Three-Dimensional Computer Mapping of Disease in Los Angeles County

KENNETH P. SATIN, DrPH
RALPH R. FRERICHS, DVM, DrPH
ELIZABETH M. SLOSS, MPH

A FUNDAMENTAL PREMISE OF EPIDEMIOLOGY is that disease is not randomly distributed in a population. Therefore epidemiologic studies are often conducted to ascertain the distribution and frequency of disease within a population. In such studies attention is routinely focused on determining which subgroup is affected and where and when the disease is occurring. Study of this trilogy of factors, commonly referred to as person, place, and time, can be a useful way of discovering the health profile of a community.

Dr. Satin, now an epidemiologist, California Department of Health Services, Berkeley, at the time of the study was a graduate student in the Division of Epidemiology, University of California at Los Angeles. Dr. Frerichs is associate professor of epidemiology, UCLA School of Public Health, Los Angeles, Calif. 90024. Ms. Sloss is a graduate student in the Division of Epidemiology, UCLA.
The study was funded by the Sanitation Districts of Los Angeles County (CSD No. 2402) and by the U.S. Environmental Protection Agency (EPA No. R806399010). Tearsheet requests to Dr. Frerichs.

The frequency of a disease may be related to geographic location, that is, to the specific environmental characteristics of a region. These environmental characteristics may be of natural origin, like the fungus *Coccidioides immitis* which causes San Joaquin Valley fever in the population of arid regions of the southwestern United States. Or they may arise because of the actions of man, as did the air pollution in London that was responsible in 1952 for an increase in deaths related to respiratory impairment (*1,2*). Countless other diseases are also known to be related to the environment.

Because the geographic distribution of a disease may stimulate hypotheses regarding its etiology, medical geographers and epidemiologists have commonly displayed disease patterns in maps. For example, in the outbreak of cholera in London in 1848–54, John Snow mapped the exact location of those who died of cholera, noting that the sites were clustered around the Broad Street water pump and thus were related to the water supply (*3*). Since then, maps have often been used to display epidemiologic data—usually mortality or morbidity rates—for the geographic areas of interest. On a shaded map of Pittsburgh, Patno (*4*) showed the prevalence of cancer in that city in 1947. Greenburg (*5*) presented 1959–61 mortality rates on a map of 10 subareas of Staten Island, New York. Chapman and Coulson (*6*) prepared colored maps showing the 1966 mortality patterns in Los Angeles County by census tract for infant deaths, total mortality, and mortality from cancer, ischemic heart disease, stroke, and accidents.

More recently, Mason and McKay (*7*) have published the results of one of the largest studies of disease occurrence by geographic area. The maps for the study, showing the cancer mortality rates for each county in the United States ($N = 3{,}056$), were produced by an automated cartographic (computer mapping) system. With these data, many investigators have arrived at new hypotheses about the etiology of certain cancers (*8–10*).

Two-dimensional maps such as those mentioned in the preceding paragraphs are often difficult to interpret and summarize, however, because they contain numerous subareas of multiple colors or patterns. We describe here a computer technique for displaying spatial data on disease frequency that minimizes these difficulties by the introduction of a third dimension to maps.

Methods

Preparation of data for mapping. The data to be mapped were obtained from computer files of vital statistics records for Los Angeles County (LAC) for 1969–71 and from the second-count data computer tapes of the 1970 U.S. Census of the Population for LAC. Two measures of health outcome were analyzed: death from all causes and morbidity related to potential waterborne diseases. These diseases were selected from the list of notifiable diseases reported to the Los Angeles County Health Department and include infectious hepatitis, shigellosis, meningitis of undetermined etiology, typhoid, salmonellosis, amebiasis, meningococcal infections, meningitis of etiology other than mumps, and leptospirosis. These measures of health outcome were selected to coincide with those used in the analyses of a larger study being conducted in Los Angeles County to assess the impact on human health of consumption of reclaimed wastewater (*11*). These broad health outcome measures were used because in the absence of a complete inventory of waterborne contaminants, hypotheses regarding specific health effects are difficult to formulate.

