Multilevel Analysis of Infectious Diseases

Ana V. Diez Roux and Allison E. Aiello

Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor

Traditional study designs, such as individual-level studies and ecological studies, are unable to simultaneously examine the effects of individual-level and group-level factors on risk of disease. Multilevel analysis overcomes this limitation by allowing the simultaneous investigation of factors defined at multiple levels. Areas in which multilevel modeling can be applied to sexually transmitted infection (STI) research include examining how both group-level and individual-level factors are related to individual-level STI outcomes, assessing interactions between individual-level and group-level constructs, and exploring how factors at multiple levels contribute to group-to-group differences in rates of disease. In this article, we review the fundamentals of multilevel modeling, the applications of multilevel models for the examination of STIs, and the key challenges associated with using multilevel modeling for infectious-disease research.

Over the past few decades, epidemiological studies have focused, for the most part, on the identification of individual-level risk factors for disease. The underlying assumption in this approach has been that the causes of disease can be found at the level of individuals. This individual-centered approach has been reflected in behavioral and biomedical models of disease causation and reemerges today in the notion that genetic factors play a major role in the etiology of disease. Populations are usually viewed as collections of individuals, rather than as meaningful entities with inherent properties that may be related to the likelihood that individuals within them develop disease. Recently in epidemiology, however, interest has been increasing in recovering the population or group dimension and in reconsidering the types of variables, types of study designs, and types of analytical approaches needed to develop explanations of the causes of disease that incorporate individual-level and population-level factors [1–4]. The growing consensus is that, with regard to both scientific validity and the practical implications for prevention of disease, in-vestigations of the causes of diseases need to include factors defined at multiple levels. The importance of multilevel determinants has been especially highlighted in recent reviews of the epidemiology of sexually transmitted infections (STIs), for which the influence of population-level factors has long been recognized [5–8].

Multilevel analysis recently has emerged as an analytical strategy that may be useful in incorporating factors defined at multiple levels in epidemiological analyses [9–11]. This article will review the role of population-level factors in infectious-disease epidemiology, summarize the fundamentals of the multilevel approach, and briefly review recent applications of this approach in infectious-disease and STI research. We conclude with a discussion of the strengths and limitations of multilevel analysis in the study of infectious diseases and the challenges raised by its use.

POPULATION-LEVEL (OR GROUP-LEVEL) FACTORS IN INFECTIOUS-DISEASE EPIDEMIOLOGY

The notion that population-level factors or, more generically, group-level factors are important in understanding the distribution of disease has long been present in the study of infectious-disease epidemiology. The classic example is the notion of herd immunity—that is, the idea that an individual's risk of contracting an infectious disease depends, in part, on the level of

immunity in the group or population to which he or she belongs [12]. Herd immunity, a group-level property, is important in understanding the reasons for not only group-level differences in the incidence of disease but also an individual's risk of disease. Thus, group-level factors are important even when the objective is to draw inferences regarding the causes of disease in individuals. In 1916, in the context of studying malaria, Ross [13] alluded to the relevance of group-level factors in infectious diseases in his theory of dependent happenings or events, in which the frequency of the event depends on the number of individuals already affected. More recently, Halloran and Struchiner [14] have discussed the implications of dependent happenings, with regard to study design and inferences in epidemiology.

The phenomenon that the probability of an individual developing an outcome depends in part on the prevalence of the outcome in the group to which he or she belongs also may apply to health-related behaviors. For example, an adolescent's likelihood of smoking may depend on the extent of smoking in his or her peer group. On a larger scale, a person's likelihood of smoking may depend on smoking levels in the society to which he or she belongs. The prevalence of the behavior may generate and, in turn, be influenced by social norms (a group-level factor) regarding its acceptability and desirability. Norms and the prevalence of the behavior may, in turn, influence and be influenced by group- or population-level factors, such as cigarette advertising and legislation regarding smoking in public places. The influence of social norms also applies to behaviors directly linked to the transmission of infectious diseases, such as sexual behaviors. However, the population- or group-level factors relevant to infectious diseases are not limited to those involving the contagion or contagion-like processes described above. Examples of other population-level factors that may be relevant to STIs include the availability and cost of services (e.g., sex education, condoms, or treatment clinics), legislation and enforcement of legislation regarding commercial sex workers, and the educational and occupational opportunities of women. Group-level factors may affect each of the components of the reproductive ratio of an infectious disease (i.e., the rate of transmission, the duration of infectiousness, and the number of contacts per unit of time). For example, the rate of transmission of HIV may be influenced by the prevalence of condom use in the group, the duration of infection may be influenced by the availability and cost of treatment clinics within a geographic area, and the number of contacts per unit of time may be influenced by norms and patterns of sexual contacts at the group level.

