Are deaths from liver cancer, kidney cancer, and leukemia clustered in San Antonio?

F. BENJAMIN ZHAN, PHD

High mortality rates per square mile associated with liver cancer, lung cancer, kidney cancer, and leukemia have been reported in San Antonio. Without taking into account the underlying at-risk population, cancer mortality rates per square mile alone may give the public a misleading picture. This study conducts a geographic cluster analysis of mortality of the four types of cancer mentioned above, considering both mortality cases and the at-risk population. The analysis uses statewide cancer mortality data over an 8-year period from 1990 through 1997. Results from the study indicate that only one statistically significant cluster of liver cancer mortality cases exists in Bexar County and its adjacent counties compared with the rest of Texas. No clusters of lung cancer, kidney cancer, or leukemia exist in San Antonio.

Send reprint requests to F. Benjamin Zhan, PhD, Associate Professor, Department of Geography, Southwest Texas State University, San Marcos, TX 78666; e-mail: fbzhan@swt.edu.

INTRODUCTION

Concerns have arisen that cancer mortality rates in San Antonio may be elevated (1,2). For example, Bradshaw (2) reported high mortality rates of liver cancer, lung cancer, kidney cancer, and leukemia per square mile in some areas of San Antonio. However, without consideration of the underlying at-risk population and its associated demographic characteristics, mortality rates per square mile alone may give a misleading picture of cancer mortality rates in San Antonio. In addition, to avoid preselection bias, cancer mortality in the entire state of Texas should be analyzed first to determine whether cancer mortality rates are indeed excessively higher in San Antonio than in the rest of the state. Although considerable research has been reported about the characteristics of cancer mortality and incidences in Texas (3–7), the current literature does not provide a clear picture of the geographic concentration of cancer mortality in Texas; hence, findings in these reported studies cannot be used to determine whether or not mortality rates of these four types of cancer are in fact excessive in San Antonio.

On the basis of these observations, I conducted cluster analyses of cancer mortality cases for the entire state of Texas using cancer mortality data over the 8-year period from 1990 through 1997. The analyses aim to determine whether statistically significant mortality clusters of liver cancer, lung cancer, kidney cancer, or leukemia exist in Bexar County (where San Antonio is located). A cluster in this context is defined as a collection of adjacent area units whereby cancer mortality rates are excessive compared with the rest of the study area, and this excessiveness is statistically significant. The goal of the study was achieved through a geographical cluster analysis of cancer mortality cases that used the spatial scan statistic available from the National Cancer Institute. This method has been used in geographic cluster analysis of cancer mortality and diseases in a variety of situations (8,9).

Four separate cluster analyses, one for each of the four types of cancer, were conducted. County was the area unit used for the analyses. Because San Antonio accounts for most of the population in Bexar County, use of Bexar County to represent San Antonio in the analyses seems appropriate. Another important reason to use county as the area unit for the analyses is that reliable yearly estimations of the *at-risk population* at the county level were readily available from the US Census Bureau. Although conducting the analysis at an area unit smaller than a county is desirable, no accurate estimation of annual population data of Texas cross-tabulated by race, age group, and sex for smaller area units was available as of this writing. Needless to say, accurate at-risk population data relate directly to the reliability of the clusters detected.

DATA AND METHODS

Three types of data were needed for the analyses: geographic location, at-risk population, and cancer mortality. Geographic coordinates of county polygon centroids were obtained within a Geographic Information System and were used to represent the locations of the 254 counties in Texas. The at-risk population data at the county level from 1990 through 1997 were obtained from the US Census Bureau. These data were reformatted so that the at-risk population data were cross-tabulated by 4 races (white, African American, Hispanic, and other), 18 age groups (5-year interval), and 2 sexes. This resulted in a total of 144 combinations of race, age group, and sex, and each combination had a population count associated with it for each year.

Statewide cancer mortality data from 1990 through 1997 were obtained from the Bureau of Vital Statistics of the Texas Department of Health. Four cancer mortality data sets were used in the analyses. The first data set contains 6607 deaths from liver cancer. The second set is composed of 71,678 deaths resulting from lung cancer. The third data set includes 5554 deaths related to kidney cancer. The final data set consists of 9639 deaths caused by leukemia. All cases in the 4 data sets were cross-tabulated by the 4 races, 18 age groups, and 2 sexes.

