

The Study of Group-Level Factors in Epidemiology: Rethinking Variables, Study Designs, and Analytical Approaches

Ana V. Diez Roux

From the Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI.

Received for publication December 16, 2003; accepted for publication March 11, 2004.

A key notion that has received much attention in epidemiology over the past few years has been that not all disease determinants can be conceptualized as individual-level attributes, hence the need to consider features of the groups to which individuals belong when studying the causes of ill health. This has led epidemiologists and public health researchers to rethink the ideas on ecologic studies and ecologic variables traditionally espoused in epidemiology (1–6). This reconceptualization of ecologic or group-level variables has been manifested, for example, in recent interest and debate on the possible health effects of group-level constructs, such as income inequality (7, 8), social capital (9, 10), and neighborhood characteristics (11–14). In this context, the advent of the statistical technique of multilevel models has been viewed as especially promising because of its ability to incorporate both group-level and individual-level predictors in the study of health (4, 15–17).

The idea that factors beyond individuals, referred to as group-level, ecologic, macro-level, or population-level factors (1, 3, 5, 18, 19), are important to health is not new. Two well-known examples include the concept of herd immunity in infectious diseases and Rose's distinction between the causes of cases and the causes of incidence rates in chronic diseases. Herd immunity implies that a person's likelihood of contracting an infectious disease depends in part on the level of immunity in the population to which he or she belongs (20). In his seminal paper, "Sick Individuals and Sick Populations," Geoffrey Rose (18) discusses a related concept: the idea that studies that focus on what distinguishes sick individuals from healthy individuals *within* a population or group may miss important disease determinants. This is because population-level factors are invariant within a population and, hence, cannot be investigated in studies restricted to comparisons of individuals within a population (21). To detect these factors, researchers need studies that compare different populations (or groups) and investigate population-level (or group-level) factors.

Discussions of group-level and individual-level factors in epidemiology are sometimes interpreted as implying that population-level factors are important in understanding

between-population differences and that individual-level factors are important in understanding between-individual differences. A key point, however, is that factors at multiple levels may be important to understanding the causes of variability within a level. For example, both individual-level and group-level factors are important in understanding the causes of between-population differences in disease rates. Likewise, both population-level and individual-level factors are important in understanding the causes of disease in individuals. For example, herd immunity, a group-level property, is important in understanding not only the reasons for group differences in the incidence of disease but also an individual's probability of contracting the disease. Group-level or population-level factors, such as the mass production of foods, may be important in understanding not only between-country differences in rates of hypertension but also the causes of hypertension in an individual. Of course, only individual-level factors will explain interindividual differences in outcomes within groups (18, 21).

Although discussion of the importance of group-level or population-level factors has long been present in epidemiology, the interest in empirically testing for group effects in epidemiologic studies is relatively new. This interest has recently motivated many methodological discussions on the uses and misuses of ecologic variables and the strengths and limitations of multilevel models (1, 2, 4, 15, 22–24). This review will summarize selected issues related to the use of ecologic variables, ecologic studies, and multilevel studies in epidemiology. It will conclude with a discussion of the new challenges raised by the emerging multilevel paradigm. The focus will be on basic conceptual and methodological issues, rather than on specific empirical applications.

RECONSIDERING GROUP-LEVEL VARIABLES

Several different terms, including macro-level, ecologic, population-level, and group-level factors, have been used to refer to factors defined above the level of individuals (1, 3, 5, 18, 19, 25, 26). In the remainder of this paper, the terms "groups" and "group-level variables" will be used to refer

generically to higher-level units (units defined above the lowest level at which the outcome is measured) and the variables that characterize them. These groups can be families, friendship groups, neighborhoods, schools, states, countries, and so on. The term “individuals” will be used generically to refer to the lower-level units nested within the higher-level units. Individuals can be persons nested within neighborhoods, families, or countries. The lower-level units can also be smaller groups nested within larger groups (e.g., neighborhoods nested within states).

A possible use of group-level variables in epidemiology is as proxies for unavailable individual-level data. Thus, for example, mean area (neighborhood) income may be used as a proxy for unavailable individual-level income (27, 28). The greater the heterogeneity within the group, the greater the measurement error introduced by using the aggregate for the group as a proxy for the individual-level construct of interest. Sometimes, when the exposure of interest has large within-individual variability (such as diet), the aggregate measure may be used because it is believed to be a better measure of “true” individual-level exposure than a single individual-level measure (29). Recent discussions of the use of group-level variables have noted that, in many cases, the aggregate group-level measure may be tapping into a different construct than its individual-level namesake (1, 2, 17). For example, mean neighborhood income may be of interest per se, as a neighborhood-level attribute that may be related to health over and above the income of individuals. Thus, the limitations of using an aggregate measure as a proxy for its individual-level namesake relate not only to measurement error but also to construct validity, that is, whether it is indeed the same construct that is measured by both variables (2, 30).