The smallest geographic unit for which both vital statistics data and census data are coded is the census tract. Since the vital statistics data had been coded with 1960 census tract codes and the 1970 population data were available only by 1970 census tracts, the census tract unit had to be modified to account for the changes in census tract boundaries between the 1960 and 1970 censuses. To link the 1960 and 1970 census tracts so that the health data and the census data would relate to the same geographic area, a "correspondence tract" was formed that maintained the same geographic boundary for both census periods. Thus, in Los Angeles County, 1,159 correspondence tracts linked the 1,297 census tracts of 1960 to the 1,576 census tracts of 1970.

Measurement of morbidity and mortality. Our analysis of the pattern of disease in Los Angeles County was based on a standardized mortality ratio (SMR) for each correspondence tract. Similarly, a standardized morbidity ratio (SMR) was used in analyzing the LAC waterborne disease distribution. The SMR is the ratio of the observed number of events in an area (determined from the vital statistics records) divided by the expected number of events in the same area. To obtain the expected number of events, the rate of the condition of interest as observed in a standard population is applied to the population of the local area. For example, in the analyses we performed, the standard population comprised residents of Los Angeles County in 1970, and the local population consisted of the residents of each correspondence tract. If an SMR was equal to 1.0, the local area was experiencing the same number of events as would be expected, given the average rates for the standard area. If an SMR was less than 1.0, fewer events were occurring than expected. If an SMR was greater than 1.0, more events were occurring in the local area than would be expected based on the experience of the county as a whole. An advantage of using the SMR rather than a crude disease or mortality rate is that the SMR can always be viewed in terms of its reference value, 1.0. Thus, regardless of the absolute disease rate, the health experience of the local area can be compared with a reference area by noting the deviation of the SMR from unity.

Figure 1. Centroids of census tracts in Los Angeles County, 1970

To determine if the observed number of events differed significantly from the expected number (that is, if the SMR for each tract differed from 1.0), we assumed that the observed number of deaths in a given tract followed a Poisson distribution with a mean equal to the expected number of deaths. The standardized ratio was tested against the hypothesized value of 1.0 (no difference from the standard) by using the Poisson distribution for small numbers of observed deaths—less than 10—and the standard normal distribution for observed deaths of 10 or more (*12*). For a small number of events, the probability that the observed results were due to chance was estimated from a Poisson distribution by using the following formula for the probability that a particular value, X, would occur:

$$P(X) = \frac{e^{-\mu} \mu^{x}}{X!}$$

where $\mu =$ the expected number of events. For 10 or more events, the normal approximation to the distribution was used. We derived the probability that the observed results were due to chance from the normal distribution, using the standard normal deviate (z) given by Armitage (*12*),

$$z = \frac{O - E}{\sqrt{E}}$$

where O is the observed number of events and E the expected number. The choice of 10 events as a cutpoint is arbitrary, although in accord with the procedure followed by Haenszel and associates (*13*). In addition, for the mapping procedure, tracts with less than 1.0 expected events were assigned an SMR of 0.0 because it was assumed that a small number of events would result in unstable and potentially large or misleading ratios. Tracts with no observed events were also assigned the value 0.0 for their SMR.

Computer mapping procedure. We used the SYMAP and ASPEX computer programs developed by the Laboratory for Computer

