Socio-Econ. Plan. Sci. Vol. 18, No. 4, pp. 293–299, 1984
Printed in the U.S.A.

0038–0121/84 $3.00 + .00

GUIDELINES AND AN ILLUSTRATION OF THE USE OF ECOLOGICAL DATA FOR SEEKING CLUES OF EXCESS RISK[†]

MICHAEL GREENBERG
Graduate Program in Public Health, and School of Urban and Regional Policy, Rutgers University and the University of Medicine and Dentistry of New Jersey, New Brunswick, NJ 08903, U.S.A.

and

MARK BURRINGTON and CHARLES SMITH
Department of Geography, University of Illinois, 607 S. Mathews, Urbana, IL 61801, U.S.A.

(*Received* 5 *February* 1984)

Abstract—Ecological studies have acquired a bad reputation as weak scientific studies suffering from the ecological fallacy and many intractable limitations of data and method. In order to reduce these problems, it is suggested that analyses be focused on specific populations-at-risk, diseases, risk factors and places. In addition, seven guidelines are suggested which will limit the possibility of false positive results and provide the best clues for expensive, follow-up research. These guidelines are testing to make sure that the results are consistent across different times, places and methods. The guidelines were applied to finding the best clues for occupational-related cancers among white males aged 35–64 in the State of Illinois during 1950–1975. The most interesting clue found was that most coal mining areas in Illinois during the late 1960s and 1970s manifested the highest cancer mortality rates of trachea, bronchus, and lung and among the highest rates of increase of this type of cancer. This finding was unexpected because many studies have shown low lung cancer rates among coal miners.

1. INTRODUCTION

The goals of ecological studies are to find clues about the causes of disease and to test intervention programs[1].[‡] A typical ecological study focuses on the causes of disease at the census tract or county scales using correlation and regression methods to measure the extent of statistical association between indicators of disease (e.g. usually death rates, less frequently incidence and prevalence rates) and factors thought to contribute to increased or reduced risks of disease (e.g. smoking, ethnicity, socioeconomic status, air pollution). Some authors have questioned the efficacy of ecological studies[2, 3]. Many researchers with whom the senior author speaks will no longer undertake ecological studies because such studies are viewed as weak science. This paper argues that ecological studies are needed and that many of the weaknesses can be minimized by setting realistic research goals and by following guidelines. These guidelines are illustrated with data from a recently completed ecological study of occupational-related cancer in Illinois.

2. AN OVERVIEW OF MAJOR CRITICISMS OF ECOLOGICAL STUDIES

The goals, data and methods of ecological studies have all been questioned. The problem with goals is that ecological studies often are mistakingly used to imply cause-and-effect and to quantify risk. There are three major criticisms of data. The first of these is that the total population is studied rather than high risk subgroups, which are the populations that should be studied. The second is that a single disease rate is calculated (e.g. 1950–1969) so that clues that can be obtained from studying time series (e.g. 1950–1954, 1960–1964, 1970–1974) are missed. The third criticism of data is that the data sets about factors thought to increase or decrease the risk of disease are grossly inadequate. The major criticism of method is that results are typically unstable. This is because the results are dependent upon the method and change when the method changes; relationships assumed to be linear are not really linear; the data contain outliers; and ecological correlations suffer from multicolinearity (intercorrelation among the independent variables).

These criticisms and methods of limiting their impact are described in the next three sections of the paper.

3. REALISTIC GOALS FOR ECOLOGICAL STUDIES

First, and most important, it should not be the goal of ecological studies to find cause-and-effect relationships because of data and method limitations. Second, as laudible as their designers goals might be, a major reason that much of the ecological research is not as useful as it might be is that many ecological studies are conducted as fishing expeditions of every available population at risk and possible risk factor instead of as studies of specific diseases and subpopulations. Fishing expeditions with ecological data sets may find some interesting clues, but will also find

[†] This research was supported by a George A. Miller Fellowship from the University of Illinois and a Faculty leave stipened from Rutgers.

[‡] Morgenstern[1] carefully distinguishes between these two use of ecological data. This paper focuses on the first.

many spurious leads which cannot be distinguished from the real clues because such a wide net has been cast that the data and methods cannot separate the good from the bad clues.

