Published in final edited form as:

Am J Prev Med. 2006 February ; 30(2 Suppl): S37–S49.

Cancer Map Patterns:

Are They Random or Not?

Martin Kulldorff, PhD, Changhong Song, MS, David Gregorio, PhD, Holly Samociuk, BS, and Laurie DeChello, MPH

From the Department of Ambulatory Care and Prevention (Kulldorff), Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts, Department of Statistics (Song), University of Connecticut, Storrs, Connecticut; and Department of Community Medicine and Health Care (Gregorio, Samociuk, DeChello), University of Connecticut, Farmington, Connecticut

Abstract

Background—Maps depicting the geographic variation in cancer incidence, mortality or treatment can be useful tools for developing cancer control and prevention programs, as well as for generating etiologic hypotheses. An important question with every cancer map is whether the geographic pattern seen is due to random fluctuations, as by pure chance there are always some areas with more cases than expected, or whether the map reflects true underlying geographic variation in screening, treatment practices, or etiologic risk factors.

Methods—Nine different tests for spatial randomness are evaluated in very practical settings by applying them to cancer maps for different types of data at different scales of spatial resolution: breast, prostate, and thyroid cancer incidence; breast cancer treatment and prostate cancer stage in Connecticut; and nasopharynx and prostate cancer mortality in the U.S.

Results—Tango's MEET, Oden's Ipop, and the spatial scan statistic performed well across all the data sets. Besag-Newell's R, Cuzick-Edwards k-NN, and Turnbull's CEPP often perform well, but the results are highly dependent on the parameter chosen. Moran's I performs poorly for most data sets, whereas Swartz Entropy Test and Whittemore's Test perform well for some data sets but not for other.

Conclusions—When publishing cancer maps we recommend evaluating the spatial patterns observed using Tango's MEET, a global clustering test, and the spatial scan statistic, a cluster detection test.

Introduction

Cancer maps can provide important clues concerning geographic variability in the incidence, mortality, tumor characteristics, treatment or survival, etiology, prevention, screening, or treatment of cancer. The cause of such variation is not restricted to etiologic risk factors, such as occupational exposures, dietary habits, or the natural environment. Equally important are various cancer control factors such as cancer screening, medical practice and access to health care, as well as genetic and socioeconomic variables. One example of the use of cancer maps is the 1975 finding of high oral cancer mortality rates in women in southeastern U.S.,¹ which lead to the identification of snuff dipping as a primary risk factor for oral cancer.² Another example is the finding of low in-situ breast cancer incidence rates in northeastern Connecticut,

Address correspondence and reprint requests to: Martin Kulldorff, PhD, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, 6th Floor, Boston MA 02215. E-mail: Martin_Kulldorff@hms.harvard.edu..

^{3,4} after which the Connecticut State Health Department intensified mammography screening efforts in that area.

Atlases of cancer incidence or mortality, or more general disease atlases that incorporate maps for several cancer sites, have been produced in many countries including Canada,⁵ Denmark,⁶ Estonia,⁷ Finland,⁸ Germany,^{9,10} Great Britain,^{11,12} Norway,¹³ Poland,¹⁴ Spain,¹⁵ Sweden,¹⁶ Switzerland,¹⁷ and the U.S.^{18–20} In the future, cancer atlases may also incorporate maps concerning the geographic variation in disease stage, histology, mode of detection, treatment, survival, and such studies are already being published in scientific journals.^{21–24}

Even if there are no true geographic differences in risk, there will always be some geographic patterns apparent to the naked eye. As in all medical research, it is important to evaluate whether observed patterns/results are likely to be due to chance or not. For geographic data, this is done using a test for spatial randomness, adjusting for the geographic distribution of the population at risk, as well as covariates such as age. Such tests are not a replacement for the cancer maps, but an important complement. If the null hypothesis of spatial randomness is rejected, it means that there are likely to be predictors that are geographically unevenly distributed, and the map pattern may then give clues as to what those predictors are. Some test statistics also provide information as to the nature of the nonrandom pattern indicating, for example, cluster locations or estimates of the geographic scale of clustering. If, on the other hand, the null hypothesis is not rejected, the geographic pattern observed is less likely to provide important information, and we should watch out that we do not spend a lot of time interpreting random spatial noise.

Most cancer atlases do not present tests for spatial randomness in conjunction with the presented maps, but there are some exceptions. Three examples are the New York State maps of cancer incidence (www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm); the Canadian cancer incidence atlas⁵; and the Spanish cancer mortality atlas.¹⁵ The former used a cluster detection test, the spatial scan statistic,²⁵ whereas the latter two used different global clustering tests, Moran's I ²⁶ and Smans' D ,²⁷ respectively. Tests for spatial randomness are more commonly used in articles published in scientific journals, where the focus is often on a single cancer site in a very specific geographic area.^{22–24,28–32}

Many different tests for spatial randomness have been proposed, but there is a lack of comparative evaluations using different types of actual disease data. In this study, nine different tests for spatial randomness are evaluated using nine different data sets, looking at the incidence, mortality, staging, and treatment of breast, prostate, thyroid, or nasopharynx cancer in either Connecticut or the U.S. The choice of the nine tests was based on a combination of their past use for cancer maps, publicly available software, personally available software, our general familiarity with the tests, and our own curiosity about their performance.