Figure 2. Geographic orientation to three-dimensional view of Los Angeles County

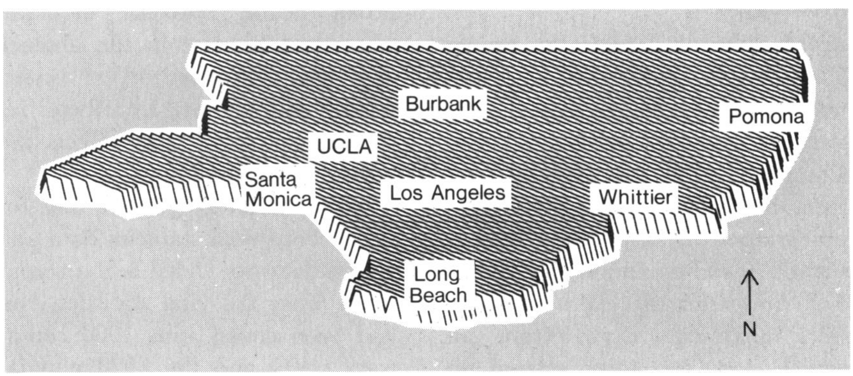
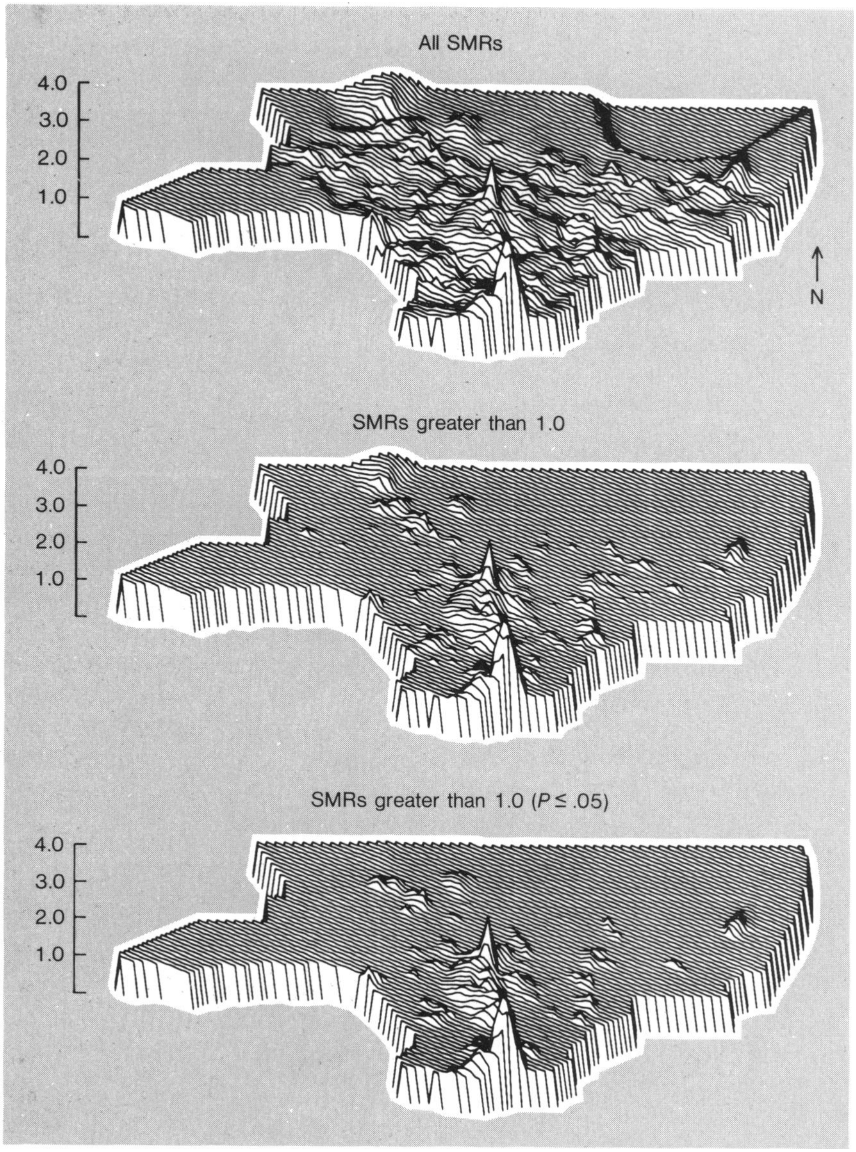


Figure 3. Age- and sex-standardized mortality ratios (SMRs) for all causes of death, Los Angeles County, 1969-71



Graphics and Spatial Analysis, Graduate School of Design, Harvard University, to produce three-dimensional maps for the SMRs. The SYMAP program was used first to assign geographic coordinates to the raw data, in this case a standardized ratio corresponding to the centroids of the census tracts in LAC. For the SYMAP program, both the geographic coordinates that represent the outline of the study area and the centroids of the basic geographic unit are needed. In some instances, a computer file containing the coordinates can be purchased from private purveyors. In general, however, the set of coordinates can be determined by tracing an outline of the study area onto a rectangular grid labeled with Cartesian coordinates, from which the outline and centroid points can be directly identified. These coordinates are then entered into the SYMAP program.

