al., 1984; Goldstein, 1987; Bryk & Raudenbush, 1992). From our perspective, the widespread (ab)use of the model is due to the recent introduction of user-friendly software, especially HLM and MlWin, and an accessible translation for SAS users by Verbeke and Molenbergs (1997) and Singer (1998). See also Kreft, de Leeuw, and van der Leeden (1994) and De Leeuw and Kreft (2001).

social scientific perspective.[12] In short, multilevel models are useful because they overcome the within-group dependency problem whilst permitting estimation of ecological effects.[13] But there is nothing magical about multilevel models; the principle difference between them and simple OLS regression models is that multilevel models permit complex error terms (i.e., variance components) by using sophisticated computational algorithms. As a result, more elaborate, if not more intuitive and accurate, inferences *may* be drawn.

A causal multilevel model for neighborhood effects

We now conceptually develop the *standard* multilevel model for estimating neighborhood effects with observational data. The section is important because despite the vast and growing literature cited above, no one appears to have *conceptually developed* the multilevel model with an eye on causal inference.[14] Only through a causally motivated development will the benefits and limitations of a model, if any, reveal themselves. We see four steps in an etiologic multilevel contextual effects analysis.

Step 1: Examine neighborhood-to-neighborhood differences in the endpoint

The first step is to determine if there are any between-neighborhood differences with respect to any endpoint, y_{ig}. This is accomplished by fitting what Raudenbush and Bryk (2002) call a fully unconditional model, which is simply a one-way random-effects analysis of variance (RANOVA). Let the *micro-level* model be a "regular" ANOVA that says each person's health outcome, y_{ig}, is equal to their neighborhood's mean health outcome, a_g, plus some residual error e_{ig}:

$$y_{ig} = a_g + e_{ig}. \tag{4}$$

We now specify a *macro-level* model (sometimes called a level-2 model) of the neighborhood-specific intercept (i.e., mean) with an indicator for the grand mean and a separate random (error) term for neighborhoods,

$$a_g = a^* + u_g, \tag{5}$$

where a^* identifies the grand mean, and u_g is a random effect (i.e., error or deviate from a^*) for each neighborhood g. Literally substituting Eq. (5) into Eq. (4) yields the familiar RANOVA model:

$$y_{ig} = a* + u_g + e_{ig}. \tag{6}$$

This model says that each person's CVD risk, y_{ig}, is equal to a grand mean of all people's CVD risk, denoted a^*, plus an effect, u_g, which is merely a grand-mean deviate specific to a person's own neighborhood, plus some residual or random error e_{ig}. In essence, the random effect provides for a separate intercept for each neighborhood, and thus overcomes the problem of people "clustering" within neighborhoods discussed above.

It is critical to understand that each neighborhood has its own specific grand-mean deviate, or random effect, u_g, for the endpoint of interest y_{ig}. If people were randomly assigned to neighborhoods, any variation in health outcomes between neighborhoods could only be attributable to neighborhoods themselves. Like model 1, our job would be easy under such circumstances: the neighborhood effect for any neighborhood g would be equal to g's random effect, $u_g = \bar{\Delta}$. Notice also that the variance of random-effects, denoted $\mathrm{VAR}(u_g) = \sigma^2_G$, may be viewed as the maximum *between* neighborhood variation in our endpoint that is explainable by the characteristics/aspects of neighborhoods themselves.[15] This means that if we randomized and the estimated value of σ^2_G was zero, then there would be no (detectable) between-neighborhood variability in our endpoint and hence no neighborhood effects.

Things, however, are not so easy. As mentioned with respect to model 2, the "selection" of people to neighborhoods induces systematic difference in the background composition of residents across neighborhoods.

[12] Searle, Casella, and McCulloch (1992) and McCulloch and Searle (2001) contributed noteworthy statistically oriented presentations.

[13] Another approach is the so-called fixed-effects or within-group model. This model essentially adds an indicator (i.e., dummy) variable for each group/neighborhood, and thereby overcomes within-group clustering. The principal disadvantage of this model is that is does not yield between-group variances, and so cannot offer independent effects of group-level characteristics; the fixed-effect model is thus not multilevel in the traditional sense. For more, see Winship and Morgan (1999).

[14] Of the five major texts focussed on multilevel modeling, per se (Goldstein, 1995; Kreft & De Leeuw, 1998; Snijders & Bosker, 1999; Heck & Thomas, 2000; Raudenbush & Bryk, 2002), only Goldstein (1995) has an index citation for "cause." Goldstein devotes very little attention to the issue and essentially notes multilevel models are not adequate for causal analysis. Hausman and Taylor (1981), Heckman and Vytlacil (1998) and Greenland (2002) carefully address these issues but their papers are written at a level of statistical sophistication typically inaccessible to social epidemiologists. As discussed below, the experimental approach of Murray (1998), Donner and Klar (2000) and Murray (2001) are notable exceptions to this criticism.

[15] As with the model name, nomenclature varies dramatically across discipline and even from author to author. Ours is derivative of McCulloch and Searle (2001), which we think readers will find most intuitive for purposes here. Social scientists familiar with such models should note that our $\sigma^2_G = \tau_{00}$ employed by Raudenbush, Goldstein, and other social scientists.

Since our definition of neighborhood effect excludes such compositional effects, we shall say that u_g, and its variance σ_G^2, and all estimates related to them, are confounded. The following step aims to free the estimates from "selection" confounding.

Step 2: Free random effects from "selection" confounding bias

Adjusting for "selection" bias is the most difficult and largely overlooked step in contextual effects research to date (Achen & Shively, 1995). Focus now is therefore turned toward identifying background factors related to people moving to or residing in their neighborhoods. These will be used to specify a "selection" equation. Relevant variables might include individual-level SES measured by total household-income and educational achievement (see Oakes & Rossi, 2003). Other factors are probably important too, but for purposes here we shall assume SES is the only meaningful background variable in the selection equation.[16]

Of secondary concern to our "selection-model" is the inclusion of other known disease-specific person-level confounders, such as the age and sex of residents. When added, the resultant micro-level equation could be written,

$$y_{ig} = a_g + b_1\,\mathrm{SES}_{ig} + b_2\,\mathrm{AGE}_{ig} + b_3\,\mathrm{SEX}_{ig} + e_{ig}. \quad (7)$$

Leaving the macro-level model (i.e., model 5) unchanged, we can substitute it into (7) and get

$$y_{ig} = a^* + b_1\,\mathrm{SES}_{ig} + b_2\,\mathrm{AGE}_{ig} + b_3\,\mathrm{SEX}_{ig} + u_g + e_{ig}, \quad (8)$$

which is sometimes referred to as the Laird-Ware or SAS notation for a multilevel model (Laird & Ware, 1982; Littell, Milliken, Stroup, & Wolfinger, 1996; Ferron, 1997). The benefit of this single-equation substituted presentation of a multilevel model is that it makes plain that a multilevel regression model is a regression model, with, in this case, just one extra error term. Notice that model 8 yields predicted person-level outcomes for our endpoint adjusted for differences in neighborhood composition due to "selection."

We can now more easily state the most critical assumption of multilevel (i.e., random effect) models for causal neighborhood effects:

$$E(u_g|\mathrm{SES\ AGE\ SEX}) = E(u_g|\text{'selection-model'}) \approx E(u_g|\mathbf{X_{ig}}) = 0. \quad (9)$$

Like Eq. (3), Eq. (9) is a causal assumption and simply says that conditional on our "selection" model, there is no confounding between people in neighborhoods. Again, unless this assumption is met parameter estimates will generally be confounded (Hausman & Taylor, 1981; Clogg & Haritou, 1997; Heckman & Vytlacil, 1998; Woolridge, 2000). It is now clear that the covariates in model 8 are a nuisance: they would not be necessary if we could randomly assign subjects to neighborhoods. Under perfect randomization, model 8 would, Ceteris paribus, yield the same estimate of σ_G^2, as model 6, the simple RANOVA. This view differs from the usual way multilevel models are conceived and presented, especially in their two-level social scientific form. What is more, the constraint of Eq. (9) means that the micro-model serves as a simplified propensity score model (Rosenbaum & Rubin, 1983; Rosenbaum, 1984, 2002; D'Agostino, 1998; Robins et al., 2000; see Harding, 2002) that statistically equates the background confounding characteristics of the persons residing in neighborhoods. It follows that interest in, or interpretation of, level-1 "effects" is inconsistent with an etiological neighborhood effect methodology.[17]

Step 3: Estimate context effects

Assuming Eq. (9) holds and that the adjusted between neighborhood variance, σ_G^2, is positive, the next step is to explain why. This is accomplished by including neighborhood-level (i.e., ecological) variables as fixed-effects, which changes the essence of model 5 from an ANOVA into a multivariable regression model. Any theoretically consistent neighborhood-level variable, such as neighborhood SES, the presence of a hazardous waste facility, average health knowledge, or collective efficacy would be an appropriate explanatory variable. But only theoretically informed *mutable* variables should be of interest to social epidemiologists seeking to make policy recommendations.

Consider the following neighborhood-level model, which specifies a socio-structural relationship to *explain* the (random) variation in a_g:

$$a_g = a^* + b_4\,\mathrm{NSES}_g + u_g. \quad (10)$$

Let NSES_g be some neighborhood-level SES measure: it may be the mean SES of people within a neighborhood defined by educational attainment, household income, or other commonly used composite scores. It may also tap some aggregate notion of SES, such as percent working poor. In any case, substituting model 10 into model 7 yields

$$y_{ig} = a* + b_1\,\mathrm{SES}_{ig} + b_2\,\mathrm{AGE}_{ig} + b_3\,\mathrm{SEX}_{ig} + b_4\,\mathrm{NSES}_g + u_g + e_{ig}, \quad (11)$$

[16] Because they do not affect conclusions, this paper ignores centering issues which appear to be useful and important to applications. See Kreft and De Leeuw (1995) and Raudenbush and Bryk (2002).

[17] Level-1 coefficients indicate social selection or inequality, not group effects. In a personal email communication (December 2002), distinguished statistician David A. Freedman insightfully noted that the level-1 equation has no causal content because it describes how nature chose values for the manipulable variable(s) in the level-2 (causal) equation.

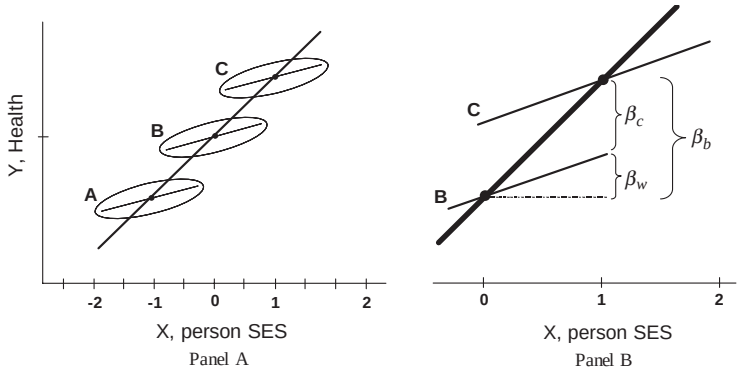

Fig. 1. Multilevel neighborhood structure: Panel A illustrates within and between relationship of SES and some health outcome, Panel B illustrates within, between, and contextual coefficients for same measures.

which estimates parameters associated with a person's health conditional on their neighborhood SES context, adjusted for "selection" and medically relevant but exogenous factors.

Much attention has focused on the magnitude and precision of b_4, the coefficient for neighborhood-level SES (Boyd & Iverson, 1979; Achen & Shively, 1995; Pickett & Pearl, 2001). As specified in model 11, this coefficient is a so-called *context effect*. In theory, this effect is the difference (i.e., causal contrast) in person *i*'s health between, say, low and high SES neighborhoods if, Ceteris paribus, that person could move from one to the other. The coefficient b_4 is thus typically viewed as the *independent* effect of neighborhood on health.