Formal statistical power evaluations have been carried out by comparing two, three, or more tests for spatial randomness.^{33–39} The advantage of this approach is that the alternative hypothesis of clustering is explicitly defined and known, so the performance of the tests can be compared in a very precise and explicitly defined manner. The disadvantage is that the alternative hypotheses chosen may or may not correspond to true and interesting geographic patterns of disease. When the methods are compared by applying them to real data, as in this paper, the disadvantage is that we do not know the true process generating the disease, so it is not known whether and what type of clustering may exist in the data. Then the result of one test statistic can only be compared to the results from other tests. To see what general patterns emerge, this must be done simultaneously for a large number of data sets and methods. The advantage with using real data for methods comparison is that it is known that the geographic data was generated from a real disease process. Hence, a study like this complements more

formal statistical power evaluations, and the results from both types of evaluations should be used in tandem when deciding what tests to use in practice.

Methods

Cancer Data Sets

Nine different geographic cancer data sets were used on which the different tests were applied for spatial randomness, summarized in Table 1.

Data on incidence, treatment, and staging of invasive breast, prostate, and thyroid cancer in Connecticut were obtained from the Connecticut Tumor Registry, a SEER-program site monitoring cancer events among a population of roughly 3.2 million residents. Address matching to street address, block group, or town was completed for 10,601 of 12,443 invasive breast cancer cases, 27,531 of 29,945 invasive prostate cancer cases and all 2805 thyroid cancer cases. For the breast cancer treatment analysis was based on 8795 women with early stage breast cancer, of which 5075 cases received partial mastectomy, whereas 3720 noncases received other forms of treatment. For the prostate cancer stage analysis, there were a total of 2055 men of which 209 had late stage disease. Some of these data sets have been described previously and analyzed using the spatial scan statistic only.^{4,40,41}

Data regarding prostate and nasopharyngeal cancer mortality across the U.S. were obtained from the National Center for Health Statistics' Multiple Cause of Death Public Use data files for years 1983–1988 that provided county of residence at time of death information on 222,377 and 4651 individuals, respectively.

All incidence and mortality rates and analyses were adjusted for age using the indirect age-adjustment method, based on population numbers obtained from the United States Census. Under the null hypothesis, each case is randomly and independently assigned to an area with probability proportional to the age-adjusted population at risk. For the treatment and late stage analyses, the null hypothesis is that each individual is equally likely to be a case. For all data sets, the null hypothesis is conditioned on the total number of cases observed so that the spatial distribution of the observed number of cases are being evaluated rather than the number of observed cases. All analyses were carried out using the spatial resolution listed in Table 1. All distances used are based on a straight line (“as the crow flies”) along the surface of the earth.

Notation

Denote c_i as the number of cases in location i , n_i as the population size of location i , C as the total number of cases in the data set, N as the total population, and d_{ij} as the distance between the centroids of locations i and j . Let $csum_{(i,j)}$ be the total number of cases in location i and its j closest neighbors, and let $nsum_{(i,j)}$ be the population size in location i and its j closest neighbors.

Global Clustering Tests

Tests for global clustering are used for investigating whether there is clustering throughout the study region, but when we are not interested in the specific location of clusters. Global clustering may occur through two different types of random processes.⁴² It could be a process where initial cases generate other cases with a comparatively higher probability among their closest neighbors, as when a disease is infectious. It could also be that there are a large number of health hazards scattered throughout the region, each creating an increased risk for the disease in a limited surrounding area. Examples of the latter could be gas stations or old apartment buildings with lead-based paint.

The performance of seven different global clustering tests were evaluated: Besag-Newell's R ,⁴³ Cuzick-Edwards k -NN,⁴⁴ Moran's I ,²⁶ Oden's I pop,³⁴ Swartz Entropy Test,⁴⁵ Tango's MEET,³⁷ and Whittemore's Test.⁴⁶ The first three require the specification of a parameter and were evaluated for multiple parameter values. To calculate the various test statistics, ClusterSeer (www.terraseer.com) was used for Oden's I pop, S-plus code provided by Dr. Toshiro Tango was used for Tango's MEET, and specially written C++ code was used for the remainder.

Besag-Newell's R

For each location i the Besag-Newell's method⁴³ creates a circle around it with a fixed number of cases m predetermined by the user. It then calculates the population in that circle, which is $nsum_{(i,f(i))}$, where $f(i)$ is chosen so that $csum_{(i,f(i))} - 1 < m \leq csum_{(i,f(i))}$. Besag-Newell's R ⁴³ is defined as

$$R = \sum_i c_i I \left(\sum_{s=0}^{m-1} \frac{e^{-\mu(i)} \mu(i)^s}{s!} > 0.95 \right)$$

where $\mu(i) = nsum_{(i,f(i))} C/N$, the expected number of cases in location i and its $f(i)$ nearest neighbors. $I()$ is the indicator function determining whether that particular circle has a statistically significant excess number of cases at the 0.05 level, based on the Poisson distribution. The method was evaluated for different values of $m = 5, 15, 50, 100, 500, 1500$. The null hypothesis is rejected when R is large.

Cuzick-Edwards' k -NN

Cuzick-Edwards' k -NN (k -Nearest Neighbors) test⁴⁴ was originally designed for point data. It is defined as

$$T_k = \sum_i c_i (csum_{(i,f(i))} - 1)$$

where $f(i)$ is chosen so that $nsum_{(i,f(i))} = k$ for a constant k chosen by the user. That is, for each case i the number of cases are counted that are among its k closest neighbors, and then they are added up. For aggregated data, the test is modified so that only a proportion of the cases in the last location is counted, equal to the proportion of the population from that location that is needed to reach k total neighbors. The method for different choices of k were evaluated, where k was set as a proportion of N , the total population at risk, so that $k/N = 0.0001, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2$, and 0.5 . The null hypothesis is rejected for large values of T_k .

