Research Forum

Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis

Ana V. Diez-Roux, MD, PhD

ABSTRACT

A large portion of current epidemiologic research is based on methodologic individualism: the notion that the distribution of health and disease in populations can be explained exclusively in terms of the characteristics of individuals. The present paper discusses the need to include group- or macro-level variables in epidemiologic studies, thus incorporating multiple levels of determination in the study of health outcomes.

These types of analyses, which have been called contextual or multilevel analyses, challenge epidemiologists to develop theoretical models of disease causation that extend across levels and explain how group-level and individual-level variables interact in shaping health and disease. They also raise a series of methodological issues, including the need to select the appropriate contextual unit and contextual variables, to correctly specify the individual-level model, and, in some cases, to account for residual correlation between individuals within contexts. Despite its complexities, multilevel analysis holds potential for reemphasizing the role of macro-level variables in shaping health and disease in populations. (*Am J Public Health.* 1998;88:216–222)

Introduction

Throughout the history of public health, and depending on the theory of disease causation prevalent at the time, different aspects of individuals and their environments have been considered important as potential "causes" of disease.[1–3] In its origins, public health was essentially ecological, relating environmental and community characteristics to health and disease.[4–8] With the advent of the germ theory and the associated unicausal theory of disease causation, infectious organisms became the relevant "environmental" factors.[9] Other aspects of the environment were important insofar as they were conducive to reproduction or transmission of the biological "causes" of disease.[10] In this century, the growing importance of chronic diseases led to the search for new causal factors. Emphasis shifted from environmental factors to individual-level factors, and research focused on behavioral and biological characteristics as risk factors for chronic diseases.

The study of the causes of disease thus shifted from the environment as a whole to specific factors within the environment (biological organisms) and to the behaviors of individuals. The model of disease causation shifted from a rather vague, holistic determination to the unicausal model of the germ theory and to the multicausal model (the "web of causation") prevalent today, in which a variety of biological and behavioral risk factors are presumed to interact in the causation of disease.[11] This process has been accompanied by progressive "individualization" of risk (i.e., attributing risks to characteristics of individuals rather than to environmental or social influences affecting populations).

This individualization of risk has perpetuated the idea that risk is individually determined rather than socially determined, discouraging research into the effects of macro-level or group-level variables on individual-level outcomes. "Lifestyle" and "behaviors" are regarded as matters of free individual choice and dissociated from the social contexts that shape and constrain them.[12] This tendency by which disease patterns are explained solely in terms of the characteristics of individuals is analogous to the doctrine of methodological individualism in social science.[13] According to this doctrine, "facts about society and social phenomena are to be explained solely in terms of facts about individuals."[13(p77)] Its logical correlate is that all variables are best measured at the individual level, rather than at the group or macro level, because it is the individual who is presumed to be truly important in the causation of disease. Group-level variables are included in the analyses only as rough approximations for individual-level data when the latter are unavailable. As discussed further in the sections to follow, ignoring the role of group- or macro-level variables may lead to an incomplete understanding of the determinants of disease in individuals as well as in populations. Group- or macro-level variables affect individuals directly and also constrain the choices that individuals make.

The methodological individualism prevalent in epidemiologic research today can be countered in several ways. On one hand, interpretation of individual-level effects should bear in mind their relationship to macro-level processes. Many variables measured at the individual level are strongly conditioned by social processes

The author is with the Division of General Medicine, College of Physicians and Surgeons and Division of Epidemiology, School of Public Health, Columbia University, New York.

Requests for reprints should be sent to Ana V. Diez-Roux, MD, PhD, Division of General Medicine, Columbia Presbyterian Hospital, 622 168th St, PH 9E105, New York, NY 10032.

operating at the levels of social groups or societies. Another approach is to actually include both macro- and micro-level variables in epidemiologic studies, as has been suggested by several authors.[12,14–17] The inclusion of group- or macro-level variables (together with individual-level variables) in public health research is quite challenging from a methodological point of view but is even more challenging from a theoretical point of view. If it is to be meaningful (and not reduced to the mere addition of yet another set of variables to the "web of causation"), it requires the development of models of disease causation (and testable hypotheses) that extend across levels and explain how individual- and group-level variables jointly shape health and disease.

The remainder of this paper discusses some of the basic concepts and issues involved in multilevel analysis in more detail.