An appropriate niche for ecological research is to identify places which show excessively high or extremely low disease rates and to suggest associated risk factors which should be given special attention in follow-up field studies. Toward this end, researchers should frame ecological studies with particular subpopulations, risk factors, diseases and regions in mind. If ecological studies are designed around a specific subpopulation, rather than all possible populations, then they become the best way of choosing places for detailed field surveys and environmental studies. Without ecological studies, analysts interested in field studies are left with educated guesses, locations that are convenient to scientists and suggestions based on political factors. Clues from ecological studies are imperfect, but provide a much firmer scientific basis for expensive field studies than do the above alternatives.

4. DATA

Two types of data are used in ecological studies: disease and surrogates for risk factors.

4.1 *Disease data*

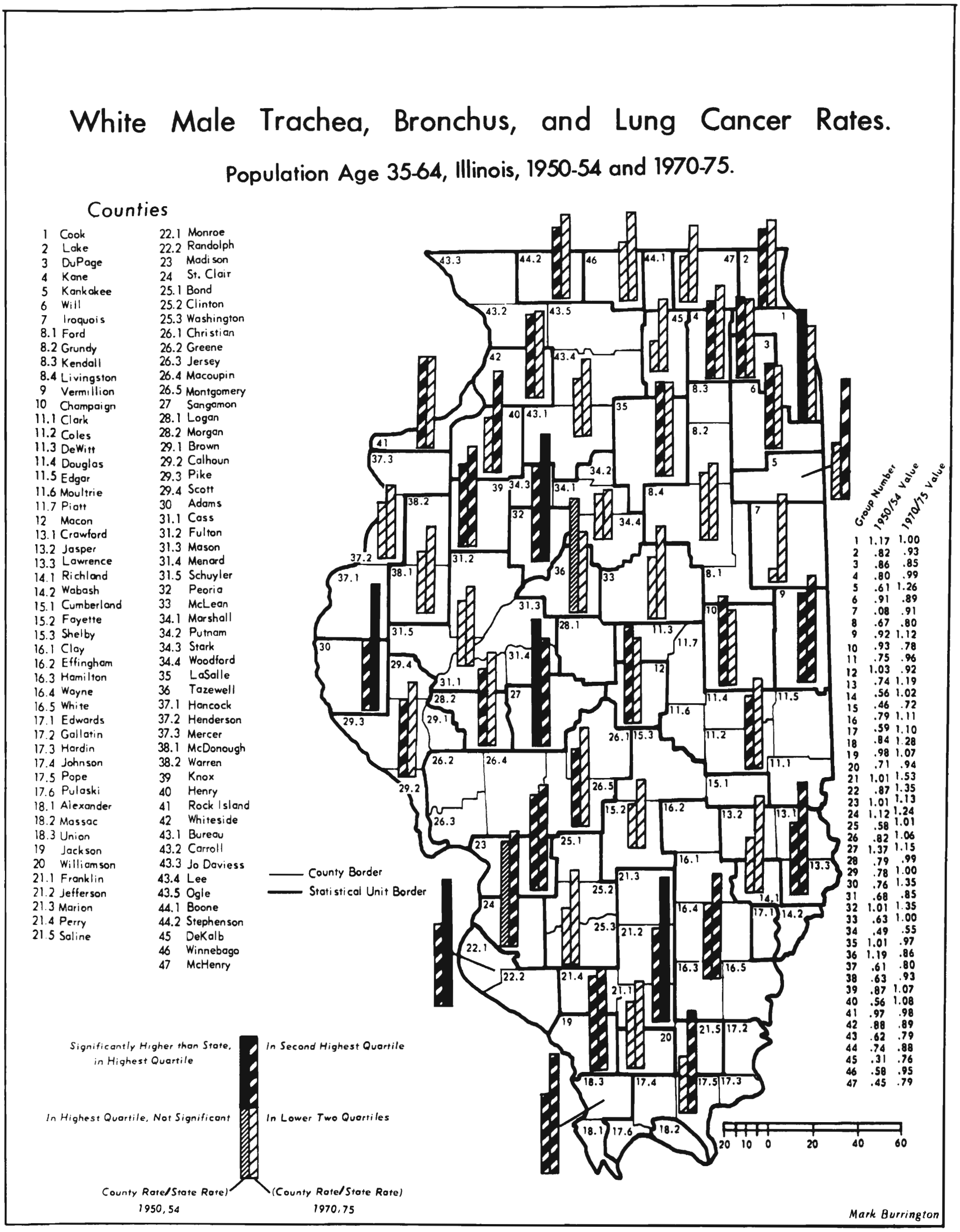
The unfortunate impression left by many ecological studies is that disease data are available only for total populations and broad time periods. This is not the case. If the initial goals of the research are specific, then the data can be carefully molded to the goals. For example, the goal of the Illinois ecological study was to seek clues about occupational-related cancers. The Illinois data set came from the data set that produced death rates that were used by the National Cancer Institute in their pioneering studies of cancer mortality in the United States[4]. In order to make the results of the Illinois study comparable with the national studies, the same methods were used in the Illinois study to calculate the populations at risk and to control for race, sex and age[4, 5]. By going back into the original data, the rates were calculated for the white male population 35–64 and separately for the years 1950–1954, 1955–1959, 1960–1964, 1965–1969 and 1970–1975. In addition, Chiang's[6] method was used to calculate confidence limits for each rate so that the rates of different places could be tested for statistical significance. In short, researchers are not restricted to published total population rates and to inflexible, broad time periods.

