The Logic in Ecological: II. The Logic of Design

ABSTRACT

The utility of ecological studies is considered in terms of the salience of their designs and is exemplified in four levels: obligate and apt; optional and apt; optional, not apt but convenient; and maladroit (neither obligate, apt, nor justifiable by convenience). Ecological studies are obligate when they are the only choice available, either because of the question asked (as in testing differences between groups and discovering group effects), or where there are "dependent happenings" (as in transactions involving more than one individual), or because individual data are lacking. Apt studies are logically appropriate; analysis and results are not extrapolated beyond necessity or without precautions. Obligate studies enforced by lack of individual data may be apt or less than apt. Optional ecological studies may be apt or, if less than apt, they may yet be convenient. Maladroit studies are neither obligate, apt, nor convenient. Each class of study is illustrated by examples ordered according to a standard design hierarchy. (*Am J Public Health.* 1994;84:830–835)

Mervyn Susser, MB, BCh, FRCP(E), DPH

Introduction

This paper considers the salience of ecological studies. Four (not exhaustive) levels of salience suggest themselves in terms of choices, aptness, and convenience: (1) Obligate (the only choice) and apt (logically appropriate); (2) Optional (not the only choice) but apt; (3) Optional, not apt but convenient; and (4) Maladroit: neither obligate, apt, nor convenient.

Ecological studies are obligate when they are the only choice available, either because of the question asked (as in testing differences between groups and discovering group effects), or because of a concern with dependent happenings (as in transactions involving more than one individual), or merely because of the lack of individual data. Apt studies are logically appropriate[1] and do not extrapolate beyond necessity. Obligate studies enforced by lack of individual data are often less than apt. Optional studies may yet be apt, or if less than apt, they may yet be convenient. Maladroit studies misconstrue purpose and design; they are neither obligate, apt, nor convenient.

Each level of salience is illustrated below by examples ordered according to a standard hierarchy of proficiency in design. Descriptive studies are treated before analytic studies. Analytic studies follow the hierarchy from cross-sectional, through observation of the sequence of routine events, to natural experiments, quasi-experiments, and controlled experiments.[2]

Obligate and Apt

Epidemiology is a utilitarian as well as a population science; it seeks to improve the health states of populations. In this spirit, Geoffrey Rose argued that concentration on the person as a unit and on a lessening of personal risk has led to the neglect of populations and of the preventive goal of reducing incidence.[3] In other words, epidemiologists cannot afford to neglect ecological studies. The pitfalls of such undertakings have overshadowed their uses; however, the comparison of groups as groups is a prerequisite for the understanding of group effects on both groups and individuals.

Descriptive ecological studies of cross-sectional type—with disease or death as numerator, and with census or a substitute as denominator—have been a staple of epidemiology for centuries. In the prototype of all these, John Graunt discovered the existence of gender and urban/rural mortality differences.[4]

Cause-specific mortality statistics introduced in England and Wales in the 1840s by William Farr[5] opened the way for longitudinal study; the analytic units are now 30 quinquennia, each with very large numbers. This cumulation yields stable results in which small changes over time are detectable and have meaning.

Analytic ecological studies of this kind have been less securely informative. The seminal efforts of Thomas McKeown to determine the causes of the changes in mortality over a century in England and Wales[6,7] led him to conclusions that are, in my view, sometimes faulty.[8]

The pervasive problem in testing hypotheses in such studies is the weakness

---

The author is the Editor of the Journal.

Requests for reprints should be sent to the Journal office.

This paper was accepted July 15, 1993.

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), Koopman and Longini (p 836), and Schwartz (p 819) and editorial by Poole (p 715) in this issue.

of measures of exposure; integral variables, such as nutrition estimated from national food supply, must depend on collateral evidence and indirect inference. The colinearity and confounding that are often present and the use of cross-sectional designs add further problems. Isolating the test variable from the complex of covariates and establishing time sequence are especially difficult.

