

# Developing Clues to Environmental Cancer: A Stepwise Approach With the Use of Cancer Mortality Data

by William J. Blot,\* Joseph F. Fraumeni, Jr.,\* Thomas J. Mason,\* and Robert N. Hoover\*

Clues to environmental determinants can often be derived from the patterns of mortality from cancer. This review focuses on the stepwise approach of using cancer mortality maps, supplemented by correlation studies linking mortality rates with demographic and industrial data at the county level, to generate hypotheses to cancer etiology which can then be pursued by analytical epidemiological studies. Advantages and limitations of this approach and its application in the study of lung cancer in the United States are described.

## Introduction

Clues to the causes of cancer often exist in the patterns of cancer mortality and incidence. Occasionally their discovery is relatively easy. The development of an extremely rare liver tumor (hepatic angiosarcoma) among three workers in a single manufacturing plant thus led to the identification of vinyl chloride as the likely causal agent (1). Such case clusters are usually detected by alert clinicians, but they can also be gleaned by examinations of routinely collected mortality and incidence statistics. The link between nasal adenocarcinoma and woodworking was first suggested by a clustering of cases in High Wycombe, England (2). In subsequent studies, the tumor was found to be excessive among furniture makers (3, 4). In the United States, an association was not made until a correlation study found that the mortality rates for nasal cancer were unusually high in counties with a heavy concentration of furniture manufacturing industries (5). Later examination of death certificates in these counties showed a fourfold excess risk among those whose usual trade, as indicated on the death certificate, was furniture manufacturing (6). Although the American

study only replicated findings first seen elsewhere, it is noteworthy that an examination of routinely collected mortality data quickly and accurately pinpointed an industrial cancer hazard. Of course, not all leads generated from such statistics prove to be good ones, but findings such as these have led to increased confidence that the mortality and incidence patterns of cancer represent a good starting point for etiologic studies.

This review will focus on our efforts to use mortality data within the United States to generate and test hypotheses to cancer etiology, stressing both the limitations and advantages of this approach.

## Cancer Mortality Data in the U.S.

Notifications of all deaths occurring in the United States are sent by the individual states to a central repository in the National Center for Health Statistics (NCHS). Copies of individual death certificates, however, are not delivered to NCHS but maintained in state offices. The NCHS prepares computer listings of mortality. The computer tapes are in single record format and contain the age and year of death, sex, race, county of usual residence, and specific cause of death for each individual who died. No identifying information, such as name, social secu-

\*Environmental Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

city number, or street address, is released.\* In addition, neither the occupation nor industry listed on the death certificates are coded. The computer tapes are updated yearly and available to the general public.

From the NCHS computer listings, we identified the numbers of deaths in the U.S. attributed to cancer during the 20-year period, 1950-69. Using age-, sex-, and race-specific county population estimates available from the decennial censuses of 1950, 1960, and 1970, it was possible to calculate mortality rates for the 3056 counties of the United States. Age-adjusted rates of mortality for 35 cancers for the individual counties were published in tabular form in 1974 (7) and are available on tape from our office. The relative distributions of these rates were then plotted in a series of computer-generated color maps published in two atlases for the white and nonwhite population (8, 9). A similar atlas plotting the distribution of mortality rates for non-neoplastic diseases (concentrating on those related to cancer) is now in press (10). A report updating mortality data through 1975, and focusing on changes in cancer patterns since 1950, is in preparation.

## Uses of Mortality Data For Etiologic Research

We have used routinely collected county mortality data in a stepwise fashion. The cancer maps (8, 9) provide the means for identifying high-risk areas where further research might pay off. Secondly, we have conducted a series of correlation studies linking the county mortality rates with demographic, socioeconomic, industrial, and environmental data at the county level.† These studies have provided additional leads that were not visually evident from

---

\*Since identifiers are not included in the NCHS tapes, it is not currently feasible to use these listings either for identifying cancer cases (and controls) for further retrospective investigation or for determining the mortality status of a cohort of known individuals. Thus there is no National Death Index in the United States and no equivalent to the "koseki" system of Japan. Progress is being made in getting such an index established, and it now appears that NCHS will initiate a national registry (with identifiers) of deaths beginning in 1979. Special studies, however, can often be accomplished by making arrangements with interested states. Although the criteria for release of data varies, states are generally willing to provide copies of death certificates for epidemiologic research purposes.