Another advantage of studying a specific group at risk is that the researcher can select specific diseases likely to be manifested in that population. For example, in the case of occupationally-related cancers, Doll and Peto[3] divide cancers into types not known to be produced by occupational hazards, possibly may be produced by occupational hazards, and cancers that definitely can be produced by occupational hazards. They ascribe percentages of cancer deaths to occupational exposures. The Doll and Peto and other lists must be adapted to the limitations of the ecological method. Ecological data sets can be controlled for age, race, sex and a few other characteristics, but cannot be controlled for the full range of risk factors that can make it difficult to determine whether one risk factor is more important than others. The implication of the inability to control for many risk factors is that it is highly unlikely that an ecological study will produce useful clues for a disease which has few deaths in a study area and in which the occupational factor is small (e.g. a county that had 20 deaths from a disease of interest and less than 10% of the deaths of this disease were thought to be related to occupational exposure).

The problem of few deaths can be controlled to some extent by combining places. This was done in the Illinois study. The State of Illinois has 102 counties. Previous trial-and-error research with New Jersey data[7] had shown that in order to have acceptable confidence in a death rate (standard error of the rate is no more than 40% of the rate) for a disease with a rate between 5 and 10/100,000 a population at risk of at least 25,000 (125,000 for a five-year study period) and preferably 50,000 (250,000 over five years) was needed. In the Illinois case a population of 25,000 or more white males aged 35–64 was the target. The 102 Illinois counties were combined into 47 aggregates to approach this target. Populous counties were not combined. The counties that were combined were primarily rural counties with relatively few residents. The aggregates were constructed by combining adjacent counties with relatively similar economic bases and levels of urbanization. The 47 counties and county aggregates are shown in Fig. 1. Combining adjacent counties has the added benefit of recognizing that what happens in one county is not independent of what happens in adjacent counties.

Even after reducing the number of working units from 102 to 47, the number of deaths for some of the cancers thought to be associated with occupational exposures was so small in so many of the aggregates that some of the aggregates would have had to have included 15–20 counties in order to obtain reliable death rates. Such a drastic reorganization of the data was not done because too much information would have been lost about individual counties. Thus, some of the diseases could not be analyzed in the Illinois case, including liver, larynx, bone, non-melanoma skin and prostate. Less than 6% of the 14,046 male cancer deaths in the United States in 1978 that were ascribed by Doll and Peto[3] to occupational exposure could not be studied in Illinois. Summarizing, one should use areas as small as possible as study units, but use the standard error of the rate, and knowledge about the demography and economy of an area as guides to the efficacy of spatial aggregation or of concluding that it is infeasible to study a disease in a region.