The recognition that group-level variables may be measures of constructs that are distinct from the characteristics of individuals has prompted recent discussion of group-level variables in the epidemiologic literature (1–5, 17). Following work done in the social sciences in the 1960s and 1970s (31, 32), group-level variables have been classified into two basic types: derived variables and integral variables (1, 3, 5, 19, 26). Derived variables (1, 19, 32) are constructed by mathematically summarizing the characteristics of individuals in the group. Some derived variables have an individual-level namesake (e.g., mean neighborhood income and individual income), but others (such as standard deviation of the income distribution or the Gini coefficient) do not. The rationale for using derived group-level variables in the investigation of group effects is that they are providing information on true group-level constructs (i.e., that they are not simply summaries of individual-level constructs). Integral variables (1, 3, 19, 32) describe group characteristics that are not derived from characteristics of its members. In the case of derived variables with an individual-level namesake, there is sometimes ambiguity regarding what the group-level variable is actually measuring: Is it characterizing a true group-level construct or is it simply an aggregate of individual-level properties? On the other hand, derived variables with no individual-level analog (such as distributional measures) and integral variables are clearly characterizing group-level attributes per se, because these variables are not defined at

the individual level. Other typologies of group-level variables based on the type of domain being measured rather than on the form of the measurement have also been proposed (33).

In considering the distinction between group-level and individual-level variables, one must note that the level of conceptualization of the construct does not always match the level at which it is practically measured. For example, it is possible to obtain a measure of a group-level construct by asking individuals to report on features of the group to which they belong. Individuals can be asked about the neighborhood in which they live, and responses across individuals residing within a given neighborhood can be aggregated up to the neighborhood level (34). Thus, a measure of a neighborhood-level construct is obtained by combining individual responses. The distinction between group-level and individual-level constructs has often been glossed over in epidemiology, because it was generally assumed that all the relevant constructs are by definition individual level. However, in any research question, a priori specification of what the relevant constructs are, and at what levels they are defined, is key.

RETHINKING THE ECOLOGIC FALLACY

Recent discussions of the role of group-level factors in epidemiology have often been linked to debates on the merits and limitations of ecologic studies in studying the causes of ill health (2, 17, 22, 23). The recognition that factors at the levels of both groups and individuals may be relevant to health sheds new light on the understanding of the ecologic fallacy. Detailed discussions of the many reasons for the ecologic fallacy (for correlation and regression coefficients) (2, 19, 35–37) can be found elsewhere, and only selected issues especially relevant from the perspective of the multilevel determinants of health will be discussed here.

Common examples of the ecologic fallacy involve situations where inferences regarding the association between an individual-level exposure and an individual-level outcome are drawn on the basis of group-level associations between the corresponding aggregate (or derived) group-level exposure and disease rates or the mean outcome for members of the group. Several years ago, Firebaugh (38) pointed out that an important reason for the ecologic fallacy (in regression coefficients) is the presence of a contextual effect of the mean X for the group on the individual-level outcome, after accounting for individual-level X (or similarly the presence of an interaction term involving mean group X). As reviewed in detail elsewhere (36), when there is a contextual effect of mean group X on the individual-level outcome (Y) (i.e., when mean group X is associated with Y independently of individual-level X), the ecologic regression coefficient will not equal the within-group individual-level relation between X and Y or the pooled individual-level relation (ignoring group membership) between X and Y . A contextual effect of mean group X will exist when the individual-level variable (e.g., individual-level income) and its group-level analog (e.g., mean neighborhood income) are tapping into distinct constructs, and both are related to the individual-level

outcome or when mean group X is associated with omitted individual-level variables (which vary from group to group). A variant of this situation is when two individual-level factors interact in causing disease, and the joint distribution of both factors is associated with mean group X (4, 37).