Contextual Analysis or Multilevel Analysis

Many sociologists have argued that the lives of individuals are affected not only by their personal characteristics but also by characteristics of the social groups to which they belong. The underlying idea is that social groups are legitimate units of analysis, that group properties are distinct from those of individual members, and that these macro-level variables may affect outcomes independently of individual characteristics or modify how individual characteristics are related to outcomes.[18,19] Consequently, in order to understand individual behaviors and outcomes, it may be useful to analyze not only the characteristics of individuals but also those of the social groups to which they belong.[15]

The study of the effects of collective or group characteristics on individual-level outcomes has been termed contextual analysis[18,20–22] or multilevel analysis.[14,23,24] The effects of group-level properties on individual-level outcomes have been referred to as contextual effects. Empirically, contextual analysis involves the incorporation of group-level variables into micro-level equations appropriate for the study of persons as units of analysis [18] (Table 1). Susser has termed these types of studies "mixed studies" because they investigate the effects of ecological independent variables on individual-level dependent variables.[15] Several different terms have been used as synonyms for group-level variables in these types of analyses, including ecological variables,[15,25] macro-level variables,[14,21] contextual variables,[18,19,21,26] and aggregate variables.[14] Because the terms "contextual variable" and "aggregate variable" have also been reserved for a special type of group-level variable,[14,15,25] their use is avoided in this paper. The terms "group-level variable," "macro-level variable," and "ecological variable" are used synonymously.

TABLE 1—Types of Studies

Independent Variable(s)	Dependent Variable	Type of Study
Group-level	Group-level	Ecological
Individual-level	Individual-level	Individual-level
Group- and individual-level	Individual-level	Multilevel or contextual

Sociologists have documented group-level effects for a variety of individual-level outcomes, including voting behavior,[27] educational attainment,[28] attitudes toward education,[29] and life satisfaction.[30] Although discussion of multilevel analysis in epidemiology is less common, several authors have emphasized the importance of including macro-level variables together with individual-level variables in epidemiologic studies. Morgenstern has argued that "although disease development is an individual biological phenomenon, it is possible that certain important disease determinants cannot be operationalized entirely at the individual level."[31(p16)] Von Korff et al.[14] have suggested that multiple levels of determination (micro- and macro-level variables) should be included in epidemiologic studies of chronic diseases. Other authors have emphasized the need to investigate the role of ecological measures of socioeconomic context in shaping the distribution of risk factors[32,33] and to articulate "the connections between the actions of individuals and the socio-ecological context in which these actions are performed."[12(p818)] In recent reviews of the use of ecological studies in epidemiology, Susser[15] and Schwartz[34] have suggested that ecological variables may provide information not captured by individual-level data. Other authors have also argued for a new epidemiologic paradigm that addresses the relations within and between structures organized in a hierarchy of levels, emphasizing the need to incorporate these multiple levels into epidemiologic analyses.[11,35]

Even within epidemiology, the idea that group-level factors may be important in shaping individual-level outcomes is not at all new. As far back as 1916, in an elegant example of epidemiologic research, Goldberger et al.[36] found that the risk of pellagra was related to the village-level availability of fresh fruits and vegetables, independently of individuals' incomes. In addition, higher household income appeared to afford less protection against pellagra in villages with poor food availability than in those with adequate food availability, illustrating the interaction between individual-level and village-level factors. Another classic example can be found in the field of infectious diseases, in which it has long been hypothesized that an individual's risk of contracting an infection depends not only on his or her own immune status but also on that of the community.

In recent years, a growing literature has highlighted the potential importance of group-level variables in understanding social inequalities in health.[16,37,38] Area-based measures of poverty and deprivation have been found to be associated with health outcomes after adjustment for individual-level factors.[39–46] Additional studies have suggested that neighborhood-level variables may also shape the distribution of health-related behaviors,[12,47–49] although other studies have found little evidence of area effects.[50,51]

Multilevel analysis has also been used in other public health fields. Neighborhood characteristics have been shown to be related to domestic violence.[52] Demographers have documented how country-level variables such as per capita gross national product are related to fertility outcomes independently of individual-level factors.[53] Community-level variables have been found to modify the effects of maternal education on the risk of infant diarrhea.[54] Multilevel analysis has been used to explore the effects of type of treatment center on childhood immunization,[55] and Koopman and Longini[56] have illustrated the interaction between group-level and individual-level variables in the transmission of dengue.

Types of Group-Level Variables Used in Multilevel or Contextual Analysis

The rationale for incorporating group-level variables in multilevel or contextual analysis is that they provide information

that is not captured by individual-level data. For example, mean neighborhood income may provide information that is not captured by individual-level income. The mean income of a neighborhood may be a marker for neighborhood-level factors potentially related to health (such as recreational facilities, school quality, road conditions, environmental conditions, and the types of foods that are available), and these factors may affect everyone in the community regardless of individual-level income. Similarly, community unemployment levels may affect all individuals living within a community, regardless of whether or not they are unemployed.