In addition, population- or group-level factors also may modify the relationship between individual-level risk factors and risk of disease [15]. For example, the increase in the risk of acquiring an STI that is associated with a given increase in the number of sex partners may be very different in groups with different structures of sexual contacts (e.g., in groups with varying degrees of assortative mating or of sexual contacts with members of other groups). The effects of individual-level variables also may differ depending on the prevalence of the outcome in the population; for example, the relative risk of infection that is associated with the number of sexual contacts may increase as the rate of infection in the population increases. In another example of the interaction between group-level and individual-level factors, the association between commercial sex work and the risk of contracting gonorrhea may differ on the basis of the availability of STI clinics in the neighborhood.

A key consequence of the presence of group-level determinants of infectious diseases is that interventions may have both individual- and group-level effects. For example, a vaccination program may affect an individual's risk of disease through an individual-level effect (the effect of vaccination on an individual's risk of disease) and through a group-level effect (the effect of community-wide vaccination on an individual's risk of disease, even if he or she is not vaccinated). Halloran and Struchiner [14] have referred to these effects as "direct" and "indirect" effects, respectively. In fact, the presence of indirect effects is the rationale for preventing an epidemic through targeted vaccination of especially vulnerable individuals. Although vaccination has group-level consequences, it is a type of intervention administered to individuals. Other types of interventions may operate directly on group-level factors affecting all individuals in the group. An example is the drainage of a swamp, to reduce the mosquito population (and, hence, the transmission of mosquito-borne diseases) in a given community. In this case, the intervention is defined at the group level, and no individual-level analogue exists. Thus, the only estimable effect is the effect of the group-level intervention. However, controlling for individual-level factors, such as the use of bed nets, may be important in estimating this group-level effect or in investigating whether the effect of the group-level intervention on the risk of disease differs by individual-level attributes.

Traditional individual-level studies and ecological studies are unable to simultaneously examine the role of individual- and group-level factors in the risk of disease. For example, an individual-level study that focused on estimating the effects of vaccination by comparing rates of disease between those vaccinated and those not vaccinated, within a given community, would completely miss the group-level effect of percent vaccinated (the indirect effects described by Halloran and Struchiner [14]). By failing to consider the effects of community-wide vaccination, erroneous conclusions could be drawn regarding the individual-level effects of vaccination that could be expected in other contexts, because the individual-level effect (direct effect) may differ depending on the prevalence of vaccination in the community (because of the indirect effect of

community-wide vaccination on the risk of disease among those not vaccinated) [16]. On the other hand, an ecological study of several communities that related the percentage of vaccinated individuals to disease rates would be unable to distinguish between the individual-level effects of vaccination and the effects of the percentage of vaccinated individuals in the communities. Multilevel analysis overcomes these limitations by allowing simultaneous examination of factors defined at multiple levels.

TYPES OF GROUP-LEVEL VARIABLES

Several different types of group-level variables may be relevant to infectious-disease epidemiology. Group-level variables characterize group-level constructs. Groups can be defined in many different ways. For example, a group can be the residents of a country or a neighborhood, classmates in a school, or a group of friends. Group-level variables have been classified into 2 types, derived variables and integral variables [17]. Derived variables are constructed by mathematically summarizing the characteristics of individuals in the group. Examples of derived variables that may be of particular relevance to STIs include the age composition of the group (which may be a marker for mixing patterns), the group prevalence of sexual behaviors known to be associated with transmission of disease (which may influence the probability of an individual adopting that behavior, owing to norms or peer pressure), and the prevalence of the STI (which may modify the effect of an individual-level behavior on the risk of disease). Integral variables differ from derived variables in that they are not summaries of the characteristics of individuals in the group. Examples of integral variables of particular relevance to STIs include community availability of STI clinics, community availability and price of condoms, and features of contact patterns at the group level [18–20]. The term "structural variables" has been used to refer to several different types of integral variables, including those that capture relationships or interactions between members of a group (e.g., characteristics of social networks within the group) [21], as well as features of communities that range from economic policies to physical environments (e.g., the availability of health services) [22].

FUNDAMENTALS OF MULTILEVEL ANALYSIS

Multilevel analysis is an analytical approach that is appropriate for data with nested sources of variability—that is, data involving units at a lower level, or micro-units (e.g., individuals), nested within units at a higher level, or macro-units (e.g., peer groups, schools, or neighborhoods) [9, 10, 23–29]. It allows the simultaneous examination of the effects of group- and individual-level variables on individual-level outcomes, while accounting for the nonindependence of within-group observations. Multilevel analysis also allows the examination of both

between- and within-group variability, as well as how group- and individual-level variables are related to variability at both levels. Thus, multilevel analysis can be used to draw inferences regarding the causes of interindividual variation (or the relationship of group- and individual-level variables to individual-level outcomes), but inferences also can be made regarding intergroup variation, whether such variation exists in the data, and to what extent it is accounted for by group- and individual-level characteristics. Moreover, in multilevel analysis, groups and contexts are not treated as unrelated but are conceived as coming from a larger population of groups about which inferences can be made. By bringing together group- and individual-level variables and examining variability both within and between groups, multilevel analysis avoids the inferential fallacies that may occur when a relevant level is ignored [4]. In infectious-disease epidemiology, multilevel analysis can be used to examine how both group- and individual-level factors are related to individual-level infectious-disease outcomes, how factors at both levels interact, and how factors at both levels contribute to group-to-group differences in disease rates.