The mortality data were analyzed by using the SaTScan (Version 2.1.3) software package. This software is based on the spatial scan statistic (10) and has been developed and distributed by the US National Cancer Institute. An article by Kulldorff (10) provides a detailed description of the spatial scan statistic, which is discussed only briefly here. In the first step in executing the software, circles are drawn, centered at specified locations in the study area. At each location, the sizes of the circles vary continuously on the basis of a prespecified maximum circle size. Second,

Table 1. Cluster analysis results of liver and lunch cancer mortality in Texas, 1990–1997. (Note: Mortality rates were adusted for race, age group, and sex.)

Cluster	County Within a Cluster	No. of Incidences	Expected No. of Incidences	Relative Risk	Log Likelihood Ratio	*P* Value
		Liver Cancer				
Most likely cluster	Comal, Blanco, Hays, Kendall, Caldwell, Guadalupe, Bexar	890	646.37	1.377	46.09	.0001
Secondary cluster 1	Tyler, Nacogdoches, San Augustine, Sabine, Houston, Angelina, Trinity, Newton, Jasper, Polk, San Jacinto, Hardin, Liberty, Orange, Jefferson	415	371.41	1.307	14.43	.0001
		Lung Cancer				
Most likely cluster	San Patricio, Victoria, Goliad, Live Oak, Bee, Refugio, Calhoun, Aransas, Duval, Jim Wells, Nueces, Kleberg	2,831	2,290.03	1.236	61.50	.0001
Secondary cluster 1	Sabine, Bowie, Titus, Franklin, Hopkins, Morris, Cass, Camp, Wood, Rains, Upshur, Marion, Kaufman, Van Zandt, Harrison, Smith, Gregg, Rusk, Panola, Henderson, Navarro, Cherokee, Anderson, Freestone, Shelby, Nacogdoches, Limestone, San Augustine, Leon, Houston, Angelina, Trinity, Robertson, Newton, Jasper, Polk, Madison, Tyler, Walker, Brazos, San Jacinto, Grimes, Montgomery, Hardin, Liberty, Waller, Orange, Jefferson, Harris, Chambers, Galveston	24,498	23,242.66	1.054	49.72	.0001
Secondary cluster 2	Wichita, Collingsworth, Childress, Hall, Hardeman, Wilbarger, Motley, Cottle, Foard, Clay, Montague, Grayson, Cooke, King, Dickens, Archer, Knox, Baylor, Jack, Wise, Denton, Haskell, Throckmorton, Young, Stonewall, Palo Pinto, Parker, Tarrant, Shackelford, Jones, Stephens, Hood, Eastland, Callahan	9,378	9,001.01	1.042	8.92	.0222

the number of cases inside and outside each circle is determined. Third, the number of cases expected inside the circle is calculated on the basis of the at-risk population in the area covered by the circle and the covariates used in the analysis.

In the fourth step of the analysis, the most likely cluster and secondary clusters are determined by computing maximum likelihood ratios (10). If a circle has the largest maximum likelihood ratio and the number of observed cases is more than expected, then the area covered by this circle is considered to be the most likely cluster. Other clusters containing more cases than expected are considered secondary clusters. These secondary clusters are ranked on the basis of their associated maximum likelihood ratios. Clearly, secondary clusters are less important than the most likely cluster. Lastly, the spatial scan statistic evaluates the statistical significance of the most likely cluster and the secondary clusters by using Monte Carlo simulations. When cases are assumed to follow the Poisson distribution in space, the null hypothesis is that no statistically significant spatial cluster exists. The simulated *P* value associated with a cluster should then be greater than a given level of significance, and the likelihood ratio should be smaller than a value determined by the software for a given significance level. Otherwise, the null hypothesis of no spatial cluster is rejected and the cluster is considered as a statistically significant cluster.

The spatial scan statistic was chosen because it does not present the problem of multiple testing found in some exploratory analysis methods (11–13). These exploratory analysis methods draw circles centered at certain locations in the study area and evaluate the significance of the cluster covered by *each* circle, one by one. This procedure introduces the problem of multiple testing, thus making these methods unsuitable for hypothesis testing. In contrast, the spatial scan statistic evaluates all circles in the study area at once and uses the value of the maximum likelihood ratio to select the most likely cluster and secondary clusters. The spatial scan statistic method evaluates the statistical significance of only the most likely cluster and secondary clusters, not every cluster, thus avoiding the problem of multiple testing (14). In addition, the spatial scan statistic does not require the user to specify the cluster size before the clustering takes place, a desirable feature needed for the analyses conducted in this study (14,15).

