Annu. Rev. Public Health 1995. 16:61–81

ECOLOGIC STUDIES IN EPIDEMIOLOGY: Concepts, Principles, and Methods

Hal Morgenstern
Department of Epidemiology and Center for Occupational and Environmental Health, University of California, Los Angeles, School of Public Health, Los Angeles, California 90024–1772

KEY WORDS: epidemiologic methods, study design, sources of bias, causal inference, aggregate studies

ABSTRACT

An ecologic study focuses on the comparison of groups, rather than individuals; thus, individual-level data are missing on the joint distribution of variables within groups. Variables in an ecologic analysis may be aggregate measures, environmental measures, or global measures. The purpose of an ecologic analysis may be to make biologic inferences about effects on individual risks or to make ecologic inferences about effects on group rates. Ecologic study designs may be classified on two dimensions: (*a*) whether the primary group is measured (exploratory vs analytic study); and (*b*) whether subjects are grouped by place (multiple-group study), by time (time-trend study), or by place and time (mixed study). Despite several practical advantages of ecologic studies, there are many methodologic problems that severely limit causal inference, including ecologic and cross-level bias, problems of confounder control, within-group misclassification, lack of adequate data, temporal ambiguity, collinearity, and migration across groups.

INTRODUCTION

An ecologic or aggregate study focuses on the comparison of groups, rather than individuals. The underlying reason for this focus is that individual-level data are missing on the joint distribution of at least two and perhaps all

0163-7525/95/0501-0061$05.00

variables within each group; in this sense, an ecologic study is an incomplete design (35). Ecologic studies have been conducted by social scientists for more than a century (14a) and have been used extensively by epidemiologists in many research areas. Nevertheless, the distinction between individual-level and group-level (ecologic) studies and the inferential implications are far more complicated and subtle than they first appear. Before 1980, ecologic studies were usually presented in the first part of epidemiology textbooks as simple descriptive analyses in which disease rates are stratified by place and/or time to generate or test hypotheses; little attention was given to statistical methods or inference (e.g. 41). The purpose of this review is to provide a methodologic overview of ecologic studies that emphasizes study design and causal inference. Although ecologic studies are easily and inexpensively conducted, the results are often difficult to interpret.

CONCEPTS AND RATIONALE

Before discussing the design and interpretation of ecologic studies, we must first define the concepts of ecologic measurement, analysis, and inference.

Levels of Measurement

The sources of data used in epidemiologic studies typically involve direct observations of individuals (e.g. age and sex), sometimes subindividual parts (e.g. intraocular pressure of each eye), and occasionally groups or regions (e.g. air pollution and social disorganization). These direct observations are then organized to measure specific variables in the study population: Individual-level variables are properties of individuals, and ecologic variables are properties of groups. To be more specific, ecologic measures may be classified into three types:

1. *Aggregate measures* are summaries (e.g. means or proportions) of observations derived from individuals in each group (e.g. the proportion of smokers or median family income).
2. *Environmental measures* are physical characteristics of the place in which members of each group live or work (e.g. air-pollution level or hours of sunlight). Note that each environmental measure has an analogue at the individual level, and these individual exposures, or doses, usually vary among members of each group, though they may remain unmeasured.
3. *Global measures* are attributes of groups or places for which there is no distinct analogue at the individual level, unlike aggregate and environmental measures (e.g. population density, level of social disorganization, or the existence of a specific law).

Levels of Analysis

The unit of analysis is the common level for which the data on all variables are reduced and analyzed. In an *individual-level analysis*, a value for each variable is assigned to every subject in the study. It is possible, even common in environmental epidemiology, for one or more variables to be ecologic measures. For example, the average pollution level of each county might be assigned to every resident of that county.

In a *completely ecologic analysis*, all variables (exposure, disease, and covariates) are ecologic measures, so the unit of analysis is the group (e.g. region, worksite, school, demographic stratum, or time interval). Thus, within each group, we do not know the joint distribution of any combination of variables at the individual level (e.g. the frequencies of exposed cases, unexposed cases, exposed noncases, and unexposed noncases); all we know is the marginal distribution of each variable (e.g. the proportion exposed and the disease rate—the T frequencies in Figure 1).

In a *partially ecologic analysis* of three or more variables, we have additional information on certain joint distributions (the M and/or N frequencies in Figure 1 and/or rarely the L frequencies); but we still do not know the full joint distribution of all variables within each group (i.e. the ? cells in Figure 1 are missing). For example, in an ecologic study of cancer incidence by county, the joint distribution of age (a covariate) and disease status within each county (the M frequencies in Figure 1) might be obtained from the census and a population tumor registry.

Multilevel analysis is a special type of modeling technique that combines analyses conducted at two or more levels (6, 71, 72). For example, an individual-level analysis might be conducted in each group, followed by an ecologic analysis of all groups using the results from the individual-level analyses. This approach is described in a later section.

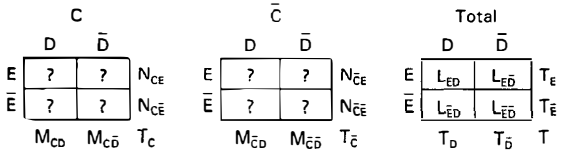

Figure 1 Joint distribution of exposure status (E vs $\bar{E}$), disease status (D vs $\bar{D}$), and covariate status (C vs $\bar{C}$) in each group of a simple ecologic analysis: T frequencies are the only data available in a completely ecologic analysis of all three variables; M frequencies require additional data on the joint distribution of C and D within each group; N frequencies require additional data on the joint distribution of E and C within each group; L frequencies require additional data on the joint distribution of E and D within each group (rarely available); and ? cells are missing in an ecologic analysis.

Levels of Inference

The underlying goal of a given epidemiologic study or analysis may be to make *biologic* (or *biobehavioral*) *inferences* about effects on individual *risks* or to make *ecologic inferences* about effects on group *rates* (45). The target level of causal inference, however, does not always match the level of analysis. For example, the purpose of an ecologic analysis may be to make a biologic inference about the effect of a specific exposure on disease risk. As we see later in this review, such *cross-level inferences* are particularly vulnerable to bias.