Fig. 1, which is adapted from Raudenbush and Bryk (2002, Figs. 5.2 and 5.1), illustrates the phenomena. Panel A shows the relationship between SES and health across a few arbitrary neighborhoods. Some preferred measure of SES is quantified on the horizontal axis. The value of a health endpoint, Y, is represented by the vertical axis. Within-neighborhood scatter plots are indicated by the ellipses. Within- and between-neighborhood OLS regression slopes are indicated by the lines. The point of Panel A is that, unlike non-hierarchical data, our data reflect an outcome variable that varies both within- and between-neighborhoods.

Panel B magnifies Panel A to reveal estimable slope coefficients for two arbitrary neighborhoods, B and C. Notice that the *within-neighborhood* variation is modeled by slopeβ_w, and the *between-neighborhood* variation is modeled by slope β_b. As shown by Raudenbush and Whillms (1995) and Raudenbush and Bryk (2002), the effect of moving from neighborhood B to C while keeping SES constant is the so-called *context effect*; which is represented by $\beta_c = \beta_b - \beta_w$. With respect to model 11, $\beta_c = b_4$.

We postpone a discussion as to whether we may interpret coefficients for neighborhood-level SES variables as *context* effects. For now, let us simply state that unless our "selection" equation is perfect, our neighborhood-level SES coefficient cannot be interpreted as such. Instead, and this is important, neighborhood-level SES coefficients will serve as further adjustments for the background composition of persons within-neighborhoods; NSES_g will capture and reflect residual confounding.

If we *assume* the coefficient for neighborhood-level SES in model 11 is a bona fide context effect, then it is conceptually easy to take the next step and model the relationship between contexts and composition by a including a so-called "cross-level" interaction term in our model (Hauser, 1970; Achen & Shively, 1995; Greenland, 2002). Consider the interaction between SES_{ig} and NSES_g. As usual (see Aiken & West, 1991), this relationship may be identified by multiplying the two variables together, such as (SES_{ig}*NSES_g), which when substituted yields

$$y_{ig} = a* + b_1\mathrm{SES}_{ig} + b_2\mathrm{AGE}_{ig} + b_3\mathrm{SEX}_{ig} + b_4\mathrm{NSES}_g + b_5(\mathrm{SES}_{ig}*\mathrm{NSES}_g) + u_g + e_{ig}. \quad (12)$$

The cross-level coefficient, b_5, is presumably a valid effect modifier, indicating the degree to which the effect of emergent contexts vary with neighborhood composition. Keen interest is understandable since, at least in theory, cross-level interactions map the relationship between the composition of neighborhoods and their emergent context effects—a phenomena formally postulated first by Durkheim in 1897.

Cross-level effect coefficients uniquely differentiate multilevel models from "standard" regression models (e.g., model 2), which ignore the hierarchical structure of data and treat observations as independent from one another, *and* optimally informed ecological regression models (e.g., models 8, 11, or 13) that estimate the effects of contextual and/or integral variables.[18] If valid,

[18] We say optimally informed because the primary trouble with ecological (i.e., between-group) regression models is that they contain no information on joint within-group covariate distributions (Morgenstern, 1995; Greenland, 2001, 2002). As

cross-level interactions may be of great interest to scientists and practitioners alike.[19] But we shall ignore cross-level interactions in the next step.

Step 4: Evaluate mutable neighborhood effects

Although it does not make much difference, for purposes of this subsection it is easiest to assume that our neighborhood-level SES measure, NSES_g, does *not* reflect a context effect but rather serves as an adjustment for a misspecified "selection" equation. So stated, recall that the preceding discussion showed that to the extent model 11 reveals extant differences between adjusted neighborhood random effects, that is $\hat{\sigma}^2_G > 0$, our etiological aim is to explain this variance through neighborhood-level variables such as hazardous waste facilities, crime rates, quality sidewalks, social cohesion, or any other such *mutable* characteristics. If we collect the corresponding "treatment" indicator variables in a vector $\mathbf{T}_g$, and include it in our model, then our fully specified neighborhood effects model becomes

$$y_{ig} = a* + b_1\,\mathrm{SES}_{ig} + b_2\,\mathrm{AGE}_{ig} + b_3\,\mathrm{SEX}_{ig} + b_4\,\mathrm{NSES}_g + \mathbf{b_T}\mathbf{T_g} + u_g + e_{ig}. \quad (13)$$

Assuming bias due to selection and other relevant health covariates are controlled by adjustment, it seems one may now test the putative causal effect of any given neighborhood effect, T_g. Recall that our counterfactual parameter of interest is the ACE of neighborhoods on health outcomes, symbolized as $\bar{\Delta}$. Because we have developed it from causally consistent first principles, we can exploit model 13, a multilevel model, and write $\hat{b}_T = \bar{\Delta}$, or more honestly, $\hat{b}_T|\text{assumptions} = \bar{\Delta}$.

Methodological obstacles

Methodological research is concerned with the logic of causal inference. The objective is to learn what conclusions can and cannot be drawn given a specified combination of assumptions (Manski, 1995). In short, we explore the extent to which an effect parameter can be "identified" through various research designs, where "identified" may be loosely defined as accurately estimated or detected (see Hsiao, 1983; Manski, 1995 for formal definitions).

We now adopt a critical methodological perspective and evaluate the social epidemiological usefulness of our "state-of-the-art" multilevel model for estimating neighborhood effects with observational data.[20] We are hardly the first to consider such things. Many able social scientists (Hauser, 1970, 1974; Hausman, 1978; Boyd & Iverson, 1979; Hausman & Taylor, 1981; Blalock, 1984; Mason, Wong, & Entwisle, 1984; Tilly, 1984; Mason, 1991; DiPrete & Forristal, 1994; Achen & Shively, 1995; see Heckman & Vytlacil, 1998) offered insights and cautionary wisdom,[21,22] but these either predated the renewed enthusiasm that came with multilevel models or were not tailored to questions addressed here. Epidemiological discussions (cf. Duncan et al., 1997; Diez-Roux, 1998; Blakely & Woodward, 2000; Diez Roux, 2001) insightfully noted many important issues, including measurement error, migration, and neighborhood definitions, but paid too little attention to causal inference.[23] Understandably, none of the more recent and rigorous discussions of causal inference in either epidemiology or social science (Susser, 1973; Greenland, 1990, 2001, 2002; Manski, 1993b; Halloran & Struchiner, 1995; Morgenstern, 1995; Sobel, 1995; Kaufman & Cooper, 1999; Kaufman & Poole, 2000; Kaufman & Kaufman, 2001; Robins, 2001; Maldonado & Greenland, 2002) addressed multilevel neighborhood effects research directly. Finally, none of the many noteworthy general discussions on causal inference with observational data (e.g., Campbell & Stanley, 1963; Cochran, 1965; McKinlay, 1975; Heckman, 1979; Leamer, 1983; Smith, 1990; Rubin, 1991; Clogg & Haritou, 1997; Copas & Li, 1997; Freedman, 1997; Winship & Morgan, 1999; Pearl, 2000; Rosenbaum, 2002) address neighborhood effects or multilevel models, which appear to present some unique issues. Only

(*footnote continued*)
shown above and by Greenland (2001, 2002), ecologically oriented multilevel models, exploit subject-level covariates to overcome the problem. They are implicitly emphasized here.

[19] In a personal email communication (October 2001), distinguished sociologist Rossi, called cross-level context effects the "Holy Grail" of social research.

[20] We readily admit that many more complex and potentially illuminating models may be fit. For example, one might model within-neighborhood SES with random slopes. In addition, some may explore alternative covariance structures and endogeneity tests (Littell et al., 1996; Greene, 1997), or different functional forms. We also acknowledge an awareness of important work on other more subtle problems, including 'boundary problems' and spatial autocorrelation within and between neighborhoods. But our developed model is no "straw man." In any case, no such advances appear sufficient to overcome the problems described in this next section.

[21] Hauser (1970, p. 662) writes "Contextual analysis [as currently conducted] provides no special insight into a process determining group differentiation…It is possible to generate contextual effects at will… Such exercises are fruitless, and the sociological literature would benefit from their absence."

[22] Tienda (1991, p. 256) writes "…the issue of *feedback* effects between individuals and their social environments [has not been properly addressed]. An adequate grasp of exposure and feedback effects is essential to sort out individual and context effects… Absent this information, it will be impossible to [identify neighborhood effects]."

[23] The Macintyre and Ellaway (2000) and the recent Greenland (2002) papers are notable exceptions.

Draper (1995) appears to have directly addressed causal inference in multilevel context effects models, but this (magnificent) paper was brief, addressed "school effects," and appeared in a journal not included in the Medline database. This evident gap probably explains the misunderstanding about the *meaning* of neighborhood effect estimates noted above and by a few informed commentators (cf. Jenks & Mayer, 1990; Achen & Shively, 1995).

Though inextricably related, we distinguish and present four *fundamental* "identification problems" for neighborhood effect research as currently applied to observational data. The first two of these appear sufficient to render dubious any *inference* about the independent causal effect of neighborhood contexts on health, at least through the use of multiple linear regression analysis of which multilevel models are a special case. The second two appear to render the causal *utility* of any important contextual effect, regardless of method, meaningless for a practicable social epidemiology. These problems are not a mere function of a multilevel statistical model, they appear to be fundamental identification problems that supersede statistical estimation issues. We worry that they are severe enough to undermine the possibility of *ever*-making sound policy recommendations from neighborhood effect studies as currently practiced. As such, our problems would render other discussions of measurement error, boundary effects, migration, and model specification, moot.

Problem 1: Social stratification confounds comparisons

The first problem is that social stratification renders the "selection" equation for any given person observed in a neighborhood nearly identical to the "selection" equation for any other person residing in the same neighborhood, but different from the "selection" equation for persons residing in other neighborhoods. It follows that there is (approximately) complete *confounding* between the background attributes of persons in a given neighborhood and (approximately) complete *separation* between the background attributes of people in other neighborhoods. The result is a paradox: we must perfectly specify our "selection" equation to get unconfounded random effects (i.e., neighborhood effects), but when we do so we reduce the value of the random effects, and their variance, to zero and so eliminate any meaningful between-neighborhood variation possibly explainable by neighborhood-level variables.[24]

One may gain an intuitive sense for the problem by considering the work on school effects by Raudenbush and Whillms (1995). Recall the purpose of this research is to estimate the effect of the actions, policies, and efforts of school teachers (administrators, staff, etc) on student achievement apart from the background characteristics (i.e., initial status) of the students they aim to develop. As a policy question, this seems only fair since one would not want to punish excellent teachers effectively developing students who, due to circumstances beyond the teachers' control, enter the school deficient; in fact, such teachers should probably receive more praise than lackadaisical teachers developing students in a school with an advantaged entering class. In any case, the effect on student achievement caused by teachers, including any resultant social synergistic effects, is by definition a context effect. Raudenbush and Whillms (1995) showed how to use a multilevel model, much like model 13, to tease apart the effect of the teachers (context effect) by controlling for the background characteristics of the entering students (composition). In effect, after controlling for the "selection" of students to schools, the remaining differences in student achievement between schools is attributable to teachers. The problem for us, to further the analogy, is that neighborhoods have no teachers. That neighborhood effects are endogenous renders any efforts to control for "selection" between neighborhood variability attributable to only chance and measurement error.