The statistical models used in multilevel analysis are referred to as "multilevel models" [26, 28, 29] or "hierarchical models" [24, 30]. These models (or variants of them) have appeared previously in the literature, under a variety of names, including "random-effects models" or "random-coefficient models" [31–33], "covariance-components models" or "variance-components models" [34, 35], and "mixed models" [36]. The following is a simplified example for the case of a normally distributed dependent variable, a single individual-level predictor, and a single group-level predictor. Analogous models can be described for dependent variables that are not normally distributed [23, 24, 28, 30], and multiple individual- and group-level predictors can be added.

For multilevel analysis involving 2 levels (e.g., individuals nested within groups), the multilevel model can be conceptualized as a 2-stage system of equations. In the first stage (level 1), a separate individual-level regression is defined for each group or higher-level unit:

$$Y_{ij} = b_{0j} + b_{1j} I_{ij} + \epsilon_{ij} \ , \ \epsilon_{ij} \sim N(0, \ \sigma^2) \ , \qquad (1)$$

where Y_{ij} is the outcome variable for the ith individual in the jth group, I_{ij} is the individual-level variable for the ith individual in the jth group, b_{0j} is the group-specific intercept, and b_{1j} is the group-specific effect of the individual-level variable. Individual-level errors (ϵ_{ij}; sometimes called "micro-errors") are assumed to be independent and identically distributed, with a mean of 0 and a variance of σ^2. Regression coefficients (b_{0j} and b_{1j}) are allowed to vary from one group to another.

In the second stage (level 2), the group- or context-specific regression coefficients defined in equation 1 (b_{0j} and b_{1j} in this

example) are modeled as functions of group-level (or higher-level) variables:

$$b_{0j} = \gamma_{00} + \gamma_{01}G_j + U_{0j} \ , \ U_{0j} \sim N(0, \ \tau_{00}) \qquad (2)$$

and

$$b_{1j} = \gamma_{10} + \gamma_{11}G_j + U_{1j} \ , \ U_{1j} \sim N(0, \ \tau_{11}) \ , \qquad (3)$$

where $\mathrm{Co}v(U_{0j}, \ U_{1j}) = \tau_{10}$ and G_j is the group-level variable, γ_{00} is the common intercept across groups, γ_{01} is the effect of the group-level predictor on the group-specific intercepts, γ_{10} is the common slope associated with the individual-level variable across groups, and γ_{11} is the effect of the group-level predictor on the group-specific slopes. In the equations for level 2, errors (U_{0j} and U_{1j}; sometimes called "macro-errors") are assumed to be normally distributed, with a mean of 0 and variances of τ_{00} and τ_{11}, respectively. The covariance between intercepts and slopes is represented by τ_{01}. Macro-errors are assumed to be independent across contexts and independent of micro-errors [27]. Thus, multilevel analysis summarizes the distribution of the group-specific coefficients in terms of 2 parts: a "fixed" part, which is common across groups (γ_{00} and γ_{01} for the intercepts and γ_{10} and γ_{11} for the slopes), and a "random" part, which is allowed to vary from group to group (U_{0j} for the intercept and U_{1j} for the slope).

By including an error term in group-level equations 2 and 3, these models allow for sampling variability in the group-specific coefficients (b_{0j} and b_{1j}) and also for the fact that the group-level equations are not deterministic (i.e., not all relevant macro-level variables may have been included in the model). The underlying assumption is that group-specific intercepts and slopes represent random samples from a normally distributed population of group-specific intercepts and slopes or, alternatively, that the macro-errors are exchangeable, that is, that the residual variation in group-specific coefficients across groups is unsystematic.

An alternative presentation of the model fitted in multilevel analysis is to substitute equations 2 and 3 into equation 1:

$$Y_{ij} = \gamma_{00} + \gamma_{01}G_j + \gamma_{10}I_{ij} + \gamma_{11}G_jI_{ij} + U_{0j} + U_{1j}I_{ij} + \epsilon_{ij} \ .$$

The model includes the effects of group-level variables (γ_{01}), individual-level variables (γ_{10}), and their interaction (γ_{11}) on the individual-level outcome Y_{ij}. These coefficients (γ_{01}, γ_{10}, and γ_{11}), which are common to all individuals regardless of the group to which they belong, are often called the "fixed coefficients" (or "fixed effects"). The model also includes a random-intercept component (U_{0j}) and a random-slope component (U_{1j}). The values of these components vary randomly across groups; hence, U_{0j} and U_{1j} are referred to as the "random coefficients" (or "random effects"). The variances in levels 1 and

2 ($\sigma^2, \tau_{00}, \tau_{11}$, and τ_{10}) are called the "(co)variance components." The parameters of the above equations (fixed coefficients, random coefficients, variances of the random coefficients, and residual variance) are estimated simultaneously, by using iterative methods.