If the objective of a study is to estimate the *biologic effect* of wearing a motorcycle helmet on the risk of motorcycle-related mortality among motorcycle riders, the target level of causal inference is biologic. On the other hand, if the objective is to estimate the *ecologic effect* of helmet-use laws on the motorcycle-related mortality rate of riders in different states, the target level of causal inference is ecologic. Note that the magnitude of this ecologic effect depends not only on the biologic effect of helmet use but also on the degree and pattern of compliance with the law in each state. Furthermore, the validity of the ecologic-effect estimate depends on our ability to control for differences among states in the joint distribution of confounders, including individual-level variables such as age and amount of motorcycle riding.

We might also be interested in estimating the *contextual effect* of an ecologic exposure on individual risk, which is also a form of biologic inference (5, 64). If the ecologic exposure is an aggregate measure, we would generally want to separate its effect from the effect of its individual-level analogue. For example, we might estimate the contextual effect of living in a poor area on the risk of disease, controlling for individual poverty level (33). Similarly, in evaluating motorcycle-helmet laws in the U.S., we might want to estimate the contextual effect of living in a state that mandates helmet use on the risk of motorcycle-related mortality in riders, controlling for individual helmet use. Contextual effects are also relevant in infectious-disease epidemiology, where the risk of disease depends on the prevalence of the disease in others with whom the individual has contact (37, 65).

Rationale for Ecologic Studies

There are several reasons for the widespread use of ecologic studies in epidemiology, despite frequent cautions about their methodologic limitations:

1. *Low cost and convenience* Ecologic studies are inexpensive and take little time because various secondary data sources, each involving different information needed for the analysis, can easily be linked at the aggregate level. For example, data obtained from population registries, vital records,

large surveys, and the census are often linked at the state, county, or census-tract level.

2. *Measurement limitations of individual-level studies* In environmental epidemiology and other research areas, we often cannot accurately measure relevant exposures or doses at the individual level for large numbers of subjects—at least not with available time and resources. Thus, the only practical way to measure the exposure may be ecologically (45, 46). This advantage is especially true when investigating apparent clusters of disease in small areas (66). Sometimes individual-level exposures, such as dietary factors, cannot be measured accurately because of substantial within-person variability; yet ecologic measures might accurately reflect group averages (31).
3. *Design limitations of individual-level studies* Individual-level studies may not be practical for estimating exposure effects if the exposure varies little within the study area. However, ecologic studies covering a much wider area might be able to achieve substantial variation in mean exposure across groups (e.g. 50).
4. *Interest in ecologic effects* As noted above, the stated purpose of a study may be to assess an ecologic effect, i.e. the target level of inference may be ecologic rather than biologic. Ecologic effects are particularly relevant when evaluating the impacts of population interventions such as new programs, policies, or legislation.
5. *Simplicity of analysis and presentation* In large, complex studies conducted at the individual level, it may be conceptually and statistically simpler to perform ecologic analyses and to present ecologic results than to do individual-level analyses. For example, data from large, periodic surveys, such as the National Health Interview Survey, are often analyzed ecologically by treating some combination of year, region, and demographic group as the unit of analysis.

STUDY DESIGNS

In an ecologic study design, the planned unit of analysis is the group. Ecologic designs may be classified on two dimensions: the method of exposure measurement and the method of grouping (35, 45). Regarding the first dimension, an ecologic design is called *exploratory* if the primary exposure of potential interest is not measured, and *analytic* if the primary exposure variable is measured and included in the analysis. In practice, this dimension is a continuum, since most ecologic studies are not conducted to test a single hypothesis. Regarding the second dimension, the groups of an ecologic study may be identified by place (multiple-group design), by time (time-trend design), or by a combination of place and time (mixed design).

Multiple-Group Study

EXPLORATORY In this type of exploratory study, we compare the rate of disease among many regions during the same period. The purpose is to search for spatial patterns that might suggest an environmental etiology or more specific etiologic hypotheses. For example, the National Cancer Institute (NCI) mapped the age-adjusted cancer mortality rates in the U.S. by county for the period 1950–69 (42). For oral cancers, they found a striking difference in geographic patterns by sex: Among men, the mortality rates were greatest in the urban Northeast, but among women, the rates were greatest in the Southeast. These findings led to the hypothesis that snuff dipping, which is common among rural southern women, is a risk factor for oral cancers (2). The results of a subsequent case-control study supported this hypothesis (70).

Exploratory ecologic studies may also involve the comparison of rates between migrants and their offspring and residents of their countries of emigration and immigration (31, 41). If the rates differ appreciably between the countries of emigration and immigration, migrant studies often yield results suggesting the influence of certain types of risk factors for the disease under study. For example, if US immigrants from Japan have rates of a disease similar to US whites but much lower than Japanese residents, the difference may be due to environmental or behavioral risk factors operating during adulthood. However, the interpretation of results from these studies is often limited by differences between countries in the classification and detection of disease or cause of death.

In mapping studies, such as the NCI investigation, a simple comparison of rates across regions is often complicated by two statistical problems. First, regions with smaller numbers of observed cases show greater variability in the estimated rate; thus the most extreme rates tend to be observed for those regions with the fewest cases. Second, nearby regions tend to have more similar rates than do distant regions (i.e. autocorrelation) because unmeasured risk factors tend to cluster in space. Statistical methods for dealing with both problems have been developed by fitting the data to an autoregressive spatial model and using empirical Bayes techniques to estimate the smoothed rate for each region (9, 44, 47). The degree of spatial autocorrelation or clustering can be measured to reflect environmental effects on the rate of disease (68, 69). The empirical Bayes approach can also be applied to data from analytic multiple-group studies (described below) by including covariates in the model (e.g. 8, 12).

ANALYTIC In this type of study, we assess the ecologic association between the average exposure level or prevalence and the rate of disease among many groups. This is the most common ecologic design; typically, the unit of analysis is a geopolitical region. For example, Hatch & Susser (29) examined the

association between background gamma radiation and the incidence of childhood cancers between 1975 and 1985 in the region surrounding a nuclear power plant. Average radiation levels for each of 69 tracts in the region were estimated from a 1976 aerial survey. The authors found positive associations between radiation level and the incidence of leukemia (an expected finding) as well as solid tumors (an unexpected finding).