Integral variables sometimes are precisely measured. In William Farr's hands, altitude and cholera afforded a strong cross-sectional association.[9] His mistaken causal interpretation bypassed another integral effect: the crucial intervening variable of contaminated water supply,[10] which was discovered in the same data by John Snow.[11] Snow created for himself the rare advantage of a well-demarcated natural experiment. Before the cholera epidemic, one of two companies supplying water to the boroughs of Lambeth and Vauxhall in London had moved upstream in the Thames for its supply. Able to identify exactly the companies that supplied each of the intimately mingled households of the two boroughs, Snow demonstrated an eightfold increased risk with the downstream supply, and an 88% reduction in mortality for the upstream supply.[10]

More recently, strong cross-sectional associations of high altitude with low birthweight have yielded fruitful hypotheses about the effect of anoxia.[12] A quasi-experimental intervention to alert obstetricians was the plausible agent in closing the gap in perinatal mortality between the heights of Colorado and lesser altitudes.[13]

Latitude also serves as an integral variable. Cross-sectional observations related low frequencies of multiple sclerosis to southern latitude.[14] In South Africa in the southern hemisphere, those affected were predominantly immigrants who had spent their youth in Europe before migration.[15] New hypotheses involving exposure to viral infections followed and cleared away initial attributions to race.[16]

Contagion or dependent happenings[1,17] make ecological analysis both obligate and apt. Transmission of infection, behavior, and other social phenomena involves direct and indirect transactions between individuals and groups. Analyis of individuals is necessary but insufficient alone to capture diffusion in a population. In my still-evolving view, analysis of context—whether of large groups, households, pairings, or all of these together—should enable one to capture the transactions and interactions among them. Dependent happenings create a major category of obligate and apt ecological studies.

Quasi-experiments at the ecological level also have a long history. In proto-analytic endeavors 150 years ago, Edwin Chadwick tested the effects of closed-circuit sanitation systems on mortality by comparing groups assembled post hoc or observed before and after.[18] Despite weakness of design and analysis by today's standards, the causal inference gains support from consistent replication and ultimate coherence.[2]

Quasi-experiments are a preferred mode for testing existing operations in health and medical care systems. The entire newborn program for New York City's low-birthweight infants over a 4-year period was tested in a quasi-experimental design to determine the effect of "intensive care" as compared with less intensive care.[19] Intensive care is an integral variable that characterizes an entire hospital; the dependent variable was therefore the mortality rates of low-birthweight infants in hospitals grouped by a predefined level of intensity of care. Because individual data were collected on the 13 500 births, the major confounders as well as contextual variables such as ethnic distribution could be controlled. These results led to an unchallenged inference of a favorable effect of the program on newborn mortality.

Controlled experiment is in the common wisdom the most decisive of tests. Yet during the past half-century, many controlled ecological experiments have failed. The recurring problem has been their grossly insufficient statistical power. When interventions are allocated to entire groups and not to individuals, the unit for the power calculation is the group. Even when this requirement was recognized, however, a sufficiency of units was often sacrificed to insufficient resources; a maximum of detail to refine measurement and control covariates was preferred. Such studies produce convincing results only with the luck of large effects or consistent patterns not at variance with expectations.

A remedy resides in the converse procedure. The strategy, advocated by Richard Peto, sacrifices detail in favor of numbers. The remedial principles include groups numerous enough to ensure statistical power; an intervention as simple and decisive as possible; an endpoint as simple and clear-cut as possible; randomization to neutralize covariates, confounding, and other differences between the intervention and comparison groups; and retention of the study population up to the measured endpoint.

An experiment to test the effect of vitamin A on child mortality[20] exemplifies this approach. Four hundred fifty Sumatran villages were randomized to treatment or no treatment with single, long-acting oral doses of vitamin A. At the time of treatment, baseline data were obtained on the children of all study villages. The endpoint, established at one follow-up visit, was mortality over the intervening 12 months. Although the favorable result is vulnerable in some respects, some subsequent tests suggest, if they do not certify, the roadworthiness of the approach.[21]

Meta-analyses. Problems of statistical power are not confined to the ecological mode. This pervasive deficiency in clinical trials has led to a surge in meta-analyses, or quantitative overviews.[22–24] These analyses are akin to the obligatory ecological approach in that no quantitative alternative to grouping the available studies exists.

