

# The Ecological Effects of Individual Exposures and Nonlinear Disease Dynamics in Populations

## ABSTRACT

To describe causally predictive relationships, model parameters and the data used to estimate them must correspond to the social context of causal actions. Causes may act directly upon the individual, during a contact between individuals, or upon a group dynamic. Assuming that outcomes in different individuals are independent puts the causal action directly upon individuals. Analyses making this assumption are thus inappropriate for infectious diseases, for which risk factors alter the outcome of contacts between individuals. Transmission during contact generates nonlinear infection dynamics. These dynamics can so attenuate exposure-infection relationships at the individual level that even risk factors causing the vast majority of infections can be missed by individual-level analyses. On the other hand, these dynamics amplify causal associations between exposure and infection at the ecological level. The amplification and attenuation derive from chains of transmission initiated by exposed individuals but involving unexposed individuals. A study of household exposure to the only vector of dengue in Mexico illustrates the phenomenon. An individual-level analysis demonstrated almost no association between exposure and infection. Ecological analysis, in contrast, demonstrated a strong association. Transmission models that are devoid of any sources of the ecological fallacy are used to illustrate how nonlinear dynamics generate such results. (*Am J Public Health*. 1994;84:836-842)

James S. Koopman, MD, MPH, and Ira M. Longini, Jr, PhD

## Introduction

Ecological analyses are subject to diverse, subtle, and strong biases when they are used to make inferences about individual effect parameters.<sup>1-11</sup> But not all risk factor effects are manifest directly upon individuals. Infectious-disease risk factor effects are commonly manifest upon or during a contact between individuals. The infectious status of the person contacted and, indirectly, the risk factors in the person contacted are thus important determinants of the rates at which individuals with any particular set of risk factors get infected. This makes the outcome of exposure in one individual dependent upon the outcome of exposure in others. It also makes the population dynamics of infectious disease highly nonlinear.

We argue that the analytic framework used by researchers who have discussed the ecological fallacy<sup>1-13</sup> is inappropriate in the case of infectious diseases because that framework implies dynamic linearity of disease development in a population. Dynamic linearity occurs when the rate of disease development in one group is independent of how many individuals there are in other groups. When the assumption of dynamic linearity does not hold, statistics estimated using models that assume dynamic linearity do not accurately reflect the effect of exposures at either the individual or the population level. More specifically, when nonlinear population processes generate dependence of outcomes between individuals, neither regression coefficients from ecological studies nor risk differences from individual studies reflect how much change in disease levels can be expected from a given change in exposure levels.

## Ecological- and Individual-Level Data

Let us first consider the usual form of data gathered from individuals that can be aggregated into ecological data. Such data specify the exposures and diseases of individuals. Usually individuals are represented by rows of data, and their exposure and disease variables are represented by columns. One of the individual variables may be the study site in which the subject resides.

For our discussion, we consider only a single dichotomous exposure variable and a single dichotomous outcome variable. That eliminates the possibility of nonlinear relationships between quantitative exposure levels and disease frequency. It also eliminates the need to specify multivariate model forms such as additive or multiplicative models. When variables are continuous rather than dichotomous and when different forms of multivariate relationships are specified, a whole new set of issues arises in the relationships between ecological- and individual-level analyses, which we wish to avoid.

We use  $k$  to specify the site in which subjects reside. The symbol  $r_{+k}$  is used to

---

James S. Koopman is with the Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor. Ira M. Longini, Jr, is with the Division of Biostatistics, School of Public Health, Emory University, Atlanta, Ga.

Requests for reprints should be sent to James S. Koopman, MD, MPH, Department of Epidemiology SPH-1, University of Michigan, 109 Observatory St, Ann Arbor, MI 48109.

This paper was accepted January 27, 1994.

**Editor's Note.** Nicole Schupf was the Editor in charge of the blind peer review for this paper. See related articles by Susser (p 825 and p 830) and Schwartz (p 819) and editorial by Poole (p 715) in this issue.

represent the proportion of subjects at site  $k$  with the disease. The + indicates that this value represents the sum across both exposed and unexposed individuals. Within any population, the proportion with the outcome can be further divided by exposure status:  $r_{0k}$  represents the proportion of unexposed subjects at site  $k$  that have the disease, and  $r_{1k}$  represents the proportion of exposed subjects with the disease. We use  $p_k$  to represent the proportion of subjects at site  $k$  that are exposed.