The conceptual distinction between group-level and individual-level variables becomes important when (1) the variable has analogues at both levels but both measure different constructs (as in the preceding examples) or (2) the variable can be measured only at the group level because it is defined only at the group level (e.g., existence of certain types of regulations, population density, degree of income inequality in a community, political regime, legal status of women).

Variables that reflect the characteristics of groups have been classified into two basic types[14,19,21,25,26,57]: derived variables and integral variables. Derived variables (also termed analytical or aggregate variables) summarize the characteristics of individuals in the group (means, proportions, or measures of dispersion; for example, percentage of persons with incomplete high school educations, median household income, standard deviation of the income distribution). Susser has also referred to derived variables as contextual variables.[15] Integral variables (also termed primary or global variables) describe characteristics of the group that are not derived from characteristics of its members (e.g., the existence of certain types of regulations, availability of health care, political systems, or population density). Integral variables do not have analogues at the individual level. A special subset within derived variables is the average of the dependent variable within the group. Susser has referred to the contextual effect of grouped dependent variables on the outcome as "contagion."[15] This special use of derived variables has been shown to be particularly relevant in infectious disease but may also be important in the study of other health outcomes. For example, the prevalence rate of disease in a group may affect the probability that a given individual acquires the infection and may modify the effects of individual-level variables on the risk of infection.[15,58,59] Likewise, the probability of adopting a certain behavior (Y_{ij}) may depend in part on the degree to which the behavior has already been adopted in the community ($\overline{Y}_j$).[21]

Although derived and integral variables are sometimes presented as conceptually distinct, they are closely interrelated. Derived variables often operate by shaping certain integral properties of the group. For example, the composition of a group may influence the predominant types of interpersonal contacts, values, and norms or may shape organizations or regulations within the group that affect all members.[57] Essentially, whenever derived variables are used in contextual analysis, they are assumed to be capturing group properties that are more than summaries of individual properties.

There is no direct correspondence between so-called "environmental variables" (a term used to refer to many different types of variables) and group-level variables. The "social environment," for example, may have expressions at both the group and individual levels. A researcher investigating how community participation (a characteristic of the social environment) is related to health outcomes could measure (1) the number of community organizations to which each individual belongs (an individual-level variable), (2) the percentage of persons in the community who belong to at least 1 organization (a group-level derived variable), and (3) the number of organizations in the community (a group-level integral variable). Each variable would be measuring a slightly different aspect of the concept of community participation. Other environmental variables, such as the availability of public recreation spaces, are group-level integral variables.

The categorization of other group-level measures of the physical or chemical environment (e.g., air pollution or hours of sunlight exposure) is more complex. It has been argued that group-level measures of physical or chemical exposure should be conceptualized as a separate category distinct from both derived and integral variables.[25] These variables are not derived by aggregating the characteristics of individuals, but they do have group-level and individual-level analogues (e.g. days of sunlight in the community and individual-level sunlight exposure information).[25] However, group-level measures of physical or chemical exposures are generally used as proxies for individual-level exposures (which may be more difficult to measure for logistic or methodological reasons) rather than as indicators of a group-level property that is conceptually different from the analogous measure at the individual level. In this sense, they differ from the other group-level variables discussed here (derived and integral), which are presumed to provide information on conceptual group-level properties. Of course, physical or chemical exposures (measured at the individual level or the group level) may still have group-level determinants, such as the presence of a toxic waste dump in an area (an integral group-level variable) or the mean income or racial/ethnic composition of the neighborhood (a derived group-level variable). For example, the mean atmospheric levels of a certain pollutant in a neighborhood may be strongly determined by the presence of a particular industry in the area. In this case, neighborhood mean pollutant level can be thought of as a marker for an integral group-level variable (presence of the industry in the neighborhood).

Levels of Analysis Problems: Fallacies in Studies Involving Multiple Levels

Most of the literature regarding levels of analysis problems in epidemiology has focused on the fallacies of inferring relationships at 1 level based on data from another level. These fallacies arise when "the methods fail to fit the conceptual model" (i.e., the conceptual model being tested corresponds to 1 level but the data are from another level).[60] Because epidemiologists have (at least recently) been concerned with drawing inferences at the individual level, the ecological or aggregative fallacy has received much more attention than its counterpart, the atomistic fallacy. Both fallacies can be thought of as variations of the problem of drawing inferences at 1 level of aggregation based on data from another level. The ecological fallacy is the fallacy of drawing inferences at the individual level based on group-level data. Suppose a researcher finds that, at the country level, increasing per capita income is associated with increasing motor vehicle–related mortality. If the researcher infers that, within countries, increasing income is associated with increasing motor vehicle–related mortality, she or he may be committing the ecological fallacy because, within countries, motor vehicle–related mortality may always be higher in low-income than in high-income persons.