Many variants of the more general model illustrated above are possible. For example, only group-specific intercepts (b_{0j}) may be modeled as random (these models have been called "random-effects models"). When covariate effects (b_{1j} in the example above) are modeled as random, these models have been called "random-coefficient models." When some of the coefficients are fixed and others are random, these models have been called "mixed-effects models" or simply "mixed models." Multilevel models also can account for multiple nested contexts (or levels) [24, 28], allowing fixed and random coefficients to be associated with variables measured at different levels of the data hierarchy being analyzed. They can be modified to allow for nonhierarchical, overlapping, or cross-classified contexts (e.g., children nested simultaneously within neighborhoods and schools) [37], and they can be fitted to binary or count outcomes, by specification of the level 1 model as a logistic or Poisson model [23, 24, 28, 30].

In contrast to ecological studies, multilevel analysis incorporates individual-level information. This allows group-level effects to be estimated after adjustment for individual-level variables and also accounts for differences in the individuals in the groups, when group-to-group variability is examined. In contrast to traditional individual-level studies, multilevel analysis incorporates group-level variables, allows the effects of individual-level variables to vary from group to group, and examines not only interindividual variability but also intergroup variability. For example, a study of the determinants of transmission of STIs among adolescents could examine how school-level factors (e.g., sex education, condom availability, or density and structure of social networks) are related to the risk of infection among students, after controlling for individual-level factors such as the number of sex partners that each student has. Differences in the relationship between number of sex partners and risk of disease across schools and the extent to which this relationship is modified by school-level factors also could be examined. The study could also investigate the extent to which school-to-school differences in incidence rates are attributable to individual-level factors (i.e., the individual characteristics of students attending the different schools) or to school-level factors. For example, multilevel analysis could be used to investigate how school-to-school variability in risk of disease changes as individual-level and school-level variables are added to the model (by investigating changes in τ_{00} [equation 2] as variables are added). In another example, a multilevel model could be used to estimate how community-to-community variability in incidence of AIDS changes as information

on individual-level risk behaviors and community-level measures of condom availability or effects of needle-exchange programs are added to the model. The same model could be used to examine the relationship between community availability of condoms and risk of developing AIDS, after controlling for individual-level confounders.

EXAMPLES OF RECENT APPLICATIONS OF MULTILEVEL ANALYSIS IN INFECTIOUS-DISEASE AND STI RESEARCH

Empirical applications of multilevel analysis in the study of STIs (or infectious diseases) remain rare. A PubMed and Medline search (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) incorporating combinations (in pairs) of the phrases "multilevel model" and "hierarchical linear model" with "STI," "HIV," and "infection" performed in September 2003 found only a few articles that reported use of multilevel models ($N = 11$) [38–48]. Most of these studies focused on behavioral or functional outcomes, rather than on biological or disease end points. They applied multilevel analysis primarily to the examination of the effects of group-level and individual-level factors on individual-level outcomes or to the analysis of multiple measures for individuals over time (for which both individual-specific and measurement occasion–specific variables are of interest). The group-level factors examined mostly referred to characteristics of intervention sites (including treatment characteristics and site-specific prevalences of infection), rather than to characteristics of potentially relevant social groups, such as peer groups or social networks. For example, Tinsman et al. [39] used multilevel models to examine the effects of community-based HIV outreach services for high-risk substance abusers and individual client characteristics on the persuasion of at-risk clients to receive HIV tests and to enter structured substance-abuse treatment centers. The analysis included clients nested within outreach projects. Characteristics of the outreach projects, such as whether the site provided HIV testing or gave referrals, and client characteristics such as age, use of cocaine or crack, and injection drug use (IDU) in the previous 30 days were examined in relation to HIV testing and entry into treatment. In a multisite study of IDU, Wang et al. [44] investigated how individual-level variables such as sex, age, and use of a shooting gallery and group-level variables such as HIV seroprevalence and mean level of IDU, by project site, were related to needle transfer. In a study examining patterns of HIV-associated risk behaviors among patients in 96 drug-treatment programs in 11 US cities, Broome et al. [46] used multilevel modeling to examine how individual-level factors (e.g., sex, use of cocaine, and antisocial personality disorder) and prevalence of AIDS at the drug-treatment site (a group-level factor) were related to HIV-associated risk behaviors. Dausey and Desai [47] used multilevel modeling to examine the impact of substance abuse, by patients with or without mental illness, on HIV-associated risk behaviors at 71 national substance-abuse treatment sites. Although multilevel models were used to adjust for clustering within the 71 treatment sites, site-level variables were not examined.