Moran's I

Moran's I ²⁶ was originally proposed to analyze continuous attribute data such as weight or income. Subsequently, this statistic has often been used to analyze count data as well, such as cancer incidence in Canada.⁵ The test statistic is defined as:

$$I = \frac{\sum_i \sum_j w_{ij} \left(\frac{c_i}{n_j} - r \right) \left(\frac{c_j}{n_i} - r \right)}{\sum_i \left(\frac{c_i}{n_j} - r \right)^2}$$

where $r = \frac{1}{L} \sum_i \frac{L_i}{n_i}$, L is the total number of locations and

$$w_{ij} = \begin{cases} 1 & \text{if } j \text{ is a neighbor of } i \\ 0 & \text{otherwise} \end{cases}$$

The concept of neighbor can be defined in different ways. In this article, j is either defined as a neighbor of i if location j is one of location i 's u nearest neighbors in terms the distances d_{ij} between them, or, j is defined as a neighbor of i if $d_{ij} < D$ for some fixed distance D . The method was evaluated using different values for the parameters, with $u = 5, 10, 20, 40, 100, 150, 500$ and $D = 50, 100, 150$ miles. The null hypothesis is rejected when I is large.

Oden's Ipop

Oden's Ipop³⁴ is defined as

$$Ipop = \frac{C^2 \sum_i \sum_j w_{ij} \left(\frac{c_i}{C} - \frac{n_i}{N} \right) \left(\frac{c_j}{C} - \frac{n_j}{N} \right) - C \left(1 - 2 \frac{C}{N} \right) \sum_i w_{ii} \frac{c_i}{C} - C^2 N \sum_i w_{ii} \frac{n_i}{N}}{\left(N^2 \sum_i \sum_j w_{ij} \frac{n_i}{N} \frac{n_j}{N} - N \sum_i w_{ii} \frac{n_i}{N} \right) \frac{C}{N} \left(1 - \frac{C}{N} \right)}$$

The weights $w_{ij} = 1$ were used when i and j are adjacent neighbors (sharing at least one common point on their boundaries) and $w_{ij} = 0$ otherwise, as available in the Cluster-Seer software. Because there are no boundaries for residential point locations, this method could not be applied to the prostate cancer stage data. The null hypothesis is rejected for large values of $Ipop$.

Swartz' Entropy Test

Swartz⁴⁵ proposed a test for spatial randomness based on the concept of entropy. The test statistic can be written⁴⁷ as

$$SET = \ln(C!) + \ln((N - C)!) - \sum_i (\ln(c_i!) + \ln((n_i - c_i)!)).$$

The null hypothesis of no clustering is rejected when SET is small. The method can only be used for spatially aggregated data, so it was not possible to apply it for the late stage prostate cancer data.

Tango's Maximized Excess Events Test

For a given parameter λ , Tango's Excess Events Test³⁵ is defined as

$$EET(\lambda) = \sum_i \sum_j e^{-\lambda d_{ij}^2} \lambda^2 \left(c_i - n_i \frac{C}{N} \right) \left(c_j - n_j \frac{C}{N} \right).$$

This is simply a weighted sum of the excess number of events (observed minus expected) in location i times the excess number of events in location j , where the weight is higher when locations i and j are close. Choosing a large λ makes the test sensitive to geographically large clusters, whereas a small λ makes it more sensitive to small clusters.

To be able to detect clustering irrespectively of its geographical scale, Tango³⁷ proposed the Maximized Excess Events Test (MEET)

$$MEET = \min_{0 \leq \lambda \leq U} P(EET(\lambda) > eet(\lambda) \mid H_0, \lambda),$$

where $eet(\lambda)$ is the observed value of the Excess Events Test statistic conditioning on λ , and U is an upper limit on λ specified by the user. The null hypothesis is rejected when MEET is small.

Whittemore's Test

Whittemore et al.⁴⁶ proposed the test statistic

$$W = \frac{1}{2} \sum_i \sum_j d_{ij} c_i c_j$$

This is the sum of all the interpoint distances between each pair of cases. The null hypothesis is rejected when W is small.

Cluster Detection Tests

Cluster detection tests are concerned with local clusters, and are used when there is simultaneous interest in detecting their location and testing their statistical significance. Two such methods were evaluated, the circular spatial scan statistic²⁵ and the Cluster Evaluation Permutation Procedure (CEPP) of Turnbull et al.⁴⁸ The calculations were carried out using the SaTScan (www.satscan.org) and ClusterSeer (www.terraseer.com) software respectively.

Spatial Scan Statistic

As applied for this study, the spatial scan statistic²⁵ uses a very large number of overlapping circles of different size and in different locations, each with a different set of neighboring locations within it, and each a possible candidate for being a cluster. The likelihood is calculated for each circle, and the definition of the test statistic is the maximum likelihood over all circles evaluated.

In mathematical notation:

$$T = \min_{i, j \in J(i)} \frac{L(i, j)}{L_0} I\left(\frac{csum(i, j)}{nsum(i, j)} > \frac{C}{N}\right)$$

where $L(i, j)$ is the likelihood under the alternate hypothesis that there is a cluster in location i and its j closest neighbors, L_0 is the likelihood under the null hypothesis and $I()$ is the indicator function, ensuring that the cluster has a higher rate than expected by chance. $J(i)$ defines the upper limit on the circle size, and is usually defined so that $nsum(i, J(i)) \leq N/2 < nsum(i, J(i)+1)$, which means that the circles contain at most 50% of the total population. It can be shown that²⁵:

$$\frac{L(i, j)}{L_0} = \left(\frac{csum(i, j)}{nsum(i, j) \frac{C}{N}} \right)^{csum(i, j)} \left(\frac{C - csum(i, j)}{C - nsum(i, j) \frac{C}{N}} \right)^{C - csum(i, j)}$$

The null hypothesis is rejected when T is large. There are usually multiple clusters found, but secondary clusters are only reported if they are statistically significant and do not overlap with a reported stronger cluster.