Meta-analysis departs from some typical scripts for the obligate ecological class in the availability of data for individual group members. Analysis of both group and individual data ensures against egregious fallacy. The weighted average of the within-group individual associations (B_w) is the measure least vulnerable to error in extrapolation [see Susser[1]].

Optional and Apt

For this class of studies, disasters or natural experiments often qualify. Natural experiments entail sharp (hence, usually short), well-defined events of distinct magnitude and, preferably, the availability of preexisting data. Two examples, the Bangladesh floods of 1988[25] and the Dutch famine of 1944 to 1945,[26] offer different points of interest.

In Bangladesh, 6 months after a representative sample of children aged 2 to 9 years had been surveyed, the area was flooded. In an a priori test of the effects of disaster, the opportunity was taken to follow the children 5 months later and to repeat questions about behavior.

Because individual data were at hand, they were first analyzed at that level—that is, xy. In individual children, behavior before and after the flood (xy) did not differ according to the danger of drowning incurred. At the ecological level (XY), on the other hand, a large before-and-after ($X1$-and-$X2$) flood effect was

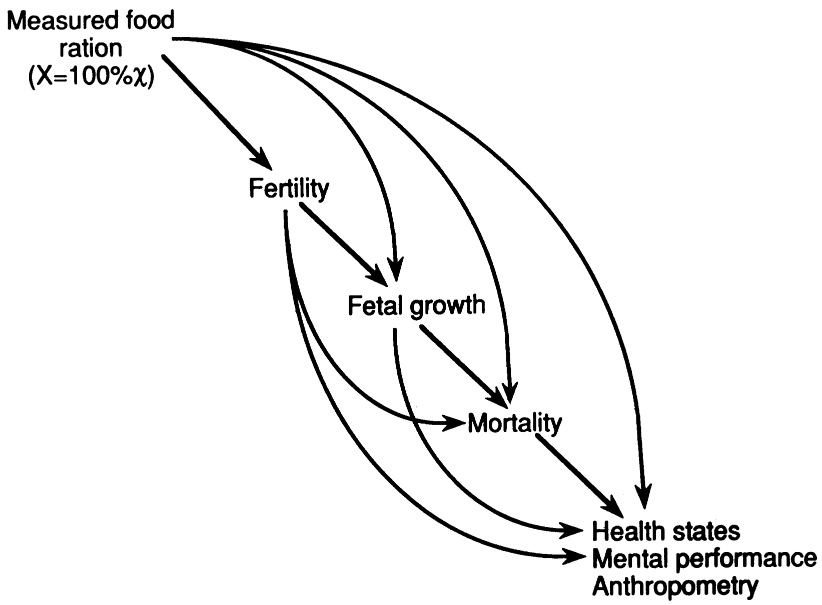

FIGURE 1—The Dutch hunger winter: scheme of famine effects on successive developmental stages culminating at age 19 in young men.

present for both aggressive behavior and enuresis. Before the flood, only 1 child of the total sample of 434 and none of the 162 children found and tested on follow-up was reportedly aggressive; after it, 16 were. Before, 21 of the 162 who were found and retested did not have sphincter control; after, 45 were enuretic. While aggression could conceivably be a function of aging over a period of 11 months, rapid normal rates of remission[27] ensure that an increase in enuresis could not be.

In this calamity (and, in the scanty literature, in others), the integral variables of before-and-after population prevalences served better than any individual measure of danger to capture the children's experience. Comparisons of successive states of a population are necessarily ecological. Conceivably, grouping enhanced weak individual measures. More likely, regardless of direct individual experience of danger, terror was at large and induced post-traumatic stress in the most vulnerable. Here, dependent happenings created an integral effect detectable by group but not by individual measures.

The Dutch famine, created by the German forces at the end of World War II, was unique among famines in the extent of recoverable detail and in the measures and demarcation of exposure. The famine was confined to the West, one of three national administrative regions; in each region, food rations were distributed to every individual in known amount. This savage, unnatural experiment allowed for the testing of hypotheses about the effects of severe prenatal food deprivation on subsequent development—in particular, mental performance, growth, and health—in young men.