Hammond (39) adds another reason for differences between ecologic and individual-level regression coefficients: grouping by the dependent variable. For example, if persons are grouped into neighborhoods on the basis of their income (due to social processes driving economic residential segregation) and we are interested in estimating the relation between race and income at the individual level, the ecologic regression coefficient between percent Black and mean neighborhood income would differ from the individual-level coefficient relating race to individual-level income because of the grouping process involved. Essentially, the grouping process generates a "group effect" analogous to the contextual effects of the mean group X described above.

The three sources of the ecologic fallacy described above all pertain to situations where there is some form of group effect. This includes situations where there is a failure to distinguish constructs at different levels (e.g., mean group X is assumed to measure the same thing and individual-level X), where something about the groups is associated with individual-level predictors of the outcomes (mean group X is associated with other individual-level factors related to Y), or where some social process results in the grouping of persons by the dependent variable. In common epidemiologic explanations of the ecologic fallacy, these group effects are a nuisance that makes it difficult for epidemiologists to draw inferences regarding individual-level associations based on group-level data. The recent interest in multilevel determinants has resulted in growing discussion among epidemiologists on the best study designs and analytical approaches to actually study these group health effects.

CONTRASTING STUDY DESIGNS

Recognizing the relevance of constructs defined at multiple levels to the health of individuals is useful in rethinking the advantages and disadvantages of studies with different units of analysis. Ecologic studies examine group-level variables as predictors of variability in group-level outcomes, such as disease rates. However, they are unable to investigate the contribution of individual factors to between-group differences. In the case of predictors that are derived variables, the ecologic analysis cannot distinguish the individual-level effect of the variable from its contextual effect. For example, a study relating mean neighborhood income to blood pressure levels could not differentiate whether differences across neighborhoods are due to the effects of individual-level income or to the contextual effects of mean neighborhood income. From a public health perspective, however, the ecologic association may itself be of interest. For example, a study may want to investigate the relation between the introduction of a mass media campaign to prevent teenage smoking and the prevalence of teenage smoking in the area. For public health purposes, it may be irrelevant whether the campaign operates through its individual-level effect (i.e., only individuals who see the adver-

tisements on television quit smoking) or through a contextual effect (the mass media campaign creates a climate conducive to quitting smoking which affects everyone regardless of whether they are individually exposed to the advertisements or not). In the absence of individual-level confounders, the ecologic association would be useful in drawing a causal inference regarding the effects of the mass media campaign.

Individual-level studies typically examine individual-level variables as predictors of variability in individual-level outcomes. Traditional individual-level studies are unable to investigate the role of group-level factors in explaining variability in the outcome across individuals. For example, a study of risk factors for depression could not examine the interactions between ethnicity and neighborhood ethnic composition if group-level data are not available. Individual-level studies that include individuals from different relevant groups and collect information on group-level properties can include group-level variables as predictors of individual-level outcomes in individual-level equations. These models have been known as contextual models in the social sciences (25, 40, 41). Special methods may be necessary to account for within-group correlations in individual-level outcomes that persist after individual-level and group-level factors are taken into account. Ignoring this correlation may lead to incorrect estimates of standard errors (42). Efficiency of estimation may also be reduced (42). One common approach to account for within-group correlations is to use marginal models (43), also referred to as "population-average models" (42) or "covariance pattern models" (44). Marginal or population-average models model the population-average response as a function of covariates without explicitly accounting for heterogeneity across groups (43). In contrast to the multilevel models described below, marginal models do not allow examination of group-to-group variability per se or of the factors associated with it. Neither do they allow decomposition of total variability in the individual-level outcome into within- and between-group components.

Multilevel studies are studies in which both groups and individuals are the units of analysis. These types of studies allow the simultaneous investigation of between-group and within-group variability in individual-level outcomes. Recent developments in the statistical technique of multilevel modeling (45–47) have stimulated interest in multilevel studies. Multilevel modeling is an analytical approach that is appropriate for data with nested sources of variability, that is, involving units at a lower level or "micro units" (e.g., individuals) nested within units at a higher level or "macro units" (e.g., groups such as schools or neighborhoods) (16, 45–50). Although the use of these models in multilevel studies is relatively new, these models (or variants of them) have many different applications and have previously appeared in different literatures under a variety of names, including random effects models or random coefficient models (42, 51, 52), covariance components models or variance components models (53, 54), and mixed models (44).