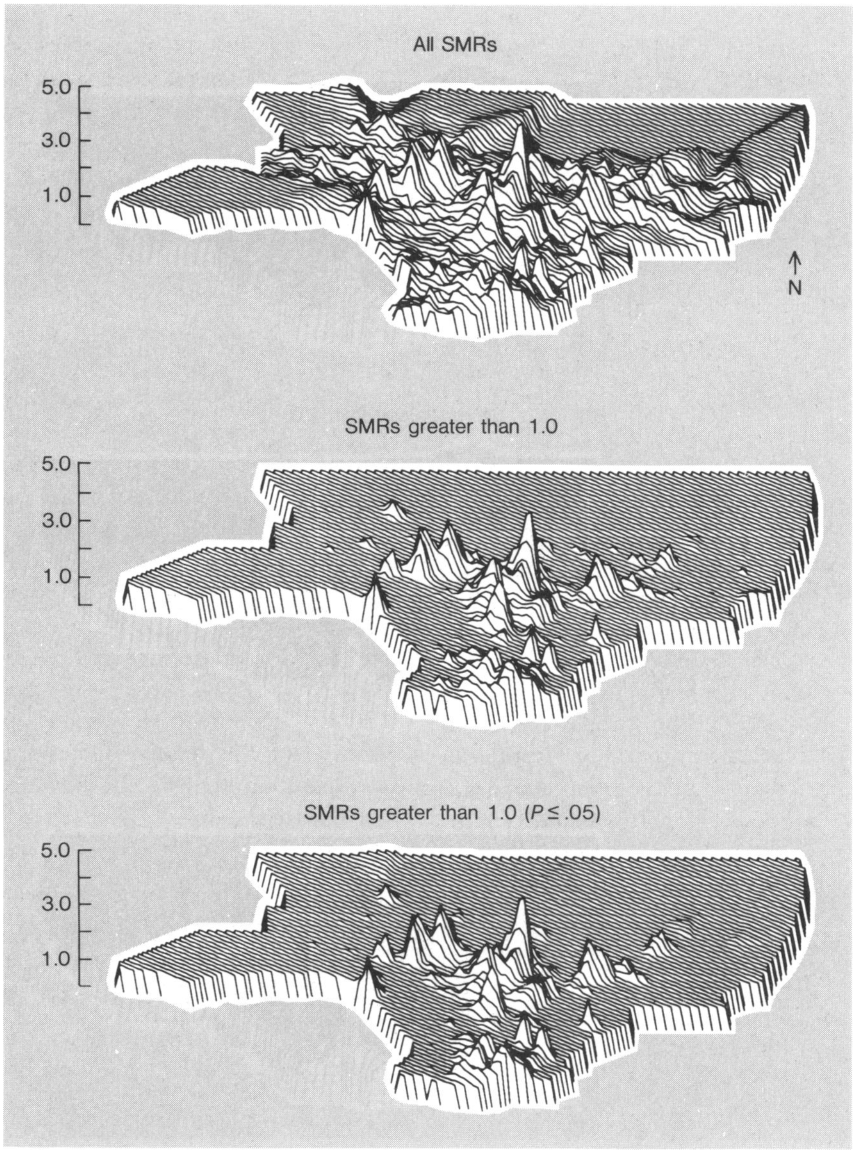
Since the correspondence tract was the geographic unit used in our analysis, all census tracts within a given correspondence tract were assigned the same SMR. For each set of data points, SYMAP automatically interpolated between centroids so that a continuous surface would be obtained. The program was then used to create a computer file containing a matrix of geographic coordinates along with the corresponding value for the "height" (interpolated standardized ratio) to be plotted at each location. This file served as the input file for the ASPEX program with which the three-dimensional maps were produced. The ASPEX program was run interactively on an IBM 3033 computer with a Tektronix 4081 cathode ray tube terminal.