The atomistic fallacy is the fallacy of drawing inferences at the group level based on individual-level data. In some cases, group-level associations (rather than individual-level associations) may themselves be the topic of interest, and associations

observed at the individual level may not always hold at the group level. For example, even if, within countries, increasing individual-level income is associated with decreasing coronary heart disease mortality, at the country level increasing per capita income may be associated with increasing rather than decreasing coronary heart disease mortality rates. The use of individual-level data to infer group-level associations may thus lead to incorrect inferences. Of course, the degree to which associations measured at the group level are biased estimates of individual-level associations (and vice versa) differs according to the problem at hand. The many circumstances that may give rise to a discordance between associations measured at the ecological level and the individual level, and their mathematical explanations, have been abundantly discussed in public health journals.[34,61–63]

Ultimately, the inferential fallacies just discussed are methodological problems that can be overcome by ensuring that the data collected match the level at which inferences are to be made. In some cases, statistical methods can be used to partly reduce the bias that may result when associations at 1 level are inferred to another.[62] A more substantial problem, however, is whether a study focused on a single level ignores information that is crucial to understanding the problem being investigated (i.e., whether the mechanisms operating at 1 level can be adequately understood without reference to other levels). As formulated by Riley,[60] although the level at which data are collected may fit the conceptual model being investigated, important facts pertaining to other levels may be ignored (i.e., the methods may fail to fit the facts).

Ignoring relevant group-level variables in a study of individual-level associations may lead to what Riley has termed the psychologistic fallacy, that is, assuming that individual-level outcomes can be explained exclusively in terms of individual-level characteristics. As a hypothetical example, a study based on individuals might find that immigrants are more likely to develop depression than natives. But suppose this is true only for immigrants living in communities where they represent a small minority. A researcher ignoring the contextual effect of community composition might attribute the higher overall rate in immigrants to the psychological effects of immigration per se or even to genetic factors, ignoring the importance of community-level factors and thus committing the psychologistic fallacy.[57,60] ("Psychologistic fallacy" is not the most appropriate term, because the individual-level factors used to explain the outcome need not only be psychological. Other authors have used the term "individualistic fallacy" [57]; however, because the term has also been used as a synonym of the term "atomistic fallacy," [22,64] it is avoided here.)

Analogously, ignoring the role of individual-level factors in a study of groups may lead to what has been termed the sociologistic fallacy.[60] Riley provides a public health–related example. Suppose a researcher finds that communities with higher rates of transient population have higher suicide rates and concludes that higher rates of transient population lead to social disorganization, breakdown of social networks, and increased risk of suicide among all community inhabitants. It is possible that individual-level data will show that most suicides occur in transient residents and that suicide rates in transient residents and permanent residents are constant across communities. If this is the case, the researcher would be committing the sociologistic fallacy in attributing the higher suicide rates to social disorganization affecting all community members rather than to differences across communities in the percentage of transient residents, who may be more likely to commit suicide as a result of individual-level factors. Both the psychologistic and the sociologistic fallacies can be thought of as special types of confounding: relevant variables pertaining to other levels have been excluded from the model. By incorporating multiple levels of determination in the study of individual outcomes, multilevel analysis allows for the effects of macro- and micro-level variables as well as their interactions, avoiding the psychologistic as well as the sociologistic fallacy.

The two groups of fallacies (ecological and atomistic vs sociologistic and psychologistic) are not unrelated. When a group-level variable is related to the outcome independently of the same variable measured at the individual level, or when the group-level variable modifies the effects of the individual-level variable on the outcome (i.e., when there are contextual effects), ecological regression coefficients will differ from the corresponding individual-level coefficients.[62,65,66] This occurs because the individual- and the group-level effects of the variable are confounded in the ecological regression coefficient. As illustrated by Greenland and Robins,[62] the existence of contextual effects is one (but not the only) source of the ecological fallacy. The different types of fallacies are schematically summarized in Table 2.

TABLE 2—Types of Fallacies

Unit of Analysis	Level of Inference	Type of Fallacy
Group	Individual	Ecological
Individual	Group	Atomistic[a]
Individual: relevant group-level variables excluded	Individual	Psychologistic[a]
Group: relevant individual-level variables excluded	Group	Sociologistic

[a]Also called individualistic by some authors.