†We use the term "correlation" very loosely to refer to analyses which attempt to measure the association of county mortality rates with demographic and other variables. The statistical method most often employed to measure and test association is weighted linear multiple regression, supplemented by ridge regression analyses (11). Such studies in which aggregate (not individual) data sources are linked are often referred to as "ecological" studies.

the maps, and have helped to refine and narrow the hypotheses suggested by the geographic patterns of various cancers. Finally, we have embarked upon a series of analytic field studies, mostly case-control (retrospective) interview studies in parts of the country where mortality rates are high for particular cancers. As an intermediate step between correlation studies and field interview studies, we have often examined death certificates from certain localities for people who died of cancer relative to other causes of death. Although only limited information is available on the certificate, the statements with respect to birthplace, marital status, duration of residence within the county, occupation, and industry among cancer cases and controls can be compared.

There are limitations as well as strengths in each of these steps in uncovering new information about the causes of cancer. Computer technology has now advanced to the point where it is relatively easy and inexpensive to construct maps of cancer mortality or, for that matter, maps of any county characteristic for which data are available. The input into our computer mapping routine are  $x$ - $y$  coordinates of the county boundaries in the United States and classifications as to which color to shade each county. In our system the maps are initially produced on microfiche, being in effect photographed from a cathode-ray tube screen, at a cost of less than \$10.00 per film. Given a data file which lists, for example, median income for 1975 for each county of the U.S., we can readily produce a map of the geographic distribution of income across the country by specifying the quintile rank (or other categorization) for each county.

The primary advantage to mapping mortality or other variables is that a large volume of data can be visually comprehended almost instantaneously. Thus, it is immediately evident that the highest stomach cancer mortality rates in the United States congregate in counties in the upper north central region of the country, and that skin cancer mortality is elevated throughout the South (8). Since the distribution of cancer across the U.S. is not random, with the patterns for each cancer differing and with clusters of high mortality not delimited by state boundaries, environmental influences seem likely. Clusterings of high rates in certain communities serve as "smoke signals" to environmental exposure that may be uncovered through further study. By revealing the distribution of cancer over the country, the maps help generate etiologic hypotheses, thus raising questions about cancer causes rather than answering them.

Also useful for hypothesis generation is the correlational approach, relating cancer mortality rates

with demographic and industrial data at the county level. Taking into account the potential pitfalls of multiple regression analyses (11), the technique can help determine whether the geographic variation in cancer rates is related to urbanization, socioeconomic status, ethnicity, industrial exposures, or other indices. The associations identified by correlation studies, despite their refinement over the maps, still serve only as clues rather than hard evidence of causation. This is illustrated by the correlation of high bladder cancer rates with the presence of chemical manufacturing plants, regardless of regional location or urbanization level of the county (12, 13). A cause-and-effect relationship seems plausible, since it is known that  $\beta$ -naphthylamine and other aromatic amines in these plants can induce bladder tumors in workers (14, 15). However, it is impossible to make a causal inference based on the county data alone, since we have no information on the individuals who died in the counties, their occupations, duration of residence and employment, or personal habits (e.g., smoking) that may affect bladder cancer risk.

Ecological studies based on county data provide a compromise between studies comparing characteristics of larger geographic units such as nations or states and studies comparing characteristics of individuals. Indeed, counties seem to be particularly good units for study, since they are usually small enough to be relatively homogeneous with respect to environmental features, yet large enough to yield stable mortality rates. Even smaller geographic units in the U.S., such as zip codes or census tracts (defined only within standard statistical metropolitan areas), could also be used in correlation studies, but age-, race-, and sex-specific mortality and census data are not routinely recorded at levels below the county and even where available, the numbers of cancer deaths in these subdivisions might be too small, particularly if subdivided by age or time period, for meaningful analysis.