Multilevel modeling also has been used to examine repeated measures in HIV, STI, and infectious-disease research. In these applications, the individual level and the measurement occasions nested within it (which is analogous to groups and the individuals nested within them) are of interest [38, 40–43, 48]. In another application of multilevel modeling with repeated measures, Rosel et al. [45] predicted future changes in incidence of AIDS into 2005, using multiple historical measures of incidence of AIDS in 19 communities from 1983 to 1986. In summary, applications of multilevel analysis have only recently begun to emerge in the infectious-disease literature. Extending the use of multilevel analysis, however, will require careful consideration of several key issues. Two of these issues (defining relevant groups and group-level variables and problems related to study design and data sources) are common to the study of both infectious and chronic diseases. Two additional issues (modeling dependencies and dynamic relationships) also are present in the study of both infectious and chronic diseases but are of particular relevance to infectious diseases.

DEFINING GROUPS AND GROUP-LEVEL VARIABLES

Key challenges in the use of multilevel analysis in infectious-disease epidemiology are identifying the levels that are relevant to the research question of interest, specifying the relevant constructs or variables at each level, operationalizing the relevant groups, and measuring the relevant group-level variables. Defining the theoretical model guiding the research, including the relevant levels and the constructs at each level, is fundamental if meaningful information is to be obtained from multilevel analysis. Even after the theoretical model has been developed, defining or operationalizing the relevant groups is often a challenge. In some cases, such as the investigation of school-level effects, the demarcation of groups may be straightforward. In other cases, such as for peer groups or neighborhoods, boundaries may be fuzzy and difficult to define. In some situations, the group relevant to a particular health outcome may be difficult to determine. For example, in studying group effects on behaviors among adolescents, is the relevant "group" the school, the classroom, the friend network, or the neighborhood? In addition, groups of varying sizes may be relevant for different research questions. The couple may be the relevant group or context for some research questions, whereas the neighborhood or even the city as a whole may be relevant for other questions. Defining relevant levels may be especially complex for an outcome like transmission of STIs, for which a variety of nested and nonnested

groups or contexts of varying sizes may be relevant. For example, the group of friends, the school, the neighborhood, and the city of residence are 4 contexts of possible relevance to incidence of STIs in adolescence. Neighborhoods are nested within cities; however, schools may or may not be nested within neighborhoods, and groups of friends may or may not be nested within schools or neighborhoods.

Conceptualizing and measuring the group-level variables of interest are other key challenges. What features of schools or peer groups are relevant to the behaviors of adolescents, and how should these features be measured? Some measures may be relatively simple, such as the availability of condoms in the school or the presence of a sex-education program. Others may be much more complex. For example, how should social norms or the structure of contacts within a group be measured? Although health research has become very sophisticated at measuring individual-level properties, the measurement of group-level or population-level attributes potentially related to health is still in its infancy. Two approaches recently used in the social sciences to measure attributes of one type of "group" (a neighborhood) included systematic observation of neighborhoods by trained raters and surveys of neighborhood residents [49–51]. The observations of multiple raters and multiple survey respondents for each neighborhood were aggregated, to construct neighborhood-level measures by using multilevel techniques. Geographic information systems can be used to develop area-based measures of the availability of health-related resources, such as health clinics, which then can be used as predictors in multilevel analyses of neighborhood health effects. Other methodological approaches may be necessary to capture complex group properties of special relevance to STI research, such as the density and structure of networks within a group [52].

STUDY DESIGN AND DATA SOURCES

To date, most applications of multilevel analysis in the health field have used data collected for other purposes, usually during traditional individual-level studies to which other sources of data on group characteristics have been appended. As a result, both the structure of the data and the quality of the group-level information available are often limited with regard to multilevel analysis and may hamper the ability to detect group-level effects. The use of multilevel analysis with data originally collected with only individual-level studies in mind has 2 important limitations. The first is related to the number of groups represented in the data and the number of observations per group. Power in multilevel analysis depends on both the number of groups and the number of observations per group. The relative importance of the number of groups and the number of individuals per group depends on the specific question of interest [26, 53]. The second limitation is related to the availability of relevant group-level data and the validity and reli-

ability of the measures available. As we have noted above, the measurement of group-level constructs is still in its infancy. In contrast, individual-level measurements are often very sophisticated. As a result, gross misspecification of group-level attributes occurs (owing to misspecification of both groups and group-level variables), compared with the specification of individual-level factors for which group-level effects are often adjusted. The growing interest and familiarity with multilevel analysis are likely to promote the design of studies with multilevel analysis in mind, in which special attention would be given to the most appropriate data structure for the multilevel question of interest and to the definition and characterization of groups.