Additional Issues Relevant to Multilevel or Contextual Analysis

One of the main criticisms leveled at contextual analysis is related to what has been referred to as "misspecification of the model at the individual level": the observed "contextual or group effect" may be due to the omission of individual-level variables related to the outcome and to the group characteristic investigated.[21,67,68] This is essentially a problem of confounding (or residual confounding) that is common to epidemiologic studies in general, including those based exclusively on individual-level data. For example, suppose that neighborhood violence level (measured by mean number of violent crimes committed in the neighborhood each year) is found to be associated with increased risk of hypertension after adjustment for age and gender. These results may be interpreted as suggesting that neighborhood violence, possibly through its effects on the stress levels experienced by individuals, is related to the development of hypertension. On the other hand, it is also possible that relevant individual-level variables have been excluded from the model and that the observed neighborhood effects are due to confounding. If persons living in more violent neighborhoods tend to be of low income and low-income persons have an increased risk of hypertension (as a result of income-related differences in diet, obesity, or other factors), the observed neighborhood effect will disappear when individual-level income is included in the model.

There has been considerable debate on the issue of confounding by individual-level variables in multilevel analyses.[67–71] In cases in which the excluded individual-level variable is related to the outcome and is not an intermediary variable in the causal pathway linking group-level characteristics to disease, its exclusion will clearly lead to a confounded estimate of the group-level effect. However, if the individual-level variable is in the causal pathway linking group-level variables to the outcome, unadjusted group-level effects may be more appropriate estimates of group effects than adjusted estimates. Obviously, if it affects health, "context" must somehow "get into the body," and it ultimately operates through individual-level processes; thus, if one were to include enough individual-level variables in the regression models, the group-level effect would disappear. However, this does not imply that individual-level variables, rather than group-level variables, are the "true" cause of the outcome. An analogous problem arises in studies of individuals when intermediary variables in the causal pathway leading from exposure to disease are included in regression models. Ultimately, the issue of deciding whether a given variable is an independent cause or is an intermediate step in a causal chain involving less proximate variables cannot be resolved by statistical methods alone. It needs to be addressed by specifically studying the pathways involved.

In addition to confounding by individual-level variables, multilevel analysis incorporates yet another dimension for potential confounding: confounding may also occur at the group level if relevant group-level variables have been omitted from the model.[21] In the preceding example, increased neighborhood violence may be associated with another group-level variable, such as weaker and less cohesive social networks, which may be the relevant group-level factor actually related to hypertension in individuals.

As in the case of confounding, the problem of multicollinearity between variables is not specific to multilevel analysis and is often present in studies limited to individual-level data. In multilevel analysis, multicollinearilty may also exist between group-level variables and between group-level and individual-level variables. In some cases, correlations between group-level variables (e.g., mean neighborhood income and neighborhood unemployment) may be substantial. In small and homogeneous groups, derived variables (constructed by aggregating the characteristics of individuals) may be highly correlated with their individual-level analogues (e.g., mean neighborhood income and individual-level income).[21] If two variables are strongly collinear, it may be extremely difficult (and perhaps not very meaningful) to tease apart their independent effects. Of course, the degree to which these different types of multicollinearity are present may vary greatly depending on the research question.

Two additional issues that arise in contextual analysis are the selection of the appropriate contextual unit and the selection of the contextual variables to be included in the analyses. Both issues are largely dependent on the specific research question being investigated and on the theoretical model underlying the research. Selecting the appropriate contextual unit may be problematic because "contexts" are often not clearly defined and may have imprecise limits. For example, if the neighborhood is the contextual unit of interest, researchers need to specify exactly what constitutes a neighborhood in operational terms. The boundaries of a neighborhood, as perceived by its inhabitants, may not always coincide with the geographical units (e.g., census tracts or block groups) for which data are available. Of course, many "contexts" relevant to public health research (such as workplaces or schools) may not be based on geographically defined area of residence, and still others (such as professional organizations or extended families) may not be geographically defined at all. In addition, individuals may form part of a variety of nested or overlapping contexts,[21] and teasing apart their independent effects may be extremely difficult. For example, if neighborhoods are strongly segregated on the basis of people's jobs, the effects of the neighborhood context and the work context may be confounded and difficult to isolate. Selecting the appropriate contextual variables requires identifying the construct that one wishes to measure and deciding what variables will be used to measure it. Often, even when the relevant group-level constructs can be identified, the variables needed to measure them may not be available, and indirect indicators must be used.

An important methodological issue that may arise in some studies including data from multiple levels is the possibility that individual-level outcomes within groups may be correlated, even after group- and individual-level variables are taken into account. For example, in a study of the effects of school-level factors on adolescent drug use in which 20 individuals per school are sampled in 100 schools, the outcomes for adolescents within a school may be correlated even after individual- and group-level variables are taken into account (because children within a school may share other school-level variables not accounted for in the analyses). This residual correlation violates the assumption of independent observations that underlies usual modeling strategies. Several different strategies have been developed to deal with this issue, which is analogous to the problem that arises in longitudinal data analysis when repeated measures are taken on an individual over time. They include hierarchical linear or random effects models [23,28,53,72–76] and marginal or population-average models.[77] The statistical details and assumptions, as well as the situations under which inferences from these models will differ from standard regression models, have been extensively discussed in the literature.