The leads produced by the cancer maps and correlation studies can be pursued by cohort (prospective) studies, if a particular high-risk group (e.g., industrial exposure) appears to be influencing the distribution of cancer. It is usually more appropriate to carry out case-control (retrospective) investigations in areas with high rates. As an intermediate step, a comparison of death certificates of persons who died of cancer relative to those who died of other causes is often a quick and inexpensive means of bringing the analysis from the aggregate (county) to the individual level. In particular, the death certificate statements on occupation and industry may be scanned for case-control differences. Although

they represent only a crude description of the decedent's usual work, with no information as to detail, variety, or duration of employment, a comparison of the statements can help test, or at least sharpen hypotheses about occupational factors, and thus aid in the decision-making process of whether to invest in more costly and time-consuming studies in the field.

In the next section we will outline how we have followed this stepwise approach, from the maps to correlation analyses to death certificate examinations to field interview studies in a search to identify the environmental determinants of lung cancer, the most common cancer among men in the United States.

## **An Example — Lung Cancer in the U.S.**

The geographic pattern of lung cancer mortality, 1950-69, among white males was unexpected. The cancer maps showed high rates for this tumor in metropolitan areas of the Northeast and Great Lakes region, but the highest mortality clustered in coastal areas of the South (Fig. 1). Mortality was elevated in counties along the Gulf of Mexico from Texas eastward to the Florida panhandle (with high rates especially concentrated in southern Louisiana), and along a strip of counties on the Atlantic Coast from below Jacksonville, Florida northward through Charleston, South Carolina. A similar pattern was seen for lung cancer among white females, but the coastal phenomenon was less intense.

To seek explanations for the unusual distribution of this tumor in the United States, we correlated the county rates with a variety of county indices (16). Mortality increased with urbanization (Table 1) throughout the country, and tended to be inversely related to socioeconomic status as measured by median education level, median income, or a linear combination of both. Controlling for these factors as well as broad regional location, we examined the relationships between the lung cancer rates and measures of industrial activity. From special manufacturing censuses, the county location and approximate work force of every manufacturing plant in the U.S. was known. Using the manufacturing data for 1963 (17), the earliest year for which county level data were available to us on computer tape, we classified each county as to its relative involvement in each of 18 major industrial categories (defined by two-digit standard industrial classification codes). Relating these indices to the mortality rates revealed positive associations between lung cancer and the chemical, petroleum, and paper and pulp manufac-