ACCOUNTING FOR AND MODELING DEPENDENCIES

A common critique of standard epidemiological approaches in infectious-disease research is that studies do not account for dependencies between outcomes for individuals [14, 54, 55]. Dependencies can occur in several ways. The probability of contracting disease for a given individual may be dependent on the frequency of disease among other individuals in the same group, because group prevalence or incidence of infection influences the likelihood that a given individual will come into contact with an infected person. The degree of contagiousness of infected individuals in the group also may influence an individual's likelihood of acquiring the infection. Complex patterns of dependencies may arise because a given individual's risk of infection may be dependent on the risk of those with whom he or she regularly interacts, as well as on the type and frequency of the interactions. Thus, the degree of dependency between outcomes may differ for different pairs or sets of individuals within a group, depending on contact patterns. Although dependencies also may be present for chronic diseases (through the "contagion" of behaviors discussed above), the presence of complex dependencies at multiple levels is a defining feature of infectious-disease outcomes.

Multilevel analysis accounts for dependencies between outcomes for individuals within groups through the incorporation of group-level variables that individuals in a group share and by allowing for random effects and random coefficients at the group level. To the extent that dependencies between outcomes within a group are adequately captured by the individual- and group-level variables and the random coefficients included in the model, multilevel models will yield valid estimates of group- and individual-level effects and their SEs. However, in some infectious-disease applications, the patterns of dependencies may be complex (e.g., the pattern of dependencies may vary from person to person within the group, on the basis of each person's unique location within a sexual network) and may not be fully captured by the relatively simple multilevel models

described in this article. Work that examines the extent to which the types of dependencies present in infectious diseases can be adequately captured in multilevel models is needed. The extent to which ignoring certain types of dependencies leads to invalid results may differ on the basis of the research question.

Multilevel models are very flexible in the incorporation of random effects and coefficients to account for dependencies (e.g., random effects and coefficients can be different for individuals with different individual-level characteristics within a group). However, they do not explicitly model the contact patterns and dynamic relationships that give rise to these dependencies within a group or how these dependencies influence the course of an epidemic in a population over time, as do simulation models of infection transmission [56].

MODELING DYNAMIC RELATIONSHIPS

Like usual regression approaches commonly used in epidemiology, multilevel analysis does not easily allow investigation of dynamic and reciprocal relations between predictors or between predictors and outcomes. For example, multilevel models do not model the possibility that individual-level properties (or individual-level relations between variables) may influence group characteristics and that these emergent group characteristics may, in turn, shape individual-level independent variables. These types of effects may be particularly important when examining changes over time. For example, an individual-level intervention, such as vaccination at time 1, may result in modifications of group properties (herd immunity) at time 2, which may, in turn, modify the effect of the individual-level intervention (vaccination) on risk of disease at a future time. Modifications to multilevel models that allow for these dynamic processes have been proposed [57, 58], but their investigation remains a challenge.

Multilevel analysis does not provide the same information as approaches that employ mathematical models of transmission of disease to study the dynamics over time of an infectious disease in a population. Approaches based on transmission dynamics study the changes that occur in a dynamic system (a population) when features of the system change. Patterns of disease in populations, as opposed to the determinants of disease in individuals, are the main focus of interest [54, 55]. Mathematical models of the transmission dynamics of infectious diseases (based on different scenarios regarding, for example, contact patterns and transmission probabilities) can be used to estimate how a specific change at the population level modifies the dynamics of transmission of disease and, hence, the incidence or prevalence of disease [59, 60]. In contrast, in multilevel analysis, interest centers on estimating group and individual effects on individual-level outcomes, on studying interactions between individual and group characteristics, and on estimating between- and within-group variability in outcomes. Although the effects of group-level variables such as the structure of social contacts in a group (which may be predictors of the dynamics of transmission of disease) can be examined, the actual temporal dynamics of transmission itself are not modeled. The appropriateness of multilevel analysis depends on the question of interest. Strategies that directly model the nonlinear dynamics of transmission systems may be more appropriate for research questions that pertain to predictors of changes in patterns of disease in populations over time [54]. Both methods can be thought of as complementary. For example, simulation models of transmission dynamics may yield insights into group-level or population-level variables that are relevant to infectious diseases and that can be examined in empirical multilevel analyses, including both individual-level and group-level predictors. Further work that contrasts the data requirements for multilevel analysis versus transmission models and that explicates differences in the types of questions that can be addressed with each method and the insights that each approach can yield would be of great value to infectious-disease epidemiology.