Eighty-seven per cent of the types of cancers that Doll and Peto ascribed to occupational exposure are of three types that could be studied in the Illinois case: lung (including trachea and bronchus); bladder; and leukemia. In addition to these three, pancreas, brain and central nervous system, and other lymphatic cancers including Hodgkin's Disease and non-Hodgkin's lymphoma were included in the research for two reasons. First, these types are considered as cancers that possibly may be produced by occupational hazards[3]. Second, each has many cases, is increasing among white males 35–64, and had higher rates in the State of Illinois than the United States. Without doubt, however, lung cancer was the key disease of interest. Doll



Group Number	1950/54 Value	1970/75 Value
1	1.17	1.00
2	.82	.93
3	.86	.85
4	.80	.99
5	.61	1.26
6	.91	.89
7	.08	.91
8	.67	.80
9	.92	1.12
10	.93	.78
11	.75	.96
12	1.03	.92
13	.74	1.19
14	.56	1.02
15	.46	.72
16	.79	1.11
17	.59	1.10
18	.84	1.28
19	.98	1.07
20	.71	.94
21	1.01	1.53
22	.87	1.35
23	1.01	1.13
24	1.12	1.24
25	.58	1.01
26	.82	1.06
27	1.37	1.15
28	.79	.99
29	.78	1.00
30	.76	1.35
31	.68	.85
32	1.01	1.35
33	.63	1.00
34	.49	.55
35	1.01	.97
36	1.19	.86
37	.61	.80
38	.63	.93
39	.87	1.07
40	.56	1.08
41	.97	.98
42	.88	.89
43	.62	.79
44	.74	.88
45	.31	.76
46	.58	.95
47	.45	.79

Fig. 1.

and Peto ascribe three-fourths of male occupational cancer to the lung. Overall, the mortality data could be molded because the Illinois research was aimed at a specific population.

4.2 *Surrogates for risk factors*

We are in complete agreement with those who contend that the worst data sets are the surrogates for risk or as they are sometimes called etiological factors. There are three main reasons for this conclusion. One is that there are no data for some important factors. For example, smoking, alcohol, nutrition and migration data are almost always missing from ecological studies. It is sometimes possible to use a disease to represent an etiological factor (e.g. lung cancer rates to represent smoking[8]) and to estimate these factors from other data sets (e.g. smoking related to demographic characteristics of the population[9]). But these are leaps of faith that can lead to erroneous conclusions. Second, surrogates may be measured at one scale, the disease data at another scale. The disease data may reflect one period, the surrogate data another. For example, the disease data are for areas and periods of 3–20 yr, whereas the environmental data are for a limited number of points in the area and are samples of a limited number of times, typically more recent than is desirable to include the latency of chronic diseases like cancer.

The Illinois surrogate data set is typical. There were no direct measures for many confounding factors such as smoking, diet and other lifestyle factors. The only available surrogates for these factors were U.S. Bureau of the Census data such as foreign born and stock, educational achievement, white collar occupation and income. Available indicators of air and water quality were totally inadequate in geographical and temporal coverage. The State of Illinois[10] published an atlas series which yielded data on many variables including the locations of different types of industry, agricultural production and mining. The atlases allowed us to test some variables that are not usually available. For example, geological data were included because some types of bedrock contain higher concentrations of hazardous substances than others and that difference may lead to differences in cancer risk[11–13]. Statistical tests of association between the selected diseases and more than 50 variables were made. The methodological section will describe the guidelines that were followed to make these tests.

5. METHOD

There are no magic tricks that can overcome poor data and the limitations imposed by working with ecological data. However, the following seven guidelines are suggested as methods of avoiding false positive errors (accepting a result that is false). They should help in the selection of the best clues for follow-up studies.

(1) Test more than one surrogate of the same risk factor (e.g. air quality, air emissions, fossil fuel generation, automobile traffic density) to be sure that an association between a disease and a risk factor is not spurious. If the results are consistent across a set of surrogates, the user can be more confident that the clue is not false than the user can be if only one surrogate is tested.

(2) Use general control variables (e.g. urbanization, all manufacturing workers) to screen out weakly correlated risk factors. When the general variables produce results that are as strong as the more specific surrogates (e.g. all manufacturing workers as strong as chemical workers, iron and steel workers) fallback to the general indicators. This guideline has to be carefully used because it can lead to false negative results.

(3) Use other diseases that should not be correlated with the occupational risk factors as controls. According to Doll and Peto[3] cancer of the pharynx, tongue, myeloma, and a few others are not known to be produced by occupational hazards. The finding of an occupational-lung cancer association would be strengthened by a simultaneous finding of no association between the occupation and these control cancers.