### Components of the Ecological Regression

Greenland and Morgenstern analyzed how the association of exposure to disease at the individual level is related to the regression of exposure on disease at the ecological level, given the type of data specified above.<sup>4</sup> Consider a linear regression of the average disease level on the average exposure level in communities. The model of the fitted relationships between exposure and disease is

$$r_{+k} = a + bp_k. \quad (1)$$

Greenland and Morgenstern demonstrated that for dichotomous exposure and outcome variables, the ecological regression coefficient,  $b$ , can be divided into three components: individual effects, confounding, and effect modification.<sup>4</sup> Mathematically, the three components are as follows:

$$b = E(r_{1k} - r_{0k}) + \frac{\text{cov}(p_k, r_{0k})}{\text{var}(p_k)} + \frac{\text{cov}([p_k - E(p_k)]p_k, [r_{1k} - r_{0k}])}{\text{var}(p_k)}. \quad (2)$$

The first component of the ecological regression parameter is the expected value of the risk difference at the individual level. This risk difference could be due to cause, confounding, or biases acting at the individual level. It is generally presumed that the portion of the risk difference at the individual level due to cause would be the best parameter to use in predicting how much disease results from the exposure of a specified number of individuals. But this is true only for dynamically linear systems and not for infectious diseases.

The second component is present when the background rate in the unexposed subjects is associated with the level of exposure in the population. Thus, the second component represents confounding. It is generally presumed that such

confounding arises because some third variable causing the disease is more frequent in populations that have a higher level of the exposure of interest. But in the case of infectious diseases, we will see that this second component arises even when there are no other causes of the disease and therefore no confounding variables.

The third component is present when the risk difference in a population is associated with the level of exposure in a population. Thus, the third component represents effect modification. A major point in the Greenland and Morgenstern paper is that the ecological fallacy arises not only from confounding but also from effect modification.<sup>4</sup> As with confounding, it is often presumed that a third variable, which varies across populations in parallel with the frequency of exposures in those populations, accounts for this third component of the ecological regression coefficient. We will again demonstrate, however, that for infectious diseases, no third variable is needed to get an association of the risk difference and the proportion exposed.

Equation (2) assumes no misclassification. Brenner et al. demonstrate that unbiased misclassification that decreases the strength of association at the individual level will increase the strength of association in an ecological analysis.<sup>3</sup> Thus, unbiased misclassification will generate a significant difference in the association between exposure and disease at the individual and the ecological levels.

One commits an ecological fallacy if one assumes that the ecological effect  $b$  reflects the individual effect  $r_{1k} - r_{0k}$  when, in fact,  $b$  is significantly determined by the second and third components in equation (2) or  $b$  is increased and  $r_{1k} - r_{0k}$  is decreased by unbiased misclassification. Another type of fallacy arises, however, when one assumes that the individual-level effect  $r_{1k} - r_{0k}$  is the effect of interest when, as in the infectious disease example that follows, it is not.

### An Infectious Disease Example

A dengue fever study in which both individual- and ecological-level associations between exposure and disease were assessed illustrates the phenomenon on which we will concentrate. Dengue fever is a virus infection transmitted only by *Aedes* mosquitoes. Koopman et al.<sup>14</sup> studied 50 individuals under age 29 from each of 70 villages in Mexico with regard to their dengue antibody levels and dengue risk factors. Only one individual

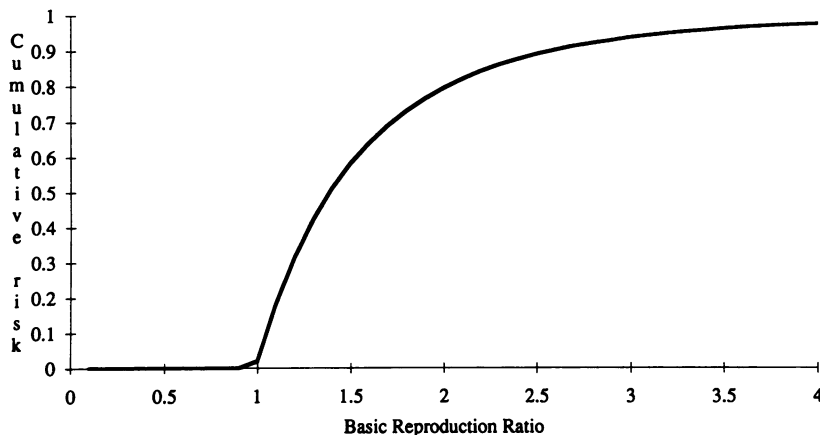
per household was studied. The study was carried out during an annual inter-epidemic period after 5 years of epidemics that had followed 25 years without infection.