CONCLUSION

Epidemiologists increasingly recognize the need to examine both macro- and microlevel factors in studying the causes of infectious and chronic diseases. Researchers studying STIs have long been aware of the role that population-level factors play in shaping epidemics and risk of disease among individuals. Multilevel analysis has recently emerged as a powerful analytical tool that allows the simultaneous investigation of how population-level (or group-level) and individual-level factors contribute to disease outcomes. The presence of complex dependencies between outcomes may make the application of multilevel analysis in infectious-disease research more complex than it is in chronic-disease research, because of the need to adequately capture these dependencies in the model. In addition, there may be questions of particular relevance to infectious diseases (such as factors associated with transmission of disease) that may not be adequately investigated through multilevel analysis alone. In thinking about the multilevel determinants of infectious diseases, distinguishing between the statistical technique of multilevel analysis and the more general issues of investigating determinants of outcomes at multiple levels is important. The latter is likely to require a multiplicity of analytical approaches, of which multilevel analysis is one example.

References

1. Susser M. The logic in ecological. II. The logic of design. Am J Public Health **1994**; 84:830–5.
2. Diez Roux AV, Schwartz S, Susser E. Ecologic studies and ecological variables in public health research. In: Detels R, McEwen J, Beaglehole

R, Tanaka H, eds. The Oxford textbook of public health. Vol. 2. London: Oxford University Press, **2002**:493–507.

3. 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.

4. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. Am J Public Health **1998**; 88:216–22.

5. Aral SO, Holmes KK. Sexually transmitted diseases in the AIDS era. Sci Am **1991**; 264:62–9.

6. Aral SO, Holmes KK. Social and behavioral determinants of the epidemiology of STDs: industrialized and developing countries. In: Wasserheit J, ed. Sexually transmitted diseases. 3rd ed. New York: McGraw-Hill, **1999**:39–76.

7. Padian NS, Aral SO, Holmes KK. Individual, group, and population approaches to STD/HIV prevention. In: Wasserheit J, ed. Sexually transmitted diseases. 3rd ed. New York: McGraw-Hill, **1999**:1231–8.

8. Shiboski S, Padian NS. Population- and individual-based approaches to the design and analysis of epidemiologic studies of sexually transmitted disease transmission. J Infect Dis **1996**; 174(Suppl 2):S188–200.

9. Diez Roux AV. Multilevel analysis in public health research. Annu Rev Public Health **2000**; 21:171–92.

10. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Soc Sci Med **1998**; 46: 97–117.

11. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. J Epidemiol Community Health **2000**; 54:367–74.

12. Fine PE. Herd immunity: history, theory, practice. Epidemiol Rev **1993**; 15:265–302.

13. Ross R. An application of the theory of probabilities to the study of a priori pathometry: part I. Proc R Soc Series A **1916**; 92:204–30.

14. Halloran ME, Struchiner CJ. Study designs for dependent happenings. Epidemiology **1991**; 2:331–8.

15. Koopman JS, Longini IM Jr. The ecological effects of individual exposures and nonlinear disease dynamics in populations. Am J Public Health **1994**; 84:836–42.

16. Halloran ME, Struchiner CJ, Longini IM Jr. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. Am J Epidemiol **1997**; 146:789–803.

17. Diez Roux AV. A glossary for multilevel analysis. J Epidemiol Community Health **2002**; 56:588–94.

18. Aral SO. Sexual network patterns as determinants of STD rates: paradigm shift in the behavioral epidemiology of STDs made visible. Sex Transm Dis **1999**; 26:262–4.

19. Manhart LE, Aral SO, Holmes KK, Foxman B. Sex partner concurrency: measurement, prevalence, and correlates among urban 18–39-year-olds. Sex Transm Dis **2002**; 29:133–43.

20. Gorbach PM, Stoner BP, Aral SO, H Whittington WL, Holmes KK. "It takes a village": understanding concurrent sexual partnerships in Seattle, Washington. Sex Transm Dis **2002**; 29:453–62.

21. Lazarsfeld P, Menzel H. On the relation between individual and collective properties. In: Etzioni A, ed. A sociological reader on complex organizations. New York: Holt, Rinehart, and Winston, **1971**:499–516.

22. Sumartojo E. Structural factors in HIV prevention: concepts, examples, and implications for research. AIDS **2000**; 14(Suppl 1):S3–10.

23. DiPrete TA, Forristal JD. Multilevel models: methods and substance. Annu Rev Sociol **1994**; 20:331–57.

24. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. London: Sage, **2002**.

25. Von Korff M, Koepsell T, Curry S, Diehr P. Multi-level research in epidemiologic research on health behaviors and outcomes. Am J Epidemiol **1992**; 135:1077–82.

26. Snijders TAB, Bosker R. Multilevel analysis: an introduction to basic and advanced multilevel modeling. London: Sage, **1999**.

27. Mason W, Wong G, Entwisle B. Contextual analysis through the multilevel linear model. In: Leinhardt S, ed. Sociological methodology. San Francisco: Josey-Bass, **1983**:72–103.

28. Goldstein H. Multilevel statistical models. New York: Halsted Press, **1995**.

29. Kreft I, deLeeuw J. Introducing multilevel modeling. London: Sage, **1998**.

30. Wong G, Mason W. The hierarchical logistic regression model for multilevel analysis. J Am Stat Assoc **1985**; 80:513–24.