(4) Test to see if statistical associations are consistently significant in different time periods and in different regional configurations. This means testing the association between, e.g. chemical manufacturing employment and bladder cancer during more than one time period. A finding that the association was significant during only one time period does not mean that the clue should be discarded, but it does mean that a logical explanation would have to be found to explain why the association was significant during one period and not during another. The best way of assuring that a statistical association is not being unduely influenced by a few counties is to draw a random sample of the counties and retest for relationships. The finding that an association disappears does not necessarily mean that the relationship is spurious. It may mean, however, that the occupation-disease clue is restricted to a limited number of places, which may be an important clue to follow-up.

(5) Try to make the results independent of any single method. This can be done in two ways. One is to use a variety of parametric and nonparametric statistical tests. If a clue is good, it should reappear whether the statistical measure of association is the Pearson r, Kendall's Tau, a Spearman rank correlation, or a weighted Pearson r.

The second way is to use methods that do not require correlations between diseases and surrogates for risk factors. National Cancer Institute researchers have picked out places, diseases and etiological factor combinations for more detailed research by comparing the disease rates in target areas to rates for the nation as a whole and/or to nearby areas with similar demographic characteristics with the exception of the etiological factor of interest[14–17]. Another, related method, compares changes in disease rates in a local area of interest to changes in the region surrounding the county of interest and to the nation in a four-equation model[18]. Called the local component method, it assigns initial components of change in the local county cancer rate to national and then to regional trends. The difference between the actual change and the change in the county rate that would have occurred if the county rate had followed the national and regional patterns is considered to be the local county component of change. Any county that has a rate of increase that is higher than the national rate of increase and the rate of increase of its surrounding region should receive attention whether there are or there are not any ecological correlations.

(6) Look for and analyze exceptions to an association. For example, if one finds a consistent association between bladder cancer and the location of industry X, try to find and account for the few countries containing industry X and exhibiting low rates of bladder cancer.

(7) In order to seek further clues which may have been missed, study the geographical patterns of regression residuals.

There is some question about the relative utility of the correlation coefficient and the regression coefficient. In a recent paper supporting ecological analyses, Morgenstern[1] argued that the correlation coefficient but not the regression coefficient(s) may be biased where groups tend to be homogeneous with respect to one of the independent variables. Our experience with Illinois and New Jersey data and the experience of others[19] who have tested the impact of scale of the study unit on the correlation coefficient shows that it does change as the scale changes. But we are more interested in the consistency of the association across different times and places and methods than we are in the precise size of the correlation coefficient. With respect to regression coefficients, we have found that the independent variables are almost always correlated, and the only means of getting unbiased regression coefficients are by using one variable to represent a group of variables or to use a statistical method (e.g. principal components analysis, ridge regression) that will eliminate this source of bias. Overall, there is little question that studying regression residuals is potentially valuable. There is some question about the relative utility of the regression and correlation coefficients.

6. RESULTS

All of the data and method guidelines were applied to the Illinois data. A full presentation of these is beyond the scope of this paper. Results are presented insofar as they illustrate the application of the guidelines and are important results of the Illinois study. The major result is that the ecological correlation method produced one clue that seems worthy of further research; the local component method suggested four counties and county aggregates that seem worthy of follow-up. The results of the two approaches overlap.

6.1 *Ecological correlation: trachea, bronchus and lung cancer and underground coal mining*

The 1950–1954 tests with Kendall's Tau rank correlation, Pearson unweighted correlation and weighted correlations (weighted by total population and by the square root of total population) showed the expected association with urbanization (correlation ranged from 0.45 to 0.71, $p < 0.001$). The only other noteworthy statistically significant correlations at $p < 0.05$ in 1950–1954 were with the control variable total manufacturing employment (Kendall's Tau 0.57, $p < 0.001$) and the coal mining variables (Kendall's Tau ranged from 0.21 to 0.29, $p < 0.05$).

In 1970–1975 the coal mining variables were the strongest correlates in the rank and unweighted Pearson correlation tests (0.38 to 0.50, $p < 0.01$). When the rates were weighted by population size, the urbanized northern part of Illinois dominated the relationships (r was 0.65 and 0.66, $p < 0.001$).