In multilevel studies, multilevel models allow the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes while accounting for the nonindependence of observations within

groups. They also allow the examination of both between-group and within-group variability, as well as how group-level and individual-level variables are related to between-group, within-group, and total interindividual variability in the outcome. Thus, multilevel models can be used to draw inferences regarding the causes of interindividual variation (or the relation of group-level and individual-level variables to individual-level outcomes), but inferences can also be made regarding intergroup variation, whether it exists in the data, and to what extent it is accounted for by group-level and individual-level characteristics. Groups are not treated as unrelated but are conceived as coming from a larger population of groups about which inferences may be made. Multilevel models thus allow researchers to deal with the microlevel of individuals and the macrolevel of groups or contexts simultaneously (16). Multilevel models can also be used in situations involving multiple nested contexts (45, 46) (e.g., multiple measures over time on individuals nested within neighborhoods), as well as overlapping or cross-classified contexts (e.g., children nested within neighborhoods and schools) (55). Reviews of multilevel modeling, with specific focus on public health applications, have been published over the past few years (15, 16, 24).

In contrast to the marginal models described above, in multilevel models heterogeneity across groups is explicitly modeled (45). Multilevel models investigate and explain the source of group-to-group variation (and of the within-group correlation) by modeling group-specific regression coefficients as a function of group-level variables plus random variation. These differences between multilevel and marginal models have consequences for the interpretation of regression coefficients: In the multilevel model, the regression coefficient estimates how the response changes as a function of covariates *conditional* on group-specific random effects or random coefficients; in the marginal model, the coefficient expresses how the response changes as a function of covariates “averaged” over group-to-group heterogeneity (or group random effects) (42, 43).

Although the possibility of multilevel studies has been greeted with great enthusiasm, causal inference in a multilevel context raises numerous challenges. Selected issues related to the incorporation of group-level and individual-level factors in epidemiology are summarized below.

DESIGN AND SAMPLE SIZE

To date, the majority of multilevel studies have been cross-sectional (12, 34, 56, 57). The longitudinal studies that exist have generally investigated features of groups measured at a single point in time in relation to changes in health over time or incidence of disease or death (58–61). Studies that follow both groups and persons over time to examine if changes in group-level factor cause changes in individual-level outcomes are extremely rare. Careful consideration of the time lags to be expected between exposure to a certain group-level factor and its health consequences is crucial in these studies (57). Better attention to the time dimension in multilevel studies is likely to improve the ability to draw causal inferences regarding group-level effects.

Sample size and power calculations in multilevel studies are complex and remain an area of active research (47, 62, 63). In general, the power for estimating the individual-level regression coefficients depends on the total sample size (63). The power for higher-level (group-level) effects and cross-level interactions (interactions between group-level and individual-level variables) depends more strongly on the number of groups than on the total sample size (63). However, the power to estimate the ratio of between-group to total variability (the intraclass correlation coefficient) is affected by the number of groups and the number of persons per group in a different manner than the power to detect associations of group-level variables with individual-level outcomes (the fixed effects of group properties) (47). Snijders and Bosker (47) show that, for a fixed total sample size, the standard error of the association between a group-level variable and an individual-level outcome (e.g., the association between neighborhood availability of recreational spaces and the physical activity of individuals) may be minimized by sampling many groups with relatively few observations per group. On the other hand, for relatively low intraclass correlations (common in epidemiologic and social science research), small group sizes may result in large standard errors for the intraclass correlation coefficient estimated (47). Thus, a given study may have insufficient power to detect between-group variance and yet have sufficient power to detect the fixed effect of a specific group-level attribute. Power and sample size calculations need to specify the key multilevel parameters of interest, and tradeoffs may be involved.

GROUPS AND GROUP-LEVEL VARIABLES

The “groups” relevant to a specific health outcome may be difficult to define (e.g., neighborhoods) or have fuzzy and changing boundaries (e.g., friendship groups). Data are often unavailable for the theoretically relevant group of interest so a crude proxy is used (e.g., census tracts for neighborhoods) (34, 56). This results in substantial misspecification of the group and the group-level construct of interest. Whereas epidemiology has become very sophisticated at measuring individual-level attributes, the measurement of group-level attributes remains in its infancy. In some cases, the measurement of group-level constructs may be very simple (e.g., the presence of a certain law), but in others (e.g., social capital, the structure of social networks, or features of neighborhoods related to physical activity or stress), it is not. Recent methodological developments in the measurement of group-level constructs (some of which have been termed “ecometrics” (64)) are a welcome sign. For example, Raudenbush and Sampson (64) have proposed statistical methods to assess the reliability and validity of measures of group-level constructs created by combining information obtained from several observers or survey respondents per group. This approach allows assessment of the agreement among respondents within a group, the reliability of the group-level measure for discriminating between groups, and the construct validity of the group-level measure (by relating the measure obtained to other sources of data) (65). Other measures of group-level constructs may involve approaches

that do not necessarily involve aggregation of individual measures (e.g., the structure of connections between individuals within a group or the use of geographic information systems to develop measures of neighborhood availability and accessibility of resources). As the study of group-level factors becomes more common, other approaches to measuring attributes of groups are likely to emerge.