Data analysis in this type of multiple-group study usually involves fitting the data to a mathematical model. For example, Prentice & Sheppard (51) proposed a linear relative rate model using iteratively reweighted least-squares procedures to estimate the model parameters. Prentice & Thomas (52) also considered an exponential relative rate model, which, they argue, may be more parsimonious than the linear-form model for specifying covariates. These methods can be applied to data aggregated by place and/or time (to be discussed below). Use of ecologic modeling to estimate exposure effects is described in the next section.

Time-Trend Study

EXPLORATORY An exploratory time-trend or time-series study involves a comparison of the disease rates over time in one geographically defined population. In addition to providing graphical displays of temporal trends, time-series data can also be used to forecast future rates and trends. This latter application, which is more common in the social sciences than in epidemiology, usually involves fitting the outcome data to autoregressive integrated moving average (ARIMA) models (30, 48). The method of ARIMA modeling can also be extended to evaluate the impact of a population intervention (43), to estimate associations betweens two or more time-series variables (7, 48), and to estimate associations in a mixed ecologic design (60; see below).

A special type of exploratory time-trend analysis often used by epidemiologists is age-period-cohort (or cohort) analysis. Through graphical displays or formal modeling techniques, the objective of this approach is to estimate the separate effects of three time-dependent variables on the rate of disease: age, period (calendar time), and birth cohort (year of birth) (32, 35). Because of the linear dependency of these three variables, there is an inherent statistical limitation (identification problem) with the interpretation of age-period-cohort results. The problem is that each data set has alternative explanations with respect to the combination of age, period, and cohort effects; there is no unique set of effect parameters when all three variables are considered simultaneously. The only way to decide which interpretation should be accepted is to consider the findings in light of prior knowledge and, possibly, to constrain the model by ignoring one effect.

Lee et al (40) conducted an age-period-cohort analysis of melanoma mortality among white males in the U.S. between 1951 and 1975. They concluded

that the apparent increase in the melanoma mortality rate was due primarily to a cohort effect. That is, persons born in more recent years experienced throughout their lives a higher rate than did persons born earlier. In a subsequent paper, Lee (39) speculated that this cohort effect might reflect increases in sunlight exposure or sunburning during youth.

ANALYTIC In this type of time-trend study, we assess the ecologic association between change in average exposure level or prevalence and change in disease rate in one geographically defined population. As with exploratory designs, this type of assessment can be done by simple graphical displays or by time-series regression modeling (e.g. 48). With either approach, however, the interpretation of findings is often complicated by two problems. First, changes in disease classification and diagnostic criteria can produce very misleading results. Second, the latency of the disease with respect to the primary exposure may be long, variable across cases, or simply unknown. Thus, employing an arbitrary lag between observations—or an empirically defined lag that maximizes the estimated association between the two trends—can also produce misleading results (28).

Darby & Doll (13) examined the associations between average annual absorbed dose of radiation fallout from weapons testing and the incidence rate of childhood leukemia in three European countries between 1945 and 1985. Although the leukemia rate varied over time in each country, they found no convincing evidence that these changes were attributable to changes in fallout radiation.

Mixed Study

EXPLORATORY The mixed ecologic design combines the basic features of the multiple-group study and the time-trend study. Time-series (ARIMA) modeling or age-period-cohort analysis can be used to describe or predict trends in the disease rate for multiple populations. For example, to test Lee's (39) hypothesis that changes in sunlight exposure during youth can explain the observed increase in melanoma mortality in the U.S., we might conduct an age-period-cohort analysis, stratifying on region according to approximate sunlight exposure (without measuring the exposure). Assuming the amount of sunlight in the regions has not changed differentially over the study period, we might expect the cohort effect described above to be stronger for sunnier regions.

ANALYTIC In this type of mixed ecologic design, we assess the association between change in average exposure level or prevalence and change in disease rate among many groups. Thus the interpretation of estimated effects is en-

hanced because two types of comparisons are made simultaneously: change over time within groups and differences among groups. For example, Crawford et al (11) evaluated the hypothesis that hard drinking water (i.e. water with a high concentration of calcium and magnesium) is a protective risk factor for cardiovascular disease (CVD) mortality. They compared the absolute change in CVD mortality rate between 1948 and 1964 in 83 British towns, by water-hardness change, age, and sex. In all sex-age groups, especially for men, the authors found an inverse association between water-hardness change and CVD mortality. In middle-aged men, for example, the increase in CVD mortality was less in towns that made their water harder than in towns that made their water softer.

EFFECT ESTIMATION

A major quantitative objective of most epidemiologic studies is to estimate the effect of one or more exposures on disease occurrence in a well-defined population at risk. A measure of effect in this context is not just any measure of association, such as a correlation coefficient; rather, it reflects a particular causal parameter, i.e. a counterfactual contrast in disease occurrence (21, 24, 27, 46, 58). In studies conducted at the individual level, effects are usually estimated by comparing the rate or risk of disease, in the form of a ratio or difference, for exposed and unexposed populations. In multiple-group ecologic studies, however, we cannot estimate effects directly in this way because of the missing information on the joint distribution within groups. Instead, we regress the group-specific disease rates (Y) on the group-specific exposure prevalences (X). For example, fitting the data to a linear model produces the following prediction equation: $\hat{Y} = B_0 + B_1X$, where B_0 and B_1 are the estimated intercept and slope, using ordinary least-squares methods. The estimated biologic effect of the exposure (at the individual level) can be derived from the regression results (1, 19). The predicted disease rate ($\hat{Y}$) in a group that is entirely exposed is $B_0 + B_1(1) = B_0 + B_1$, and the predicted rate in a group that is entirely unexposed is $B_0 + B_1(0) = B_0$. Therefore, the estimated rate difference is B_1 and the estimated rate ratio is $1 + B_1/B_0$. Note that this ecologic method of effect estimation requires rate predictions be extrapolated to both extreme values of the exposure variable (i.e. $X = 0$ and 1), which are likely to lie well beyond the observed range of the data. It is not surprising, therefore, that different model forms (e.g. log-linear vs linear) can lead to very different estimates of effect (22). Fitting a linear model, in fact, may lead to negative, and thus meaningless, estimates of the rate ratio.