Thirty-six monthly national birth cohorts were grouped in relation to famine exposure by trimester of pregnancy. Two-pronged comparisons of the exposed and unexposed—before and after the famine and between famine and nonfamine regions—complemented each other: historical confounding with time change in the period mode is absent in the regional mode; demographic confounding is largely absent in the historical mode.

Atypically in such circumstances, individual exposure and outcome measures at several developmental stages (fertility; birth outcome; mortality; and, in young men, morbidity, physique, and mental performance) were available (Figure 1). Nonetheless, the predominant analyses presented use grouped variables for two main reasons.

First, with regard to exposure, during the famine the official ration given to all individuals (x) was a reasonable index of individual intake, with unofficial foraging causing some underestimate; outside of the famine, this ration was less accurate. The ration went to everyone, so that in truth it was an integral group exposure (X). Especially outside of famine, X is a better measure of cohort intake than x is of individual intake.

The main hazard is that differences in Y might be attributed to greater or lesser change in X than was actually the case. Lean prefamine rations underestimated unofficially supplemented intake and, hence, also the change in X that followed the depletion of both official and unofficial food sources during the famine. Thus, the effect on Y of a unit change in X was likely to be overestimated. Generous postfamine rations probably did just the opposite.

Second, with regard to outcomes, requirements for confidentiality precluded the use of individual units: records of events at birth and into young adulthood ($y_1, y_2, y_3, \ldots$) could not be linked. Thus, the impact of each stage upon subsequent ones could be elicited only from grouped cohort data. A logical obstacle to individual units was the need either to extrapolate from grouped X to individual y or to accept an individual x that within each region was the same for all.

Individual units were preferred in two circumstances: to analyze intra-individual relations (as among different anthropomorphic indices), and to test the consistency of the results for $X \rightarrow Y$ against those for $x \rightarrow y$ (or $X \rightarrow y$, which, in deriving rates in this study, is the same as $x \rightarrow y$). The grouped associations, $X \rightarrow Y$, were almost always greater than the individual ones. This enlargement was to be expected. Exclusive of exposures and their health effects, between birth cohorts the characteristics of groups remained much the same, a homogeneity offset by the marked variation in both their exposure to famine and the effects of famine. In contrast, within cohorts the characteristics of individuals retained their normal heterogeneity. It follows that the proportion of variance in outcomes available to be accounted for by exposure was greater for groups than for individuals. In these analyses, the data were arrayed in four different ways to test the same hypotheses, and XY and xy were generally not in conflict in terms of direction and statistical significance.

Several results from this study about prenatal food deprivation either survive subsequent challenge at the individual level or have proven to be coherent with targeted animal experiments or other studies.[28–35] Likely factors in the apparent robustness of such ecological results are a

number of design features. The hypotheses were established a priori, and hence the grouped units (cohorts) were assembled around *X*, the famine exposure, as in the proper conduct of any such test. Unusually precise and well-demarcated indices of exposure were able to be applied. Not least, a very large number of individuals and a sufficient number of cohorts ensured adequate power. Finally, several different ways of assembling, elaborating, and testing the data safeguarded sensible interpretation.

Optional, Not Apt but Convenient

To enter an ecological study under this class is to trade logical requirements against opportunity and available resources. We consider two studies of the carcinogenic effects of radiation from nuclear plants. The charge that a choice is for convenience over aptness depends on the objective of the studies. An ecological approach is apt for determining effects on entire communities.

The first example is an investigation of the 1979 Three Mile Island nuclear accident for possible effects on the incidence of and mortality from cancers sensitive to radiation.[36–38] Although actual emissions were known to be low and (in 1985) the elapsed latent period was short, a federal court deemed the work to be in the public interest. In the dispersed and relatively small population typical for such sites, it is also typical that no individual exposures (*x*) were known; indeed, the only data initially available were estimates from measures at the plant.