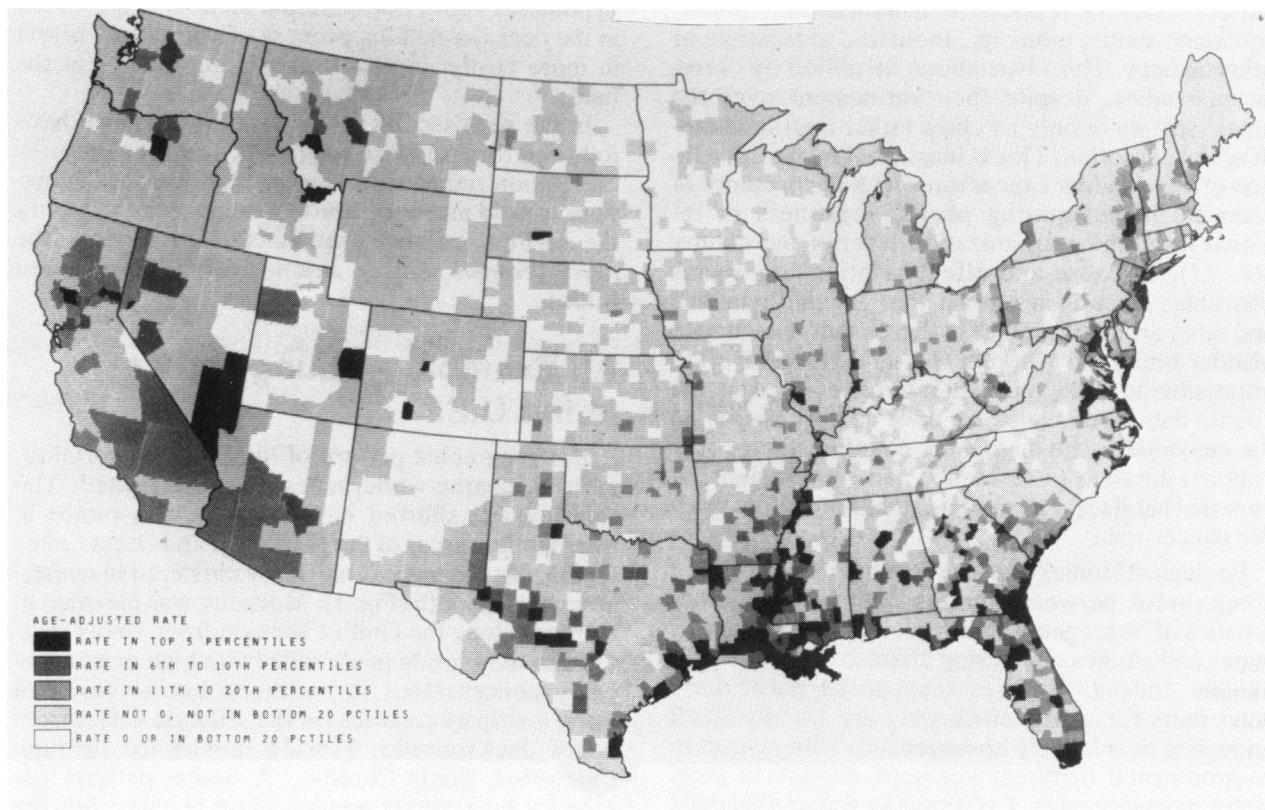


FIGURE 1. U.S. county lung cancer mortality for white males, 1950–68.

Table 1. Average annual age-adjusted mortality rates (per 100,000 population) for lung cancer, 1950–69, among whites, according to sex, geographic region, and urbanization.<sup>a</sup>

Sex	Geographic region	Urban, %			
		0-24.9	25-49.9	50-74.9	75+
Males	Northeast	32.8	35.2	37.7	43.5
	Southeast	29.1	32.4	38.7	46.0
	Midwest	26.1	29.2	32.6	39.6
	South Central	31.2	34.4	36.0	40.3
	North Central	21.9	23.9	27.2	34.5
	Mountain	23.0	23.9	26.8	30.4
	Far West	32.0	32.8	34.2	40.7
Females	Northeast	5.9	5.6	5.6	6.7
	Southeast	5.1	5.1	6.0	6.8
	Midwest	4.9	5.0	5.4	6.3
	South Central	5.4	5.3	5.9	6.9
	North Central	4.2	4.7	4.8	5.4
	Mountain	4.3	4.6	5.1	5.6
	Far West	5.7	6.2	6.2	7.7

<sup>a</sup>Adopted from Blot and Fraumeni (16).

turing industries. The correlations were seen among males, but not females, further implying an occupational link. Since the petrochemical industry is prominent along the Texas and Louisiana coast and the paper industry is the largest employer along the northern Florida-Georgia coast, it seemed as if the excess rates in these areas may have an industrial component.

To further investigate the Georgia cluster, we obtained copies of death certificates of approximately 1700 while male residents of coastal Georgia who died during 1961-74, half due to primary lung cancer, half to other causes. Comparison of the occupational statements showed a higher proportion of lung cancer than control certificates mentioning work in the wood-paper industry, but the excess was limited to residents of rural coastal counties and not found among residents of the three major cities in the area (18). The lack of uniformity suggested that a single explanation for the area's high rates was unlikely. Nationally, the lung cancer rates in paper mill counties were also inconsistent: high in the East and South, but unremarkable in the north central states and far west (16).

Concurrently with the death certificate analysis, we initiated a case-control study interviewing all recent cases (or their next-of-kin in the event they had died) with lung cancer diagnosed at the four large hospitals located in coastal Georgia. Hospital controls of similar age, sex, race, and county of residence were also identified. Approximately 90% of both cases and controls could be located. Of these, 97% consented to be interviewed for information on their, or their next-of-kin's lifetime residence, occupation and smoking histories.