Few of the statistical associations between lung cancer and specific manufacturing groups were statistically significant at $p < 0.05$; or if they were significant, they were strongly correlated with the control variables and the fallback guideline was employed. Furthermore, no interesting correlations were found between the ecological variables and the other types of cancer. The few that were statistically significant were between the cancers and the control variables.

Is the statistical association between lung cancer and coal mining in Illinois a real clue or a spurious statistical association? There is a substantial literature on the subject of which the following are exemplary and the last is the best summary of lung diseases among coal miners [20–25]. The majority of ecological and case studies of British, Australian and American coal miners and coal mining areas show deficits of lung cancer. The usual reason given is that workers die from pneumoconiosis and other lung diseases before they can die of lung cancer. But there are other explanations including that the load of dust in the lungs from working in coal mines and from burning soft coal at home leads to immunological enhancement of the lungs which leads to rejection of cells which had undergone some changes associated with neoplastic transformation. Another explanation [17] is that the overall cancer pattern in coal mining counties suggests a low social status lifestyle, one which is manifested in relatively high lung, cervical and stomach cancer rates and relatively low rates of leukemia, colon and breast cancer mortality.

The last often seen explanation is that cigarette smoking is the key factor, not exposure to coal dust. The evidence for this hypothesis is not conclusive. Fox [26] and Fox and Adelstein [27] found a Pearson correlation of 0.72 between lung cancer standard mortality ratios (SMR) and smoking scores among 25 occupational categories in Britain. The highest smoking score was 137 and was for miners and quarrymen. Their lung cancer SMR was 116. Green and Laquer [25] report research which shows that coal miners who smoked cigarettes had an eightfold excess of lung cancers deaths when compared with non-smoking coal miners.

The most recent major case-control effort in the United States was a study of almost 23,000 coal miners covered by the UMWA health and retirement funds during the mid-1970s [23]. The research found lung cancer to be moderately higher in the miner cohort than in the total U.S. male population. With respect to the Illinois research, the district results are intriguing. The U.S. was divided into 10 districts for the UMWA study. Illinois was included in the St. Louis district along with Arkansas, Iowa, Kansas, Missouri and Oklahoma. Standard mortality ratios for ten causes of death were reported including respiratory cancer, pneumoconiosis, emphysema without bronchitis. The St. Louis region had a relative risk of 0.98 for all causes of death, a result that was not statistically significant. The highest SMR for the St. Louis region was 1.22 for respiratory cancer (not significant at 0.05); the lowest was other respiratory diseases, particularly pneumoconiosis. Compared to the other 9 UMWA medical areas, the St. Louis region showed the most marked differences between respiratory cancer and other respiratory-related causes of death.

The lung cancer results for Illinois and the UMWA study suggested that Illinois could be an exception to the usual pattern of deficit of lung cancer and high rates of other chronic respiratory diseases in coal mining areas. But examination of 1959–1961 and 1973–1976 mortality data did not support this hypothesis. During 1959–1961 the average rate of mortality from non-neoplastic lung diseases in the 10 coal counties and coal county aggregates was 15% higher ($p < 0.20$) than the average rate in the remaining counties of Illinois. In 1973–1976, part of the period when the coal mining-lung cancer relationship was the strongest and most consistent ecological relationship in Illinois, the difference of means between the coal and non-coal counties was 12% ($p < 0.10$). Thus, lung cancer rates in the Illinois coal counties are high, but they do not seem to be high because of low death rates due to other lung diseases.

Almost all of the Illinois coal counties have high trachea, bronchus and lung cancer mortality rates. Ten of the 47 counties and aggregates were classified as underground coal mining areas. In 1970–1975, 5 of the 10 highest white male trachea, bronchus and lung cancer mortality rates in Illinois were in these counties, and 8 of the 10 (80%) had rates higher than the State of Illinois compared to 11 of the 37 for the remaining counties (30%) (see Fig. 1).