With choice of study variables limited to *Xy* or *XY*, the more economical first-stage study was to opt for *XY*, as was done. In either case, individual cancer data had to be collected, but for an ecological study researchers could dispense with laborious individual follow-up. In view of both limited funds and the limited power inherent in any design so soon after the accident, the effort of collecting greater detail on individuals about their location, movements, and attributes was of doubtful advantage. The public interest, moreover, could be reasonably framed by the overall question of whether the accident had an effect on the surrounding communities rather than on individuals.

About 5500 incident cancers were identified, by postal address, among people living within 10 miles of the plant. The exposure measure was refined by two

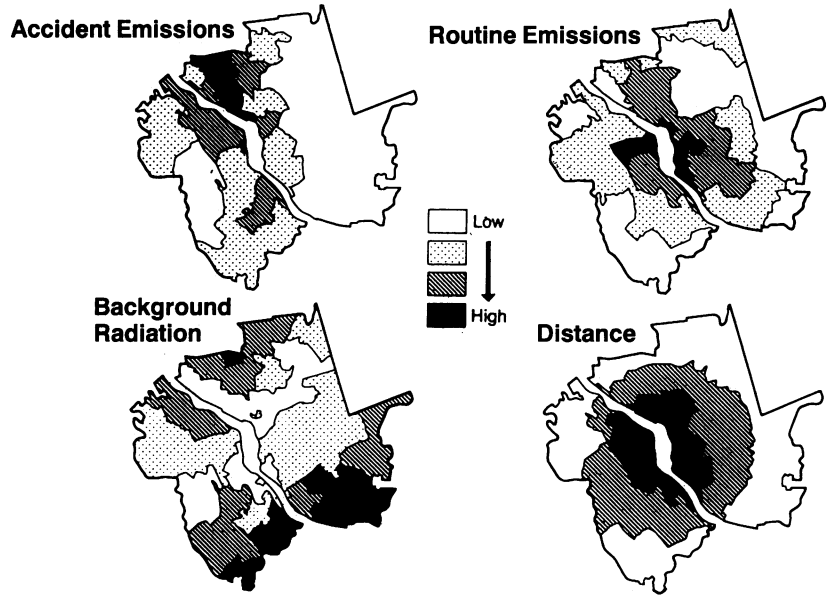

FIGURE 2—Models of four potential exposures—accident and routine emissions, background radiation, and distance—at the Three Mile Island nuclear plant.

steps. First, the area was divided into 69 blocks (average 2300 persons)—small but still usable units in terms of rates and statistical power—aligned with the previous census.

Second, special effort went into estimating exposure. From weather and topographical data, three mathematical models of the dispersion of radiation across the 69 blocks were built: one for the accident, one for repeated workday routines, and one for background radiation (Figure 2). A fourth model was simply the one most often used in similar circumstances—namely, distance from the plant. (Ultimately, the estimates of accident emissions were validated [$r = .92$] by monitoring data that came to hand.) Accident exposure could thus be refined both by controlling for confounding modeled for three other sources and by adjusting the demographic variation detectable from the census. Cancer incidence in high-exposure areas was tested against that in low-exposure areas.

Evidence for accident effects emerged but was not persuasive. Even for other evidence that appeared persuasive, such as the association of childhood concerns with background radiation, inference and interpretation must be guarded if the leap is made to such individual matters as cumulative risk. However, provided that other criteria, including replications, are met, causal inferences about period incidence or incidence density could be somewhat less cautious. They gain protection from a priori hypotheses, exposure measures refined and hedged against confounding to the degree possible, and a sufficient number of the smallest usable group units.

The second example in this class is a National Cancer Institute study of the effects of nuclear plants on cancer in the surrounding populations.[39] Sixty-two nuclear plants across the nation were the focus; the county around each plant provided the populations at risk. For each of these counties, three counties without plants provided the comparisons. The measure of exposure was presence of a plant, an integral variable. Measures of effect were mortality ratios for all cancers, all leukemia, and childhood leukemia (with US rates as the standard) over periods from 1950 to 1984, when the plants were operating. The design rested on cross-sectional comparisons—that is, *XY*. Periods before and after plants began activity could be compared wherever length of operations permitted time-lagging of effects in relation to exposure. No indication of radiation effects was found.