SELECTING THE RELEVANT GROUPS OR "LEVELS"

A crucial point in investigating multilevel determinants of health is selecting the relevant levels for analysis. The methodological issues discussed above with respect to the study of individuals nested within groups (e.g., persons nested within neighborhoods) apply across a continuum of nested levels (e.g., neighborhoods nested within regions and regions nested within countries). Thus, it is possible to envision populations (e.g., of individuals, of neighborhoods, of states) at each level. Multilevel studies can be used to draw inferences about variability at different levels and also to investigate how factors at multiple levels affect outcomes.

A multiplicity of different nested (or nonnested) groups or levels may be relevant for a particular research question. Specifying the relevant levels is part of the development of the theory that should precede the data collection and statistical analysis. An important methodological complexity is that the variance apportioned to a given level in multilevel models may be over- or underestimated if a relevant level is ignored in the analysis (24). In addition, misspecification of the relevant level may result in incorrectly concluding that groups (or higher-level) effects are absent. For example, if the research question pertains to the impact of availability of healthy foods on diet, and neighborhoods are specified as the higher-level unit for which food availability is measured, the absence of an effect of neighborhood food availability could be entirely consistent with a large effect of country-level food availability. In this case, the absence of the neighborhood effect could result from low variability in healthy food availability across neighborhoods within the country. If this is the case, neighborhood food availability would not be detected as an important predictor of individual-level diet. The failure to include the country level in the analyses would lead the researcher to miss the country-level food availability effect (this situation is directly analogous to the inability of studies restricted to individuals from a single group to detect group effects (21)).

DISTAL AND PROXIMAL FACTORS

Because group-level factors must ultimately affect individuals in order to influence health, their effects must necessarily be mediated through more proximate individual-level processes. At the same time, some individual-level factors may be confounders of group-level effects either because individuals are selected into groups based on their individual-level attributes (e.g., persons with low income are selected into disadvantaged neighborhoods) or because individual-level factors and group-level factors are associated for other reasons (e.g., persons living in countries characterized by mass production of processed foods may also be less

physically active). Indeed, much of the effort in the estimation of group-level effects in multilevel studies goes into controlling appropriately for individual-level confounders of group effects (12, 16). Residual confounding by mismeasured or unmeasured individual-level variables has long been a critique of studies of group effects (11, 34, 66). On the other hand, many of these individual-level factors may be mediators of group effects, raising questions regarding whether or not group-level effects should be adjusted for these factors (57). Even more complex situations may arise when a factor is both a mediator and a confounder of the higher-level effect. For example, neighborhood availability of healthy foods may affect the cardiovascular health of individuals through its influence on the diet of individuals (diet is a mediator of neighborhood effects on cardiovascular disease). On the other hand, living in a neighborhood with poor availability of healthy foods and individual diet may be associated simply because they have a common antecedent (individual-level income). Thus, individual-level diet may be both a mediator and a confounder of neighborhood effects on cardiovascular risk.

It has been noted that the use of multiple regression approaches to partition indirect and direct effects (e.g., the portion of a group effect that is mediated through a given variable and the portion that is not) may lead to incorrect conclusions regarding the presence and strength of direct effects (67, 68). The extent to which the approach of estimating a direct effect by comparing a group-level effect before and after adjusting for a mediator results in substantial bias in real life situations (as opposed to hypothetical examples) and remains to be fully determined (69). It is likely to vary from research problem to research problem, depending on the extent to which adjustment for the mediator actually introduces substantial confounding by other unmeasured variables related to the mediator and the outcome. When individual-level variables may be both confounders and mediators of the effect of interest, special estimation procedures may be necessary to correctly estimate the effect (70, 71). The extension of these emerging methods to a multilevel data structure is likely to be quite complex. Estimating "direct" effects in a multilevel context may be rendered even more complicated by the fact that a given variable may mediate both group-level and individual-level effects; that is, a "mediating" variable may be the common effect of variables at both the individual level and the group level.