Turnbull's Cluster Evaluation Permutation Procedure

Turnbull's Cluster Evaluation Permutation Procedure (CEPP)⁴⁸ uses a large number of overlapping circles. For each location i the method creates a circle around it with a fixed pre-determined population size k . It then calculates the number of observed cases within the circle, and the test statistic is the maximum number of cases over all circles. For point data, the mathematical definition is:

$$CEPP = \max_i csum(i, f(i))$$

where $f(i)$ is chosen so that $nsum_{(i,f(i))} = k$.

When the data is aggregated, it is not always possible to find a circle with the exact population required and the method then counts only a proportion of the cases in the last location corresponding to the proportion of the population needed to reach the required population size. The method for different values of k were evaluated. For the incidence and mortality data, it was set so that the expected counts in the circles were fixed at a selection of $kN/C = 1, 5, 10, 50, 100, 500, 1000, 5000$. For the treatment and stage data, it was set as $k = 10, 50, 100, 500$, and 1000. The null hypothesis is rejected for large values of CEPP. The method can detect multiple clusters, but because ClusterSeer only reports the top three and these are usually almost identical, only the top cluster is reported for each parameter value.

Results

The results for the global clustering tests are presented in Tables 2 and 3. The results for the cluster detection tests are shown in Tables 4, 5, 6, and 7 and in Figures 1 through 4. There is strong evidence of spatial clustering for all data sets except late stage prostate cancer in Connecticut. It should be noted though, that these results are partly due to the geographical nature of the disease and partly due to sample size.

For the eight data sets for which there is strong evidence of clustering, Tango's MEET, the spatial scan statistic, and Oden's Ipopt all presented p -values of < 0.05 .

For those tests requiring a parameter choice to define the scale of clustering, the results are dependent on the parameter chosen. This is still the case even if we ignore the parameter choices at the lower or upper end of the scale. For Besag-Newell's R and Cuzick-Edward's kNN there is no parameter value for which the results are consistent across the data sets in the sense that the p -value is always low when other tests show strong evidence for spatial clustering. Turnbull's CEPP tend to perform better for large parameter values, at least for the data sets considered in this paper. The most extreme parameter sensitivity is for prostate cancer incidence when using the distance based Moran's I, where $p = 0.0002$ for $D = 50$ miles, $p = 1.0$ for $D = 100$ miles and $p = 0.0001$ for $D = 150$ miles. When examining parameter values in between, it was found that there was a gradual but very steep increase in the p -values as a function of the distance D , with $p = 0.004$ for $D = 55$, $p = 0.49$ for $D = 60$, and $p = 1.0$ for $D = 80$, and an equally steep decrease, with $p = 0.015$ for $D = 110$ and $p = 0.0001$ for $D = 120$.

The "odd test" is Whittemore's method, which has a significant p -value for the prostate cancer stage data ($p = 0.02$), for which other tests show little evidence for clustering, while failing to reject the null hypothesis ($p = 0.49$) for prostate cancer incidence for which there is overwhelming evidence of spatial clustering from all the other methods.

The spatial scan statistic found clusters in areas where Turnbull's CEPP did not, but that is an artifact of the software used, related to the number of secondary clusters reported, rather than a difference in the methods themselves. By using different parameter settings for Turnbull's CEPP it is possible to pinpoint the part of a larger cluster with the highest rate. The same information can be obtained from the spatial scan statistic by limiting the size on reported clusters, although this was not done.

Discussion

Maps depicting the geographic variation of cancer can provide important clues for cancer control and prevention, as well as for cancer etiology. When producing such maps, whether individually or as part of a cancer atlas, it is important to carefully select appropriate mapping techniques.^{49–54} It is also important to complement the maps with a test for spatial

randomness to determine whether the geographical patterns seen are random or not. The purpose of this paper is to evaluate a large number of methods to help in the selection of what statistical test to use for that purpose.

When evaluating a particular test statistic, there are always some data sets for which the null hypothesis will be rejected and other data sets for which it will not. To get a good picture of test performance when using real data, it is important to evaluate the test statistics using many different data sets. For example, the fact that both cluster detection tests and five of seven global clustering tests reject the null for the Nasopharynx mortality data is a very strong indication that there truly is spatial clustering in that data, and the inability of Swartz Entropy Test and Moran's I to reject the null hypothesis is, hence, somewhat troubling. This is the only data set for which Swartz Entropy Test does not reject when most of the other do, so it could be a rare problem. The problems with Moran's I are much more consistent, and hence, it is suggested that Moran's I should not be used for cancer count data such as incidence or mortality, although it may be better for continuous attribute data for which it was originally proposed.

The p -values of Besag-Newell's R , Cuzick-Edwards k -NN, Moran's I , and Turnbull's CEPP depend a lot on the specific parameter chosen, and from a practical point of view, it is not always easy to select an appropriate parameter a priori. One option is to do the test for multiple parameter values, and although that gives potentially useful information about the scale of clustering, it also leads to multiple testing and if some p -values are significant whereas others are not, it is hard to draw any definite conclusions. To do a Bonferroni type adjustment is inefficient because tests with different but similar parameter values are highly correlated. Oden's I_{pop} also requires the choice of a spatial scale parameter, although only one option was used, provided to us by the Cluster-Seer software, so it is not known how sensitive this method is to the choice of parameter.