In this large undertaking, parsimony and convenience were well served by the design. Economies in determining exposure, populations at risk, and effects were

considerable. Data on plant location, county populations, and cancer deaths among them were all to be had from existing records.

A reasonable question is whether the logic of the design requirements for a fair test were equally well served. Appropriately, the target and comparison groups were assembled around the exposure. The exposure itself, however, is crudely specified: the exposure models for the Three Mile Island plant (Figure 2) reveal how weak the assumption of dispersion in concentric rings can be. Dilution of effect is likely to be severe with units as large as counties; large numbers of persons assumed to be exposed will not have been.

The outcome measures are also problematic. The question of time sequence of cause and effect is only partially met by the time-lagging of plant operations against mortality periods. In addition, while childhood leukemia actually had a lesser risk after plant operations began, effective treatment had by then been introduced. This confounding was perhaps mitigated by the use of US mortality for the appropriate period as a standard, but doubt remains.

On the other hand, this ecological design is not handicapped as many are by small numbers of the appropriate group units, and statistical power is not a problem. Hence, it might be argued that even effects diluted by weak measures should be detectable. Whether these null results should be regarded as contributory rather than indeterminate and uninformative is debatable. The authors recognized that the case for indeterminacy could not be brushed aside.

Maladroit Studies

Studies under this head are neither necessary, logically appropriate, nor justifiable by convenience and economy. The problems involved are precisely those of the ecological fallacy and are well seen in a paper that examined the nature of the fallacy.[40] The authors analyze the same pairs of variables within several levels of aggregation and compare the differences in their associations. Unlike the preceding examples, it follows from the use of the same variables that any aggregation effect must be contextual and not integral.

The authors illustrate their empirical points with data from the second National Health and Nutrition Survey. Correlations and linear regression coefficients were compared for 13 pairs of variables related to diet and physique. A randomly selected representative national sample (with some groups oversampled) was composed of 13 820 adults. The primary sampling units, 64 counties or groups of counties, served for a first level of aggregation; these were then combined into larger aggregates of 34 states and 6 regions. The test for consistent association is made more severe by still another level unspecified by the authors; some variable pairs, such as height and weight, were intra-individual.

Although the trends of the individual correlation and regression coefficients for the 13 pairs of variables were not intolerably dissimilar, the authors were disturbed to find reversals of direction and of statistical significance. Further, the ordering by size of the associations to be found in individual and grouped data could not be predicted. In only 4 of the 13 pairs of associations did their magnitude enlarge as the aggregate was enlarged; in 3 pairs, the order was inverse, and in 6, the order was inconsistent. This result, the authors note, contrasts with many reports in the literature that indicate, as in the Dutch study, that group associations are commonly larger than individual ones.

The divergence in all likelihood resides in the constraint placed on analysis by the random selection of the study population. This mode of constructing a sample is apt for describing accurately what is to be found in a population. It is not apt for isolating and eliciting the strongest accessible association between a test variable and an outcome in challenging a hypothesis. For that purpose, the sample must be assembled around the independent variable, X or x, and groups that are to be compared need to be homogeneous save for the study variables.

Conclusion

In this explication, my aim has been to clarify the nature, utility, and problems of grouped data as the unit of analysis, whether used of necessity or by preference, and to treat ecological studies not merely as a resort of second-class research or a sandbox for methodologists. Such studies, as I have tried to show, have their own obligations and legitimacies as a public health tool. Some ecological studies have produced knowledge that has stood uncontested for many years.

Nonetheless, hazards reside in ecological studies, and still more can be found in extrapolation across aggregate and individual levels. Thus, protective rules are needed. While certain features are special,[41] the elements are those of epidemiological research regardless of level: accurate measures of the independent variable, of the dependent variable, and of associated variables, supported by multiple indices; the strongest design within the constraints of possibility; a sufficiency of numbers to ensure adequate statistical power; and rigorous analysis that does not neglect elaboration as a mode of a priori hypothesis testing.[10,42] Among these elements, weak measures of exposure, weak design, neglect of obtainable individual measures, and insufficiency of grouped units and inadequate analysis are the besetting sins of ecological studies.