Now consider the problem with respect to the multilevel model developed above. Eq. (9) showed that we had to perfectly specify our "selection" equation to get unbiased estimates of the random-effect variance $\hat{\sigma}^2_G$ and all the neighborhood-level effects that model it. But because of social stratification doing this makes the value of $\hat{\sigma}^2_G$ approach zero.[25] In other words, in order to estimate unbiased neighborhood effect estimates, $T_g = \bar{\Delta}$, we must necessarily make adjustments until there is nothing for the neighborhood-level variables to

[24] Very recent work by Harding (2002) shows that complete overlap may not obtain if we condition on 'neighborhood of adolescence' and limit comparisons to very and moderately poor neighborhoods. Although the public health benefit of such comparisons remains to be demonstrated, Harding's approach reveals the benefit of rigorous "dependency-based" (i.e., fixed effect) estimation and theorizing.

[25] There are of course exceptions: some rich people do live in less wealthy areas and some people spend all of their income on "good" neighborhoods. But, these people are exceptions to the rule and should not be given that same level of statistical credence as the majority. In fact, this very point is being made in recent discussions of "weighted propensity score" analyses, where group-specific "odd balls" are down weighted relative to other subjects in the data set (Hullsiek & Louis, 2002; Robins et al., 2000).

explain.[26] Consequently, observed differences in health outcomes between neighborhoods cannot be separated from the "selection" or background information of individuals residing there. This holds for "context effects" emerging from social interactions or "integral effects" from things like hazardous waste facilities.[27]

Technically, the crux of this matter lies in the assignment mechanisms (see Rosenbaum & Rubin, 1983; Rosenbaum, 1984, 2002; Rubin, 1991; Copas & Li, 1997), in this case both the sampling and treatment mechanisms, which appear uniquely distinct in multilevel studies. Efforts to make our *sampling* mechanism conditionally independent end up making our *treatment* mechanism perfectly dependent or endogenous.[28] The better we specify our (high-dimension) sample-selection model, the more limited our treatment model becomes. Social stratification results in almost perfect separation in propensity scores across subclasses, which renders treatment effect estimates meaningless.[29] Neighborhood effect research suffers a damning trade-off of efforts to meet conditional independence and spherical disturbance assumptions.

Problem 2: Emergent contexts are endogenous

Another way to think about neighborhood effects is that they are emergent properties of the social interactions of the residents.[30] Examples might include (Durkheimian) social norms or political activism. As such, neighborhood effects would appear less dependent on "selection" equations and thus identifiable. Not so, any emergent effects of neighborhoods on a person's health are *by definition* completely endogenous to the composition of neighborhoods; there is no exogenous intervention causing them. As shown by Manski (1993a, b, 1995), endogenous effects are not identifiable.[31,32]

In this case an epidemiological intuition is easy to convey. The trouble with estimating endogenous neighborhood effects is that, like prototypical infections, incidence depends on prevalence (Halloran & Struchiner, 1991, 1995). Estimating emergent effects with either standard or multilevel linear regression models is like trying to estimate incidence by controlling for prevalence in a cross-sectional study. This task is obviously impossible since multivariable regression models (i.e., *conditional* mean models) assume that once we condition on some confounder, Z, we can estimate the conditionally *independent* effect of some X on Y. Endogenous context effects are by definition conditionally *dependent* and thus violate model *assumptions*. Again, we cannot identify the causal effect of interest.

One might consider incorporating the dependent aspects of the problem into a regression model or research design. It *appears* reasonable to control for time-one background characteristics (composition) to estimate the contextual effects that emerge in the period before they are observable at time-two. Even if we ignore regression to the mean (Campbell & Kenny, 1999), the fact is neighborhood compositions change over time and such change is presumably related to state of affairs (i.e., composition and any emergent context effects, or prevalence) at time-one (see Talih & Fricker, 2002). This means that contexts *indirectly* affect composition, or in other words incidence *indirectly* affects prevalence (Halloran & Struchiner, 1991, 1995). The upshot is grim news for studies aiming to estimate

[26] This statement assumes no instruments exist; see footnote 33 below.

[27] The correlation between social composition and so-called locally undesirable land uses (LULUs) conforms to our common sense ability to identify disadvantaged neighborhoods in the first place and, at least within our social system, why without exogenous intervention they become more disadvantaged over time (Molotch, 1976; Logan, 1978; Shlay & Rossi, 1981; White, 1987; Wilson, 1987; Kohlhase, 1991; Massey & Denton, 1993; Anderton, Anderson, Oakes, & Fraser, 1994; Anderton, Oakes, & Egan, 1997; Massey et al., 1994; Romo & Schatz, 1995; Massey, 1996; Oakes, Anderton, & Anderson, 1996; Evans & Kantrowitz, 2002).

[28] A still deeper point is made by Draper (1995), who lays plain how the sampling mechanism determines difference between specific and general causal inference. Generalizability is not enhanced by specifying neighborhoods as random effects; doing so does not make them a random sample from the larger population. See Berk (1991) and Berk et al. (1995) for accessible discussion.

[29] Manski approaches the general problem another way. He demonstrates that strong dependence/collinearity arises when attributes defining reference groups (e.g., neighborhoods) are a subset of attributes that affect treatments (composition). See "case (a)" in Manski (1995, p. 132).

[30] Here is where precise definitions begin to matter and much of the confusion can be traced. Manski (1995, p. 127–8) defines *endogenous effects* as those in which an individual's behavior varies as a function of the active behaviors of others. *Context effects* are those in which an individual's behavior varies as a function of some structural aspect of their group, such as percentage white. Both effects are social effects but only the former exhibits (non-Markovian) feedback.

[31] Manski (1995, p. 1) offers a global intuition: "Suppose you observe that almost simultaneous movements of a man and his image in a mirror. Does the image cause the man's movements or reflect them?" Manski called this the "reflection" problem.

[32] Durlauf (Brock & Durlauf, 2002; Durlauf, 2001) extends Manski's work and notes a few sufficient conditions for identification of nearly endogenous effects. Recent interdisciplinary work on the effects of conditioning treatment effects with putative confounders affected by treatments illuminates the problem and, given definitions, seems to lay plain the issue (cf. Robins et al., 1992, 2000; Robins, 1989).

neighborhood context effects with (multilevel) regression models.[33,34]

Problem 3: Extrapolation

We now briefly discuss a second set of identification problems. These are relevant to not just neighborhood effects but context effects more generally, and severely undermine the *utility* of context-related causal inference in a practicable social epidemiology. The first issue (third overall) is that multilevel neighborhood effect models assume people are "exchangeable" *between* neighborhoods, which defies real-world logic.[35] Within our causal framework this means that estimated treatment effects (i.e., neighborhood-level parameter estimates) must be practicable. Investigators routinely make an exchangeability assumption and in most cases it makes good sense. But shall we assume that we can statistically equate (i.e., exchange) person i from neighborhood g_0 in Appalachia with person j from g_1 in Beverly Hills by merely including some covariates in a model? The answer is no, and the reason, which is illustrated in Fig. 2, appears somewhat unique to multilevel neighborhood effects models.[36]

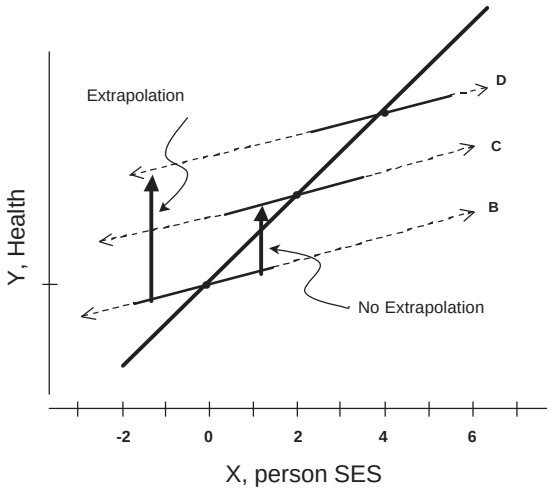

Fig. 2. Inferential and practicability errors due to extrapolation in arbitrary multilevel neighborhood structure. Note: diagram magnifies Fig. 1 to highlight the "within slopes" of three arbitrary neighborhoods, B, C, and D.

Fig. 2 shows that our model essentially estimates both within-neighborhood SES slopes for each neighborhood as well as a between-neighborhood SES slope across neighborhoods. These slopes are based on observed data but, because it is a model, extend infinitely in both directions. It is the (linear) *extrapolation* of the within-neighborhood slopes that causes the problem. Most neighborhood effect studies implicitly propose to (hypothetically) move people of constant SES from one neighborhood to another—the definition of a context effect. Yet without completely altering the social structure of society, the effect of such moves (i.e., interventions) are based on pure extrapolations. That the data do not support the inference means there is identification error.[37] With respect to Fig. 2, "treating" a person of low SES in neighborhood B with the benefits of neighborhood D, a *context effect*, is meaningless in the real world.

Problem 4: Disequilibria

The last problem again goes to the utility of neighborhood effect models and concerns dynamic

[33] Another approach might be to employ the instrumental variable technique often used in econometrics (Case & Katz, 1991; Blundell & Windmeijer, 1997; Heckman & Vytlacil, 1998; Rice, Jones, & Goldstein, 1998; Spencer & Fielding, 2000). The idea here is to adjust our model for the endogenous (i.e., dependent) explanatory variables by finding "instrumental variables" that are (a) uncorrelated with the error term and (b) correlated strongly with the explanatory variables (Greene, 1997; Newhouse & McClellan, 1998; Woolridge, 2000). But even this approach cannot overcome the complete endogeneity that neighborhoods appear to exhibit, and even if it could the prospects for finding good "instruments" are slim.

[34] Yet another alternative is to apply techniques that capitalize on spatial autocorrelation, such as those used in spatial statistics (Doreian, 1981; Haining, 1990; Lawson, 2001). But among problems is one that appears to undermine the approach: spatial modeling implicitly assumes away any notion of neighborhood as a discrete socio-political unit. The models instead rely on some distance decay function, which implicitly assumes a model of social interaction. Unless tested or at least testable, the social interaction model cannot be justified which renders spatially induced contextual parameters unidentified. Manski (1993a, 1995) named this reference group problem as the 'reflection' problem.

[35] By "exchangeable" we mean the *possibility* of equivalent response types across groups. Without the possibility of exchangeability, modeled treatment (i.e., causal) effects are meaningless. See also Draper, Hodges, Mallows, and Pregibon (1993) and Draper (1995) for profound insight into this issue.

[36] The MTO study avoids this issue by, to be simple, moving subjects from bad neighborhoods to less bad ones—there were no extreme differences between treatment and control groups. Overlap of experimental neighborhoods would seem to forestall extrapolation errors. But such an effect may explain why only

(*footnote continued*)
about 40% of those given vouchers to move actually did so. See Goering and Kraft (1999) and Katz et al. (2001) for more. In any case, marginal effects from two similar neighborhoods appear to have little public health value.

[37] Methodologists say that such a point is "off the support" of the data. Identification of on-support inferences is detectable in the data. Off-support identification is qualitatively different, requiring substantive reasoning for plausibility.

social behavior. Our causal framework requires that a treatment given to one person does not affect (the treatment given to) another person. Rubin (1976) and Little and Rubin (2000) named this the stable unit-treatment assumption (SUTVA) and it remains of great interest to epidemiologists (cf. Halloran & Struchiner, 1995; Kaufman & Cooper, 1999). Along with other topics of interest to social epidemiologists, SUTVA is violated by most notions of neighborhood effects. The reason is that disequilibria obtain when relocating, say, a poor person to a wealthier neighborhood. Indeed, by definition, moving a large number of poor people to a wealthy neighborhood reduces the wealth of the target neighborhood (Schelling, 1971). With respect to Fig. 2, although an intervention (i.e., relocation) of people from B to C does not suffer the extrapolation problem discussed above, when such a move is done en masse SUTVA is violated. The impact is that a there will be post-intervention disequilibria as neighborhood C is compositionally transformed into B. Having the same composition as the original neighborhood, B, any causal estimate inferable from the move is lost.[38] Such an identification problem undermines the practicability of the MTO study and all others relying on its (hypothetical) treatment effect.