Conclusion

Perhaps the most challenging aspect of multilevel analysis is that it requires a theory of causation that integrates micro- and macro-level variables and explains these relationships and interactions across levels. How are group-level variables presumed to operate? How do individuals interact with their contexts? These new models may need to transcend the rather simplistic notion of causation that underlies much epidemiologic research,[11] incorporating other levels of determination such as structural or holistic determination.[78] As defined by Bunge, structural determination refers to "the process by which the behavior of an individual (a molecule in a fluid, a person in a social group) is determined by the overall structure of the collection to which it belongs."[78(p19)] A similar notion has been alluded to by Loomis and Wing, who referred to the need to think of "cause not as a property of agents, but one of systems in which the population phenomena of health and disease occur," [79(p2)] and to conceive populations as organized groups with relational properties rather than mere aggregates of individuals.[79] The complexity of developing theoretical formulations that relate multiple levels is an important difficulty in multilevel analyses,[21,22] but this complexity is likely to be a better reflection of reality than the simpler multicausal models prevalent today.

Multilevel analysis is no simple task. It raises numerous theoretical and methodological problems that have yet to be resolved. In addition, it should be emphasized that the inclusion of contextual variables is not indispensable in research on the social determinants of health. In fact, many

hypotheses regarding social influences on health can and should be tested using individual-level data. However, because the processes of disease causation extend across levels and are likely to involve the interaction of individual-level and macrolevel variables,[14,79] it seems appropriate to begin to develop models of disease causation that integrate these levels and to include both types of variables in epidemiologic studies.

Paradoxically, epidemiology, the study of disease in populations, has largely been reduced to the study of individual-level risk factors for disease. Multilevel analysis is one way to begin to restore a population or societal dimension to epidemiologic research (i.e., the idea that factors operating at the levels of groups or societies affect the health of individuals within them). It challenges epidemiologists to develop models of disease causation that integrate macro- and micro-level determinants. Together with other analytic strategies and study designs, multilevel analysis may contribute to strengthening and revitalizing research into the social and collective determinants of health. □

Acknowledgments

I would like to thank Carles Muntaner, Javier Nieto, and Jose Tapia for their many contributions to the ideas developed in this manuscript.