As an illustration of rate-ratio estimation in an ecologic study, consider Durkheim's (16) examination of religion and suicide in four groups of Prussian provinces between 1883 and 1890 (see Figure 2). The groups were formed by

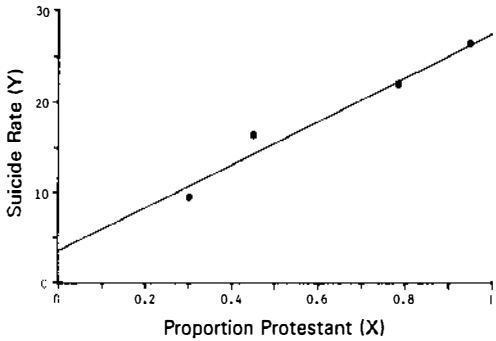

Figure 2 Suicide rate (Y, per 10^5/year) by proportion Protestant (X) for four groups of Prussian provinces, 1883–90. The four observed points (X, Y) are (0.30, 9.56), (0.45, 16.36), (0.785, 22.00), and (0.95, 26.46); the fitted line is based on unweighted least-squares regression [Source: Adapted from Durkheim (16)].

ranking 13 provinces according to the proportion (X) of the population that was Protestant. Using ordinary least-squares linear regression, we estimate the suicide rate ($\hat{Y}$, per 10^5/year) in each group to be $3.66 + 24.0(X)$. Therefore, the estimated rate ratio, comparing Protestants with other religions, is $1 + (24.0/3.66) = 7.6$. Note in Figure 2 that the fit of the linear model is excellent ($R^2 = 0.97$).

There are two methods used to control for confounders in multiple-group ecologic analyses. The first is to treat ecologic measures of the confounders as covariates ($\mathbf{Z}$) in the model, e.g. percent male and percent white in each group. If the individual-level effects of the exposure and covariates are additive (i.e. if the disease rates follow a linear model), then the ecologic regression of Y on X and $\mathbf{Z}$ will also be linear with the same coefficients (22, 38). That is, the estimated coefficient for the exposure variable can be interpreted as the rate difference adjusted for other covariates, analogously to the crude estimate discussed above.

The second method used to control for confounders in ecologic analyses is rate standardization for these confounders (57), followed by regression of the standardized rates as the outcome variable. Note that this method requires additional data on the joint distribution of the covariate and disease within each group (i.e. the M frequencies in Figure 1). Nevertheless, it cannot be expected to reduce bias unless all predictors in the model (X and $\mathbf{Z}$) are mutually standardized for the same confounders (22, 25, 56). Standardization of the exposure prevalences, for example, requires data on the joint distribution of the covariate and exposure within groups (i.e. the N frequencies in Figure 1); however, this information is not often available in ecologic studies.

As in individual-level analyses, product terms (e.g. XZ) are often used in ecologic analyses to model interaction effects, i.e. to assess effect modification. In ecologic analyses, however, the product of X and Z (both group averages) is not, in general, equal to the average product of the exposure (x) and covariate (z) at the individual level within groups. Assuming a linear model, XZ will be equal to the mean xz in each group only if x and z are uncorrelated within groups (22). Thus, as pointed out in the next section, interaction (nonadditive) effects at the individual level complicate the interpretation of ecologic results.

METHODOLOGIC PROBLEMS

Despite the many practical advantages of ecologic studies mentioned previously, there are several methodologic problems that may severely limit causal inference, especially biologic inference.

Ecologic Bias

The major limitation of ecologic analysis for making causal inferences is ecologic bias, which is the failure of expected ecologic effect estimates to reflect the biologic effect at the individual level (18, 19, 25, 45, 54). In addition to the usual sources of bias that threaten individual-level analyses (35, 57), the underlying problem of ecologic analyses for estimating biologic effects is heterogeneity of exposure level and/or covariate levels within groups; as noted earlier, this heterogeneity is not fully captured with ecologic data because of missing information on joint distributions (see Figure 1). Robinson (55) was the first to describe mathematically how ecologic associations could differ from the corresponding associations at the individual level within groups of the same population. He expressed this relationship in terms of correlation coefficients; this relationship was later extended by Duncan et al (15) to regression coefficients in a linear model. The phenomenon became widely known as the *ecologic(al) fallacy* (61), and the magnitude of the ecologic bias may be severe in practice (10, 17, 54, 62, 63).

As an illustration of ecologic bias, consider again Durkheim's data on religion and suicide (Figure 2). The estimated rate ratio of 7.6 in the ecologic analysis may not mean that the suicide rate was nearly eight times greater in Protestants than in non-Protestants. Rather, since none of the regions were entirely Protestant or non-Protestant, it may have been non-Protestants (primarily Catholics) who were committing suicide in predominantly Protestant provinces. It is certainly plausible that members of a religious minority might have been more likely to take their own lives than were members of the majority. The implication of this alternative explanation is that living in a predominantly Protestant area has a contextual effect on suicide risk among

non-Protestants, i.e. there is an interaction effect at the individual level between religion and religious composition of one's area of residence.

Interestingly, Durkheim (16) compared the suicide rates (at the individual level) for Protestants, Catholics, and Jews living in Prussia. From his data, we find that the rate was about twice as high in Protestants as in other religious groups. Thus, there appears to be substantial ecologic bias (i.e. comparing rate-ratio estimates of about 2 vs 8). Durkheim, however, failed to notice this quantitative difference because he did not actually estimate the magnitude of the effect in either analysis.

Greenland & Morgenstern (25) showed that ecologic bias can arise from three sources when using simple linear regression to estimate the crude exposure effect: The first may operate in any type of study; the latter two are unique to ecologic studies (i.e. *cross-level bias*), but are defined in terms of individual-level associations.