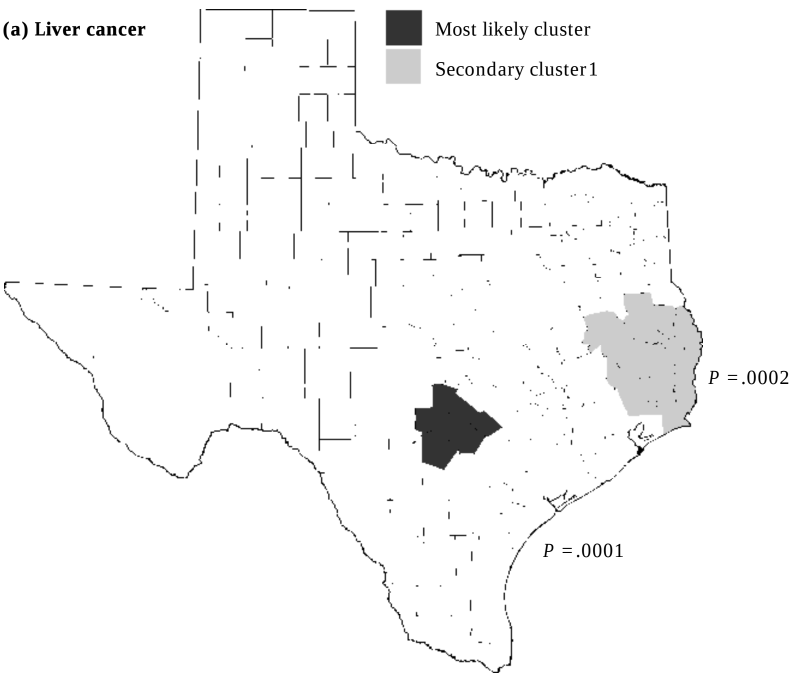
The analyses of the 4 data sets were performed on a Dell Inspiron 7500 notebook computer. For each of the 4 cancer mortality data sets, exactly 9999 Monte Carlo simulations were conducted to test the significance of the clusters detected.

Table 2. Cluster analysis results of kidney cancer and leukemia mortality in Texas, 1990–1997. (Note: Mortality rates were adjusted for race, age group, and sex.)

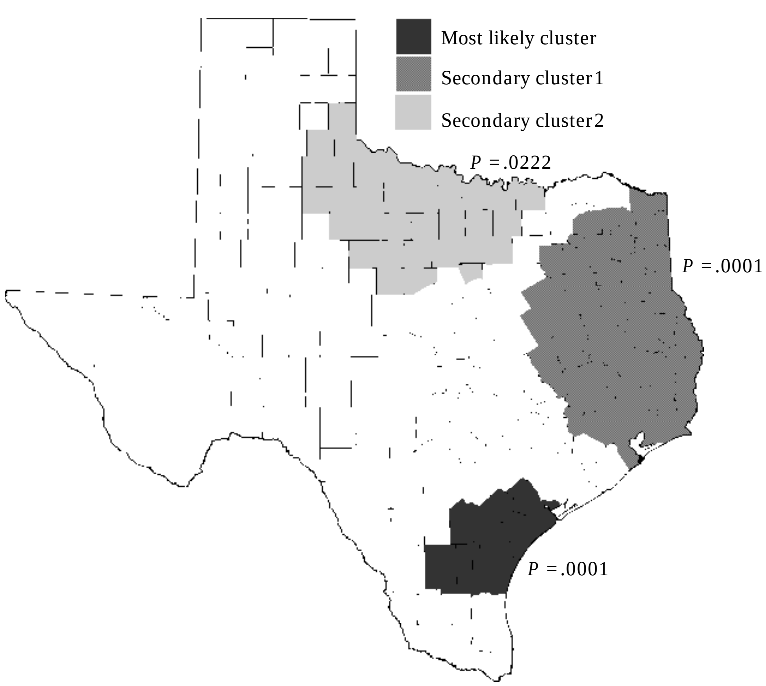
Cluster	County Within a Cluster	No. of Incidences	Expected No. of Incidences	Relative Risk	Log Likelihood Ratio	*P* Value
		Kidney Cancer				
Most likely cluster	Karnes, Bastrop, Hays, Kendall, Cladwell, Comal, Guadalupe, Gonzales, Bexar, Medina, Lavaca, Wilson, De Witt, Jackson, Atascosa, Victoria, Frio, Goliad, Live Oak, Bee, La Salle, McMullen, Refugio, Calhoun, San Patricio, Aransas, Nueces	831	751	1.107	4.79	.5740
Secondary cluster 1	Orange, Hardin, Jefferson	175	141	1.241	3.91	.8494
		Leukemia				
Most likely cluster	Chambers, Hardin, Liberty, Jefferson, Galveston	442	371.54	1.19	6.56	.1689
Secondary cluster 1	Robertson, Freestone, McLennan, Limestone, Leon, Falls, Milam, Madison, Brazos, Grimes, Burleson	369	310.58	1.188	5.36	.4093