Issues related to the limitations of investigating complex causal chains using multiple regression methods do not apply only to studies involving group-level or even social factors. In fact, distal and proximal factors (as well as complex interactions between factors) will be present even when the focus is on a limited and more proximate section of the full causal process (e.g., the link at the individual level between physical activity and the development of atherosclerosis). It has been argued that different analytical approaches, such as systems-based approaches, may be necessary to understand complex causal processes like these. Interestingly, the call for systems-based approaches has been made by those interested in the more distal social determinants of health (72, 73) and by those interested in under-

standing the more proximal biologic processes leading to disease (74, 75). Although intellectually appealing, the empirical application of systems approaches to specific empirical questions in public health remains a challenge.

OBSERVATIONAL STUDIES AND CAUSAL INFERENCE

Observational multilevel studies face the same problems as other observational studies in estimating causal effects from observational data. The ability to draw causal inferences is based on the extent to which the methods used approximate the counterfactual comparison of interest. One important limitation of past work in this regard especially prevalent in research on neighborhood health effects has been the reliance on group-level derived variables (e.g., neighborhood mean income) as proxies for the relevant integral group-level variable of interest (12, 33, 34). This has limited the extent to which the data available allow researchers to approximate the counterfactual contrast of interest, even within the limitations of observational studies. Better specification of the group-level factors of interest (e.g., moving from crude proxies to specific neighborhood-level attributes) and the testing of specific hypotheses will improve the ability to draw causal inferences.

The extent to which group-level effects can be validly estimated through the use of multiple regression methods (including multilevel models) to control for individual-level confounders has been questioned (11). The adjusted comparison requires assumptions regarding the effects of the individual-level variable on the outcome across groups, and it may involve extrapolations beyond the support in the data if there is little overlap in the distribution of the individual-level variable across groups (e.g., individual-level income across levels of neighborhood disadvantage). The extent to which this is a problem is an empirical question and may vary from research problem to research problem. This is no different from similar situations involved in adjusted comparisons in individual-level studies. The use of propensity score (76) matching has recently been proposed as an alternative to traditional regression approaches in estimating neighborhood effects (one type of group-level effect) (77). One of the advantages of this approach is that it restricts comparisons to individuals matched on propensity scores and hence comparable on individual-level variables. This eliminates problems associated with extrapolating beyond the support in the data. However, it implies restricting analyses to a subset of the sample that may be different from the full sample, and it does not resolve the problem of mismeasured or omitted individual-level confounders. In addition, matching on variables that are the *consequence* of the group-level variable creates problems analogous to those of adjusting for a mediator in regression approaches (77). In a promising application of propensity score matching coupled with sensitivity analyses of results to unmeasured confounders, Harding (77) has recently demonstrated that neighborhood effects on teenage pregnancy and dropping out are unlikely to result from confounding by individual-level variables.

Another complexity of observational studies of group effects is that certain group-level properties may be at least partly endogenous to the characteristics of the individuals that

make up the group (11, 24). This makes the identification of these group-level effects from observational data problematic. The extent to which group-level properties are endogenous to individual-level properties is likely to vary for different group-level constructs and different research questions (13, 14). Endogeneity may appear more of a problem in the case of derived group-level variables (e.g., mean neighborhood income) that are constructed by aggregating the characteristics of individuals within a group. However, as noted above, these variables are often used as proxies for a more clearly exogenous integral group-level property. Endogeneity is also a possibility for some integral group-level variables (e.g., dietary habits of residents may influence neighborhood availability of healthy foods). However, it is unlikely that all group-level attributes are fully endogenous to the individual characteristics of persons of which the group is composed. Strategies to at least partly deal with the problem of endogeneity in the multilevel context have been proposed (13).

The pathways linking group-level constructs to individual health are likely to be complex and involve reciprocal causation (or jointly dependent or endogenous variables) and feedback loops. The standard analytical methods used in observational epidemiologic studies may not allow full elucidation of these pathways; other complementary approaches (both quantitative and qualitative) will be necessary. Ultimate certainty regarding group-level effects can only be achieved in a randomized experimental study. However, experimental approaches are obviously not always feasible for the types of group-level variables that might be of interest, and extrapolating from experimental studies may be no less problematic than extrapolating from observational studies (78). Thus, it is likely that researchers will have to continue to rely on hopefully improved observational studies as a source of information regarding group-level effects.

