2001

Comparing Three Spatial Cluster Tests from Rare to Common Spatial Events

Ge Lin

Comparing Three Spatial Cluster Tests from Rare to Common Spatial Events

By

Ge Lin

RESEARCH PAPER 2001-24

Assistant Professor
Department of Geology and Geography and the Regional Research Institute
West Virginia University
P.O. Box 6825
Morgantown, WV 26506-6825
Email: glin@wvu.edu

Paper presented at the 48th Annual North American Meetings of the *Regional Science Association International*, Chicago, IL, November 2001

Abstract: In the past few years, several new tests for spatial clustering have been proposed. With ever increasing capability of GIS and wider availability of spatial statistic functions, spatial analysts are likely to face challenge of properly using these tests. Seemingly gaps also exist between the development of new tests and follow up evaluations against various assumptions. In this research note, I compare three cluster tests along a range of distribution from rare to common spatial events. The results not only revealed sensitive data feature that each test is designed to detect, but also clarified the interpretation based on the nature of the test.

1. INTRODUCTION

The classical spatial autocorrelation statistics, such as Moran's *I* and Geary *C*, assume that attribute values (e.g., disease prevalence) are either in equal probability among all the geographic units or from a single parent distribution. However, many researchers have noted that population sizes often vary substantially among area units (Basag and Newell, 1991; Oden, 1995; Bao and Henry, 1996). When the traditional permutation test of equal probability is applied to this situation, substantially large variation often occurs in sparsely populated areas. There are two ways to account for population size in spatial statistics (Dutilleual, 1993). One is to modify existing statistical methods to account for population effects; the other is to deal with the underlying distribution of spatial dependency of population counts. For the former approach, Oden (1995), and Bao and Henry (1996) applied regional population weights to adjust respectively the global (adjusted I) and local (GLISA) of Moran's I tests. For the latter approach, Rogerson (1999) extended Oden (1995) and Tango (1995) and developed the spatial Chi-square test *R* based on a categorical approach to spatial dependency. The test statistic treats spatial events as independent Poisson variables under the null hypothesis and by definition accounts for population size of each area unit. Lin and Zeng (1999) noted that *R* might treat a negative spatial association or an area with both high and low values as a cluster, and current Moran's *I* and *G* statistics may not be able to correct this problem. They, therefore, provided a simple likelihood ratio test to check if neighborhood values around an area are positively correlated, thus eliminating negative spatial associations while maintaining the general categorical approach and Poisson assumptions.

Currently, there have been some assessments and applications of adjusted *I* (Bao & Henry 1995; Rogerson, 1999), and population weighting schemes have been incorporated to some statistic package, such as S-plus (xxx). However, there have been few evaluations, if any, for Rogerson *R* and Lin-Zeng likelihood ratio tests. Hence, if one simply applies a cluster test without fully understanding its underlying assumption, there are some potential dangers of reaching an erroneous conclusion. To better understand the intent and suspected mode of each method (Wartenberg and Greenberg, 1990), I compare two categorical oriented tests--*R* and Lin-Zeng likelihood ratio test (*LR-T*)-- in terms of the general or global test. I also include the Getis-Ord *G* (Getis and Ord, 1992) test, because unlike the local Gs, the global G has not been evaluated extensively, and it might provide a viable alternative to R or the LR-T even the sample size is sufficient large but not normally distributed. To clarify some of issues, I compare the three tests along a continuum of spatial events from relative rare, presumably following the Poisson distribution, to relatively common, presumably following the normal distribution. In the following, I briefly introduce each method, and then compare testing results using simulated data on a lattice and a case study of West Virginia Lung Cancer.

2. THREE CLUSTER TESTS

2.1 Rogerson R

Tango (1995) proposed a general test of clustering, C_g, for rare disease based on a quadratic form. Using the χ^2 approximation with the degree of freedom being adjusted by a Gamma function for skewness, C_g, tests whether spatially distributed disease rates are independent or clustered based on a spatial weight (*W*) matrix. Similar to Moran's *I*, an expected

rate is derived for each region within a study area and this rate is compared with the observed rate within each region. Similar to Oden, Tango's C_g also adjusts for population size in each region. A special case of Tango C_g is the Rogerson R (Rogerson, 1999), a spatial version of the Chi-square goodness-of-fit statistic.