For each disease condition of interest, a series of three maps was prepared. The top map in the series shows the SMRs for each correspondence tract in Los Angeles County. The peaks represent SMRs greater than 1.0; valleys, SMRs less than 1.0; and flat planes, SMRs of 1.0. The middle map shows those areas with SMRs greater than 1.0, a value indicating there were more deaths than expected. In this map and the bottom one, tracts with SMRs of less than 1.0 had their SMR set at 1.0 to provide a reference plane that would facilitate evaluation of the magnitude of the surrounding ratios. The bottom map shows only census tracts with SMRs greater than 1.0 at the 0.05 level of statistical significance.

Results

Reference maps of Los Angeles County (figs. 1 and 2) are presented to aid in the interpretation of the analysis. Figure 1, a two-dimensional map, illustrates the location of the centroids of the census tracts in the county, based on the 1970 U.S. Census. Approximately 5,000 people live in each tract; thus, a high density of census tract centroid markers in a given area implies a relatively large population. The sparse population of the western peninsula and along the northern border of LAC is reflected in the diminished density of census tracts in those areas. Figure 1 also suggests that census tracts in these areas tend to be geographically large. As a consequence, a three-dimensional plot of the SMR for each tract in these areas will appear as a broad plane or as a gradually sloping hill or valley, depending on the value of the SMR. The most densely populated area of LAC is the central region; somewhat smaller population centers are located northwest, east, and south of this central region. In the central region, each census tract covers a relatively small area; hence, its SMR will appear as a sharp peak or valley on the three-dimensional map. Figure 2 serves as a reference to the approximate location of some large cities in LAC as they would appear in the three-dimensional orientation used in our analyses.

Figure 4. Age- and sex-standardized morbidity ratios (SMRs) for potential waterborne diseases, Los Angeles County, 1969-71



Deaths from all causes. The geographic distribution of deaths from all causes is shown in figure 3. The top map indicates that there is considerable variation in this distribution throughout the county. Along the northern and western borders, the SMRs tend to be 1.0 or less; in the central portion, small positive and negative deviations from the expected number of deaths can be seen. However, large peaks in the downtown regions of Los Angeles and Long Beach indicate much higher zones of excess mortality as compared with the rest of the county. In these areas, the mortality experience is about three times greater than expected. Looking at the center map, we see that a zone of excess mortality from all causes of death extends from Los Angeles south to Long Beach. The bottom map in the figure shows that with statistical criteria as the basis, mortality is significantly greater than expected.

Potential waterborne diseases. All the potential waterborne diseases selected for this analysis were combined, and a single standardized ratio was calculated for each correspondence tract. The geographic pattern of morbidity in the county from these diseases is shown in figure 4. Considerable variation in morbidity is apparent within the densely populated areas, that is, in the central and southern parts of the county. There are zones of excess morbidity from the central area south to Long Beach as well as in communities in the eastern part of the county. The bottom map in the figure reveals that the statistically significant excess morbidity in some correspondence tracts is almost five times that expected based on the entire county as the standard.

Discussion

Three-dimensional maps (figs. 3 and 4) demonstrate the patterns of excess mortality from all causes of death and of morbidity attributable to potential waterborne diseases in Los Angeles County in the period 1969–71. Although mortality and morbidity varied considerably throughout the county, the maps demonstrate that several areas experienced significantly higher rates of mortality and morbidity than the county. The zone of excess mortality from all causes in the central corridor includes the skid row areas of Los Angeles and Long Beach as well as the socioeconomically depressed area of Watts. Socioeconomic conditions may explain why the mortality ratios for these areas were relatively high. Three-dimensional maps can be useful in revealing not only the health status of a community but also can serve as a means of generating hypotheses regarding which population characteristics or environmental conditions may be responsible for the geographic distribution of disease in a community.