Exposure classification of individuals was made on the basis of whether *Aedes aegypti* larvae were found in the household. Infection history was dichotomously classified using a complement fixation antibody assay of sera. At the ecological or village level, the frequency of *Aedes* larvae per household ranged from 0% to 69%. The proportion of individuals infected in the different communities ranged from 0% to 90%. These broad ranges of exposure and outcome frequencies do not represent a situation with high risk of coming to an erroneous conclusion about individual effects because of the ecological fallacy.<sup>10,11</sup>

The ecological analysis was performed with BMDPLR.<sup>15</sup> This routine essentially replaces all individual-level exposure variables with the ecological-level exposure variables and then carries out a logistic regression. The odds ratio of a 0% larva level vs a 100% larva level was 12.7, with a 95% confidence interval from 8.3 to 17.8. The individual-level analysis was performed with a pooled Mantel-Haenszel odds ratio across the different communities. There was no significant heterogeneity of odds ratios across communities. The pooled Mantel-Haenszel odds ratio across different communities was 1.1, with a 95% confidence interval from 0.81 to 1.48.

The *Aedes* mosquitoes identified in the above study are the only genus of mosquito transmitting dengue in Mexico. These mosquitoes breed mainly around houses, and there is generally strong clustering of infection by household.<sup>16</sup> If one concluded from the lack of an individual-level association that the ecological association of household infestation with infection was an artifact, one would have missed a most important cause of dengue infection—namely, household infestation with *Aedes*.

Transmission dynamics, confounding, and exposure misclassification could each contribute to the dramatic difference in the individual and ecological measures of association we observed. Confounding would have to be greater at the ecological level than at the individual level. It thus seems unlikely that confounding could account for the dramatic differences observed. On the other hand, exposure misclassification is significant because we are characterizing exposure by using an



**FIGURE 1—Theoretically derived cumulative risks during an SIR (“susceptible–infectious–recovering”) epidemic as a function of the basic reproduction ratio.**

inter-epidemic level of exposure perhaps some years after an infection was experienced during an epidemic year.

Transmission decreases the individual-level effects and increases the ecological effects as follows: transmission from infected individuals with household mosquitoes to individuals without household mosquitoes raises the rate of infection in unexposed individuals. This effect is amplified through continuing chains of infection. The more exposed individuals there are in a community, the more the rate of infection in the unexposed population will be raised by this secondary transmission and the smaller the individual-level association will be. At the ecological level, however, all of this secondary transmission will amplify the effect of increasing exposure levels in a community.

In terms of equation (2), this secondary transmission will raise the background rate of infection in the unexposed population and thus contribute to the second component on the right-hand side of the equation. As circulation of infection increases, the proportion of infected mosquitoes will rise, and therefore a given exposure to household mosquitoes will increase in risk. Thus, the third component of equation (2) will also make a significant contribution to the difference between the ecological and the individual effect. Note that this increasing infection rate of mosquitoes will also increase the risk of mosquito exposure outside of the household. This amplifies the difference between individual and ecological levels that is attributable to the second compo-

nent on the right-hand side of equation (2). Although equation (2) is relevant to linear regression rather than to logistic regression, the qualitative phenomenon discussed above still apply.

### *Quantitative Models of Transmission Effects on Individual and Ecological Associations*

In the real-life situation discussed above, confounding and misclassification are possible explanations for the divergence in individual- and ecological-level associations. To isolate the effects of transmission dynamics on this divergence, we now present results from a theoretical model in which no third variables are introduced that could generate confounding or effect modification and in which classification is 100% accurate.

Consider an SIR infection process in which individuals start out susceptible (S), then become contagious to others (infectious, I), and finally control their infection through an immune response that eliminates their contagiousness and leads to recovery (R). To make our point, we do not need the added complication of considering an insect vector in our model. We just consider a simple SIR infection in which (1) all individuals in each village are equally susceptible at the beginning of an epidemic, (2) all individuals mix randomly within their village, (3) the populations are large both in absolute terms and in relationship to the number of individuals that introduce infection into the commu-

nity, and (4) the level of exposure is constant over the period of the epidemic.