Fig 1. Statistically significant clusters of cancer mortality in Texas, 1990–1997: (a) liver cancer, (b) lung cancer. (Note: Mortality rates were adjusted for race, age group, and sex.)

(a) Liver cancer



(b) Lung cancer



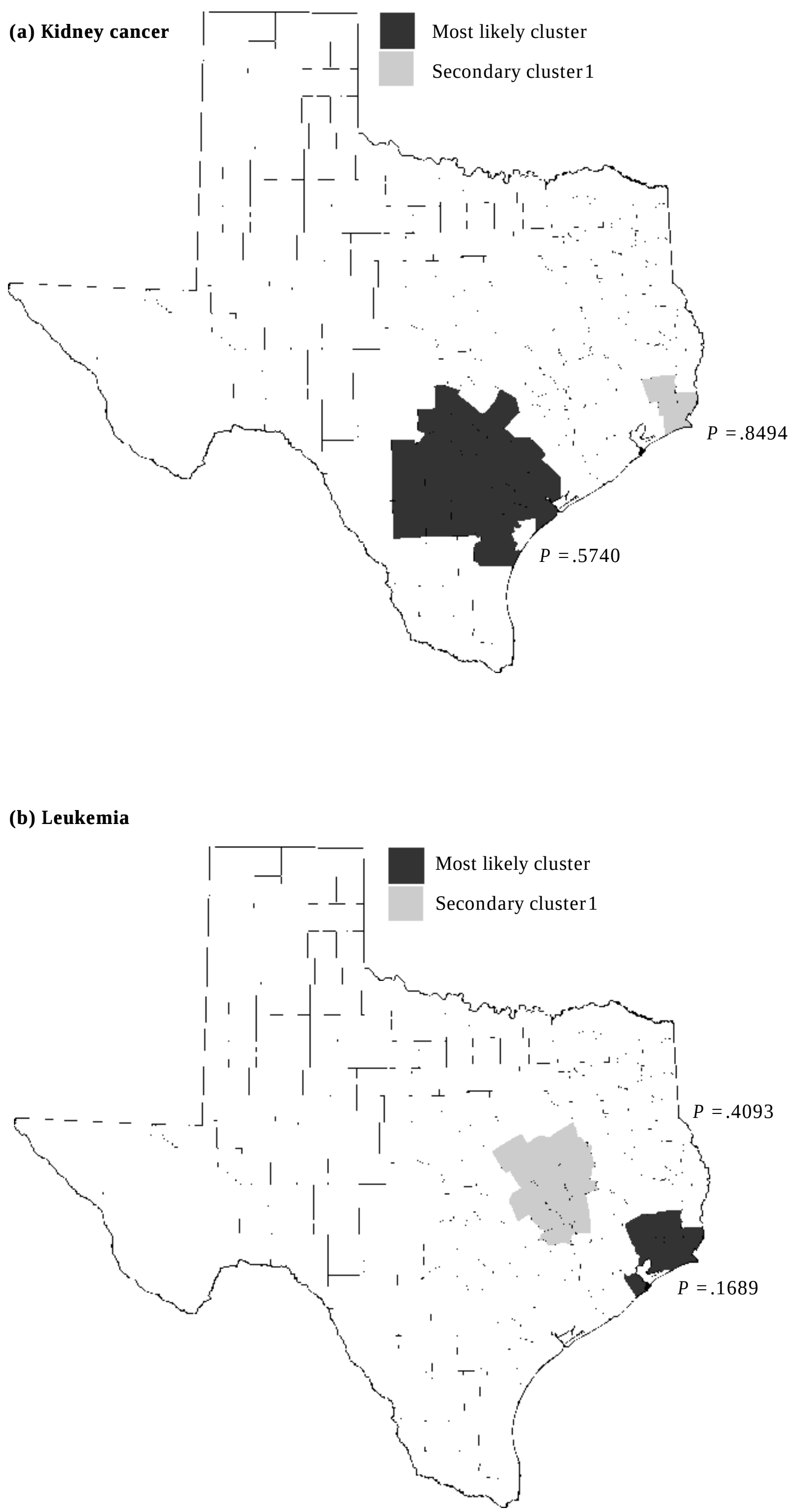
RESULTS

Analysis results related to the first and second data sets are reported in Table 1 and Fig 1. Statistically significant clusters are found in deaths resulting from liver and lung cancer (Table 1 and Fig 1). For deaths caused by liver cancer, the most likely cluster ($P = .0001$) includes the 7 San Antonio area counties of Comal, Bexar, Blanco, Hays, Kendall, Guadalupe, and Caldwell (Fig 1a). The other statistically significant cluster ($P = .0002$) of liver cancer mortality is located in an area northeast of Houston. This cluster embraces 15 counties, including Tyler, Nacogdoches, San Augustine, Sabine, Houston, Angelina, Trinity, Newton, Jasper, Polk, San Jacinto, Hardin, Liberty, Orange, and Jefferson (Fig 1a).

Three statistically significant clusters of lung cancer mortality emerged (Table 1 and Fig 1b). The most likely cluster ($P = .0001$) is located in South Texas, in the Victoria and Corpus Christi area, including 12 counties: San Patricio, Victoria, Goliad, Live Oak, Bee, Refugio, Calhoun, Aransas, Duval, Jim Wells, Nueces, and Kleberg. The first secondary cluster ($P = .0001$) is a large cluster containing 51 counties in East Texas. These 51 counties include Sabine, Bowie, Titus, Franklin, Hopkins, Morris, Cass, Camp, Wood, Rains, Upshur, Marion, Kaufman, Van Zandt, Harrison, Smith, Gregg, Rusk, Panola, Henderson, Navarro, Cherokee, Anderson, Freestone, Shelby, Nacogdoches, Limestone, San