References

1. Cassel J. Social science theory as a source of hypotheses in epidemiological research. *Am J Public Health.* 1964;54:1482–1488.
2. Susser M, Susser E. Choosing a future for epidemiology: I. eras and paradigms. *Am J Public Health.* 1996;86:668–673.
3. Pearce N. Traditional epidemiology, modern epidemiology, and public health. *Am J Public Health.* 1996;86:678–683.
4. Catalano R. Paradigm succession in the study of public health. In: *Health, Behavior, and the Community.* New York, NY: Pergamon Press; 1979:87–137.
5. Brockington CF. The history of public health. In: Hobson W, ed. *Theory and Practice of Public Health.* 5th ed. London, England: Oxford University Press Inc; 1979:1–8.
6. Taylor R, Rieger A. Medicine as social science: Rudolph Virchow on the typhus epidemic in upper Silesia. *Int J Health Serv.* 1985;15:547–559.
7. Fee E. The origins and development of public health in the United States. In: Holland W, Deteb R, Knox G, eds. *The Oxford Textbook of Public Health.* 2nd ed. London, England: Oxford University Press Inc; 1991;1:3–22.
8. Kennedy DA. Community health and the urban environment. In: Hinkle LE, Loring WC, eds. *The Effect of the Man Made Environment on Health and Behavior.* Atlanta, GA: Centers for Disease Control, Public Health Service; 1977:7–44. DHEW Publication CDC 77-8318.
9. Loring WC. Introduction. In: Hinkle LE, Loring WC, eds. *The Effect of the Man Made Environment on Health and Behavior.* Atlanta, Ga: Centers for Disease Control, Public Health Service; 1979:vii–xxxvi. DHEW Publication CDC 77-8318.
10. Dodge DL, Martin WT. The root of the probem: changing causes of death. In: *Social Stress and Chronic Illness.* Notre Dame, Ind: University of Notre Dame Press; 1970:3–25.
11. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med.* 1994;39:887–903.
12. Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med.* 1996;42:817–830.
13. Lukes S. Methodological individualism reconsidered. In: Emmet D, Macintyre A, eds. *Sociological Theory and Philosophical Analysis.* New York, NY: Macmillan Publishing Co Inc; 1970:76–88.
14. Von Korff M, Koepsell T, Curry S, Diehr P. Multi-level research in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol.* 1992;135:1077–1082.
15. Susser M. The logic in ecological: I. the logic of analysis. *Am J Public Health.* 1994;84: 825–829.
16. Krieger N, Rowley DL, Herman A, et al. Racism, sexism, and social class: implications for studies of health, disease, and well-being. *Am J Prev Med.* 1993;9(suppl 6):82–122.
17. Jones K, Duncan C. Individuals and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework. *Health Place.* 1995;1:27–40.
18. Blalock HM, Wilken PH. *Intergroup Processes: A Micro-Macro Perspective.* New York, NY: Free Press; 1979.
19. Dogan M, Rokkam S. Introduction. In: Dogan M, Rokkam S, eds. *Social Ecology.* Boston, Mass: MIT Press; 1969:1–15.
20. Iversen G. *Contextual Analysis.* Newbury Park, Calif: Sage Publications; 1991.
21. Blalock HM. Contextual-effects models: theoretical and methodological issues. *Annu Rev Sociol.* 1984;10:353–372.
22. Scheuch EK. Social context and individual behavior. In: Dogan M, Rokkam S, eds. *Social Ecology.* Boston, Mass: MIT Press; 1969: 133–155.
23. Hermalin AI. The multilevel approach: theory and concepts. In: *The Methodology for Measuring the Impact of Family Planning Programmes on Fertility.* New York, NY: United Nations; 1986:15–31. Population Studies 66. Addendum Manual IX.
24. Hox J, Kreft I. Multilevel analysis methods. *Sociol Methods Res.* 1994;22:283–299.
25. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health.* 1995;16:61–81.
26. Lazarsfeld PF, Menzel H. On the relation between individual and collective properties. In: Etzioni A, ed. *A Sociological Reader on Complex Organizations.* New York, NY: Holt, Rinehart & Winston Inc; 1971:499–516.
27. Przeworski A. Contextual models of political behavior. *Political Methodology.* 1974;1:27–61.
28. Garner C, Raudenbush S. Neighborhood effects on educational attainment: a multilevel analysis. *Sociol Educ.* 1991;64:251–262.
29. Robson BT. *Urban Analysis: A Study of City Structure with Special Reference to Sutherland.* Cambridge, England: Cambridge University Press; 1969.
30. Fernandez RM, Kulik JC. A multilevel model of life satisfaction: effects of individual characteristics and neighborhood composition. *Am Sociol Rev.* 1981;46:840–850.
31. Morgenstern H. Socioeconomic factors: concepts, measurement, and health effects. In: *Measuring Psychosocial Variables in Epidemiologic Studies of Cardiovascular Disease: Proceedings of a Workshop.* Washington, DC: National Institutes of Health; 1985:3–35. NIH Publication 85-2270.
32. Wing S, Casper M, Riggan W, et al. Socioenvironmental characteristics associated with the onset of decline of ischemic heart disease mortality in the United States. *Am J Public Health.* 1988;78:923–926.
33. Wing S, Barnett E, Casper M, Tyroler HA. Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women. *Am J Public Health* . 1992;82:204–209.
34. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and its consequences. *Am J Public Health.* 1994;84: 819–824.
35. Susser M, Susser E. Choosing a future for epidemiology: II. from black box to Chinese boxes and ecoepidemiology. *Am J Public Health.* 1996;86:674–677.
36. Goldberger J, Wheeler GA, Sydenstrycker E. A study of the relation of family income and other economic factors to pellagra incidence in seven cotton mill villages of South Carolina in 1916. *Public Health Rep.* 1920;35: 2673–2714.
37. Macintyre S, Maciver S, Sooman A. Area, class and health: should we be focusing on places or people? *J Soc Polit.* 1993;22:213–234.
38. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation.* 1993;88:1973–1998.
39. Carstairs V, Morris R. Deprivation and mortality: an alternative to social class? *Community Med.* 1989;11:210–219.
40. Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland, England and Wales. *BMJ.* 1989;299:886–889.
41. Morgan M, Chinn S. ACORN group, social class and child health. *J Epidemiol Community Health.* 1983;37:196–203.
42. Fox AJ, Jones DR, Goldblatt PO. Approaches to studying the effect of socioeconomic circumstances on geographic differences in mortality in England and Wales. *Br Med Bull.* 1984;4:309–314.
43. Haan M, Kaplan G, Camacho T. Poverty and health: prospective evidence from the Alameda County Study. *Am J Epidemiol.* 1987;125:989–998.
44. Hochstim JR. Poverty area under the microscope. *Am J Public Health.* 1968;58: 1823–1827.
45. Anderson RT, Sorlie P, Backlund E, Johnson N, Kaplan GA. Mortality effects of community socioeconomic status. *Epidemiology.* 1997;8:42–47.
46. Humphreys K, Carr-Hill R. Area variations in