An alternative: community trials

The preceding has begged the nihilistic question: Are multilevel neighborhood effect studies of any use to a practicable social epidemiology? Our answer is "yes, but…"

Estimation of independent neighborhood effects from observational data and multilevel models, as described above, appears as if they *will always be wrong*, but some might be useful for theory development, preliminary testing, and provisional conclusions when experiments are not possible. There is no question that neighborhood effect studies have spurred unwitting epidemiologists to seriously consider social contexts as a central variable in public health research. But the meaning attributable to any such estimates is clearly dubious. In the absence of evidence to the contrary, it seems that all prior neighborhood effect estimates, whether from a single or multilevel model, reflect measurement error and misspecification rather than causal effects. Given this problem's long history, we worry that continued indifference to causal inference may divert scarce resources away from fruitful social epidemiological investigations aiming to improve the public's health. Or worse, such indifference might yield whimsical inferences that lead to policies that literally harm people. Proposals for and findings from carefully conducted observational multilevel neighborhood effect studies need not be rejected, but they do merit careful scrutiny by investigators, funding agencies, policymakers, and the public.

What can social epidemiologists interested in the effect of neighborhood contexts on health do? First, "less sophisticated" methods requiring greater immersion into neighborhood contexts and dependencies, perhaps through anthropological methods and social theory, hold promise (see Tilly, 1984; King, Keohane, & Verba, 1994; Trostle & Sommerfeld, 1996; Macintyre & Ellaway, 1998; Ferman & Kaylor, 2001; Frohlich, Potvin, Chabot, & Corin, 2002). Second, the work of Kaufman and Cooper (1999, 2001), Kaufman and Poole (2000) and Kaufman and Kaufman (2001) provides another possibility that merits greater attention. In short, these workers are developing methods based on explicit conditional probabilities to infer the effects of complex and inter-related distal causes, such as SES status, on health outcomes. Since the preceding discussion essentially showed that neighborhood effects are dependent happenings,[39] Kaufman's approach may prove invaluable.[40] So emphasized, we have two worries: (1) dependency-oriented theory and methodology is in its infancy and will require enormous advances to bear practicable fruit; (2) the approach may *unintentionally* limit social epidemiology to observational data and "file-cabinet" analyses. We take note of Cochran (1955),

[38] Such unintentional effects were cause for concern in efforts to desegregate schools in the USA. Investigators showed that when black children were "bused" to suburban predominantly "white" schools, white families moved out, dramatically altering neighborhood composition (Farley, Richards, & Wurdock, 1980). Such phenomena may be viewed as 'Not in My Backyard' (NIMBY) disequilibria.

[39] The term "dependent happenings" has been attributed to Sir Ronald Ross and seems to have been popularized by Halloran and Struchiner (1991). The gist of the idea is that the magnitude of events (e.g., disease incidence) depends on or is fully endogenous to background conditions or prevalence.

[40] Note well that social epidemiologists need not break ground in this regard. Statistical methodology is rapidly advancing useful tools (cf. Greenland, 2001, 2002; Rosenbaum, 2002; Robins et al., 2000; Heckman & Vytlacil, 1998; Manski & Nagin, 1998; Newhouse & McClellan, 1998; Copas & Li, 1997; Heckman, 1997, 1979; Manski, 1995; Robins et al., 1992; Robins & Greenland, 1986). And there is a vast but often overlooked literature on social interaction, exemplified by early dynamic theories of social contagion (e.g., Crane, 1991; Grannovetter & Soong, 1983; Grannovetter, 1978; Schelling, 1971; Simon, 1952) and more recent cutting-edge theories of social interaction that are linked, not surprisingly, to evolutionary (economic) game theory (e.g., Bowles & Gintis, 2002; Brock & Durlauf, 2002; Bendor & Swistak, 2001; Durlauf, 2001; Gintis, 2000; Ostrom, 2000; Bowles, 1998; Wilson & Sober, 1994; Boyd & Richardson, 1985; Maynard Smith, 1982; Cavalli-Sforza & Feldman, 1973).

who insisted that when considering methods for studying human populations we *always* ask if an experiment is possible. We herewith ask, and enthusiastically answer "yes."

Just as the randomized clinical trial is the canonical design in pharmacological testing (Pocock, 1983; Freidman, Furberg, & DeMets, 1998), the randomized community trial is canonical design for neighborhood effect studies in particular, and social epidemiology more generally.[41] Detailed presentation is beyond our scope, but the design is well understood and firmly established in the health promotion, evaluation science, and statistical literatures (cf. McGraw et al., 1989; Feldman & McKinlay, 1994; Feldman, 1997; Hannan, Murray, Jacobs, & McGovern, 1994; Feldman, McKinlay, & Niknian, 1996; Murray, 1998; Donner & Klar, 2000; Raudenbush & Liu, 2000).[42] The gist is that community-level "treatments," such as mass-media campaigns to improve health knowledge, the repair of bad sidewalks, or community policing initiatives, are randomly assigned to a randomly selected set of entire existing communities (Charlton, D'Souza, Tooley, & Silver, 1985; Meyer, Job-Spira, Bouyer, Bouvet, & Spira, 1991; Shipley, Hartwell, Austin, Clayton, & Stanley, 1995; Holder, Saltz, Treno, Grube, & Voas, 1997; Feldman et al., 1998; LeFort, Gray-Donald, Rowat, & Jeans, 1998; Persky et al., 1999; Biglan, Ary, Smolkowski, Duncan, & Black, 2000; Luepker et al., 2000).

Unlike (multilevel) observational studies, community trials may be *designed* to control the so-called assignment mechanisms between the composition of neighborhoods and contexts and, by dint of randomization, render them strongly ignorable.[43] This means the *independence* of neighborhood effects may be assumed, which means our desired average causal contrasts, $\bar{\Delta}$, will be readily estimable—sometimes without adjustment. In essence, community trial designs turn MTO and such experimental studies on their head. Instead of randomizing people to neighborhoods, we randomize "contexts" to neighborhoods. What is more, unlike prototypical clinical trials that assume people, like Robinson Crusoe, are unaffected by social relations, the primary focus of community trials is on social relationships and interactions. Finally, unlike clinical trials that (necessarily) study people in laboratory settings, community trials study effects in the so often complicated, messy, and dynamic *real-world*.

We emphasize that community focused treatments (i.e., interventions) need not be limited to groupwise implementations of individual-level (i.e., behavioral) interventions—though many of these could prove vital to reducing disparities and otherwise improving the public's health. *Social* interventions, such as efforts to politically mobilize neighborhood members, alter norms, change local policies, fix sidewalks, and clean parks, are natural fits (cf. Perry et al., 1993; Forster et al., 1998; Wagenaar et al., 2000).[44] Fitting well in community trial designs are community-based participatory interventions gaining favor in many areas (Nichter, 1984; Lefebvre, Lasater, Carleton, & Peterson, 1987; Travers, 1997; Israel, Schulz, Parker, & Becker, 1998; Higgins, Maciak, & Metzler, 2001; O'Fallon & Dearry, 2001, 2002; cf. Cockburn & Trentham, 2002; Koch, Selim, & Kralik, 2002; Krieger et al., 2002; Schulz, Krieger, & Galea, 2002). The fact is that researchers have barely scratched the surface of possible interventions for enhancing population health, especially with respect to *social* interventions presumably of interest to social epidemiologists (Moffitt, 2001).[45] The good news is that since community trial interventions must be developed a priori, it is likely that the design will promote the development of theory for contextual (i.e., social or endogenous feedback) interventions and ways to better measure them (Rossi, 1972; Marsden, 1990; Raudenbush & Sampson, 1999b; Sampson & Raudenbush, 1999; Schweingruber & McPhail, 1999; Caughy, O'Campo, & Patterson, 2001; Doreian, 2001; Glaeser & Scheinkman, 2001; Krieger, 2001). Accordingly, the design is *constructive* of causal inference in a practicable social epidemiology.

But the unsurpassable advantage of community trial designs is that they test practicable public health interventions—be they good or bad. The design

[41] Community trials are a special case of group trials, which are sometimes called cluster trials. Given this paper's topic, we use the former term but prefer "group randomized trial" for the more general design.

[42] Sorensen et al. (1998) provides an excellent and accessible contemporary overview.

[43] Randomization is optimal, but may under the right circumstances not be necessary. There is an abundant literature of effect estimators for quasi-experimental designs (Heckman & Hotz, 1989; Moffitt, 1991; Heckman, 1992; Friedlander & Robins, 1995). Exogenous treatments are necessary, and constructive.

[44] It should be no surprise that cutting-edge community trials incorporate infectious-disease interventions and analytic strategies (see Longini et al., 2002; Hayes et al., 2000; Halloran et al., 1999, 1997; Halloran & Struchiner 1995, 1991). Such efforts show that the insights of Kaufman and Cooper (1999) are readily, if not more easily, addressed with experimental data. Moreover, it is possible to nest a "traditional" multilevel neighborhood effects analysis in a community trial to examine heterogeneity of treatment effects, among other phenomena (Seltzer, 1994).

[45] Moffitt (2001) defines two classes of interventions: those that operate through private incentives and those that operate through social interactions. Private incentives are described as *fundamental*, for they are based on an individual's calculus ignoring the characteristics or actions of others. Social interventions may exploit social feedback, group-leaders, norms, multiple equilibria to affect change.

discourages consideration of what *could* be done "*if the world were just a little different*". Real world programs and their implementations are literally tested, hopefully with dispassionate and unforgiving rigor. The development, implementation, evaluation, refinement and perhaps dissemination of successful strategies can only enhance and add credibility to a practicable social epidemiology seeking to address health disparities and otherwise improve the public's health. If nothing else, such efforts will conform to the mutability requirement so critical to causal inference (Kaufman & Cooper, 1999).

Enthusiasm aside, it is important to recognize that community trials suffer many limitations too. Among them, first and foremost is that many questions cannot be answered by experiments; there may be ethical, practical, and sheer intellectual constraints on what can be done.[46] Second, the community trials completed to date have produced disappointing results (Susser, 1995; Fishbein, 1996; Winkleby, Feldman, & Murray, 1997; Murray, 1998; Sorensen, Emmons, Hunt, & Johnson, 1998). It is not easy to develop effective interventions, alter behavior, and credibly estimate treatment effects (Lytle et al., 1994a, b; Lytle, 1998).[47] Nor are there sufficient data on variance components needed to design new studies (Hannan et al., 1994; Feldman, 1997; Murray & Short, 1997; Murray, Clark, & Wagenaar, 2000; Murray, Phillips, Bimbaum, & Lytle, 2001). Third, community trials are very expensive, often requiring tens of millions of dollars, and require skilled interdisciplinary investigators willing to fully commit to the process, often for several years. Too often studies are woefully under-powered or otherwise poorly designed (Murray, 1998; Donner & Klar, 2000; Varnell, Murray, & Baker, 2001). Fourth, many diseases and their risk factors have extensive latency periods, which when mixed with secular trends and suboptimal designs inhibit inference (Carleton, Lasater, Assaf, Lefebvre, & McKinlay, 1987; Carleton, Lasater, Assaf, Feldman, & McKinlay, 1995; Bauman, Suchindran, & Murray, 1999). Fifth, as in other intervention designs, casual effect estimates from trials may suffer Hawthorne effects (Mayo, 1933; Roethlisberger & Dickson, 1938), which are typically conceived as changes in behavior due simply to being the focus of attention. Such social placebo effects (Cook & Campbell, 1979) violate SUTVA and muddle the construct validity of causal effects.[48] Sixth, there are unanswered ethical concerns, such as how to gain community consent (Howard-Jones, 1982; Edwards, Braunholtz, Lilford, & Stevens, 1999; Oakes, 2002). Finally, knowledgeable commentators have expressed valid concerns about the lack of intervention theory and its effect on the development and analysis of community trials (Baranowski, Lin, Wetter, Resnicow, & Hearn, 1997; Walker, 1997).[49] In sum, social experiments of any sort pose a whole set of issues not encountered in a classic agricultural experiment upon which statistical inference is too often based (Hausman & Wise, 1985; Heckman & Hotz, 1989; Smith, 1990; Heckman, 1992; Heckman & Smith, 1995; Manski, 1995; Rossi, 1997; see especially, Rossi, Freeman, & Lipsey, 1999).