Augustine, Leon, Houston, Angelina, Trinity, Robertson, Newton, Jasper, Polk, Madison, Tyler, Walker, Brazos, San Jacinto, Grimes, Montgomery, Hardin, Liberty, Waller, Orange, Jefferson, Harris, Chambers, and Galveston. The other secondary cluster of lung cancer mortality ($P = .0222$) is also large and contains 34 counties: Wichita, Collingsworth, Childress, Hall, Hardeman, Wilbarger, Motley, Cottle, Foard, Clay, Montague, Grayson, Cooke, King, Dickens, Archer, Knox, Baylor, Jack, Wise, Denton, Haskell, Throckmorton, Young, Stonewall, Palo Pinto, Parker, Tarrant, Shackelford, Jones, Stephens, Hood, Eastland, and Callahan (Fig 1b).

For mortality cases resulting from kidney cancer and leukemia, no statistically significant clusters were detected in the entire state. Parameter values associated with the most likely cluster and the first secondary cluster of mortality related to both kidney cancer and leukemia are reported in Table 2. The locations of the clusters are given in Fig 2. For example, the most likely cluster of deaths related to kidney cancer contains 27 counties. But this cluster has a *P* value of .5740 and, thus, is not statistically significant (Fig 2a). The first secondary cluster of deaths related to kidney cancer contains only 3 counties and is not statistically significant ($P = .8494$). The most likely cluster of deaths due to leukemia is in the 4 counties northeast of Houston (Chambers, Hardin, Liberty, and Jefferson) and in Galveston County, south of Houston (Fig 2b). This cluster is also not statistically significant ($P = .1689$). The first secondary cluster of leukemia-related mortality contains 11 counties in an area containing cities like Waco, Bryan, and College Station, and is not statistically significant ($P = 0.4093$).