health outcomes: artefact or ecology. *Int J Epidemiol.* 1991;20:251–258.
47. Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. *J Epidemiol Community Health.* 1991;45:35–42.
48. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health.* 1992;82: 703–710.
49. Diehr P, Koepsell T, Cheadle A, Psaty BM, Wagner E, Curry S. Do communities differ in health behaviors? *J Clin Epidemiol.* 1993;46: 1141–1149.
50. Duncan C, Jones K, Moon G. Do places matter?: a multi-level analysis of regional variation in health-related behavior in Britain. *Soc Sci Med.* 1993;37:725–733.
51. Duncan C, Jones K, Moon G. Psychiatric morbidity: a multilevel approach to regional variations in the UK. *J Epidemiol Community Health.* 1995;49:290–295.
52. O'Campo P, Gielen AC, Faden RR, Xue X, Kass N, Wang M. Violence by male partners against women during the childbearing year: a contextual analysis. *Am J Public Health.* 1995;85:1092–1097.
53. Entwisle B, Mason WM. Multilevel effects of socioeconomic development and family planning programs on children ever born. *Am J Sociol.* 1985;91:616–649.
54. Dargent-Molina P, James SA, Strogatz DS, Savitz DA. Association between maternal education and infant diarrhea in different households and community environments of Cebu, Philippines. *Soc Sci Med.* 1994;38:343–350.
55. Jones K, Moon G, Clegg A . Ecological and individual effects in childhood immunisation uptake: a multilevel approach. *Soc Sci Med.* 1991;33:501–508.
56. Koopman JS, Longini IM. The ecological effects of individual exposures and nonlinear disease dynamics in populations. *Am J Public Health.* 1994;84:836–842.
57. Valkonen T. Individual and structural effects in ecological research. In: Dogan M, Rokkam S, eds. *Social Ecology.* Boston, Mass: MIT Press; 1969:53–68.
58. Koopman JS, Longini IM, Jacquez JA, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol.* 1991;133:1199–1209.
59. Halloran ME, Struchiner CJ. Study design for dependent happenings. *Epidemiology.* 1991;2:331–338.
60. Riley MW. Special problems of sociological analysis. In: *Sociological Research I: A Case Approach.* New York, NY: Harcourt, Brace & World Inc; 1963:700–725.
61. Piantadosi S, Byar DP, Green SN. The ecological fallacy. *Am J Epidemiol.* 1988;127: 893–904.
62. Greenland S, Robins S. Ecologic studies—biases, misconceptions, and counter-examples. *Am J Epidemiol.* 1994;139:747–760.
63. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol.* 1989;18:269–274.
64. Alker HR. A typology of ecological fallacies. In: Dogan M, Rokkam S, eds. *Social Ecology.* Boston, Mass: MIT Press; 1969:69–86.
65. Hammond JL. Two sources of error in ecological correlations. *Am Sociol Rev.* 1973;38: 764–777.
66. Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am Sociol Rev.* 1978;43:557–572.
67. Hauser RM. Contextual analysis revisited. *Sociol Methods Res.* 1974;2:365–375.
68. Hauser RM. Context and consex: a cautionary tale. *Am J Sociol.* 1970;75:645–664.
69. Hauser RM. Reply to Allen Barton. *Am J Sociol.* 1970;76:517–520.
70. Barton A. Comments on Hauser's "Context and consex." *Am J Sociol.* 1970;76:514–517.
71. Farkas G. Specification, residuals, and contextual effects. *Sociol Methods Res.* 1974;2: 333–363.
72. Bryk A, Raudenbush S. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, Calif: Sage Publications; 1992.
73. Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. *J Am Stat Assoc.* 1985;80:513–524.
74. Wong GY, Mason WM. Contextually specific effects and other generalizations of the hierarchical linear model for comparative analysis. *J Am Stat Assoc.* 1991;86:487–503.
75. Entwisle B, Mason WM, Hermali HI. The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. *Demography.* 1986; 23:199–216.
76. Mason WM, Wong GW, Entwisle B. Contextual analysis through the multilevel linear model. In: Leinhardt S, ed. *Sociological Methodology 1983–1984.* San Francisco, Calif: Jossey-Bass; 1983:72–103.
77. Zeger S, Liang KY, Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44:1049–1060.
78. Bunge M. Causation and determination, causalism and determinism. In: *Causality in Modern Science.* New York, NY: Dover Publications; 1979:3–30.
79. Loomis D, Wing S. Is molecular epidemiology a germ theory for the end of the twentieth century? *Int J Epidemiol.* 1990;19:1–3.