31. Diggle PJ, Liang KY, Zeger SL. Analysis of longitudinal data. New York: Oxford University Press, **2002**.

32. Laird N, Ware H. Random effects models for longitudinal data. Biometrics **1982**; 38:963–74.

33. Longford N. Random coefficient models. Oxford: Clarendon, **1982**.

34. Dempster A, Rubin D, Tsutakawa R. Estimation in covariance component models. J Am Stat Assoc **1981**; 76:341–56.

35. Searle S. Variance components. New York: Wiley, **1992**.

36. Brown H, Prescott R. Applied mixed models in medicine. New York: Wiley, **1999**.

37. Goldstein H. Multilevel cross-classified models. Sociol Methods Res **1994**; 22:364–75.

38. Morrison DM, Gillmore MR, Hoppe MJ, Gaylord J, Leigh BC, Rainey D. Adolescent drinking and sex: findings from a daily diary study. Perspect Sex Reprod Health **2003**; 35:162–8.

39. Tinsman PD, Bullman S, Chen X, Burgdorf K, Herrell JM. Factors affecting client response to HIV outreach efforts. J Subst Abuse **2001**; 13: 201–14.

40. Breitbart W, Rosenfeld B, Kaim M, Funesti-Esch J. A randomized, double-blind, placebo-controlled trial of psychostimulants for the treatment of fatigue in ambulatory patients with human immunodeficiency virus disease. Arch Intern Med **2001**; 161:411–20.

41. Dancy BL, Marcantonio R, Norr K. The long-term effectiveness of an HIV prevention intervention for low-income African American women. AIDS Educ Prev **2000**; 12:113–25.

42. Crystal S, Sambamoorthi U. Functional impairment trajectories among persons with HIV disease: a hierarchical linear models approach. Health Serv Res **1996**; 31:469–88.

43. Cohen S, Hamrick N, Rodriguez MS, Feldman PJ, Rabin BS, Manuck SB. Reactivity and vulnerability to stress-associated risk for upper respiratory illness. Psychosom Med **2002**; 64:302–10.

44. Wang J, Siegal HA, Falck RS, Carlson RG. Needle transfer among injection drug users: a multilevel analysis. Am J Drug Alcohol Abuse **1998**; 24:225–37.

45. Rosel J, Oliver JC, Jara P, Caballer A. A multilevel time-series model for the incidence of AIDS cases in Spain. Health Place **2000**; 6:309–17.

46. Broome KM, Joe GW, Simpson DD. HIV risk reduction in outpatient drug abuse treatment: individual and geographic differences. AIDS Educ Prev **1999**; 11:293–306.

47. Dausey DJ, Desai RA. Psychiatric comorbidity and the prevalence of HIV infection in a sample of patients in treatment for substance abuse. J Nerv Ment Dis **2003**; 191:10–7.

48. Nozyce M, Hittelman J, Muenz L, Durako SJ, Fischer ML, Willoughby A. Effect of perinatally acquired human immunodeficiency virus infection on neurodevelopment in children during the first two years of life. Pediatrics **1994**; 94:883–91.

49. Sampson R, Raudenbush SW. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol **1999**; 105:603–51.

50. Raudenbush SW, Sampson RJ. Ecometrics: towards a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. Sociol Methodol **1999**; 29:1–41.

51. Sampson R, Raudenbush S, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science **1997**; 277:918–24.

52. Ghani AC, Garnett GP. Measuring sexual partner networks for transmission of sexually transmitted diseases. J R Stat Soc (Ser A) **1998**; 161: 227–38.

53. Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. Psychol Methods **2000**; 5:199–213.

54. Koopman JS. Emerging objectives and methods in epidemiology. Am J Public Health **1996**; 86:630–2.

55. Koopman JS, Lynch JW. Individual causal models and population system models in epidemiology. Am J Public Health **1999**; 89:1170–4.

56. Koopman JS. Modeling infection transmission—the pursuit of complexities that matter. Epidemiology **2002**; 13:622–4.

57. Entwisle B. Micro-macro theoretical linkages in social demography: a commentary. In: Huber J, ed. Macro-micro linkages in sociology. Newbury Park, CA: Sage, **1991**:280–6.

58. Mason W. Problems in quantitative comparative analysis: ugly ducklings are to swans as ugly scatter plots are to…? In: Huber J, ed. Macro-micro linkages in sociology. Newbury Park, CA: Sage, **1991**:231–43.

59. Hethcote HW. Three basic epidemiological models. Biomathematics. Vol 18. In: Levin SA, Hallam TG, Gross LG, eds. Applied mathematical ecology. New York: Springer-Verlag, **1989**:119–44.

60. Hughes JP, Garnett GP, Koutsky L. The theoretical population-level impact of a prophylactic human papilloma virus vaccine. Epidemiology **2002**; 13:631–9.