Fig 2. Statistically *not* significant clusters of cancer mortality in Texas, 1990–1997: (a) kidney cancer, (b) leukemia. (Note: Mortality rates were adjusted for race, age group, and sex.)



CONCLUSIONS AND DISCUSSION

The analyses show clearly that only deaths from liver cancer form the most likely and statistically significant cluster in Bexar County (in which San Antonio is located) and in its 6 surrounding counties (Comal, Blanco, Hays, Kendall, Caldwell, and Guadalupe). These data are presented in Fig 1. No statistically significant cluster exists in Bexar County or its adjacent counties for mortality related to lung cancer, kidney cancer, or leukemia (Figs 1 and 2). Therefore, we can conclude that only liver cancer mortality is an area of critical concern for Bexar County and, hence, in San Antonio. In addition, counties in the Houston area, particularly counties northeast of Houston, and some counties in south Texas in the Victoria and Corpus Christi areas, warrant attention (Fig 1).

Notice that the counties in the clusters mentioned above are in urbanized areas, which suggests that higher cancer mortality rates in these areas may be due largely to urbanization. While this observation may be partly true, the detected clusters cannot be attributed completely to urbanization because no cluster was detected in other urban areas such as the Dallas Fort-Worth area, the Austin area, or the El Paso area. Therefore, we can conclude that factors other than urbanization also played a role in the formation of the clusters of liver and lung cancer mortality in Texas.

Further studies are needed to determine the environmental, socioeconomic, behavioral, and genetic factors associated with the cluster of liver cancer mortality in San Antonio and surrounding counties. This suggestion applies also to the clusters of deaths related to liver cancer and lung cancer in the greater Houston area and in counties in the Victoria and Corpus Christi areas.

ACKNOWLEDGMENT

The author thanks the Texas Department of Health, Bureau of Vital Statistics, for providing the Texas cancer mortality data, and Forrest Wilkerson for bringing the author's attention to an article published in the *San Antonio Express-News* (October 20, 2000, p 15A). Ellen Lewis' assistance with the preparation of this article is also greatly appreciated.

REFERENCES

1. Bradshaw B. Cancer deaths in San Antonio. *San Antonio Express-News.* October 20, 2000:15A.
2. Tumiel C. Health issues unclear: studies inconclusive over the reasons for cancer cases among residents (in San Antonio). *San Antonio Express-News.* October 20, 2000:15A.
3. Martin J, Suarez L. Cancer mortality among Mexican Americans and other whites in Texas, 1969–80. *Am J Public Health.* 1987;77: 851–853.
4. Suarez L, Martin J. Primary liver cancer mortality and incidence in Texas Mexican Americans, 1969–80. *Am J Public Health.* 1987;77: 631–633.
5. Risser DR. Cancer incidence and mortality in urban versus rural areas of Texas, 1980 through 1985. *Tex Med.* 1996;92(1):58–61.
6. Weiss NS, Katz JA, Frankel LS, et al. Incidence of childhood and adolescent cancer in Texas. *Tex Med.* 1996;92(7):54–60.
7. Cooper SP, Sigurdson A, Labarthe D, et al. Assessing the burden of cancer in Texas using vital statistics data. *South Med J.* 1998;91(2): 173–181.
8. Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Stat Med.* 1996;15(7–9): 707–715.
9. Kulldorff M, Athas WF, Feuer EJ, Miller BA, Key CR. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am J Public Health.* 1998;88:1377–1380.
10. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods.* 1997;26:1481–1496.
11. Openshaw S, Charlton M, Wymer C, Craft AW A mark 1 geogrpahical analysis machine for the automated analysis of point data sets. *Int J Geogr Inf Syst.* 1987;1:335–358.
12. Fotheringham AS, Zhan F B. A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geogr Anal.* 1996;28(3):200–218.
13. Rushton G, Lolonis P. Exploratory spatial analysis of birth defect rates in an urban area. *Stat Med.* 1996;7:717–726.
14. Kulldorff M. Statistical methods for spatial epidemiology: tests for randomness. In: Gatrell A, Loytonen M, eds. *GIS and Health.* London, England: Taylor & Francis; 1998: 49–62.
15. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *Am J Epidemiol.* 1990;132(1 suppl):S136–S143.