There are two exceptions. Williamson (county 20, Fig. 1) and an aggregate of five counties (Cass, Fulton, Mason, Menard and Schuyler) (county 31) are coal areas with trachea, bronchus and lung cancer mortality rates below the state rate in 1970–1975. Both exceptions potentially undermine the relationship between coal mining and lung cancer in Illinois because they contain major coal producing areas. Indeed, Williamson county produced more coal than any other county in Illinois during much of the 1950s. Yet in 1970–1975, Williamson's trachea, bronchus and lung rate was lower than the state average. However, in 1965–1969, Williamson had the highest age-adjusted rate in the State of Illinois for the white population 35–64 (102.9/100,000) and the second highest rate in the state in 1960–1964. Thus, it appears that Williamson is not an exception, but rather the cases appeared earlier than elsewhere in the State of Illinois. Perhaps, the second exception is explained by the fact that Fulton, a coal county, was aggregated with four other counties that were less oriented than Fulton to coal mining.

6.2 *Local component method*

Eight instances were found of a county or county aggregate with a cancer mortality rate that was significantly higher than the state rate at $p > 0.20$ in the period 1970–1975. Five of the 8 were cancer of the trachea, bronchus and lung; 2 were bladder cancer; and 1 was cancer of the brain and central nervous system. Four of the 8, all of the trachea, bronchus and lung, also had rates of increase from 1950–1954 to 1970–1975 that were markedly higher than the rate of increase in the nation and in the their surrounding regions. These 4 warrant further attention: Peoria (county 32); Adams (county 30); Franklin, Jefferson, Marion, Perry and Saline (county 21); and Kankakee (county 5).

There is nothing obviously in common to all four with respect to urbanization, industrialization, ethnicity, or socioeconomic status. The 5-county aggregate (Franklin, Jefferson, Marion, Perry and Saline) is also a major coal-mining area. It is the most interesting combination of disease, risk factors and place found in this study.

7. SUMMARY

The intent of this paper was to show that ecological data should and could be used to generate clues for case-control and environmental studies. The first key to making appropriate use of ecological data is to target specific populations rather than to try to find the veritable needle in the haystack. If realistic research goals are set, then the disease data can be configured around specific high-risk populations, multiple time periods, and the smallest possible homogeneous regions. While the limitations of poor data sets for etiological factors and the ecological fallacy cannot be dismissed, a set of seven guidelines were offered for avoiding false positive results and thereby of finding the most consistent clues for follow-up studies. They are:

(1) test more than one surrogate for the same etiological factor to be sure that statistical associations are not based on inadequate data;

(2) use general control variables to screen out weakly correlated, more specific surrogates for etiological factors;

(3) use control diseases that should not be associated with the etiological factors;

(4) test statistical associations for consistency in different time periods and in different regional configurations;

(5) try to make the results independent of method by using nonparametric and parametric weighted and unweighted correlation methods, and other methods such as the local component method;

(6) try to account for exceptions, if any, to the best clues;

(7) use the regression results to seek further clues.

When applied to the study of white male occupational cancers in the State of Illinois, the procedures presented in the paper provided two types of clues. One, based on ecological correlation, is that coal mining counties have unexpectedly high rates of cancer mortality of the trachea, bronchus and lung, especially in 1970–1975. The second, based on the local component method, isolated four counties and county aggregates which manifested high rates of cancer of the trachea, bronchus and lung and marked rates of increases in these rates during the study period. Since one of the these four areas is also a major coal mining area, the results of the two methods are complementary.

If the suggestions and guidelines offered in this paper were followed, we believe that ecological studies would be perceived as constructive and scientific contributions rather than as tenuous and uncertain scientific undertakings.