Conclusion

Ever since Durkheim empirically "demonstrated" that emergent properties of groups influenced the behavior of individuals independent of their background characteristics, social scientists have aimed to estimate them. But the quantification of such effects has proved both elusive and vexing. Exploiting recent theoretical and statistical advances, social epidemiologists recently have joined the quest by examining neighborhood effects with sophisticated multilevel models and observational data. Evident motivation and enthusiasm are understandable (cf. Kaplan, 1996; Lynch, Kaplan, & Salonen, 1997; Diez-Roux, 1998; Yen & Syme, 1999; Berkman et al., 2000;

[46] In light of the public's acceptance for placebo-controlled clinical experiments, where lives are crashingly at stake, it is astonishing that the public does not embrace a experimental method for social reform (Campbell, 1973; Freeman & Rossi, 1981; Rossi et al., 1999). Marketing the methodology may prove to be the most difficult, but useful, aspect of the design.

[47] A notable example is found in the recent MPH thesis of Shierman (2002), which showed the "dose" of intervention-specific health promotion media messages did not exceed substantively similar messages unrelated to the intervention in all (paired) community settings of the recent REACT trial (Feldman et al., 1998; Simons-Morton et al., 1998; Luepker et al., 2000). In effect, the REACT dose did not exceed background noise.

[48] Although muddling the science, such effects may not be a bad thing if they actually improve health in control groups. So long as they are recognized, such "placebo effects" should not hinder a practicable social epidemiology. But in any case, the evidence for such effects is dubious in social settings (Cook & Campbell, 1979). In fact, Franke and Kaul (1978) more closely examined data from the original Hawthorne experiments and found that the cause of the observed efficiency effects appears more related to implied threats of employee termination than with a humanistic sense of connection between managers and laborers. Such work is yet another example of how careful analysis is requisite in the analysis of emergent group effects.

[49] A causal theory of how neighborhoods might affect health merits immediate and vigorous attention. Specific putative causes, such as good sidewalks, high crime rates, political activism, or diesel exhaust, must be identified and evaluated. Richly descriptive studies and advances in methods to measure neighborhood characteristics are essential. In the absence of significant advances in both theory and measurement, the prospects for estimating useful causally independent neighborhood effects appear slim (Macintyre et al., 1993).

Krieger, 2001; Kawachi, 2002), for since it is obvious that health varies across neighborhood contexts, a potentially manipulable force might be at work.

This paper highlighted the growing prominence of multilevel neighborhood effects studies in social epidemiology and the fact that few have paid careful attention to casual inference. Oversight is understandable since the technical literature on causal inference is not only daunting and disparate, but fluid, as shown by our abundant footnotes and citations. In hopes of illuminating some complex and subtle issues for social epidemiologists, we translated and tailored a great deal of interdisciplinary work in epidemiology, statistics, and the social sciences so as to motivate and develop a multilevel causal model for identifying and estimating neighborhood effects. Upon scrutiny, however, we showed that (multilevel) regression models, no matter how sophisticated, appear unable to identify useful neighborhood effects from observational data.

The recent and enthusiastic adoption of the multilevel model for neighborhood effects research appears to be a case of *statisticism*, a term used to describe an almost ritualistic appeal to significance testing and both sampling and measurement error when they are not the problem (Berk, 1991; Freedman, 1991, 1997; Clogg & Arminger, 1993; Manski, 1993a b; Berk, Western, & Weiss, 1995; Clogg & Haritou, 1997;). Nevertheless, recent attention to how neighborhood contexts affect health has been unquestionably fruitful, if for no other reason than having enriched the debate on the etiology of health and illness. There can be no question that social structures and relations impact health and that disturbing disparities exist. And it is patently obvious that health varies with neighborhood. The ''problem'' is that such phenomena are, *per force*, dependent happenings and as such render ineffective (multilevel) regression models aiming to identify independent effects.

With respect to neighborhood effects research, social epidemiologists should consider rich anthropological methods, develop a dependency-oriented approach as suggested by Kaufman, and carefully examine cutting-edge socioeconomic theories. But we insist that an experimental methodology is possible and superior for causal inference in a practicable social epidemiology. Such an approach not only ties multilevel neighborhood effect studies to both the rich history of experimental methods in human populations (cf. Cochran, 1955; McKinlay, 1981), but links social epidemiology to the rich history of active *social* experiments, exemplified by the works of Campbell (1957, 1973), Campbell and Stanley (1963), Cook and Campbell (1979), Rossi (1980), Freeman and Rossi (1981), Hausman and Wise (1985), Heckman (1992), Cook and Shadish (1994), Friedlander and Robins (1995), Heckman and Smith, (1995) and Rossi et al. (1999). This work continues to advance methods for intervening on, informing and literally improving social policy—goals consistent with social epidemiology's (Pearce, 1996; Kawachi, 2002). In any case, continued indifference to causal inference (i.e., whimsical inference) in social epidemiology threatens our success.

What is the independent causal effect of neighborhood contexts on health and how might we use this information to improve the public's health? We do not yet know, but now we know our methodology must change if we wish to find out. There is no better way to close than by noting that in one of his many delightful papers, Cochran (1955) anticipated the overarching problem: ''If nature mixes things up thoroughly, as she sometimes seems to do, statistical methods will not sort them out very well.'' Creative approaches are required.

Acknowledgements

This paper was supported by grant HL61573 from the National Heart, Lung and Blood Institute (NHLBI/NIH). Beyond the helpful recommendations of two anonymous reviewers, the comments and criticisms of several colleagues improved this paper. Thanks to Andre Araujo, Henry Blackburn, Heather R. Britt, Henry A. Feldman, David A. Freedman, Pamela Jo Johnson, Jay S. Kaufman, Ichiro Kawachi, David M. Murray, Stephen W. Raudenbush, Peter H. Rossi, Ruth N. Lopez Turley, and members of the Social Epi Workgroup, University of Minnesota. Special thanks to Ken Kleinman for abundant assistance on this and related endeavors. Formal research for this paper began whilst the author was at the New England Research Institutes, Watertown, MA, USA. The usual caveats apply.

References

Achen, C. H., & Shively, W. P. (1995). *Cross-level inference*. Chicago: University of Chicago.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, *149*, 1–43.

Akerlof, G. (1970). The market for 'Lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*, 488–500.

Akerlof, G. (1976). The economics of caste and of the rat-race and other woeful tales. *Quarterly Journal of Economics*, *90*, 599–617.

Akerlof, G. (1980). A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, *95*, 749–775.

Akerlof, G. (1997). Social distance and social decisions. *Econometrica*, *65*(5), 1005–1027.

Anderton, D. L., Anderson, A. B., Oakes, J. M., & Fraser, M. R. (1994). Environmental equity: The demographics of dumping. *Demography*, *31*, 229–248.

Anderton, D. L., Oakes, J. M., & Egan, K. L. (1997). Environmental equity in superfund. Demographics of the discovery and prioritization of abandoned toxic sites. *Evaluation Review*, *21*, 3–26.

Arrow, K. J. (1971). Political and economic evaluation of social effects and externalities. In M. D. Intrilligator (Ed.), *Frontiers of quantitative economics* (pp. 3–23). Amsterdam: North-Holland.

Arrow, K. J. (1994). Methodological individualism and social knowledge. *American Economic Review*, *84*, 1–9.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Baranowski, T., Lin, L. S., Wetter, D. W., Resnicow, K., & Hearn, M. D. (1997). Theory as mediating variables: Why aren't community interventions working as desired? *Annals of Epidemiology*, *7*, S89–S95.

Barr, D. A. (1995). The effects of organizational structure on primary care outcomes under managed care. *Annals of Internal Medicine*, *122*, 353–359.

Bauman, K. E., Suchindran, C. M., & Murray, D. M. (1999). The paucity of effects in community trials: Is secular trend the culprit? *Preventive Medicine*, *28*, 426–429.

Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, *106*(6), 1493–1545.

Berk, R. A. (1991). Toward a methodology for mere mortals. In P. V. Marsen (Ed.), *Sociological methodology*, Vol. 20 (pp. 315–324). Cambridge, MA: Blackwell.

Berk, R. A., & Ray, S. C. (1982). Selection biases in sociological data. *Social Science Research*, *11*, 351–398.

Berk, R. A., Western, B., & Weiss, R. E. (1995). Statistical inference for apparent populations (with discussion). In P. V. Marsden (Ed.), *Sociological methodology*, Vol. 25 (pp. 421–485). San Francisco: Jossey-Bass.

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, *51*, 843–857.

Beyea, J., & Greenland, S. (1999). The importance of specifying the underlying biologic model in estimating the probability of causation. *Health Physics*, *76*, 269–274.

Biglan, A., Ary, D. V., Smolkowski, K., Duncan, T., & Black, C. (2000). A randomised controlled trial of a community intervention to prevent adolescent tobacco use. *Tobacco Control*, *9*, 24–32.

Blakely, T. A., & Woodward, A. J. (2000). Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health*, *54*, 367–374.

Blalock, H. M. (1984). Contextual-effects models: Theoretical and methodological issues. *Annnual Review of Sociology*, *10*, 353–372.

Blau, P. M. (1960). Structural effects. *American Sociological Review*, *25*, 178–193.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.

Blundell, R., & Windmeijer, F. (1997). Cluster effect and simultaneity in multilevel models. *Health Economics*, *6*, 439–443.

Bowles, S. (1998). Endogenous preferences: the cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, *36*, 75–111.

Bowles, S., & Gintis, H. (1977). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. New York: Basic Books.

Bowles, S., & Gintis, H. (2000). Walrasian economics in retrospect. *Quarterly Journal of Economics*, *115*(4), 1411–1439.

Bowles, S., & Gintis, H. (2002). The inheritance of inequality. *Journal of Economic Perspectives*, *16*(3), 3–30.

Boyd, L. H., & Iverson, G. R. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.

Boyd, R., & Richardson, P. J. (1985). *Cultural and the evolutionary process*. Chicago: University of Chicago Press.

Brock, W. A., & Durlauf, N. (2002). Interaction-based models. In J. J. Heckman, & E. Leamer (Eds.), *Handbook of econometrics*, Vol. 5 (pp. 3295–3380). New York: Elsevier.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park: Sage.

Cavalli-Sforza, L., & Feldman, M. W. (1973). Models for cultural inheritance: group mean and group variation. *Theoretical Population Biology*, *4*(42), 42–55.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312.

Campbell, D. T. (1973). Reforms as experiments. In L. L. Roos (Ed.), *Quasi-experimental approaches: Testing theory and evaluating policy* (pp. 187–225). Evanston: Northwestern University Press.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guildford.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Hopewell, NJ: Houghton Mifflin.

Carleton, R. A., Lasater, T. M., Assaf, A., Lefebvre, R. C., & McKinlay, S. M. (1987). The Pawtucket heart health program: I. An experiment in population-based disease prevention. *Rhode Island Medical Journal*, *70*, 533–538.