Six rules special to grouping are useful:

1. Ensure a sufficiency of groups for testing hypotheses with groups as units.
2. Measure and take account of integral and contextual effects.
3. Prefer small groups, which have fewer interrelations and hence simpler and more manageable properties than large ones. But weigh the disadvantage that, with rare outcomes, small numbers within groups may sacrifice the stability of outcome measures.
4. To elicit maximum association in hypothesis-testing research, aim for similarity (homogeneity) between groups in all respects other than the independent and dependent study variables.
5. In hypothesis-testing research, again, aim for the greatest difference (heterogeneity) between groups on the independent variable; that is, assemble the groups around X. A corollary is that if a study is designed around Y on the case-control analogy, the design will benefit from the greatest differences between groups in Y.
6. If perforce the number of groups is limited, alleviate lack of power by increasing the number of strata or antecedent conditions a priori; this is done most simply by demographic grouping. The usual considerations of power apply—namely, the number of groups, the frequency or standard deviation of the outcome, and the size of the differences in outcome between groups. □

Acknowledgments

This work was supported by National Institutes of Health grant 5-P50-MH43520. This paper was first undertaken as the opening for a special symposium convened by Nancy Padian, King Holmes, and Sevgi Aral at the meeting of the International Society for the Study of Sexually Transmitted Diseases in Banff, British Columbia, Canada, November 1991; it was

further developed for the meeting of the American Epidemiological Society in Ann Arbor, Mich, March 1992, and for a meeting of the National Research Council of Italy on Models and Epidemiologic Methods of Research on HIV Infection, convened by Alfredo Nicolosi in Capri, September 4–7, 1992.

I owe thanks for helpful comments to Sander Greenland, David Jacobs, James Koopman, Bruce Levin, Nicole Schupf, and Zena Stein.