1. *Within-group bias* The exposure effect within groups may be biased by confounding, selection methods, or misclassification (35, 57). Thus, for example, if there is positive net bias in every group, we would expect the ecologic estimate to be biased as well.
2. *Confounding by group* Ecologic bias may result if the background rate of disease in the unexposed population varies across groups, specifically if there is a nonzero ecologic (linear) correlation between mean exposure level and the background rate.
3. *Effect modification by group* Ecologic bias may also result if the rate difference for the exposure effect at the individual level varies across groups.

Confounding and effect modification by group (the sources of cross-level bias) can arise in three ways: (*a*) Extraneous risk factors (confounders or modifiers) are differentially distributed across groups; (*b*) the ecologic exposure variable has an effect on risk separate from the effect of its corresponding individual-level analogue, e.g. living in a predominantly Protestant area vs being Protestant (in the suicide example); or (*c*) disease risk depends on the prevalence of that disease in other members of the group, which is true of many infectious diseases (37).

Unfortunately, those conditions that produce ecologic bias cannot be observed in ecologic data. Furthermore, the fit of the ecologic regression model, in general, gives no indication of the presence, direction, or magnitude of ecologic bias. Thus, a model with excellent fit may yield substantial bias (e.g. Figure 2), and one model with a better fit than another model may yield more bias.

A potential strategy for reducing ecologic bias is to use smaller units in an ecologic study (e.g. counties instead of states) in order to make the groups

more homogeneous with respect to the exposure. On the other hand, this strategy might not be feasible because of the lack of available data aggregated at the same level, and it might lead to two other problems: greater migration between groups (see below) and less precise estimation of disease rates (45, 67).

Problems of Confounder Control

As already indicated, covariates are included in ecologic analyses to control for confounding, but the conditions for a covariate being a confounder are different at the ecologic and individual levels (25, 26). At the individual level, a risk factor must be associated with the exposure to be a confounder. In a multiple-group ecologic study, in contrast, a risk factor may produce ecologic bias (i.e. it may be an ecologic confounder) even if it is unassociated with the exposure in every group, especially if the risk factor is ecologically associated with the exposure across groups (22, 25). Conversely, a risk factor that is a confounder within groups may not produce ecologic bias if it is ecologically unassociated with the exposure across groups.

Control for confounders is more problematic in ecologic analyses than in individual-level analyses (22, 25, 26). Even when all variables are accurately measured for all groups, adjustment for extraneous risk factors may not reduce the ecologic bias produced by these risk factors. In fact, it is possible for such ecologic adjustment to increase bias. It follows from the principles presented in the previous section (25) that there will be no ecologic bias in a multiple-linear-regression analysis if the following conditions are met:

1. There is no residual within-group bias in exposure effect in any group because of confounding by unmeasured risk factors, selection methods, or misclassification.
2. There is no ecologic correlation between the mean value of each predictor and the background rate of disease in the joint reference (unexposed) level of all predictors.
3. The rate difference for each predictor is uniform across levels of the other predictors within groups (i.e. the effects are additive), and each rate difference is uniform across groups (i.e. group does not modify the effect of each predictor at the individual level).

These conditions are sufficient, but not necessary, for the ecologic estimate to be unbiased, i.e. there might be little or no bias even if none of these conditions are met. On the other hand, minor deviations from these conditions can produce substantial ecologic bias (22). Since the sufficient conditions for no ecologic bias cannot be checked with ecologic data alone, the unpredictable and potentially severe nature of such bias makes biologic inference from ecologic analyses particularly problematic. Prentice & Sheppard (51) have

suggested that ecologic data be supplemented with individual-level data from each group (or a representative sample) to enhance biologic inference.

Lack of additivity at the individual level (see #3 above) is common in epidemiology, but unmeasured modifiers do not bias results at the individual level if they are unrelated to the exposure (21). Furthermore, interactions may be handled readily at the individual level by including product terms as predictors in the model (e.g. xz). In ecologic analyses, however, lack of additivity within groups is a source of ecologic bias, and this bias cannot be eliminated or reduced by the inclusion of product terms (e.g. XZ) unless the effects are exactly multiplicative and the two variables are uncorrelated within groups (53).

Another source of ecologic bias is misspecification of confounders (26). Although this problem can also arise in individual-level analyses, it is more difficult to avoid in ecologic analyses because the relevant confounder may be the distribution of covariate histories for all individuals within each group. In ecologic studies, therefore, adjustment for covariates derived from available data (e.g. proportion of current smokers) may be inadequate to control confounding. It is preferable, whenever possible, to control for more than a single summary measure of the covariate distribution (e.g. the proportions of the group in each of several smoking categories). In addition, since it is usually necessary to control for several confounders (among which the effects may not be linear and additive), the best approach for reducing ecologic bias is to include covariates for categories of their joint distribution within regions. For example, to control ecologically for race and sex, the investigator might adjust for the proportions of white women, nonwhite men, and nonwhite women (treating white men as the referent), rather than the conventional approach of adjusting for the proportions of men (or women) and whites (or nonwhites).

Within-Group Misclassification

The principles of misclassification bias with which epidemiologists are familiar when interpreting the results of analyses conducted at the individual level do not apply to ecologic analyses. At the individual level, for example, nondifferential misclassification of exposure nearly always leads to bias toward the null. In multiple-group ecologic studies, however, this principle does not hold when the exposure variable is an aggregate measure. Brenner et al (4) have shown that nondifferential misclassification of a binary exposure within groups usually leads to bias away from the null and that the bias may be severe. Greenland & Brenner (23) have provided a simple method to correct for nondifferential misclassification of exposure or disease in ecologic analyses, based on estimates of sensitivity and specificity.

In studies conducted at the individual level, misclassification of a covariate, if nondifferential with respect to both exposure and disease, will usually reduce

our ability to control for that confounder (20, 59). That is, adjustment will not completely eliminate the bias due to the confounder. In ecologic studies, however, nondifferential misclassification of a binary confounder within groups does not affect our ability to control for that confounder, provided there is no cross-level bias (3).