CONCLUSION

Recent recognition of the need to empirically examine the role of group-level factors in epidemiology is a welcome sign. A valuable outcome of the advent of multilevel models is that it has become increasingly common for epidemiologists to theorize regarding the possible health effects of group-level factors. Empirical testing for these effects, however, remains a challenge. Nevertheless, the recognition that a hierarchy of levels may be relevant to any health problem is a fundamental shift in the dominant biomedical and individual paradigm. Systems at multiple levels of organization are present across the continuum from societies to molecules. Thus, developing ways to better investigate these multilevel systems and to understand how interactions within and between levels affect health is a challenge for all health researchers.

REFERENCES

1. Susser M. The logic in ecological. I. The logic of analysis. *Am J Public Health* 1994;84:825-9.
2. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and its consequences. *Am J Public Health*

- 1994;84:819–24.
3. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216–22.
4. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. *J Epidemiol Community Health* 2000;54:367–74.
5. Von Korff M, Koepsell T, Curry S, et al. Multi-level research in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol* 1992;135:1077–82.
6. Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman L, Kawachi I, eds. *Social epidemiology*. New York, NY: Oxford University Press, 2000.
7. Lynch J, Harper S, Davey Smith G. Commentary: plugging leaks and repelling boarders—where to next for the SS income inequality? *Int J Epidemiol* 2003;32:1029–36.
8. Subramanian S, Kawachi I. Response: in defence of the income inequality hypothesis. *Int J Epidemiol* 2003;32:1037–40.
9. Kawachi I, Kennedy BP, Glass R. Social capital and self-rated health: a contextual analysis. *Am J Public Health* 1999;89:1187–93.
10. Lynch J, Due P, Muntaner C, et al. Social capital—is it a good investment strategy for public health? *J Epidemiol Community Health* 2000;54:404–8.
11. Oakes MJ. The (mis)estimation of neighborhood effects. *Soc Sci Med* (in press).
12. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;55:111–22.
13. Subramanian S. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;58:1961–7.
14. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 2004;58:1953–60.
15. Diez Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;21:171–92.
16. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;46:97–117.
17. Diez Roux AV, Schwartz S, Susser E. Ecologic studies and ecologic variables in public health research. In: Detels R, McEwen J, Beaglehole R, et al, eds. *The Oxford textbook of public health*. London, United Kingdom: Oxford University Press, 2002.
18. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14:32–8.
19. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health* 1995;16:61–81.
20. Fine PE. Herd immunity: history, theory, practice. *Epidemiol Rev* 1993;15:265–302.
21. Schwartz S, Carpenter KM. The right answer for the wrong question: consequences of type III error for public health research. *Am J Public Health* 1999;89:1175–80.
22. Susser M. The logic in ecological. II. The logic of design. *Am J Public Health* 1994;84:830–5.
23. Greenland S. A review of multilevel theory for ecologic analyses. *Stat Med* 2002;21:389–95.
24. Subramanian SV. Multilevel methods for public health research. In: Kawachi I, Berkman L, eds. *Neighborhoods and health*. New York, NY: Oxford University Press, 2003.
25. Blalock H. Contextual-effects models: theoretical and methodological issues. *Annu Rev Sociol* 1984;10:353–72.
26. Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health* 2002;56:588–94.
27. Soobader M, LeClere FB, Hadden W, et al. Using aggregate geographic data to proxy individual socioeconomic status: does size matter? *Am J Public Health* 2001;91:632–6.
28. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. *Am J Epidemiol* 1998;148:475–86.
29. Hiller J, McMichael A. Ecological studies. In: Margetts B, Nelson M, eds. *Design concepts in nutritional epidemiology*. Oxford, United Kingdom: Oxford University Press, 1991:323–53.
30. Geronimus A, Bound J, Neidert L. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *J Am Stat Assoc* 1996;91:529–37.
31. Valkonen T. Individual and structural effects in ecological research. In: Dogan M, Rokkam S, eds. *Social ecology*. Boston, MA: MIT Press, 1969:53–68.
32. Lazarsfeld P, 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, and Winston, Inc, 1971:499–516.
33. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002;55:125–39.
34. Diez Roux AV. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;91:1783–9.
35. Robinson W. Ecological correlations and the behavior of individuals. *Am Sociol Rev* 1950;15:351–7.
36. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol* 1988;127:893–904.
37. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989;18:269–74.
38. Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am Sociol Rev* 1978;43:557–72.
39. Hammond J. Two sources of error in ecological correlations. *Am Sociol Rev* 1973;38:764–77.
40. Iversen G. *Contextual analysis*. Newbury Park, CA: Sage Publications, 1991.
41. Scheuch E. Social context and individual behavior. In: Dogan M, Rokkam S, eds. *Social ecology*. Boston, MA: MIT Press, 1969:133–55.
42. Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. New York, NY: Oxford University Press, 2002.
43. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44:1049–60.
44. Brown H, Prescott R. *Applied mixed models in medicine*. New York, NY: Wiley, 1999.
45. Raudenbush SW, Bryk AS. *Hierarchical linear models: applications and data analysis methods*. London, United Kingdom: Sage, 2002.
46. Goldstein H. *Multilevel statistical models*. New York, NY: Halsted Press, 1995.
47. Snijders TAB, Bosker R. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London, United Kingdom: Sage, 1999.
48. DiPrete TA, Forristal JD. Multilevel models: methods and substance. *Annu Rev Sociol* 1994;20:331–57.
49. Mason W, Wong G, Entwistle B. Contextual analysis through the multilevel linear model. In: Leinhardt S, ed. *Sociological methodology*. San Francisco, CA: Josey Bass, 1983:72–103.
50. Kreft I, deLeeuw J. *Introducing multilevel modeling*. London, United Kingdom: Sage, 1998.
51. Laird N, Ware H. Random effects models for longitudinal data. *Biometrics* 1982;38:963–74.
52. Longford N. *Random coefficient models*. Oxford, United Kingdom: Clarendon, 1982.