Carleton, R. A., Lasater, T. M., Assaf, A. R., Feldman, H. A., & McKinlay, S. (1995). The Pawtucket Heart Health Program: Community changes in cardiovascular risk factors and projected disease risk. *American Journal of Public Health*, *85*, 777–785.

Case, A., & Katz, L. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. NBER Working Paper 3705, Cambridge, MA.

Cassel, J. (1976). The contribution of the social environment to host resistance: The fourth Wade Hampton Frost Lecture. *American Journal of Epidemiology*, *104*, 107–123.

Catalono, R., & Pickett, K. E. (2000). A taxonomy of research concerned with place and health. In S. C. Scrimshaw (Ed.), *The handbook of social studies in health & medicine* (pp. 64–83). Thousand Oaks: Sage.

Caughy, M. O., O'Campo, P. J., & Patterson, J. (2001). A brief observational measure for urban neighborhoods. *Health Place*, *7*, 225–236.

Charlton, J. R., D'Souza, M. F., Tooley, M., & Silver, R. (1985). A community trial strategy for evaluating treatment for symptomatic conditions. *Statistics in Medicine*, *4*, 11–21.

Clark, J., Potter, D. A., & McKinlay, J. B. (1991). Bringing social structure back into clinical decision making. *Social Science & Medicine*, *32*, 853–866.

Clogg, C. C., & Arminger, G. (1993). On strategy for methodological analysis. In P. V. Marsden (Ed.), *Sociological methodology*, Vol. 23 (pp. 57–74). San Francisco: Jossey-Banks.

Clogg, C. C., & Haritou, A. (1997). The regression method for causal inference and a dilemma confronting this method. In S. P. Turner (Ed.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 83–112). Notre Dame, IN: Notre Dame University Press.

Cochran, W. G. (1955). Research techniques in the study of human beings. *Milibank Memorial Fund Quarterly*, *33*, 121–136.

Cochran, W. G. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society, Series A*, *128*, 243–265.

Cockburn, L., & Trentham, B. (2002). Participatory action research: Integrating community occupational therapy practice and research. *Canadian Journal of Occupational Therapy*, *69*, 20–30.

Coleman, J. S. (1958). Relational analysis: The study of social organizations with survey methods. *Human Organization*, *17*, 28–36.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, *45*, 545–580.

Copas, J. B., & Li, H. G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B*, *59*, 55–95.

Cornfield, J. (1978). Randomization by group: A formal analysis. *American Journal of Epidemiology*, *108*, 100–102.

Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, *96*(5), 1126–1159.

D'Agostino Jr., R. B. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, *17*, 2265–2281.

De Leeuw, J., & Kreft, I. G. G. (2001). Software for multilevel analysis. In H. Goldstein (Ed.), *Multilevel modelling of health statistics* (pp. 187–204). New York: Wiley.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–8.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health*, *88*, 216–222.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, *21*, 171–192.

Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, *91*, 1783–1789.

Diez-Roux, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health*, *56*, 588–594.

Diez-Roux, A. V., Merkin, S. S., Arnett, D., Chambless, L., Massing, M., Nieto, F. J., Sorlie, P., Szklo, M., Tyroler, H. A., & Watson, R. L. (2001). Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine*, *345*, 99–106.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology*, *20*, 331–357.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomized trials in health research*. New York: Oxford.

Doreian, P. (1981). Estimating linear models with spatially distributed data. In S. Leinhardt (Ed.), *Sociology Methodology* (pp. 359–388). San Francisco: Jossey–Bass.

Doreian, P. (2001). Causality in social network analysis. *Sociological Methods and Research*, *30*, 81–114.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, *20*, 115–147.

Draper, D., Hodges, J. S., Mallows, C. L., & Pregibon, D. (1993). Exchangeability and data analysis (with discussion). *Journal of the Royal Statistical Society, Series A*, *156*, 9–38.

Duncan, G. J., Connell, J. P., & Klebanov, P. K. (1997). Conceptual and methodological issues in estimating causal effects of neighborhoods and family conditions on individual development. In J. L. Aber (Ed.), *Neighborhood poverty*, Vol. 1. New York: Russell Sage.

Durkheim, E. 1951 (1897). *Suicide*. Glencoe, IL: Free Press.

Durlauf, S. N. (2001). A framework for the study of individual behavior and social interactions (with comment). In M. E. Sobel, & M. P. Becker (Eds.), *Sociological methodology* (pp. 47–128). Washington, DC: American Sociological Association.

Edwards, S. J. L., Braunholtz, D. A., Lilford, R. J., & Stevens, A. J. (1999). Ethical issues in the design and conduct of cluster randomised controlled trials. *British Medical Journal*, *318*, 1407–1409.

Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental exposure risk. *Annual Review of Public Health*, *23*, 303–331.

Farley, R., Richards, T., & Wurdock, C. (1980). School desegregation and white flight: An investigation of competing models and their findings. *Sociology of Education*, *53*, 123–139.

Farley, R., Steeh, C., & Krysan, M. (1994). Stereotypes and segregation: Neighborhoods in the Detroit area. *American Journal of Sociology*, *100*, 750–780.

Feldman, H. A. (1997). Selecting end point variables for a community intervention trial. *Annals of Epidemiology*, *7*, S78–S88.

Feldman, H. A., & McKinlay, S. M. (1994). Cohort versus cross-sectional design in large field trials: Precision, sample size, and a unifying model. *Statistics in Medicine*, *13*, 61–78.

Feldman, H. A., McKinlay, S. M., & Niknian, M. (1996). Batch sampling to improve power in a community trial. Experience from the Pawtucket Heart Health Program. *Evaluation Review*, *20*, 244–274.

Feldman, H. A., McKinlay, J. B., Potter, D. A., Freund, K. M., Burns, R. B., Moskowitz, M. A., & Kasten, L. E. (1997). Nonmedical influences on medical decision making: An

experimental technique using videotapes, factorial design, and survey sampling. *Health Services Research*, *32*, 343–366.

Feldman, H. A., Proschan, M. A., Murray, D. M., Goff, D. C., Stylianou, M., Dulberg, E., McGovern, P. G., Chan, W., Mann, N. C., & Bittner, V. (1998). Statistical design of REACT (Rapid Early Action for Coronary Treatment), a multisite community trial with continual data collection. *Controlled Clinicial Trials*, *19*, 391–403.

Ferman, B., & Kaylor, P. (2001). Building the Spatial Community: A case study of neighborhood institutions. *Policy Studies Review*, *18*, 53–70.

Ferron, J. (1997). Moving between hierarchial modeling notations. *Journal of Educational and Behavioral Statistics*, *22*, 119–123.

Fishbein, M. (1996). Great expectations, or do we ask too much from community-level interventions? (Editorial). *American Journal of Public Health*, *86*, 1075–1076.

Forster, J. L., Murray, D. M., Wolfson, M., Blaine, T. M., Wagenaar, A. C., & Hennrikus, D. J. (1998). The effects of community policies to reduce youth access to tobacco. *American Journal of Public Health*, *88*, 1193–1198.

Franke, R. H., & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. *American Sociological Review*, *43*, 623–643.

Freedman, D. (1991). Statistical models and shoe leather (with discussion). In P. V. Marsden (Ed.), *Sociological methodology*, Vol. 21. San Francisco: Jossey-Banks.

Freedman, D. (1997). From association to causation via regression (with comments). In S. P. Turner (Ed.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 113–182). Notre Dame, IN: Notre Dame University Press.

Freedman, D. A. (2002). *On specifying graphical models for causation, and the identification problem*. Technical Report No. 601, Department of Statistics, UC, Berkeley.

Freeman, H. E., & Rossi, P. H. (1981). Social experiments. *Milibank Memorial Fund Quarterly/Health and Society*, *59*, 340–373.

Freidman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials*. New York: Springer.

Friedlander, D., & Robins, P. (1995). Evaluating program evaluations: New evidence on commonly used nonexperimental methods. *American Economic Review*, *85*, 39–52.

Frohlich, K. L., Potvin, L., Chabot, P., & Corin, E. (2002). A theoretical and empirical analysis of context: Neighbourhoods, smoking and youth. *Social Science & Medicine*, *54*, 1401–1417.

Geschwind, S. A., Stolwijk, J. A., Bracken, M., Fitzgerald, E., Stark, A., Olsen, C., & Melius, J. (1992). Risk of congenital malformations associated with proximity to hazardous waste sites. *American Journal of Epidemiology*, *135*, 1197–1207.

Gintis, H. (2000). *Game theory evolving: A problem-centered introduction to modeling strategic interaction*. Princeton, NJ: Princeton University Press.

Glaeser, E. L., & Scheinkman, J. A. (2001). Measuring social interactions. In P. Young (Ed.), *Social dynamics* (pp. 83–131). Cambridge, MA: MIT Press.

Goering, J., & Kraft, J. (1999). *Moving to opportunity for fair housing demonstration program: Current status and inital findings*. Washington, DC: US Department of Housing and Urban Development.

Goldstein, H. (1987). *Multilevel statistical models in educational and social research*. London: Edward Arnold.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Grannovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*, 1420–1443.

Grannovetter, M., & Soong, R. (1983). Threshold models of diffusion and collective behavior. *Journal of Mathematical Sociology*, *9*, 165–179.

Gray, J. (1989). Multilevel models: Issues and problems emerging from their recent application in British studies of school effectiveness. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 127–145). New York: Academic Press.

Greene, W. H. (1997). *Econometric analysis*. NJ: Prentice Hall.

Greenland, S. (1990). Randomization, statistics, and causal inference. *Epidemiology*, *1*, 421–429.

Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology*, *25*, 1107–1116.

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*, 158–167.

Greenland, S. (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology*, *30*, 1343–1350.

Greenland, S. (2002). A review of multilevel theory for ecological analyses. *Statistics in Medicine*, *21*, 389–395.

Greenland, S., & Morgenstern, H. (2001). Confounding in health research. *Annual Review of Public Health*, *22*, 189–212.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37–48.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. New York: Cambridge University Press.

Halloran, M. E., & Struchiner, C. J. (1991). Study designs for dependent happenings. *Epidemiology*, *2*, 331–338.

Halloran, M. E., & Struchiner, C. J. (1995). Causal inference in infectious diseases. *Epidemiology*, *6*, 142–151.

Halloran, M. E., Longini Jr., I. M., & Struchiner, C. J. (1999). Design and interpretation of vaccine field studies. *Epidemiology Reviews*, *21*(1), 73–88.

Halloran, M. E., Struchiner, C. J., & Longini Jr., I. M. (1997). Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American Journal of Epidemiology*, *146*(10), 789–803.

Hannan, P. J., Murray, D. M., Jacobs Jr., D. R., & McGovern, P. G. (1994). Parameters to aid in the design and analysis of community trials: Intraclass correlations from the Minnesota Heart Health Program. *Epidemiology*, *5*, 88–95.

Harding, D. J. (2002). *Counterfactual models of neighborhood effects*. Paper presented at the annual meetings of the American sociological association, Chicago, IL.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology*, *74*, 645–664.

Hauser, R. M. (1974). Contextual analysis revisited. *Sociological Methods & Research*, *2*, 365–375.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *52*, 1219–1240.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*, 1377–1398.

Hausman, J. A., & Wise, D. (1985). *Social experimentation*. Chicago: University of Chicago Press.

Hayes, R. J., Alexander, N. D., Bennett, S., & Cousens, S. N. (2000). Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research*, *9*(2), 95–116.

Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–162.

Heckman, J. J. (1992). Randomization and social policy evaluation. In I. Garfinkle (Ed.), *Evaluating welfare and training programs*. Cambridge, MA: Harvard University Press.

Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, *84*, 862–877.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, *9*, 85–110.

Heckman, J. J., & Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, *33*, 974–987.