The interview data uncovered a significantly increased lung cancer risk associated with work in area shipyards during World War II (19). The association was not accounted for by smoking habits or other occupations, and was seen in blacks and whites and in both Savannah and Brunswick where the shipyards were located during the war. The summary relative risk was 1.6 with 95% confidence limits of 1.1 to 2.1. There was some suggestion of an interaction with cigarette smoking, with relative risk being higher among heavier smokers. The findings suggest that asbestos exposures during wartime employment in shipyards may be responsible, at least in part, for the excessive mortality from lung cancer in coastal Georgia.

Our review of the "usual" occupations listed on the death certificates in the same area of coastal Georgia (18) revealed no association with shipbuilding. However, the shipyards in Brunswick and Savannah, which together employed over 35,000 persons in late 1943, closed down after the war.

Although over 20% of the lung cancer cases in the interview study reported working in shipyards at some time during their careers, hardly any listed this as their usual industry of employment. The correlation study (16) did find that rates tended to be high in counties with shipbuilding industries in 1963, but the number of large shipyards by then was small. Indeed, from a peak of 1.7 million employees in 1943, the shipbuilding industry work force rapidly declined to about 150,000 by 1950 and has not changed much since.

The findings of the Georgia interview data together with reports of asbestosis among shipyard workers elsewhere in the U.S. (20) and mesothelioma among shipyard workers in Europe (21) prompted an examination of cancer mortality in counties throughout the U.S. with large shipyards during World War II. This correlation study revealed that the rates for lung and laryngeal cancers were consistently high, with some suggestion that the rates for oral-pharyngeal, esophageal, and stomach cancers were elevated as well (22). This constellation of tumors affecting the airways and the upper gastrointestinal tract has been described previously in asbestos workers (23), suggesting that shipyard exposures to asbestos may have contributed to the clustering of several cancers in coastal areas of the country.

To further evaluate the role of shipyard exposures in the U.S., we have initiated a case-control interview study of lung cancer in the Norfolk-Newport News area of Virginia, the site of large Navy and private shipyards. This region also shows an excess mortality from lung cancer among white males. The shipyards employed over 70,000 workers during the war, but unlike the situation in Georgia, continue to be a producer of large naval ships today. We have also examined hospital records in a county in coastal Maine, site of the oldest shipbuilding company in the U.S., and preliminary analyses indicate a higher frequency of shipyard employment listed on the records for lung cancer compared to other discharge diagnoses.

Analytic studies in other areas of the U.S. where lung cancer mortality rates are high are in progress. Death certificates in southern Louisiana are now being scanned for case-control differences with respect to occupation and Acadian ancestry (as judged by family surname), with interview surveys to begin this fall. An earlier correlation study examining lung cancer mortality in U.S. counties where nonferrous metal smelters are located showed elevated lung cancer rates among both sexes, raising the possibility of a community wide effect of arsenical air pollution (24). Death certificates from ten copper, lead, or zinc smelter counties are now being examined, with

case-control interview studies in three locations set to begin soon. The higher lung cancer rates observed in counties with petroleum industries (16, 25) provided an additional clue that is being pursued in the Louisiana study, as well as in a proportional mortality analysis now being conducted among members of the Oil, Chemical, and Atomic Workers' union in this country.

Thus, despite the fact that cigarette smoking is the major cause of lung cancer in the U.S., the county-by-county maps, correlation analyses, and analytic studies suggest that other environmental determinants (especially industrial exposures) are involved to an extent greater than previously thought.