The remaining test statistics have the advantage of not requiring the specification of a particular parameter to define the scale of clustering, even though Tango's MEET and the spatial scan statistic both require an upper bound. The most consistent performer of the global clustering tests is Tango's MEET, in that it always rejects the null hypothesis for the eight data sets for which the majority of the test statistics reject it.

The nine test statistics were evaluated on data sets at different spatial resolution (aggregation levels), from exact residential coordinates to U.S. counties, and with different size and shape of the aggregation areas. It is expected that most tests will perform better at a finer spatial resolution because more information is then retained for the analysis, but that has only been thoroughly investigated for the spatial scan statistic.⁵⁵ Data is often only available at a certain aggregate level though, and it is then important to know that the methods perform well for aggregated data.

Some global clustering tests do not provide any clues to the true underlying clustering process when the null is rejected. Other tests, such as Cuzick-Edwards k -nearest neighbor test, Besag-Newell's R , Moran's I , and Tango's MEET do, in the sense that the parameter for which the p -value is smallest provides an estimate of the scale of clustering. That is an added advantage of these tests.

A different way to evaluate tests for spatial randomness is to estimate the statistical power for different known clustering models, by using simulated data sets. Song and Kulldorff³⁹ did such a study with many of the same test statistics, finding that overall, the spatial scan statistic did well for local clusters whereas Tango's MEET performed best for clustering spread throughout the map. Hence, the results for the simulated power study and this evaluation based on real data sets are fairly consistent with each other, which is reassuring.

Conclusion

When presenting cancer maps, one would normally not use a large number of tests for spatial randomness as was done in this evaluation study. Rather, we recommend using one global clustering test and one cluster detection test. Only a small selection of tests were evaluated in this paper on a few data sets and it is not known if these are the best among all methods available. Based on this study though and an earlier formal power evaluation,³⁹ it is safe to recommend the use of Tango's MEET and the spatial scan statistic for evaluating the random nature of cancer maps, although there are likely some data sets and clustering features for which other methods may be better.

Acknowledgements

This work was funded by the National Cancer Institute, grant number RO1-CA95979.

No financial conflict of interest was reported by the authors of this paper.

References

1. Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF. Atlas of cancer mortality for U.S. counties: 1950–1969. DHEW Publ. No. (NIH) 75-780. Washington, DC: U.S. Govt. Printing Office, 1975.
2. Winn DM, Blot WJ, Shy CM, Pickle LW, Toledo A, Fraumeni JF Jr. Snuff dipping and oral cancer among women in the southern United States. *N Engl J Med* 1981;304:745–9. [PubMed: 7193288]
3. Gregorio DI, Mahooti-Brooks N, Samociuk H. Breast cancer needs assessment for Connecticut. Farmington, CT: Department of Community Medicine and Health Care, University of Connecticut, 2001.
4. Gregorio DI, Kulldorff M, Barry L, Samociuk H. Geographic differences in invasive and in situ breast cancer incidence according to precise geographic coordinates, Connecticut, 1991–1995. *Int J Cancer* 2002;100:194–8. [PubMed: 12115569]
5. Le ND, Marrett LD, Robson DL, Semenciw RM, Turner D, Walter SD. Canadian cancer incidence atlas, Vol. 1. Ottawa, Canada: Canada Communication Group Publishing, 1995.
6. Carstensen B, Jensen OM. Atlas of cancer incidence in Denmark 1970–79. Copenhagen: Danish Cancer Society, Environmental Protection Agency, 1986.
7. Baburin A, Gornoi K, Leinsalu M, Rahu M. Eesti suremusatlas—Atlas of mortality in Estonia. Tallin, Estonia: Institute of Experimental and Clinical Medicine, 1997.
8. Pukkala E, Gustavsson N, Teppo L. Atlas of cancer incidence in Finland 1953–82. Publ. No 37. Helsinki: Cancer Society of Finland, 1987.
9. Menhart MW. Atlas Der Krebsinzidenz in Der Ehemaligen Deutschen Demokratischen Republik 1978–1982 Atlas of cancer incidence in the former German Democratic Republic 1978–1982. Oxford: Oxford University Press, 1993.
10. Becker N, Frentzel-Beyme R, Wagner G. Atlas of cancer mortality in the Federal Republic of Germany. Berlin: Springer Verlag, 1984.
11. Swerdlow A, Dos Santos Silva I. Atlas of cancer incidence in England and Wales 1968–85. Oxford: Oxford University Press, 1994.
12. Atlas of cancer in Scotland, 1975–1980: Incidence and epidemiological perspective. IARC Scientific Publ. No. 72. Lyon: IARC, 1985.
13. Glatte E, Finne TE, Olesen O, Langmark F. Atlas over kreft-insidens i Norge 1970–79. Oslo: Norwegian Cancer Society, 1985.
14. Zatoński W, Pukkala E, Didkowska J, Tyczynski J, Gustavsson N. Atlas umieralności na nowotwory złośliwe w polsce w latach 1986–1990. Atlas of cancer mortality in Poland 1986–1990. Warszawa: Instytut im. Marii Skłodowskiej-Curie, 1993.
15. López-Abente Ortega G, Pollán Santamaría M, Escolar Pujolar A, Errezola Saizar M, Abairra Santos V. Atlas de mortalidad por cancer y otras causas en España 1978–1992. Atlas of cancer mortality and other causes of death in Spain 1978–1992. Madrid: Fundación Científica de la Asociación Española Contra el Cáncer, 1996.