Higgins, D. L., Maciak, B., & Metzler, M. (2001). CDC urban research centers: Community-based participatory research to improve the health of urban communities. *Journal of Womens Health and Gender-Based Medicine*, *10*, 9–15.

Holder, H. D., Saltz, R. F., Treno, A. J., Grube, J. W., & Voas, R. B. (1997). Evaluation design for a community prevention trial. An environmental approach to reduce alcohol-involved trauma. *Evaluation Review*, *21*, 140–165.

Hook, E. B. (2001). (Letter to Editor) RE: Neighborhood social environment and risk of death: Multilevel evidence from the Alameda County Study. *American Journal of Epidemiology*, *151*, 1132–1133.

Howard-Jones, N. (1982). Human experimentation in historical and ethical perspectives. *Social Science & Medicine*, *16*, 1429–1448.

Hsiao, C. (1983). Identification. In Z. Griliches, & M. D. Intriligator (Eds.), *Handbook of econometrics*, Vol. 1 (pp. 223–283). New York: Elsevier.

Hullsiek, K. H., & Louis, T. A. (2002). Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics*, *2*(4), 179–193.

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, *19*, 173–202.

Jenks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In M. McGeary (Ed.), *Inner-city poverty in the United States* (pp. 111–186). Washington, DC: National Academy Press.

Kaplan, G. A. (1996). People and places: Contrasting perspectives on the association between social class and health. *International Journal of Health Service*, *26*, 507–519.

Katz, L. F., Kling, J. R., & Liebman, J. B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics*, *116*, 607–654.

Kaufman, J. S., & Cooper, R. S. (1999). Seeking causal explanations in social epidemiology. *American Journal of Epidemiology*, *150*, 113–120.

Kaufman, J. S, & Cooper, R. S. (2001). Commentary: Considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, *154*, 291–298.

Kaufman, J. S., & Kaufman, S. (2001). Assessment of structured socioeconomic effects on health. *Epidemiology*, *12*, 157–167.

Kaufman, J. S, & Poole, C. (2000). Looking back on causal thinking in the health sciences. *Annual Review of Public Health*, *21*, 101–119.

Kawachi, I. (2002). What is social epidemiology. *Social Science & Medicine*, *54*, 1739–1741.

King, G. (1997). *A solution to the ecological inference problem*. Princeton, NJ: Princeton University.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NY: Princeton University Press.

Koch, T., Selim, P., & Kralik, D. (2002). Enhancing lives through the development of a community-based participatory action research programme. *Journal of Clinical Nursing*, *11*, 109–117.

Kohlhase, J. E. (1991). The impact of toxic waste sites on housing values. *Journal of Urban Economics*, *30*, 1–26.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modelling*. Thousand Oaks: Sage.

Kreft, I. G. G., & De Leeuw, J. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1–22.

Kreft, I. G. G., de Leeuw, J., & van der Leeden, R. (1994). Review of five multilevel analysis programs: BMDP-5, GENMOD, HLM, ML3, VARCL. *American Statistician*, *48*, 324–335.

Krieger, N. (2001). Theories for social epidemiology in the 21st century: An ecosocial perspective. *International Journal of Epidemiology*, *30*, 668–677.

Krieger, J., Allen, C., Cheadle, A., Ciske, S., Schier, J. K., Senturia, K., & Sullivan, M. (2002). Using community-based participatory research to address social determinants of health: Lessons learned from Seattle Partners for Healthy Communities. *Health Education and Behavior*, *29*, 361–382.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lawson, A. B. (2001). *Statistical methods in spatial epidemiology*. New York: Wiley.

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 Wintson.

Leamer, E. (1978). *Specification searches*. New York: Wiley.

Leamer, E. (1983). Let's take the con out of econometrics. *American Economic Review*, *73*, 32–43.

Lefebvre, R. C., Lasater, T. M., Carleton, R. A., & Peterson, G. (1987). Theory and delivery of health programming in the community: The Pawtucket Heart Health Program. *Preventive Medicine*, *16*, 80–95.

LeFort, S. M., Gray-Donald, K., Rowat, K. M., & Jeans, M. E. (1998). Randomized controlled trial of a community-based psychoeducation program for the self-management of chronic pain. *Pain*, *74*, 297–306.

Lindley, D., & Smith, A. (1972). Bayes estimation for linear models. *Journal of the Royal Statistical Society, Series B*, *34*, 1–41.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior,* (Special issue), 80–94.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, *21*, 121–145.

Logan, J. R. (1978). Growth, politics, and the stratification process. *American Journal of Sociology*, *84*, 404–417.

Longini Jr., I. M., Halloran, M. E., & Nizam, A. (2002). Model-based estimation of vaccine effects from community vaccine trials. *Statistics in Medicine*, *21*(4), 481–495.

Luepker, R. V., Raczynski, J. M., Osganian, S., Goldberg, R. J., Finnegan Jr., J. R., Hedges, J. R., Goff Jr., D. C., Eisenberg, M. S., Zapka, J. G., Feldman, H. A., Labarthe, D. R., McGovern, P. G., Cornell, C. E., Proschan, M. A., & Simons-Morton, D. G. (2000). Effect of a community intervention on patient delay and emergency medical service use in acute coronary heart disease: The rapid early action for coronary treatment (REACT) trial. *JAMA*, *284*, 60–67.

Lynch, J. W., Kaplan, G. A., & Salonen, J. T. (1997). Why do poor people behave poorly? Variation in adult health behaviors and psychosocial characteristics by state of the socioeconomic lifecourse. *Social Science & Medicine*, *44*, 809–819.

Lytle, L. A. (1998). Lessons from the child and adolescent trial for cardiovascular health (CATCH): Interventions with children. *Current Opinion in Lipidology*, *9*, 29–33.

Lytle, L. A., Davidann, B. Z., Bachman, K., Edmundson, E. W., Johnson, C. C., Reeds, J. N., Wambsgans, K. C., & Budman, S. (1994a). CATCH: Challenges of conducting process evaluation in a multicenter trial. *Health Education Quarterly,* (Suppl. 2), S129–S142.

Lytle, L. A., Johnson, C. C., Bachman, K., Wambsgans, K., Perry, C. L., Stone, E. J., & Budman, S. (1994b). Successful recruitment strategies for school-based health promotion: Experiences from catch. *Journal of School Health*, *64*, 405–409.

Macintyre, S., & Ellaway, A. (1998). Social and local variations in the use of urban neighborhoods: A case study of Glasgow. *Health & Place*, *4*, 91–94.

Macintyre, S., & Ellaway, A. (2000). Ecological approaches: Rediscovering the role of the physical and social environment. In I. Kawachi (Ed.), *Social epidemiology* (pp. 332–348). New York: Oxford University Press.

Macintyre, S., Maciver, S., & Sooman, A. (1993). Area, class, and health: Should we be focusing on places or people? *Journal of Social Policy*, *22*, 213–234.

Maldonado, G., & Greenland, S. (2002). Estimating causal effects. *International Journal of Epidemiology*, *31*, 422–438.

Manski, C. F. (1993a). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.

Manski, C. F. (1993b). Identification problems in the social sciences. In P. V. Marsden (Ed.), *Sociological methodology*. San Francisco: Jossey-Banks.

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.

Manski, C. F., & Nagin, D. S. (1998). Bounding disagreements about treatment effects: A case study of sentenciy and recidivism. In A. Raftery (Ed.), *Sociological methodology*, Vol. 28 (p. Boston). MA: Blackwell.

Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology*, *16*, 435–463.

Mason, W. M. (1991). Problems in quantitative comparative analysis: Ugly ducklings are to swans as ugly scatter plots are to…? In J. Huber (Ed.), *Macro-micro linkages in sociology* (pp. 231–243). Newbury Park, CA: Sage.

Mason, W. M., Wong, G. Y., & Entwisle, B. (1984). Contextual analysis through the multilevel Linear Model. In S. Leinhardt (Ed.), *Sociological methodology: 1983–1984*. San Francisco: Jossey-Bass.

Massey, D. S. (1996). The age of extremes: Concentrated affluence and poverty in the twenty-first century. *Demography*, *33*, 395–412.

Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.

Massey, D. S., Gross, A. B., & Shibuya, K. (1994). Migration, segregation, and the geographic concentration of poverty. *American Sociological Review*, *59*, 425–445.

Mayer, S. E., & Jenks, C. (1999). Growing up in poor neighborhoods: How much does it matter? *Science*, *243*, 1441–1445.

Mayo, E. (1933). *The human problems of an industrial civilization*. New York: Viking.

Maynard Smith, J. (1982). *Evolution and the theory of games*. New York: Cambridge.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

McGraw, S. A., McKinlay, S. M., McClements, L., Lasater, T. M., Assaf, A., & Carleton, R. A. (1989). Methods in program evaluation: The process evaluation system of the Pawtucket Heart Health Program. *Evaluation Review*, *13*, 459–483.

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. (1996). Some contributions from the social system to gender inequalities in heart disease. *Journal of Health and Social Behavior*, *37*, 1–26.

McKinlay, S. M. (1975). The design and analysis of the observational study—a review. *Journal of the American Statistical Association*, *70*, 503–523.

McKinlay, S. M. (1981). Experimentation in human populations. *Milibank Memorial Fund Quarterly/Health and Society*, *59*, 308–323.

McMichael, A. J. (1999). Prisoners of the proximate: Loosening the constraints on epidemiology in an age of change. *American Journal of Epidemiology*, *149*, 887–897.

Merton, R. K. (1949). *Social theory and social structure*. Glencoe, IL: Free Press.

Meyer, L., Job-Spira, N., Bouyer, J., Bouvet, E., & Spira, A. (1991). Prevention of sexually transmitted diseases: A randomised community trial. *Journal of Epidemiology and Community Health*, *45*, 152–158.

Moffitt, R. (1991). Program evaluation with nonexperimental data. *Evaluation Review*, *15*, 291–314.

Moffitt, R. A. (2001). Policy interventions, low-level equilbria, and social interactions. In P. Young (Ed.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.

Molotch, H. (1976). The city as a growth machine: Toward a political economy of place. *American Journal of Sociology*, *82*, 302–332.

Morgenstern, H. (1995). Ecologic studies in epidemiology: Concepts, principles, and methods. *Annual Review of Public Health*, *16*, 61–81.

Murray, D. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Murray, D. M. (2001). Statistical models appropriate for designs often used in group-randomized trials. *Statistics in Medicine*, *20*, 1373–1385.

Murray, D. M., Clark, M. H., & Wagenaar, A. C. (2000). Intraclass correlations from a community-based alcohol prevention study: The effect of repeat observations on the same communities. *Journal of Studies on Alcohol*, *61*, 881–890.

Murray, D. M., Phillips, G. A., Bimbaum, A. S., & Lytle, L. A. (2001). Intraclass correlation for measures from a middle school nutrition intervention study: Estimates, correlates, and applications. *Health Education and Behavior*, *28*, 666–679.

Murray, D. M., & Short, B. J. (1997). Intraclass correlation among measures related to tobacco use by adolescents: Estimates, correlates, and applications in intervention studies. *Addictive Behavior*, *22*, 1–12.

Newhouse, J. P., & McClellan, M. (1998). Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health*, *19*, 17–34.

Nichter, M. (1984). Project community diagnosis: Participatory research as a first step toward community involvement in primary health care. *Social Science & Medicine*, *19*, 237–252.

Oakes, J. M. (2002). Risks and wrongs in social science research: an evaluator's guide to the IRB. *Evaluation Review*, *24*(6), 443–478.

Oakes, J. M., Anderton, D. L., & Anderson, A. B. (1996). A longitudinal analysis of environmental equity in communities with hazardous waste facilities. *Social Science Research*, *25*, 125–128.