The proportion of the population infected at the end of such an SIR epidemic—that is, the cumulative risk of infection ( $CR$ ) at the end of the epidemic process—has been analyzed theoretically and found to be independent of whether there is a noncontagious latent period of infection.<sup>17</sup> That proportion can be expressed in the following transcendental equation:

$$CR = 1 - e^{-R_0 CR}, \quad (3)$$

in which  $R_0$ , the basic reproduction ratio, equals the number of secondary cases generated by an infected individual over the entire course of that person's infection. In the global equation presented in (3), the  $R_0$  would represent some average value between exposed and unexposed individuals. The relationship between  $R_0$  and  $CR$  in equation (3) is presented graphically in Figure 1. Note in this curve that the value  $R_0 = 1$  represents a threshold below which there is no sustained transmission, and that at an  $R_0$  of 3 to 4, the maximum effect has been almost achieved so that, as exposure is increased beyond this level, little additional effect is observed.

Let us relate this to our dengue example.  $R_0$  is the number of dengue infections that an infected individual generates over the course of his or her infection. Its value depends on the number of mosquitoes that bite that infected individual, the proportion of those mosquitoes that survive until they become infectious, and the number of other individuals to whom those mosquitoes bite before they die.  $R_0$  will increase as the number of mosquitoes increases.

Villages with few mosquitoes will fall below the  $R_0$  threshold of 1, and there will be no sustained dengue transmission. Consequently, there will be no association between *Aedes aegypti* larva in the household and dengue infection in those villages. Individuals with *Aedes* larvae on their household premises may be more exposed to *Aedes aegypti* bites than individuals living at sites without such larvae, but they would not be more exposed to dengue because the mosquitoes would be uninfected.

In the villages that have enough mosquitoes to get the basic reproduction number above 4 or 5, there will again be little association between the presence of mosquito larvae in the house and the risk of infection. When circulation of dengue infection is high in the community, all

those who do not have larvae in the house will have been bitten frequently by mosquitoes that were born from larvae in their neighbors' house or in the market or at work or school. Consequently, exposure at the household level will again fail to distinguish risk levels because everyone has a very high risk level of being bitten by an infected mosquito.

If most villages fall at the high or low end of mosquito levels, dichotomizing the mosquito levels would make for an almost perfect correlation between infection level and mosquito level at the ecological level. At the individual level within these extreme communities, however, no association would be seen between exposure and infection.

Even in those villages that do not fall into these extremes of exposure, we expect a stronger association between exposure and disease in an ecological analysis than in an individual analysis. To demonstrate this, we constructed a compartmental model for the transmission of an SIR infection in which all individuals were classified as exposed or unexposed, independently of their infection status.

There are several different ways that an exposure can affect infection risk. It can (1) alter the biologically based susceptibility of the uninfected individual in the contact, (2) alter the contagiousness of the infected person in the contact through a biological action that increases the number of organisms that person excretes, (3) change the number or viability of organisms transmitted between individuals even when susceptibility or contagiousness is not biologically altered, (4) increase the number of contacts made by individuals, or (5) alter those persons whom an individual contacts even though the number of contacts that individual makes is not altered. The only effect we included in our model was the fourth: to increase the number of contacts. The mathematical formulation of the model or its implementation using STELLA II software (High Performance Systems, Lyme, NH) can be obtained from Dr Koopman.

We set model parameter values such that when no one in the population is exposed, the basic reproduction ratio is just less than 1 and exposure increases the number of contacts an individual has by 10%. We then varied the proportion of the exposed population in a series of simulations. The resulting proportion of the overall population that is infected at the end of the epidemic and the relative risk of infection between exposed and

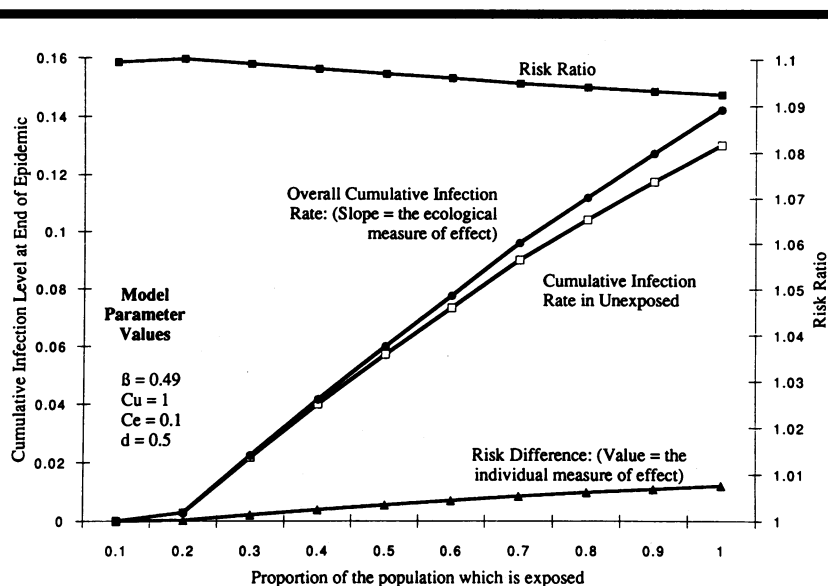


FIGURE 2—Simulated infection levels, risk differences, and risk ratios as a function of the proportion of the population that is exposed.