16. Dufinats M, Foldemo Ö, Damber L, Jonsson H, Anderson A, Möller T, et al. Atlas of cancer incidence in Sweden. Linköping: Swedish Oncological Centres, 1995.
17. Schüler G, Bopp M. Atlas der Krebsmortalität in der Schweiz, 1970–1990. Basel: Birkenhäuser.
18. Pickle LM, Mungiole M, Jones GK, White AA. Atlas of United States mortality. DHHS Publ. No. 97-1015. Hyattsville, MD: National Center for Health Statistics, 1996.
19. Devesa SG, Grauman DJ, Blot WJ, Pennello GA, Hoover RN, Fraumeni JF. Atlas of cancer mortality in the United States, 1950–94. NIH Publ. No. 99-4564. Bethesda, MD: National Cancer Institute, 1999.
20. Pickle LW, Feuer EJ, Edwards BK. U.S. predicted cancer incidence, 1999: complete maps by county and state from spatial projection models. NIH Publication No. 99-4564. Bethesda, MD: National Cancer Institute, 2003.
21. Rushton G, West M. Women with localized breast cancer selecting mastectomy treatment, Iowa, 1991–1996. *Public Health Rep* 1999;114:370–1. [PubMed: 10501140]
22. Sheehan TJ, Gershman ST, MacDougal L, Danley R, Mrosszczyk M, Sorensen AM, et al. Geographical surveillance of breast cancer screening by tracts, towns and ZIP Codes. *J Public Health Manag Pract* 2000;6:48–57. [PubMed: 18019960]
23. Roche LM, Skinner R, Weinstein RB. Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *J Public Health Manag Pract* 2002;8:26–32. [PubMed: 11889849]
24. Thomas AJ, Carlin BP. Late detection of breast and colorectal cancer in Minnesota counties: an application of spatial smoothing and clustering. *Stat Med* 2003;22:113–27. [PubMed: 12486754]
25. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997;26:1481–96.
26. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17–23. [PubMed: 15420245]
27. Smans M. Analysis of spatial aggregation. In: Boyle P, Muir CS, Grundman, eds. *Cancer mapping. recent results in cancer research*. Berlin: Springer-Verlag, 1989.
28. Jemal A, Kulldorff M, Devesa SS, Hayes RB, Fraumeni JF. A geographic analysis of prostate cancer mortality in the United States. *Int J Cancer* 2002;101:168–74. [PubMed: 12209994]
29. Jacquez GM, Greiling DA. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *Int J Health Geogr* 2003;2:3. [PubMed: 12633503]
30. Buntinx F, Geys H, Lousbergh D, Broeders G, Cloes E, Dhollander D, et al. Geographical differences in cancer incidence in the Belgian province of Limburg. *Eur J Cancer* 2003;39:2058–72. [PubMed: 12957461]
31. Sheehan TJ, DeChello LM, Kulldorff M, Gregorio DI, Gershman S, Mroszczyk M. The geographic distribution of breast cancer incidence in Massachusetts 1988–1997, adjusted for covariates. *Int J Health Geogr* 2004;3:17. [PubMed: 15291960]
32. Hsu CE, Jacobson HE, Soto Mas F. Evaluating the disparity of female breast cancer mortality among racial groups—a spatiotemporal analysis. *Int J Health Geogr* 2004;3:4. [PubMed: 14987336]
33. Vach W. Locally optimal tests on spatial clustering. In: Diday E, ed. *New approaches in classification and data analysis*. Berlin: Springer, 1994:161–8.
34. Oden N. Adjusting Moran's I for population density. *Stat Med* 1995;14:17–26. [PubMed: 7701154]
35. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med* 1995;14:2323–34. [PubMed: 8711272]
36. Rogerson P. The detection of clusters using a spatial version of the chi-square goodness-of-fit statistics. *Geogr Anal* 1999;31:130–47.
37. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Stat Med* 2000;19:191–204. [PubMed: 10641024]
38. Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Comput Stat Data Anal* 2003;42:665–84.
39. Song C, Kulldorff M. Power evaluation of disease clustering tests. *Int J Health Geogr* 2003;2:9. [PubMed: 14687424]
40. Gregorio DI, Kulldorff M, Barry L, Samociuk H, Zarfos K. Geographic differences in primary therapy for early stage breast cancer. *Ann Surg Oncol* 2001;8:844–9. [PubMed: 11776501]

41. Gregorio DI, Kulldorff M, Sheehan TJ, Samociuk H. Geographic distribution of prostate cancer incidence in the era of PSA testing, Connecticut 1984–98. *Urology* 2004;63:78–82. [PubMed: 14751353]
42. Haining R. Spatial statistics and the analysis of health data. In: Gatrell AG, Löytönen M, eds. *GIS and health*. London: Taylor and Francis, 1998.
43. Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc* 1991;A154:143–55.
44. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *J R Stat Soc* 1990;B52:73–104.
45. Swartz JB. An entropy-based algorithm for detecting clusters of cases and controls and its comparison with a method using nearest neighbours. *Health Place* 1998;4:67–77. [PubMed: 10671012]
46. Whittemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987;74:631–5.
47. Kulldorff M. Re: An entropy-based algorithm for detecting clusters of cases and controls and its comparison with a method using nearest neighbors letter. *Health Place* 1999;5:313. [PubMed: 11066903]
48. 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:S136–43. [PubMed: 2356825]
49. Brewer CA, MacEachren AM, Pickle LW, Herrmann D. Mapping mortality: evaluating colour schemes for choropleth maps. *Ann Assoc Am Geogr* 1997;87:411–38.
50. MacEachren AM, Brewer CA, Pickle LW. Visualizing georeferenced data: representing reliability of health statistics. *Environ Plan A* 1998;30:1547–61.
51. Pickle LW, Mungiole M, Jones GK, White AA. Exploring spatial patterns of mortality: the new atlas of United States mortality. *Stat Med* 1999;18:3211–20. [PubMed: 10602146]
52. Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. *Health Place* 2003;9:273–7. [PubMed: 12810334]
53. James WL, Cossman RE, Cossman JS, Campbell C, Blanchard T. A brief visual primer for the mapping of mortality trend data. *Int J Health Geogr* 2004;3:7. [PubMed: 15072581]
54. Berke O. Exploratory disease mapping: kriging the spatial risk function from regional count data. *Int J Health Geogr* 2004;3:18. [PubMed: 15333131]
55. Gregorio DI, Dechello LM, Samociuk H, Kulldorff M. Lumping or splitting: seeking the preferred areal unit for health geography studies. *Int J Health Geogr* 2005;4:6. [PubMed: 15788100]