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*(4), 769–784.

O'Fallon, L. R., & Dearry, A. (2001). Commitment of the National Institute of Environmental Health Sciences to community-based participatory research for rural health. *Environmental Health Perspectives*, *109*(Suppl 3), 469–473.

O'Fallon, L. R., & Dearry, A. (2002). Community-based participatory research as a tool to advance environmental health sciences. *Environmental Health Perspectives*, *110*(Suppl 2), 155–159.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, *14*(3), 137–158.

Ozonoff, D., Colten, M. E., Cupples, A., Heeren, T., Schatzkin, A., Mangione, T., Dresner, M., & Colton, T. (1987). Health problems reported by residents of a neighborhood contaminated by a hazardous waste facility. *American Journal of Industrial Medicine*, *11*, 581–597.

Parsons, T. (1951). *The social system*. New York: Free Press.

Pearce, N. (1996). Traditional epidemiology, modern epidemiology, and public health. *American Journal of Public Health*, *86*, 678–683.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.

Perry, C. L., Williams, C. L., Forster, J. L., Wolfson, M., Wagenaar, A. C., Finnegan, J. R., McGovern, P. G., Veblen-Mortenson, S., Komro, K. A., & Anstine, P. S. (1993). Background, conceptualization and design of a community-wide research program on adolescent alcohol use: Project Northland. *Health Education Research*, *8*, 125–136.

Persky, V., Coover, L., Hernandez, E., Contreras, A., Slezak, J., Piorkowski, J., Curtis, L., Turyk, M., Ramakrishnan, V., & Scheff, P. (1999). Chicago community-based asthma intervention trial: Feasibility of delivering peer education in an inner-city population. *Chest*, *116*, 216S–223S.

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health*, *55*, 111–122.

Pocock, S. J. (1983). *Clinical trials. A practical approach*. New York: Wiley.

Rabin, Y. (1989). Expulsive zoning: The inequitable legacy of Euclid. In J. S. Kayden (Ed.), *Zoning and the American dream*. Washington, DC: Planners Press.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1–17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods*. Newbury Park: Sage.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*, 199–213.

Raudenbush, S. W., & Sampson, R. J. (1999a). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research*, *28*, 123–153.

Raudenbush, S. W., & Sampson, R. J. (1999b). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. In M. P. Becker (Ed.), *Sociological methodology*, Vol. 29 (pp. 1–41). Boston, MA: Blackwell.

Raudenbush, S. W., & Whillms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*, 307–335.

Rice, N., Jones, A., & Goldstein, H. (1998). Multilevel models where random effects are correlated with fixed predictors: A conditioned iterative least squares estimator. *Multilevel Modelling Newsletter*, *10*, 6–12.

Robins, J. M. (1989). The control of confounding by intermediate variables. *Statistics in Medicine*, *8*, 679–701.

Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology*, *12*, 313–320.

Robins, J. M., & Greenland, S. (1986). The role of model selection in causal inference from nonexperimental data. *American Journal of Epidemiology*, *123*, 392–402.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*, 550–560.

Robins, J. M., Mark, S. D., & Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, *48*(2), 479–495.

Roethlisberger, F. J., & Dickson, W. J. (1938). *Management and the worker*. Cambridge, MA: Harvard University Press.

Romo, F. P., & Schatz, M. (1995). The structural embeddedness of business decisions: The migration of manufacturing plants in New York state, 1960–1985. *American Sociological Review*, *60*, 874–907.

Rose, G. (1985). Sick individuals and sick populations. *International Journal of Epidemiology*, *14*, 32–38.

Rosenbaum, P. R. (1984). From association to causation in observational studies: The role of tests of strongly ingnorable treatment assignment. *Journal of the American Statistical Association*, *79*, 41–48.

Rosenbaum, P. R. (2002). *Observational studies*. New York: Springer.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rossi, P. H. (1955). *Why families move: A study in the social psychology of urban residential mobility*. Glencoe, IL: Free Press.

Rossi, P. H. (1972). Community social indicators. In P. E. Converse (Ed.), *The human meaning of social change*. New York: Russell Sage.

Rossi, P. H. (1980). The presidential address: The challenge and opportunities of applied social research. *American Sociological Review*, *45*, 889–904.

Rossi, P. H. (1997). Evaluating community development programs: Problems and prospects. In W. Dickens (Ed.), *Urban problems and community development*. Washington, DC: Brookings Institute.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic/approach* (6th ed.). Thousand Oaks: Sage.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, *47*, 1213–1234.

Salsburg, D. (2001a). Cochran's motivation for multivariable regression. Personal communication.

Salsburg, D. (2001b). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York: WH Freeman.

Sampson, R. J. (1991). Linking the micro- and macrolevel dimensions of community social organization. *Social Forces*, *70*, 43–64.

Sampson, R. J., & Raudenbush, S. W. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, *105*, 603–651.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*, 918–924.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 146–186.

Schill, M. H., & Wachter, S. M. (1995). Housing market constraints and spatial stratification by income and race. *Housing Policy Debate*, *6*, 141–167.

Schulz, A. J., Krieger, J., & Galea, S. (2002). Addressing social determinants of health: Community-based participatory approaches to research and practice. *Health Education and Behavior*, *29*, 287–295.

Schweingruber, D., & McPhail, C. (1999). A method for systematically observing and recording collective action. *Sociological Methods & Research*, *27*, 451–498.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: Wiley.

Seltzer, M. H. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics*, *18*, 207–235.

Sewell, W. H. (1964). Community of residence and college plans. *American Sociological Review*, *29*, 24–38.

Sewell, W. H., & Armer, J. M. (1966). Neighborhood context and college plans. *American Sociological Review*, *31*, 159–168.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized inference*. Boston: Houghton Mifflin.

Shierman, R.A. (2002). Media information flow about myocardial infarction in 20 US communities. Unpublished MPH Thesis, Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN, USA.

Shipley, R. H., Hartwell, T. D., Austin, W. D., Clayton, A. C., & Stanley, L. C. (1995). Community stop-smoking contests in the COMMIT trial: Relationship of participation to costs. Community intervention trials. *Preventive Medicine*, *24*, 286–292.

Shively, W. P. (1974). Utilizing external evidence in cross-level inference. *Political Methodology*, *1*, 61–74.

Shively, W. P. (1987). A strategy for cross-level inference under and assumption of breakage effects. *Political Methodology*, *11*, 167–179.

Shlay, A. B., & Rossi, P. H. (1981). Keeping up the neighborhood: Estimating net effects of zoning. *American Sociological Review*, *46*, 703–719.

Simon, H. A. (1952). A formal theory of interaction in social groups. *American Sociological Review*, *17*(2), 202–211.

Simons-Morton, D. G., Goff, D. C., Osganian, S., Goldberg, R. J., Raczynski, J. M., Finnegan, J. R., Zapka, J., Eisenberg, M. S., Proschan, M. A., Feldman, H. A., Hedges, J. R., & Luepker, R. V. (1998). Rapid early action for coronary treatment: Rationale, design, and baseline characteristics. REACT research group. *Academy of Emergency Medicine*, *5*, 726–738.

Singer, J. D. (1998). Using SAS proc mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.

Smith, H. L. (1990). Specification problems in experimental and nonexperimental social research. In C. C. Clogg (Ed.), *Sociological methodology*, Vol. 20. Oxford: Basil.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks: Sage.

Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In M. E. Sobel (Ed.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 1–38). New York: Plenum.

Sorensen, G., Emmons, K., Hunt, M. K., & Johnson, D. (1998). Implications of the results of community intervention trials. *Annual Review of Public Health*, *19*, 379–416.

Spencer, N. H., & Fielding, A. (2000). An instrumental variable consistent estimation procedure to overcome the problem of endogenous variables in multilevel models. *Multilevel Modelling Newsletter*, *12*, 4–7.

Starr, P. (1982). *The social transformation of American medicine*. New York: Basic Books.

Stigler, S. M. (2002). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Stipak, B., & Hensler, C. (1982). Statistical inference in contextual analysis. *American Journal of Political Science*, *26*, 151–175.

Sullivan, L. M., Dukes, K. A., & Losina, E. (1999). Tutorial in biostatistics: an introduction to hierarchical linear modelling. *Statistics in Medicine*, *18*, 855–888.

Susser, M. (1973). *Causal thinking in health sciences: Concepts and strategies of epidemiology*. New York: Oxford.

Susser, M. (1995). Editorial: The tribulations of trials—intervention in communities. *American Journal of Public Health*, *85*, 156–158.

Susser, M. (1998). Does risk factor epidemiology put epidemiology at risk? Peering into the future. *Journal of Epidemiology and Community Health*, *52*, 608–611.

Susser, M. (1999). Should the epidemiologist be a social scientist or a molecular biologist? *International Journal of Epidemiology*, *28*, S1019–S1022.

Susser, M., & Susser, E. (1996). Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health*, *86*, 674–677.

Swaminathan, H. (1989). Multilevel data analysis: A discussion. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 277–282). New York: Academic Press.

Talih, M., & Fricker, R. D. (2002). Effects of neighborhood demographic shifts on findings of environmental injustice: A New York case-study. *Journal of the Royal Statistical Society, Series A*, *165*, 375–397.

Tilly, C. (1984). *Big structures, large processes, huge comparisons*. New York: Russell Sage.

Travers, K. D. (1997). Reducing inequities through participatory research and community empowerment. *Health Education and Behavior*, *24*, 344–356.

Trostle, J. A., & Sommerfeld, J. (1996). Medical anthropology and epidemiology. *Annnual Review of Anthropology*, *25*, 253–274.

Varnell, S. P., Murray, D. M., & Baker, W. L. (2001). An evaluation of analysis options for the one group per condition design: Can any of the alternatives overcome the problems inherent in this design? *Evaluation Review*, *25*, 440–453.

Verbeke, G., & Molenbergs, G. (1997). *Linear mixed models in practice: A SAS oriented approach*. New York: Springer.

Von Korff, M., Koepsell, T., Curry, S., & Diehr, P. (1992). Multi-level analysis in epidemiologic research on health behaviors and outcomes. *American Journal of Epidemiology*, *135*, 1077–1082.

Wagenaar, A. C., Murray, D. M., Gehan, J. P., Wolfson, M., Forster, J. L., Toomey, T. L., Perry, C. L., & Jones-Webb, R. (2000). Communities mobilizing for change on alcohol: Outcomes from a randomized community trial. *Journal of Studies on Alcohol*, *61*, 85–94.

Walker, A. M. (1997). Kangaroo Court: Invited commentary on shy's the failure of academic epidemiology: Witness for the prosecution. *American Journal of Epidemiology*, *145*, 485–486.

Weber, M. 1958 (1905). The Protestant ethic and the spirit of capitalism. New York: Scribners.

White, M. J. (1987). *American neighborhoods and residential differentiation*. New York: Russell Sage.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass and public policy*. Chicago: University of Chicago Press.

Wilson, D. S., & Sober, E. (1994). Reintroducing Group Selection to the Human Behavioral Sciences. *Behavior and Brain Sciences*, *17*, 585–654.

Winkleby, M. A., Feldman, H. A., & Murray, D. M. (1997). Joint analysis of three US community intervention trials for reduction of cardiovascular disease risk. *Journal of Clinical Epidemiology*, *50*, 645–658.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, *18*, 327–350.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–707.

Wooldridge, J. M. (2000). *Introductory econometrics: A modern approach*. New York: South-Western College Publishing.

Woolridge, J. M. (2000). *Introductory econometrics: A modern approach*. New York: South-Western College Publishing.

Yen, I. H., & Syme, S. L. (1999). The social environment and health: A discussion of the epidemiologic literature. *Annual Review of Public Health*, *20*, 287–308.