unexposed individuals are shown in Figure 2.

The top line in this figure represents the ratio of infection risk at the end of the epidemic in exposed and unexposed individuals at progressive levels of exposure in the population. This risk ratio starts off at the level of the rate ratio and decreases as the exposure level increases. The divergence in the quantitative values of risks and rates with increasing risks accounts for this falloff.

In contrast to the small differences between exposed and unexposed groups within populations, there are much larger differences in risk between populations. We see that as the exposed fraction of the population goes from 0.2 to 1.0, the infected fraction goes from 0 to 0.14. Given this increase in infection, the extrapolated increase in infection risk between a population with no exposure and one with complete exposure would be  $(0.14/0.8) = 17.5\%$ . An ecological regression would thus detect a risk difference of 0.175. At the individual level, the risk differences would range from 0 when less than 20% of the population is exposed to 0.013 when almost 100% of the population is exposed.

These simulation results demonstrate that even if there were no third variable effects, no misclassification of exposure or disease, and no extreme populations in the study (i.e., those in which almost no one got infected or those in which almost everyone got infected),

transmission phenomena still could have generated the degree of difference between individual- and ecological-level associations observed in the dengue study discussed earlier.

### *The Nature of Models Leading to Erroneous Attribution of Individual Effects*

The individual effects models that have been used to explain or justify the ecological fallacy are all models in which disease is generated by population processes that are linearly dynamic, with subsequent independence between the outcomes experienced by different individuals.<sup>1-13</sup> Greenland and Robins recognize that infectious and behavioral outcomes are excluded from such analyses,<sup>11</sup> but other authors do not acknowledge this limitation.

This deficiency in the analyses of the ecological fallacy derives from a failure to distinguish the nonlinearity of disease dynamics from the nonlinearity of exposure-disease relationships. These are two very different phenomena. Most epidemiologists and biostatisticians never consider that their analyses make assumptions relevant to the dynamic processes at the population level. They incorrectly presume that assumptions about the distributions and relationships between variables are sufficient for their analytic purposes.

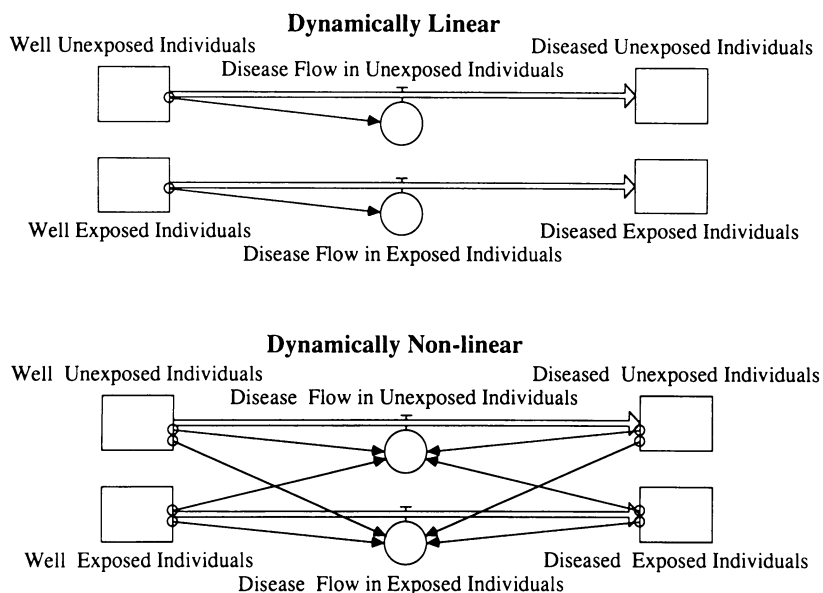


FIGURE 3—An illustration of linear and nonlinear dynamic systems.