If all but one variable (e.g. the exposure or a covariate) in a given analysis is measured at the individual level, this partially ecologic analysis may also be regarded as nonecologic with the ecologic variable misclassified. Thus, the resulting bias may be understood in terms of misclassification bias operating at the individual level.

Other Problems

LACK OF ADEQUATE DATA Certain types of data, such as medical histories, may not be available in aggregate form; or available data may be too crude, incomplete, or unreliable, such as sales data for measuring behaviors (45, 67). In addition, secondary sources of data from different administrative areas or from different periods may not be comparable. For example, disease rates may vary across countries because of differences in disease classification or case detection. Furthermore, since many ecologic analyses are based on mortality rather than incidence data, causal inference is further limited (35).

TEMPORAL AMBIGUITY In a well-designed cohort study of disease incidence, we can usually be confident that disease occurrence did not precede the exposure. In ecologic studies, however, use of incidence data provides no such assurance against this temporal ambiguity (45). The problem is most troublesome when the disease can influence exposure status in individuals or when the disease rate can influence the mean exposure in groups (through the impact of population interventions designed to change exposure levels in areas with high disease rates).

The problem of temporal ambiguity in ecologic studies (especially time-trend studies) is further complicated by an unknown or variable latent period between exposure and disease occurrence (28, 67). The investigator can only attempt to deal with this problem in the analysis by examining associations for which there is a specified lag between observations of average exposure and disease rate. Unfortunately, there may be little prior information about latency on which to base the lag, or appropriate data may not be available to accommodate the desired lag.

COLLINEARITY Another problem with ecologic analyses is that certain predictors, such as sociodemographic and environmental factors, tend to be more highly correlated with each other than they are at the individual level (10, 62).

The implication of such collinearities is that it is very difficult to separate the effects of these variables statistically; analyses yield model coefficients with very large variances, so effect estimates may be severely distorted. In general, collinearity is most problematic in multiple-group ecologic analyses involving a small number of large, heterogeneous regions (15, 64).

MIGRATION ACROSS GROUPS Migration of individuals into or out of the source population can produce selection bias in a study conducted at the individual level because migrants and nonmigrants may differ on both exposure prevalence and disease risk. Although it is clear that migration can also cause ecologic bias (36, 49), little is known about the magnitude of this bias or how it can be reduced in ecologic studies (46).

CONTEXTUAL AND MULTILEVEL ANALYSES

Knowing the severe methodologic limitations of ecologic analysis for making biologic inferences, many epidemiologists who report ecologic results argue that there can be no cross-level bias because their primary objective is to estimate an ecologic effect. For example, we might want to estimate the ecologic effect (effectiveness) of state laws requiring smoke detectors by comparing the fire-related mortality rate in those states with the law vs other states without the law (45). Although this is a reasonable objective, the interpretation of observed ecologic effects is complicated by two issues:

First, biologic inference may be implicit to the objectives of an ecologic study unless the underlying biologic and contextual effects are already known from previous research. Can smoke detectors placed appropriately in homes reduce the risk of fire-related mortality in those homes by providing an early warning of smoke? Does living in an area where most homes are properly equipped with smoke detectors reduce the risk of fire-related mortality in homes with and without smoke detectors? The first question refers to a possible biologic (biobehavioral) effect; the second question refers to a possible contextual effect. Even if these effects exist, the ecologic effect of smoke-detector laws also depends on other factors, e.g. the level of enforcement, the quality of smoke-detector design and construction, the cost and availability of smoke detectors, and their proper placement, installation, operation, and maintenance. In an ecologic study without additional information, the ecologic effect is completely confounded with biologic and contextual effects.

The second complicating issue in interpreting observed ecologic effects is the need to control for confounders measured at the individual level. Even if the exposure is a global measure, such as a law, groups are seldom completely homogeneous or comparable with respect to confounders. To make a valid comparison between states with and without smoke-detector laws, for example,

we would need to control for differences among states in the joint distribution of extraneous risk factors, such as socioeconomic status of residents, firefighter availability and access, building design, and construction (see also *Problems of Confounder Control*).

Perhaps the best solution to these problems is to incorporate both individual-level and ecologic measures in the same analysis. This approach might include different measures of the same factor; e.g. each subject would be characterized by his/her own exposure level as well as the average exposure level for all members of the group to which s/he belongs (aggregate measure). Not only would this approach help to clarify the sources and magnitude of ecologic and cross-level bias, but it would also allow us to separate biologic, contextual, and ecologic effects. It is especially appropriate in social epidemiology, infectious-disease epidemiology, and the evaluation of population interventions.

There are two statistical methods for including both individual-level and ecologic measures in the same analysis. The first method, often called *contextual analysis* in the social sciences, is a simple extension of conventional modeling such as multiple linear regression or logistic regression (5, 34). The model, which is fit to the data at the individual level, includes both individual-level and ecologic predictors. For example, suppose we wanted to estimate the effect of "herd immunity" on the risk of an infectious disease. The risk (y) of disease might be modeled as a function of the following linear component: $b_0 + b_1x + b_2\bar{x} + b_3x\bar{x}$, where x is the individual's immunity status and $\bar{x}$ is the prevalence of immunity in the group to which that individual belongs (65). Therefore, b_2 represents the contextual effect of herd immunity, and b_3 represents the interaction effect, which allows the herd-immunity effect to depend on the individual's immune status. The interaction term is needed in this application, since we would expect no herd-immunity effect among immune individuals. Note, however, that the interpretation of the interaction effect depends on the form of the model (35, 57).

An important limitation of contextual analysis is that observations for individuals within groups are not likely to be independent, which is a basic assumption of conventional modeling. If there are contextual effects, then the outcomes for individuals in the same group are more likely to resemble each other than are the outcomes for individuals in different groups. To handle this problem of within-group clustering, we treat the sampling of individuals from groups as random effects; this approach is called *multilevel modeling, hierarchical regression,* or *random-effects modeling* (6, 71, 72).