REFERENCES

1. H. Morgenstern, Uses of ecologic analysis in epidemiologic research. *Am. J. Pub. Hlth.* **72**, 1336–1344 (1982).
2. M. Hogan, P. Chi, D. Hoel and T. Mitchell, Association between chloroform levels in finished drinking water supplies and various site-specific cancer mortality rates. *J. Env. Path. Toxicol.* **2**, 873–877 (1979).
3. R. Doll and R. Peto, *The Causes of Cancer*, Oxford University Press, New York, (1981).
4. T. Mason and F. McKay, *U.S. Cancer Mortality by County:* 1950–1969. U.S. Government Printing Office, Washington D.C. (1974).
5. M. Greenberg, *Urbanization and Cancer Mortality*. Oxford University Press, New York (1983).
6. C. Chiang, Standard error of the age-adjusted death rate. *Vital Statistics* **47**, 275–285 (1961).
7. M. Greenberg, J. Caruana and R. Ziegenfus, Cancer mortality patterns in the New Jersey–New York–Philadelphia metropolitan region, 1950–1975, a final report, Part 4. New Jersey Department of Environmental Protection, Trenton, New Jersey (1980).
8. K. Cantor, R. Hoover, T. Mason and L. McCabe, Association of cancer mortality with halomethanes in drinking water. *J. Nat. Cancer Inst.* **61**, 979–985 (1978).
9. M. Greenberg, Cancer mortality in the New Jersey region, 1950–1969, 1968–1972, Part 2, Rep. to the New Jersey Department of Environmental Protection, Trenton, New Jersey (1979).
10. Division of Industrial Planning and Development, De-

partment of Registration and Education, State of Illinois. *Atlas of Illinois Resources*, 6 sections, State of Illinois, Springfield, Illinois (1959).

11. J. Bean, P. Isacson, W. Hausler Jr. and J. Kohler, Drinking water and cancer incidence in Iowa—I. Trends and incidence by source of drinking water and size of municipality. *Am. J. Epidemiol.* **116**, 912–923 (1982).
12. R. Ziskin, D. Smith, J. Hahn and G. Spivey, *Determinants of Cancer and Cardiovascular Disease Mortality in Asbestos Mining Counties of California*, National Technical Information Service, Springfield, Virginia (1978).
13. J. Bean, P. Isacson, R. Hahne and J. Kohler, Drinking water and cancer in Iowa—II. Radioactivity in drinking water. *Am. J. Epidemiol.* **116**, 924–932 (1982).
14. R. Hoover, T. Mason, F. McKay and J. Fraumeni Jr., Cancer by county: new resource for etiologic clues. *Sci.* **189**, 1005–1007 (1975).
15. R. Hoover and J. Fraumeni Jr., Cancer mortality in U.S. counties with chemical industries. *Environ. Res.* **9**, 196–207 (1975).
16. W. Blot, L. Brinton, J. Fraumeni Jr. and B. Stone, Cancer mortality in U.S. counties with petroleum industries. *Sci.* **198**, 51–53 (1977).
17. E. Creagen, R. Hoover and J. Fraumeni Jr., Mortality from stomach cancer in coal mining regions. *Arch. Env. Hlth.* **28**, 28–30 (1974).
18. M. Greenberg, A method to separate the geographical components of temporal change in cancer mortality rates. *Carcinogenesis* **1**, 553–557 (1980).
19. W. Clark and K. Avery, The effects of data aggregation in statistical analysis. *Geog. Anal.* **8**, 428–438 (1976).
20. L. Liddell, Mortality of British Coal Miners in 1961. *Brit. J. Industr. Med.* **30**, 15–24 (1973).
21. D. Ashley, Environmental factors in the aetiology of lung cancer and bronchitis. *Brit. J. Prev. Soc. Med.* **23**, 258–262 (1969).
22. J. Boyd, R. Doll, J. Faulds and J. Leiper, Cancer of the lung in iron ore (haematite) mines. *Brit. J. Industr. Med.* **27**, 97–105 (1970).
23. H. Rockette, *Mortality Among Coal Miners Covered by UMWA Health and Retirement Funds.* USDHEW, NIOSH, Rockville, Maryland (1977).
24. R. Armstrong, J. McNulty, L. Levitt, K. Williams and M. Hobbs, Mortality in gold and coal miners in western Australia with special reference to lung cancer. *Brit. J. Industr. Med.* **36**, 199–205 (1979).
25. F. Green and W. Laquer, Coal workers' pneumoconiosis (CWP). *Pathology Annual*: *Nineteen Eighty*. Part 2, pp. 333–410. Appleton-Century-Crofts, New York (1980).
26. A. Fox, Occupational mortality 1970–1972. *Population Trends*, pp. 1–8. Office of Population Censuses and Surveys, Her Majesty's Statistical Office, London (1977).
27. A. Fox and A. Adelstein, Occupational mortality: work or way of life? *J. Epidemiol. Comm. Health* **32**, 73–78 (1978).