Let us now elaborate on how transmission phenomena generate nonlinear dynamics at the population level. A dynamically linear system is illustrated on the top of Figure 3. Individuals flow through the arrow pipe from the well state to the diseased state. The spigot that controls the number of individuals flowing per unit of time is a function only of how many well individuals there are to begin with. The differential equations describing these flows will only have first-degree terms, and so the system is dynamically linear. The situation at the bottom of Figure 3 corresponds to a transmission model. The number of individuals flowing from the well to the diseased state depends on the proportion of contacted individuals who are in the infected state. The differential equations describing the flow will have second-degree terms, and so the dynamics will be nonlinear.

The underlying assumption of linear dynamics is not explicitly expressed in any of the models presented in prior work discussing the analysis of ecological data.<sup>1-13</sup> That is to be expected, since the traditions of epidemiological inquiry derive from the analysis of data expressing exposure-disease relationships rather than dynamic processes. But in every case, an implicit assumption is evident that the outcomes in any study subject are independent of the outcomes in other study subjects. That means that population dynamics are linear. The implicit nature

of such assumptions as opposed to an explicit statement of assumptions amplifies the risk from unsound assumptions.

As Halloran and Struchiner point out,<sup>18</sup> this basic difference between dynamically linear and nonlinear models was described more than 75 years ago by Sir Ronald Ross.<sup>19</sup> Unlike the usual epidemiologist of today, this physician epidemiologist regularly constructed dynamic models of the disease processes in which he was interested. He divided what we would now call events into "independent happenings" corresponding to dynamically linear models and "dependent happenings" corresponding to nonlinear models. Since Ross' time, the quantitative theory behind the methods of epidemiology have been developed mostly within the context of noncontagious diseases so that independent happenings have become an implicit assumption that is sometimes not recognized.

### ***Implications for Detecting Transmission Modes and Transmission Risk Factors***

We have demonstrated how transmission dynamics in a randomly mixing population can obscure the effect of a risk factor for transmission when an individual-level analysis with the underlying assumptions of dynamic linearity is undertaken. Recently we demonstrated an additional

mechanism whereby transmission dynamics can obscure the effects of risk factors for transmission.<sup>20</sup> That mechanism involves nonrandom patterns of contact. Since our model simulation here assumed a proportionate mixing process with random contacts, it assumed away this other mechanism.

There is a third way in which a focus on classical individual effect measures obscures the detection of risk factor effects. Such effect measures lead us to focus on only those risk factor effects that can be measured in the study subject at risk of developing infection. Risk factor effects 2 and 5 specified in the section on quantitative models of transmission effects above do not relate to exposures measured in the individuals at risk, but rather to characteristics of the individuals with whom they have contact. Those characteristics influence contagiousness and contact patterns, which are absolutely central to transmission dynamics.

The evaluation of vaccines is subject to all the pitfalls we have just outlined. The failure to assess contagiousness effects may be especially important. Some vaccines, like the Salk polio vaccine, may not significantly decrease the rate of infection in the vaccinated. They can, however, decrease the excretion of organisms in subsequently infected individuals and stop the circulation of the agent. If these effects on contagiousness are neglected in vaccine evaluations, we may make the wrong decisions regarding new vaccines like the proposed rotavirus and human immunodeficiency virus vaccines.<sup>21</sup>

The transmission dynamic effects elucidated in this paper and those due to contact patterns<sup>20</sup> are a major reason why there is still so much ignorance in epidemiology about modes of transmission. We have not evaluated in the field the relative importance of the different modes of transmission (direct skin-to-skin contact, indirect contact, airborne spread, or large droplet spread) for any of the major pathogens causing respiratory infections. What we do know for a few of these infections, like rhinoviruses, derive from experiments. Some of these experiments indicate that transmission might be controllable,<sup>22</sup> but they have not resolved whether interruption of direct hand-to-hand contact or airborne transmission will be the more efficient control strategy.<sup>23,24</sup>

Experiments that do not correspond to natural conditions cannot define public health strategies. To define the contributions of different modes of transmission to overall transmission dynamics will require

the development of new epidemiological study designs that overcome the difficulties we have outlined. That implies the use of ecological data and ecological models. Let us now discuss how that might proceed.

### Implications for Study Design

The major implication for study design of the ideas presented here is that contact patterns need to be measured in epidemiological studies of infectious diseases. The data collected need to reflect individual contact episodes or general exposure frequencies to different populations. If causally predictive effect estimates are to be made from such data, the analysis needs to employ models that parameterize causal effects upon or within contacts. A separate implication is that study designs that maximize the contrasts in what type of individuals are contacted will maximize the detection of effects that are manifest through those contacts.