53. Dempster A, Rubin D, Tsutakawa R. Estimation in covariance component models. *J Am Stat Assoc* 1981;76:341–56.
54. Searle S. Variance components. New York, NY: Wiley, 1992.
55. Goldstein H. Multilevel cross-classified models. *Soc Methods Res* 1994;22:364–75.
56. O'Campo P. Invited commentary: advancing theory and methods for multilevel models of residential neighborhoods and health. *Am J Epidemiol* 2003;157:9–13.
57. Macintyre S, Ellaway A. Neighborhoods and health: an overview. In: Kawachi I, Berkman L, eds. *Neighborhoods and health*. New York, NY: Oxford University Press, 2003:20–44.
58. Balfour JL, Kaplan GA. Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study. *Am J Epidemiol* 2002;155:507–15.
59. Diez Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med* 2001;345:99–106.
60. Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 1999;149:898–907.
61. Sloggett A, Joshi H. Deprivation indicators as predictors of life events 1981–1992 based on the UK ONS Longitudinal Study. *J Epidemiol Community Health* 1998;52:228–33.
62. Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. *Psychol Methods* 2000;5:199–213.
63. Hox JP. Multilevel analysis: techniques and applications. Mahwah, NJ: Lawrence Erlbaum, 2002.
64. Raudenbush SW, Sampson RJ. Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociol Methodol* 1999;29:1–41.
65. Raudenbush S. The quantitative assessment of neighborhood social environments. In: Kawachi I, Berkman L, eds. *Neighborhoods and health*. New York, NY: Oxford University Press, 2003:112–31.
66. Hauser R. Context and consex: a cautionary tale. *Am J Sociol* 1970;75:645–64.
67. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–55.
68. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;31:163–5.
69. Blakely T. Commentary: estimating direct and indirect effects—fallible in theory, but in the real world? *Int J Epidemiol* 2002;31:166–7.
70. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–60.
71. Robins J. The control of confounding by intermediate variables. *Stat Med* 1989;8:679–701.
72. Koopman JS, Lynch JW. Individual causal models and population system models in epidemiology. *Am J Public Health* 1999;89:1170–4.
73. Loomis D, Wing S. Is molecular epidemiology a germ theory for the end of the twentieth century? *Int J Epidemiol* 1990;19:1–3.
74. Hood L. Systems biology: integrating technology, biology, and computation. *Mech Ageing Dev* 2003;124:9–16.
75. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2001;2:343–72.
76. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757–63.
77. Harding DJ. Counterfactual models of neighborhood effects: the effect of neighborhood poverty on high school dropout and teenage pregnancy. *Am J Sociol* 2003;109:676–719.
78. Manski CF, Garfinkel I. Introduction. In: Manski CF, Garfinkel I, eds. *Evaluating welfare and training programs*. Cambridge, MA: Harvard University Press, 1992:1–24.