Figure 2. Maps of thyroid cancer standardized incidence rates (SIRs) in Connecticut, 1976–1995, by town, for those under 25 (**top**), 25 and above (**bottom**) and all ages combined (**next page**), together with the clusters detected by the spatial scan statistic and Turnbull's CEPP. The numbers refer to the clusters described in Table 5 (continued on next page).

Figure 3.

Maps of the relative risk (RR) of primary breast cancer treatment with partial mastectomy 1991–1995 (**top**), and the relative risk of late stage prostate cancer in 1998 (**bottom**), for the state of Connecticut, by census block-group, together with the clusters detected by the spatial scan statistic and Turnbull's CEPP. The numbers refer to the clusters described in Table 6.

Figure 4.

Maps of county based nasopharynx cancer standardized mortality rates (SMRs, **top**) and prostate cancer SMRs (**bottom**) in the United States by county, 1983–1988, together with the clusters detected by the spatial scan statistic and Turnbull's CEPP. The numbers refer to the clusters described in Table 7. The large arcs in the Pacific are clusters centered on Hawaii but also including small parts of Alaska or California.

Table 1
Summary of the data sets used to evaluate the various tests for spatial randomness

Cancer site	Data type	Ages	Geographic area	Spatial resolution	Locations (<i>n</i>)	Time period	Cases (<i>n</i>)
Breast	Incidence	All	Connecticut	Block groups	2905	1991–1995	10,601
Breast	Treatment	All	Connecticut	Block groups	2905	1991–1995	4,950
Thyroid	Incidence	0–24	Connecticut	Towns	169	1976–1995	209
Thyroid	Incidence	25+	Connecticut	Towns	169	1976–1995	2,596
Thyroid	Incidence	All	Connecticut	Towns	169	1976–1995	2,805
Prostate	Incidence	All	Connecticut	Block groups	2905	1984–1998	27,531
Prostate	Late stage	All	Connecticut	Residence	2038	1998	272
Nasopharynx	Mortality	All	USA	County	3136	1983–1988	4,651
Prostate	Mortality	All	USA	County	3136	1983–1988	222,377

Table 2
Results for breast cancer incidence and treatment and thyroid cancer incidence in Connecticut when using a global clustering test

Test statistic	Parameter	Breast cancer		Thyroid cancer incidence		
		Incidence	Treatment	0-24	25+	All ages
Besag-Newell's <i>R</i>	m=5	0.02	0.50	0.001	0.54	0.10
	15	0.004	0.51	0.11	0.18	0.004
	50	0.001	0.0001	0.006	0.76	0.0001
	100	0.001	0.0001	0.009	0.06	0.0001
	500	0.001	0.0001	n/a	0.009	0.0001
Cuzick-Edward's k-NN	1500	0.001	0.54	n/a	0.004	0.0001
	k/N=0.001	0.0001	0.01	0.006	0.001	0.0004
	0.005	0.007	0.001	0.0005	0.001	0.0005
	0.01	0.09	0.0001	0.0004	0.008	0.008
	0.02	0.59	0.0007	0.04	0.07	0.05
	0.05	0.37	0.02	0.33	0.05	0.05
	0.1	0.0002	0.06	0.40	0.04	0.05
	0.2	0.0001	1.00	0.05	0.09	0.05
	0.5	0.59	1.00	0.05	0.0002	0.0002
	D=50 miles	0.01	1.00	0.55	0.0004	0.0002
Moran's I Distance based	100 miles	0.03	1.00	0.17	0.06	0.04
	150 miles	1.00	0.89	0.24	0.38	0.28
	u=5	0.41	0.27	0.89	0.27	0.55
Moran's I Nearest neighbor based	10	0.42	0.33	0.57	0.21	0.30
	20	0.44	0.12	0.25	0.05	0.02
	40	0.43	0.21	0.57	0.08	0.06
	100	0.34	0.27	0.75	0.47	0.32
	150	0.33	0.24	0.83	0.99	0.99
	500	0.78	0.17	n/a	n/a	n/a
		0.0001	0.003	0.02	0.0001	0.0001
Oden's Ipop		0.0001	0.0002	0.006	0.002	0.0007
Swartz' Entropy		0.0001	0.0001	0.004	0.002	0.0005
Tango's MEET		0.0001	0.0001	0.07	0.002	0.0003
Whittemore's Test		0.0003	1.00		0.0002	

MEET, maximized excess events test.