For sexually transmitted infections, the appropriate ecological level at which data should be gathered is that of a sexual partnership. The pattern of such partnerships then needs to be modeled at an even higher ecological level. We suggest that contact patterns can feasibly be measured by means of sample designs in which statistical target population includes the majority of the partners of study subjects. The data to be collected include individual risk and infection characteristics of the study subjects, characteristics of partners that are perceptible to the study subject, and characteristics of partnerships that could theoretically be reported by either partner. We are currently developing appropriate statistics for these types of data. It is important to note that such data cannot be adequately conceptualized in terms of separate cases having several variable values determined for them. That is to say, such data cannot be fit into the standard framework for ecological data that we presented earlier in this paper. This usual conceptualization of an epidemiological data set is inadequate because it does not define what classes of individuals are in contact with each other and it has no place for data that are specific to a contact with a particular type of individual.

For non-sexually transmitted infections, the study designs most capable of describing differential contact patterns are cross-population studies. These involve different sets of individuals who might be assumed to have contact within but not across populations. The example

and simulations presented here suggest that cross-population analyses could have considerably more power to detect effects of risk factors on transmission than could analyses that examine effects within separate populations and then combine them. This runs counter to the usual epidemiological wisdom that it is always best to stratify comparisons within the most narrow and homogeneous population groups.

To go beyond detection of effects and begin to measure how risk factors affect contacts between individuals or the outcomes of those contacts, just designing cross-population studies will be insufficient. Models that parameterize those effects will have to be used in the analysis.

### Summary and Conclusions

Infectious diseases and social behaviors have a characteristic that is inconsistent with the assumptions of most statistical analyses in epidemiology. That characteristic is that the outcome in one individual influences the outcomes and alters the risk factor effects in other individuals. When dependence between outcomes is generated by a risk factor affecting transmission rather than a risk factor acting directly on the exposed individual, neither the risk difference determined from individual data nor the ecological regression coefficient reflect causal parameters in the processes generating infection or behaviors. Thus, it is meaningless to say that one or the other of them is more correct.

When a risk factor affects transmission, exposure effect statistics should estimate parameters reflecting the effects of exposure on the existence or the outcome of contacts that could transmit infection. They should reflect either the rate at which different types of individuals contact each other or the probability of transmission between infected and susceptible individuals in those contacts. These contact rates and transmission probabilities are the basic parameters of transmission models. Transmission models should thus be used in the analysis of data meant to examine the relationships between risk factors for transmission and infection levels.

More generally, we could state that for any epidemiological analysis, a key step is to define the individual, social, or ecological context within which causes are acting. Then one should collect data relevant to that level of action and analyze the data, using models that parameterize the causal action at the appropriate level.

To estimate risk factor effects, one would prefer statistics that have well-known variance characteristics and that can be shown theoretically to converge upon the underlying parameter of interest. Estimators with these characteristics are not widely available for the estimation of transmission parameters. They are available only for limited situations, such as when infection is studied in families.<sup>25</sup> The development of comparable procedures for the analysis of ecological data where the groupings involve larger populations is needed. In the meantime, to develop a practical science of transmission system analysis capable of assessing modes of transmission and estimating the contagiousness and susceptibility effects of risk factors and vaccines, we will have to use estimation procedures for the parameters of transmission models that have less desirable characteristics.

The guiding principle to avoid error that comes from our focus does not relate to grouping rules, as in the case of Susser's approach.<sup>1</sup> Rather, it is that models used in analysis should correspond as closely as possible to the dynamic processes generating the data. In the case of infectious diseases, this means that transmission models should be used. This maximizes the usefulness of parameters estimated because those parameters have predictive significance when applied in the models used to estimate them. □

### Acknowledgments

Critical reviews of this manuscript by Sander Greenland and Carl Simon significantly contributed to its composition.

### References

1. Susser M. The logic in ecological: II. the logic of design. *Am J Public Health*. 1994; 84:830-835.
2. Von Korff M, Koepsell T, Curry S, Diehr P. Multi-level analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol*. 1992;135:1077-1082.
3. Brenner H, Savitz DA, Jöckel KH, Greenland S. Effects of nondifferential exposure misclassification in ecologic studies. *Am J Epidemiol*. 1992;135:85-95.
4. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol*. 1989;18:269-274.
5. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol*. 1988;127: 893-904.
6. Richardson S, Stücker I, Hémon D. Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. *Int J Epidemiol*. 1987;16:111-120.
7. Lincoln JR, Zeitz G. Organizational properties from aggregate data: separating individual and structural effects. *Am Sociol Rev*. 1980;45:391-408.