References

1. Susser M. The logic in ecological: I. the logic of analysis. *Am J Public Health.* 1994;84:825–829.
2. Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol.* 1991;133:635–648.
3. Rose G. Sick individuals and sick populations. *Int J Epidemiol.* 1985;14:32–38.
4. Graunt J. *Natural and Political Observations Made upon the Bills of Mortality.* London, England: Roycroft; 1662. Reprinted Baltimore, Md: Johns Hopkins University Press; 1939.
5. Humphreys N. *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr.* 1885. Reprinted with an introduction by Susser M and Adelstein A. New York, NY: New York Academy of Medicine; 1975.
6. McKeown T. *The Modern Rise of Population.* London, England: Edward Arnold; 1976.
7. McKeown T. *The Role of Medicine: Dream, Mirage or Nemesis?* London, England: Nuffield Provincial Hospitals Trust; 1976.
8. Susser M. Health as a human right: an epidemiologist's perspective on the public health. *Am J Public Health.* 1993;83:418–426.
9. Farr W. *Report on the Mortality of Cholera in England in 1848–9.* London, England: His Majesty's Stationery Office; 1852.
10. Susser M. *Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology.* New York, NY: Oxford University Press; 1973.
11. Snow J. *On the Mode of Communication of Cholera.* 2nd ed. London, England: J. Churchill; 1855. Reprinted as *Snow on Cholera.* New York, NY: The Commonwealth Fund; 1956.
12. Yip R. Altitude and birth weight. *J Pediatr.* 1987;111:869–876.
13. Unger C, Weiser JK, McCullough RE, Keefer S, Moore LG. Altitude, low birth weight and infant mortality in Colorado. *JAMA.* 1988;259:3427–3432.
14. Acheson DE. The epidemiology of multiple sclerosis. In: McAlpine D, Lumsden CE, Acheson ED, eds. *Multiple Sclerosis. A Reappraisal.* Baltimore, Md: Williams and Wilkins; 1965; I:3–58.
15. Dean G. Disseminated sclerosis in South Africa. *Br Med J.* 1949;i:842–845.
16. Davenport CB. Multiple sclerosis from the standpoint of geographic distribution and race. In: Duna CL, et al., eds. *Multiple sclerosis.* New York, NY: Paul B. Hoeber; 1922;2:8–19.
17. Koopman JS, Prevots DR, Vaca Marin MA, et al. Determinants and predictors of dengue infection in Mexico. *Am J Epidemiol.* 1991;133:1168–1178.
18. Chadwick E. *Report on the Sanitary Conditions of the Labouring Populations of Great Britain, 1842.* Edinburgh, Scotland: Edinburgh University Press; 1965.
19. Paneth N, Kiely JL, Wallenstein S, Marcus M, Susser M. Newborn intensive care and mortality in low birthweight infants: a population study. *N Engl J Med.* 1982;307:149–155.
20. Sommer A, Djunaedi E, Loeden AA, Tilden R, Melle L. Impact of vitamin A supplementation on childhood mortality: a randomized controlled community trial. *Lancet.* 1986;I:1169–1172.
21. Glasziou PP, Mackerras DEM. Vitamin A supplementation in infectious diseases: a meta-analysis. *Br Med J.* 1993;306:366–370.
22. Pillemer DB. Conceptual issues in research synthesis. *J Special Educ.* 1984;18:27–40.
23. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline: I. control of bias and comparison with large cooperative trials. *Stat Med.* 1987;6:315–325.
24. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev.* 1987;9:1–30.
25. Durkin MS, Kahn N, Davidson LL, Zaman SS, Stein ZA. The effects of a natural disaster on child behavior: evidence for posttraumatic stress. *Am J Public Health.* 1993;83:1549–1553.
26. Stein Z, Susser MW, Saenger G, Marolla FM. *Famine and Human Development: The Dutch Hunger Winter of 1944–45.* New York, NY: Oxford University Press; 1975.
27. Forrester RM, Stein Z, Susser MW. A trial of conditioning therapy in nocturnal enuresis. *Dev Med Child Neurol.* 1964;6:158–166.
28. Milunsky A, Jick H, Jick S, et al. Multivitamin/folic acid supplementation in early pregnancy reduces the prevalence of neural tube defects. *JAMA.* 1989;262:2847–2852.
29. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council vitamin study. *Lancet.* 1991;338:131–137.
30. Susser ES, Lin SP. Schizophrenia after prenatal exposure to the Dutch Hunger Winter 1944/45. *Arch Gen Psychiatry.* 1992;49:983–988.
31. Susser M. The challenge of causality: human nutrition, brain development and mental performance. *Bull N Y Acad Med.* 1989;65:1032–1049.
32. Susser M. Maternal weight gain, infant birth weight, and diet: causal sequences. *Am J Clin Nutr.* 1991;53:1384–1396.
33. Jones AP, Friedman MI. Undernourished during pregnancy. *Science.* 1982;215:1518–1519.
34. Aaby P, Seim E, Knudsen K, Bukh J, Lisse IM, da Filda MC. Increased post-perinatal mortality among children of mothers exposed to measles during pregnancy. *Am J Epidemiol.* 1990;132:531–539.
35. Aaby P, Bukh J, Lisse IM, Seim E, da Filda MC. Increased perinatal mortality among children of mothers exposed to measles during pregnancy. *Lancet.* 1988;I:516–519.
36. Hatch MC, Beyea J, Nieves JW, Susser M. Cancer near the Three Mile Island Nuclear Plant. *Am J Epidemiol.* 1990;132:397–417.
37. Hatch MC, Susser M. Background gamma radiation and childhood cancers within ten miles of a US nuclear plant. *Int J Epidemiol.* 1990;19:546–552.
38. Hatch MC, Wallenstein S, Beyea J, Susser M. Cancer rates after the Three Mile Island nuclear accident and proximity to the plant. *Am J Public Health.* 1991;81:719–724.
39. Jablon S, Hrubek Z, Boice JD. Cancer in populations living near nuclear facilities. *JAMA.* 1991;265:1403–1408.
40. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol.* 1988;127:893–903.
41. 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.
42. Maclure M. Multivariate refutation of aetiological hypotheses in non-experimental epidemiology. *Int J Epidemiol.* 1990;19:782–787.