The method of indirect adjustment of mortality employed in this investigation has been used previously in studies of spatial patterns of mortality. Howe published atlases of the United Kingdom in 1963 (*14*) and 1970 (*15*) that displayed standardized mortality ratios as patterns on two-dimensional maps. Chapman and Coulson (*6*) produced several maps of Los Angeles showing SMRs in 1966 by community; they used multiple colors (printed as varied shades in black and white reproductions) on each two-dimensional map to indicate the size of the SMR. Calle (*16*) mapped cancer mortality in Columbus, Ohio, by census tract, indicating by patterns on a two-dimensional map those SMRs that were significantly higher or lower than 1.0.

Computer mapping is especially appropriate for analysis of disease patterns in small areas. Interpretation of a two-dimensional map with multiple patterns, however, requires the reader to remember a color or pattern key. To summarize the entire map or to compare areas on it is often difficult. These problems of interpretation are eliminated when a three-dimensional mapping routine is used to represent the data on disease frequency. Since three-dimensional maps show the data in a continuous fashion, these data need not be categorized. The value at a given location on the map is represented as a peak whose height is proportional to the magnitude of the ratio. Thus, three-dimensional maps can summarize large amounts of data, as well as provide an immediate visual impression of the relative magnitude and distribution of disease throughout an area.

Although computer mapping routines have not been used extensively either in the health field in general or in epidemiologic research in particular, the advances in the last two decades in computer graphics have provided systems that can be easily used by nontechnical personnel. Computer mapping provides a relatively inexpensive and simple method of displaying spatial data, which is consistently accurate when used with multiple data sets. An added advantage is that reproduction-ready copies can be made in a matter of minutes.

References

1. Logan, W. P. D.: Mortality in the London fog incident. Lancet 1: 336–338 (1953).
2. Ministry of Health: Mortality and morbidity during the London fog of December 1952. Report on public health and medical subjects No. 95. London, 1954.
3. Snow, J.: Snow on cholera. The Commonwealth Fund, New York, 1936.
4. Patno, M. E.: Geographic study of cancer prevalence within an urban population. Public Health Rep 69: 705–715 (1954).
5. Greenburg, L., Field, F., Reed, J. I., and Glasser, M.: Air pollution and cancer mortality. Arch Environ Health 15: 356–361 (1967).
6. Chapman, J. M., and Coulson, A.: Community diagnosis: an analysis of indicators of health and disease in a metropolitan area. Int J Epidemiol 1: 75–81 (1972).
7. Mason, T. J., et al.: Atlas of cancer mortality for U.S. counties: 1950–1969. DHEW Publication No. (NIH) 75–780. U.S. Government Printing Office, Washington, D.C., 1975.
8. Mason, T. J.: Geographic patterns of cancer risk: a means for identifying possible occupational factors. Ann NY Acad Sci 271: 370–376 (1976).
9. Hoover, R., and Fraumeni, J. F.: Cancer mortality in U.S. counties with chemical industries. Environ Res 9: 196–207 (1975).
10. Blot, W. J., and Fraumeni, J. F.: Geographic patterns of lung cancer: industrial correlations. Am J Epidemiol 103: 539–550 (1976).
11. Frerichs, R. R., Sloss, E. M., and Satin, K. P.: Epidemiologic impact of water reuse in Los Angeles County. Environ Res. In press 1982.
12. Armitage, P.: Statistical methods in medical research. Blackwell Scientific Publishers, Oxford, 1971, pp. 388–389.
13. Haenszel, W., Loveland, D. B., and Sirken, M. G.: Lung-cancer mortality as related to residence and smoking histories. I. White males. J Natl Cancer Inst 28: 947–1001 (1962).
14. Howe, G. M.: National atlas of disease mortality in the United Kingdom. Thomas Nelson and Sons, Ltd., London, 1963.
15. Howe, G. M.: National atlas of disease mortality in the United Kingdom. Thomas Nelson and Sons, Ltd., London, 1970.
16. Calle, E. E.: Computer mapping of cancer mortality by census tract: Columbus, Ohio, 1956–1974. Master of science thesis, Ohio State University, 1978.