Table 3

Results for prostate cancer incidence and late stage in Connecticut, as well as nasopharynx and prostate cancer mortality in the United States, when using tests for global clustering

Test statistic	Parameter	Connecticut		United States	
		Prostate incidence	Prostate late stage	Nasopharynx mortality	Prostate mortality
Besag-Newell's <i>R</i>	m=5	0.0003	0.47	0.14	0.60
	15	0.0001	0.29	0.55	0.25
	50	0.0001	0.54	0.0001	0.0001
	100	0.0001	0.27	0.0001	0.0001
	500	0.0001	n/a	0.0001	0.0001
Cuzick-Edward's k-NN	1500	0.0001	n/a	0.0008	0.0001
	k/N=0.001	0.0001	0.004	0.0001	0.0001
	0.005	0.0001	0.16	0.0001	0.0001
	0.01	0.0001	0.58	0.0001	0.0001
	0.02	0.0001	0.77	0.0001	0.0001
	0.05	0.0001	0.26	0.0001	0.0001
	0.1	0.0001	0.21	0.0001	0.0001
	0.2	0.0001	0.23	0.18	0.0001
	0.50	1.00	0.09	1.00	0.0001
	D=50 miles	0.0002	0.33	0.07	0.0001
Moran's I	Distance based	1.00	0.81	0.12	0.0001
	100 miles	0.0001	0.89	0.10	0.0001
	150 miles	0.52	0.47	0.74	0.99
Moran's I	Nearest neighbor based	0.07	0.72	0.56	0.91
	10	0.02	0.13	0.31	0.15
	20	0.02	0.43	0.92	0.13
	40	0.006	0.29	0.01	0.14
	100	0.002	0.19	0.19	0.01
Oden's Ipop	150	0.0001	0.75	0.99	0.0001
	500	0.0001	n/a	0.0001	0.0001
	Swartz' Entropy	0.0001	n/a	0.79	0.0001
	Tango's MEET	0.0001	0.18	0.0001	0.0001
	Whittemore's Test	0.49	0.02	0.0001	0.0001

MEET, maximized excess events test.

Table 4
Results for breast and prostate cancer incidence in Connecticut when using a cluster detection test

Test statistic	Breast cancer incidence				Prostate cancer incidence			
	Obs	Exp	RR	p	Obs	Exp	RR	p
Spatial scan	#1	373	1.34	0.001	6286	5252	1.20	0.0001
	#2	462	1.30	0.001	4064	3433	1.18	0.0001
	#3	252	1.40	0.006	46	7	6.16	0.0001
	#4	482	1.26	0.01	30	7	4.38	0.0001
	#5				24	5	4.82	0.0002
	#6				162	103	1.57	0.003
	#7				182	125	1.46	0.02
	#8				15	3	4.62	0.03
Turnbull's CEPP	kN/C= 1	6	4.7	0.33	7	1	7.0	0.0004
	5	20	3.6	0.003	31	5	6.2	0.0001
	10	29	2.3	0.22	48	10	4.8	0.0001
	50	111	1.7	0.01	91	50	1.8	0.0001
	100	194	1.5	0.001	173	100	1.7	0.0001
	500	778	1.2	0.0007	735	500	1.5	0.0001

CEPP, cluster evaluation permutation procedure; Obs, observed; Exp, expected; RR, relative risk.

Table 5
Results for thyroid cancer incidence in Connecticut when using a cluster detection test

Thyroid cancer incidence												
Test statistic	Ages 0–24					Ages 25+					All ages	
	Obs	Exp	RR	p		Obs	Exp	RR	p		Obs	Exp
Spatial scan Turnbull's	#1	12	4.38	0.02		988	885	1.12	0.01		1079	955
	kN/ C=1	5.3	5.3	0.05		2	1	2.00	0.97		2	1
CEPP	5	15	3	0.008		10	5	2.00	0.65		11	5
	10	22	2.2	0.02		19	10	1.90	0.25		19	10
	50	72	1.4	0.01		74	50	1.49	0.04		75	50
	100	125	1.25	0.008		135	100	1.35	0.02		139	100
		500		670		500	1,34	0.03	580		500	1,16

Obs, observed; Exp, expected; RR, relative risk.

Table 6
Results for breast cancer treatment and late stage prostate cancer in Connecticut when using a cluster detection test

Test statistic	Breast cancer treatment			Prostate cancer, late stage		
	Obs	Exp	RR	Obs	Exp	RR
Spatial scan Turnbull's CEPP	#1	775	1.21	3	0.4	7.6
	k=10	10	1.86	5	1.3	3.78
	50	44	1.64	15	6.6	2.27
	100	85	1.59	24	13.2	1.81
	500	363	1.35	83	66	1.25
	1000	674	1.26	153	132	1.16
						0.84
						0.99
						0.68
						0.61
						0.49
						0.21

CEPP, cluster evaluation permutation procedure; Obs, observed; Exp, expected; RR, relative risk.

Table 7
Results for nasopharynx and prostate cancer mortality in the United States when using a cluster detection test

Test statistic	Nasopharynx cancer mortality				Prostate cancer mortality			
	Obs	Exp	RR	p	Obs	Exp	RR	p
Spatial scan	#1	253	114	2.2	16612	13587	1.22	0.0001
	#2	242	140	1.7	27831	25781	1.08	0.0001
	#3	4	0.09	44.6	7783	6809	1.14	0.0001
	#4				1443	1147	1.26	0.0001
	#5				609	458	1.33	0.0001
	#6				757	621	1.22	0.0003
	#7				698	581	1.20	0.04
Turnbull's CEPP	kN/C= 5	21	5	4.2	13	5	2.6	0.12
	50	145	50	2.9	95	50	1.9	0.001
	100	229	100	2.3	181	100	1.8	0.001
	1000	1170	1000	1.2	1491	1000	1.5	0.001
	5000		n/a		5000	1.3	0.001	

CEPP, cluster evaluation permutation procedure; Obs, observed; Exp, expected; RR, relative risk.