8. Langbein LI, Lichtman AJ. *Ecological Inference*. London, England: Sage; 1978.
9. Green HAJ. *Aggregation in Economic Analysis*. Princeton, NJ: Princeton University Press; 1964.
10. Greenland S. Divergent biases in ecologic and individual-level studies. *Stat Med*. 1992;11:1209-1233.
11. Greenland S, Robin J. Accepting the limits of ecologic studies: Drs Greenland and Robin reply to Drs Piantodosi and Cohen. *Am J Epidemiol*. 1994;139:769-771.
12. Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health*. 1982;72:1336-1344.
13. Cohen BL. Ecological versus case-control studies for testing a linear-no threshold dose-response relationship. *Int J Epidemiol*. 1990;19:680-684.
14. Koopman JS, Prevots DR, Vaca-Marin MA, et al. Determinants and predictors of dengue infection in Mexico. *Am J Epidemiol*. 1991;133:1168-1178.
15. Dixon WJ, Brown MB, Engelman L, et al. *BMDP Statistical Software*. Berkeley, Calif: University of California Press; 1983.
16. Likosky WH, Calisher CH, Michelson AL, et al. An epidemiologic study of dengue type 2 in Puerto Rico, 1969. *Am J Epidemiol*. 1973;97:264-275.
17. Longini IM. The generalized discrete-time epidemic model with immunity: a synthesis. *Math Biosci*. 1986;81:1-23.
18. Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1992;2:331-338.
19. Ross R. An application of the theory of probabilities to the study of a priori pathometry. Part 1. *Proc R Soc Series A*. 1916;92:204-230.
20. Koopman JS, Longini IM, Jacquez JA, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol*. 1991;133:1199-1209.
21. Koopman JS, Simon CP, Jacquez JA. Assessing contagiousness effects of vaccines and risk factors for transmission. In: Kaplan EH, Brandeau ML, eds. *Modeling the AIDS Epidemic: Planning, Policy, and Prediction*. New York, NY: Raven Press; 1994:439-460.
22. D'Alessio DJ, Peterson JA, Dick CR, Dick EC. Transmission of experimental rhinovirus colds in volunteer married couples. *J Infect Dis*. 1976;133:28-36.
23. Dick EC, Jennings LC, Mink KA, Wartgow CD, Inhorn SL. Aerosol transmission of rhinovirus colds. *J Infect Dis*. 1987;156:442-448.
24. Gwaltney JM Jr, Moskalski PB, Hendley JO. Hand-to-hand transmission of rhinovirus colds. *Ann Intern Med*. 1978;88(4):463-467.
25. Longini IM, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases: risk specific household and community transmission parameters. *Am J Epidemiol*. 1988;128:845-859.

## *An Invitation for the AAAS Resource Directory of Scientists with Disabilities*

The American Association for the Advancement of Science (AAAS) Project on Science, Technology, and Disability invites scientists and engineers with disabilities to be included in the third edition of the *Resource Directory of Scientists and Engineers with Disabilities*. Potential candidates for the directory must hold, or be working toward, a degree in a scientific, engineering, or medical discipline or currently be employed in a scientific field. Funded by the National Science Foundation, the project's directory has assisted hundreds of individuals in entering and advancing in scientific disciplines. The directory helps connect persons with disabilities and their families with professors, teachers, and counselors who can serve as role models and mentors.

The directory lists scientists, mathematicians, and engineers from all parts of the country with their disciplines, degrees, and disabilities. Individuals include professionals

who were born with a disability and those who acquired their disability midcareer. Persons listed in the directory are also asked to consult for academia, government agencies, and industry, as well as serve on peer review panels and symposia.

Established in 1975, the AAAS Project on Science, Technology, and Disability has sought and shared expert advice from scientists and engineers with disabilities. Since the passage of the Americans with Disabilities Act (ADA), the directory has become a valuable source of expertise.

To be included in the directory, or for more information, please contact Laureen Summers, Program Associate, or Patricia A. Thompson, Editorial Specialist, AAAS Project on Science, Technology, and Disability, AAAS, 1333 H St, NW, Washington, DC 20005; or call (202) 326-6645 (V/TDD). Information can also be sent via fax to (202) 371-9849.