Multilevel modeling is a powerful technique with many applications; it can be used to estimate contextual and ecologic effects and to derive improved (empirical Bayes) estimates of biologic effects. At the first level of analysis, we might predict individual risk or health status within each group as a function

of several individual-level variables. At the second (ecologic) level, we predict the estimated regression parameters (e.g. the intercept and slopes) from the first level as a function of several ecologic variables. For example, Humphreys & Carr-Hill (33) used multilevel modeling to estimate the contextual effect of living in a poor area on several health outcomes, controlling for the individual's income and other covariates. In a conventional ecologic analysis, the effects of living in a poor area and income would be confounded, and ecologic estimates of effect would be susceptible to cross-level bias.

CONCLUSIONS

Several practical advantages make ecologic studies especially appealing for undertaking various types of epidemiologic research. Despite these advantages, however, ecologic analysis poses major problems of interpretation when making ecologic inferences and especially when making biologic inferences (due to ecologic bias, etc). From a methodologic perspective, it is best to have individual-level data on as many relevant nonglobal measures as possible. Just because the exposure variable is measured ecologically, for example, does not mean that other variables should be as well.

Even when the stated purpose of the study is to estimate an ecologic effect, biologic inference is usually implicit in epidemiology. Thus, to address the underlying research questions, we typically would want to estimate and/or control for biologic and contextual effects, preferably using multilevel analysis. In contemporary epidemiology, the "ecologic fallacy" reflects the failure of the investigator to recognize the need for biologic inference and thus for individual-level data.

ACKNOWLEDGMENTS

The author would like to thank Drs. Sander Greenland and Matthew Longnecker for their helpful comments.

Literature Cited

1. Beral V, Chilvers C, Fraser P. 1979. On the estimation of relative risk from vital statistical data. *J. Epidemiol. Community Health* 33:159–62
2. Blot WJ, Fraumeni JF Jr. 1977. Geographic patterns of oral cancer in the United States: etiologic implications. *J. Chron. Dis.* 30:745–57
3. Brenner H, Greenland S, Savitz DA. 1992. The effects of nondifferential confounder misclassification in ecologic studies. *Epidemiology* 3:456–59
4. Brenner H, Savitz DA, Jöckel K-H, Greenland S. 1992. Effects of nondifferential exposure misclassification in ecologic studies. *Am. J. Epidemiol.* 135: 85–95
5. Boyd LH Jr, Iversen GR. 1979. *Contex-*

tual Analysis: Concepts and Statistical Techniques. Belmont, CA: Wadsworth
6. Bryk AS, Raudenbush SW. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage
7. Catalano R, Serxner S. 1987. Time series designs of potential interest to epidemiologists. *Am. J. Epidemiol.* 126:724–31
8. Clayton DG, Bernardinelli L, Montomoli C. 1993. Spatial correlation in ecological analysis. *Int. J. Epidemiol.* 22:1193–202
9. Clayton D, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671–81
10. Connor MJ, Gillings D. 1984. An empiric study of ecological inference. *Am. J. Public Health* 74:555–59
11. Crawford MD, Gardner MJ, Morris JN. 1971. Changes in water hardness and local death-rates. *Lancet* 2:327–29
12. Cressie N. 1993. Regional mapping of incidence rates using spatial Bayesian models. *Med. Care* 31:YS60–65 (Suppl.)
13. Darby SC, Doll R. 1987. Fallout, radiation doses near Dounreay, and childhood leukaemia. *Br. Med. J.* 294:603–7
14. Dogan M, Rokkan S, eds. 1969. *Social Ecology*. Cambridge, MA: MIT Press
14a. Dogan M, Rokkan S. 1969. Introduction. See Ref. 14, pp. 1–15
15. Duncan OD, Cuzzort RP, Duncan B. 1961. *Statistical Geography: Problems in Analyzing Areal Data*, pp. 64–67. Westport, CT: Greenwood Press
16. Durkheim E. 1951. *Suicide: A Study in Sociology*, pp. 153–54. New York: Free Press
17. Feinleib M, Leaverton PE. 1984. Ecological fallacies in epidemiology. In *Health Information Systems*, ed. PE Leaverton, L Massö, pp. 33–61. New York: Praeger
18. Firebaugh G. 1978. A rule for inferring individual-level relationships from aggregate data. *Am. Sociol. Rev.* 43:557–72
19. Goodman LA. 1959. Some alternatives to ecological correlation. *Am. J. Sociol.* 64:610–25
20. Greenland S. 1980. The effect of misclassification in the presence of covariates. *Am. J. Epidemiol.* 112:564–69
21. Greenland S. 1987. Interpretation and choice of effect measures in epidemiologic analysis. *Am. J. Epidemiol.* 125: 761–68
22. Greenland S. 1992. Divergent biases in ecologic and individual-level studies. *Stat. Med.* 11:1209–23
23. Greenland S, Brenner H. 1993. Correcting for non-differential misclassification in ecologic analyses. *Appl. Statist.* 42: 117–26
24. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. 1991. Standardized regression coefficients: a further critique and review of some alternatives. *Epidemiology* 2:387–92
25. Greenland S, Morgenstern H. 1989. Ecological bias, confounding, and effect modification. *Int. J. Epidemiol.* 18:269–74
26. Greenland S, Robins J. 1994. Invited commentary: ecologic studies—biases, misconceptions, and counterexamples. *Am. J. Epidemiol.* 139:747–60
27. Greenland S, Schlesselman JJ, Criqui MH. 1986. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am. J. Epidemiol.* 123:203–8
28. Gruchow HW, Rimm AA, Hoffman RG. 1983. Alcohol consumption and ischemic heart disease mortality: are time-series correlations meaningful? *Am. J. Epidemiol.* 118:641–50
29. Hatch M, Susser M. 1990. Background gamma radiation and childhood cancers within ten miles of a US nuclear plant. *Int. J. Epidemiol.* 19:546–52
30. Helfenstein U. 1991. The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int. J. Epidemiol.* 20:808–15
31. Hiller JE, McMichael AJ. 1991. Ecological studies. In *Design Concepts in Nutritional Epidemiology*, ed. BM Margetts, M Nelson, pp. 323–53. Oxford: Oxford Univ. Press
32. Holford TR. 1991. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu. Rev. Public Health* 12:425–57
33. Humphreys K, Carr-Hill R. 1991. Area variations in health outcomes: artefact or ecology. *Int. J. Epidemiol.* 20:251–58
34. Iversen GR. 1991. *Contextual Analysis*. Newbury Park, CA: Sage
35. Kleinbaum DG, Kupper LL, Morgenstern H. 1982. *Epidemiologic Research: Principles and Quantitative Methods*, pp. 77–81, 130–34, 184–280. New York: Van Nostrand Reinhold
36. Kliewer EV. 1992. Influence of migrants on regional variations of stomach and colon cancer mortality in the western United States. *Int. J. Epidemiol.* 21:442–49
37. Koopman JS, Longini IM Jr. 1994. The ecological effects of individual exposures and nonlinear disease dynamics in populations. *Am. J. Public Health* 84: 836–42