#### REFERENCES

1. Creech, J. L., and Johnson, M. N. Angiosarcoma of the liver in the manufacture of polyvinyl chloride. *J. Occup. Med.* 16: 150 (1974).
2. Macbeth, R. Malignant disease of the paranasal sinuses. *J. Laryngol. Otol.* 79: 592 (1965).
3. Acheson, E. D., Hadfield, E. H. and Macbeth, R. G. Carcinoma of the nasal cavity and accessory sinuses in woodworkers. *Lancet* i: 311 (1967).
4. Acheson, E. D., Cordell, B. H., Hadfield, E., and Macbeth, R. G. Nasal cancer in woodworking in the furniture industry. *Brit. Med. J.* ii: 587 (1968).
5. Brinton, L. A., Stone, B. J., Blot, W. J., and Fraumeni, F. F., Jr. Nasal cancer in U.S. furniture industry counties. *Lancet* i: 628 (1976).
6. Brinton, L. A., Blot, W. J., Stone, B. J., and Fraumeni, J. F., Jr. A death certificate analysis of nasal cancer among furniture workers in North Carolina. *Cancer Res.* 37: 3473 (1977).
7. Mason, T. J., and McKay, F. W. U.S. Cancer Mortality by County: 1950-69. DHEW Publ. No. (NIH) 74-615, U.S. Government Printing Office, Washington, D.C., 1974.
8. Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., and Fraumeni, J. F., Jr. Atlas of Cancer Mortality for U.S. Counties: 1950-69. DHEW Publ. No. (NIH) 75-780, U.S. Government Printing Office, Washington, D.C., 1975.
9. Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., and Fraumeni, J. F., Jr. Atlas of Cancer Mortality among U.S. Nonwhites: 1950-69. DHEW Publ. No. (NIH) 76-1204, U.S. Government Printing Office, Washington, D.C., 1976.
10. Mason, T. J., et al. Atlas of Non-Neoplastic Diseases. In press.
11. Mosteller, F., and Tukey, J. W. Uses of regression coefficients. In: *Data Analyses and Regression*, Addison-Wesley, Reading, Mass., 1977, p. 299.
12. Hoover, R., and Fraumeni, J. F., Jr. Cancer mortality in U.S. counties with chemical industries. *Environ. Res.* 9:196 (1975).
13. Blot, W. J., and Fraumeni, J. F., Jr. Geographic patterns of bladder cancer in the United States. *J. Natl. Cancer Inst.* 61: 1017 (1978).
14. Case, R. A. M., Hosker, M. E., McDonald, D. B., and Pearson, J. T. Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuffs intermediates in the British chemical industry. *Brit. J. Ind. Med.* 11: 75 (1954).
15. Clayson, D. B. Occupational bladder cancer. *Prev. Med.* 5: 228 (1976).
16. Blot, W. J., Fraumeni, J. F., Jr. Geographic patterns of lung cancer in the United States: industrial correlations. *Am. J. Epidemiol.* 103: 539 (1976).
17. U.S. Department of Commerce. Census of Manufactures 1963. U.S. Government Printing Office, Washington, D.C., 1966.
18. Harrington, J. M., Blot, W. J., Hoover, R. N., Housworth, W. J., Heath, C. W., and Fraumeni, J. F., Jr. Lung cancer in coastal Georgia: a death certificate analysis of occupation. *J. Natl. Cancer Inst.* 60: 295 (1978).
19. Blot, W. J., Harrington, J. M., Toledo, A., Hoover, R., Heath, C. W., and Fraumeni, J. F., Jr. Lung cancer after employment in shipyards during World War II. *N. Engl. J. Med.* 299: 620 (1978).
20. International Agency for Research on Cancer. IARC monographs on the Evaluation of Carcinogenic Risk of Chemical to Man, Vol. 14, Asbestos. IARC, Lyon, France, 1977.
21. Selikoff, I. J., and Hammond, E. C. Asbestos-associated disease in United States shipyards. *CA* 28: 87 (1978).
22. Blot, W. J., Stone, B. J., Fraumeni, J. F., Jr., and Morris, L. F. Cancer mortality in U.S. counties with shipyard industries during World War II. *Environ. Res.* 18: 281 (1979).
23. Selikoff, I. J. Cancer risk of asbestos exposure. In: *Origins of Human Cancer*. H. Hiatt, J. Watson, and J. Winston, Eds., (Cold Springs Harbor Conference on Cell Proliferation, Vol. 4), Cold Springs Harbor Laboratory, Cold Spring Harbor, N.Y., 1977.
24. Blot, W. J., and Fraumeni, J. F., Jr. Arsenical air pollution and lung cancer. *Lancet* ii: 142 (1975).
25. Blot, W. J., Brinton, L. A., Fraumeni, F., Jr., and Stone, B. J. Cancer mortality in U.S. counties with petroleum industries. *Science* 198: 51 (1977).