38. Langbein LI, Lichtman AJ. 1978. *Ecological Inference*. Beverly Hills, CA: Sage
39. Lee JAH. 1982. Melanoma and exposure to sunlight. *Epidemiol. Rev.* 4:110–36
40. Lee JAH, Petersen GR, Stevens RG, Vesanen K. 1979. The influence of age, year of birth, and date on mortality from malignant melanoma in the populations of England and Wales, Canada, and the white population of the United States. *Am. J. Epidemiol.* 110:734–39
41. MacMahon B, Pugh TF. 1970. *Epidemiology: Principles and Methods*, pp. 137–98, 175–84. Boston: Little, Brown & Co.
42. Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF Jr. 1975. *Atlas of Cancer Mortality for US Counties: 1950–1969*, pp. 36, 37. DHEW Publ. No. (NIH) 75–780. Washington, DC: US GPO
43. McDowall D, McCleary R, Meidinger EE, Hay RA Jr. 1980. *Interrupted Time Series Analysis*. Beverly Hills, CA: Sage
44. Mollie A, Richardson S. 1991. Empirical Bayes estimation of cancer mortality rates using spatial models. *Stat. Med.* 10:95–112
45. Morgenstern H. 1982. Uses of ecologic analysis in epidemiologic research. *Am. J. Public Health* 72:1336–44
46. Morgenstern H, Thomas D. 1993. Principles of study design in environmental epidemiology. *Environ. Health Perspect.* 101:23–38 (Suppl. 4)
47. Moulton LH, Foxman B, Wolfe RA, Port FK. 1994. Potential pitfalls in interpreting maps of stabilized rates. *Epidemiology* 5:297–301
48. Ostrom CW Jr. 1990. *Time Series Analysis: Regression Techniques*. Newbury Park, CA: Sage. 2nd ed.
49. Polissar L. 1980. The effect of migration on comparison of disease rates in geographic studies in the United States. *Am. J. Epidemiol.* 111:175–82
50. Prentice RL, Kakar F, Hursting S, Sheppart L, Klein R, Kushi LH. 1988. Aspects of the rationale for the Women's Health Trial. *J. Natl. Cancer Inst.* 80: 802–14
51. Prentice RL, Sheppard L. 1989. Validity of international, time trend, and migrant studies of dietary factors and disease risk. *Prev. Med.* 18:167–79
52. Prentice RL, Thomas D. 1993. Methodologic research needs in environmental epidemiology: data analysis. *Environ. Health Perspect.* 101:39–48 (Suppl. 4)
53. Richardson S, Hémon D. 1990. Ecological bias and confounding (letter). *Int. J. Epidemiol.* 19:764–66
54. Richardson S, Stücher I, Hémon D. 1987. Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. *Int. J. Epidemiol.* 16:111–20
55. Robinson WS. 1950. Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* 15:351–57
56. Rosenbaum PR, Rubin DB. 1984. Difficulties with regression analyses of age-adjusted rates. *Biometrics* 40:437–43
57. Rothman KJ. 1986. *Modern Epidemiology*, pp. 41–49, 82–94. Boston: Little, Brown & Co.
58. Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Stat.* 6:34–58
59. Savitz DA, Baron AE. 1989. Estimating and correcting for confounder misclassification. *Am. J. Epidemiol.* 129:1062–71
60. Sayrs LW. 1989. *Pooled Time Series Analysis*. Newbury Park, CA: Sage
61. Selvin HC. 1958. Durkheim's "Suicide" and problems of empirical research. *Am. J. Sociol.* 63:607–19
62. Stavraky KM. 1976. The role of ecologic analysis in studies of the etiology of disease: a discussion with reference to large bowel cancer. *J. Chron. Dis.* 29: 435–44
63. Stidley C, Samet JM. 1994. Assessment of ecologic regression in the study of lung cancer and indoor radon. *Am. J. Epidemiol.* 139:312–22
64. Valkonen T. 1969. Individual and structural effects in ecological research. See Ref. 14, pp. 53–68
65. Von Korff M, Koepsell T, Curry S, Diehr P. 1992. Multi-level analysis in epidemiologic research on health behaviors and outcomes. *Am. J. Epidemiol.* 135:1077–82
66. Walter SD. 1991. The ecologic method in the study of environmental health. I. Overview of the method. *Environ. Health Perspect.* 94:61–65
67. Walter SD. 1991. The ecologic method in the study of environmental health. II. Methodologic issues and feasibility. *Environ. Health Perspect.* 94:67–73
68. Walter SD. 1992. The analysis of regional patterns in health data: I. Distributional considerations. *Am. J. Epidemiol.* 136:730–41
69. Walter SD. 1992. The analysis of regional patterns in health data: II. The power to detect environmental effects. *Am. J. Epidemiol.* 136:742–59
70. Winn DM, Blot WJ, Shy CM, Pickle LW, Toledo A, Fraumeni JF Jr. 1981. Snuff dipping and oral cancer among

women in the southern United States. *N. Engl. J. Med.* 304:745–49

71. Wong GY, Mason WM. 1985. The hierarchical logistic regression model for multilevel analysis. *J. Am. Statist. Assoc.* 80:513–24

72. Wong GY, Mason WM. 1991. Contextually specific effects and other generalizations for the hierarchical linear model for comparative analysis. *J. Am. Statist. Assoc.* 86:487–503