Given a population size (ξ_i) and disease prevalence (N_i) at region i for a study area with m regions, the random variable $\boldsymbol{r}$ is the $m \times 1$ vector of $r_i = N_i/N$, where $N=N_1+N_2+\ldots+N_m$, and the nonrandom variable $\boldsymbol{p}$ can be expressed by the $m \times 1$ vector of $p_i=\xi_i/\xi$, where $\xi=\xi_1+\xi_2+\ldots+\xi_m$. The spatial chi-square goodness-of-fit statistics is defined as:

$$R= \sum_i \sum_j w_{ij} (r_i-p_i)(r_j- p_j) \qquad (1)$$

Where w_{ij} are elements of the weight matrix (W) defined by

$$w_{ij} = a_{ij} / \sqrt{p_i p_j} \; . \qquad (2)$$

a_{ij} is a measure of geographic closeness of region i to region j. Substituting w_{ij} for $a_{ij} / \sqrt{p_i p_j}$ in equation (1) we have

$$R= \sum_j (r_i-p_i)^2/p_i + \sum \sum_{i \neq j} a_{ij} (r_i-p_i)(r_j- p_j) / \sqrt{p_i p_j} \qquad (3)$$

Rogerson further provides the expected value and variance of R. Under the sparseness (or rare disease), we can adjust the degree of freedom by the Gamma function, and then test the significance using the chi-square approximation based on the one-sided probability of the limiting distribution. The motivation for developing the R test is evident from equation (3). The first term is the usual chi-square statistic and it measures non-spatial deviation or the deviation of the diagonal elements in a spatial matrix; the second term is the spatial chi-square statistic measuring spatial deviations of nearby regions, or the off diagonal elements in a spatial matrix. R is the sum of the two terms hence accounts for both non-spatial and spatial variation.

However, equation 3 also reveals some problems, especially when dealing with a relatively common disease. First, either 'hot' or 'cool' spots would make the test significant just as the usual chi-square test does. Second, cool and hot spots could coexist in a disease pattern, and *R* cannot differentiate a cool spot from a hot spot. As a result of the second problem, this test may not be sensitive to local variation within a cluster. In other words, these tests generally only provide statistic significance for the existence of spatial associations similar to the *G* test, negative spatial association could also make the statistic significant. In this regard, *R* should be treated as a test of spatial association for count data.

2.2 Getis-Ord G

The Getis-Ord *G* test was developed along with the local version of G_i and G_i^* (Getis and Ord 1992, Ord and Getis 1995) under the assumption that the variable has a natural origin. For count data, this distribution property effectively excludes negative values in the same way as *R*. The local G_i^* is to test the deviation of local pattern (including the location *i*) from the average values of an attribute (x) of interest, and the *G*, as a global measure of spatial association tests the aggregated effect of G_i^* from attribute values for spatial concentration or lack of concentration. Although many have evaluated the G-test, most of the evaluations are on local statistics (Graglia, et al. 2000; Sokal et al 1998). The global G has not been subject to extensive evaluations. For any two points (i,j), $G(d)$ is formulated as a proportion between the sum of multiples of attribute values x_i with all other *x*s within distance *d* (e.g., neighboring regions) and the sum of all multiples of all *x* pairs in the study area:

$$G(d) = \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i \sum_j x_i x_j} \qquad (4)$$

The expected value [$E(G(d)) = \sum_i \sum_j w_{ij}(d) / [n(n-1)]$. In essence, it is a test for spatial associations, in which spatially included pairs within *d* are measured for their similarity by referring to the sample mean, and this similarity is then contrasted with a measure of spatial similarity via spatial weight matrix W (Hubert, 1977). Even though the G test is based on the properties of the normal distribution, when the variable is not normally distributed, the test is asymptotically normal as long as the sample units within *d* are sufficient (for similar comments on Moran's *I*, see Cliff and Ord , 1971). Similar to Moran's *I* statistics, a positive and significant *Z* value for *G* or *Z(g*) suggests an overall tendency of high attribute values clustered spatially. Unlike Moran's *I*, a significant *Z(g)* suggests the existence of low-value clusters, although it may also suggests juxtapositions of high and low values. Unlike *R*, sample events in *G* are converted to rate, and therefore not sensitive to population size as long as the event/population ratio is retained, which is the very the concern that motivates the development of *R*.

2.3 Spatial likelihood ratio test (*LR-T*)

Noticing potential problems with *R*, Lin and Zeng (1999) proposed a spatial likelihood ratio test for common diseases similar to the parameterized simultaneous autoregressive (SAR) test (Basag, 1974; Haining, 1990) against an existing pattern. The likelihood ratio test is a general test of clustering designed to supplement other tests, and because of this role, the test at the global level is sufficient, as local versions of R_i is the uniformly most powerful (UMP) score tests of the R.

Unlike the traditional approach of testing no difference in mean between localized values and the expected value in spatial autocorrelation, the LR_T focuses on similarity in values

around a neighborhood. It assesses the spatial dependence with a built-in cluster component for the existing data, and then tests the significance of the cluster component against the null hypothesis of no cluster. If values in neighboring regions for a given location are positively correlated, then regions around that location are clustered.

Following Rogerson's notation, *m* denotes number of area units in the study area; ξ_i and N_i denote, respectively, the population size and the number of incidences (e.g., diseases) at region *i*. Assuming *p* as the probability of getting the disease in the study area, we have a set of random Poisson variables N_i, each with a mean of $p\xi_i$. When ξ is large, these Poisson variables asymptotically follow a multinormal distribution (Johnson, 1982), i.e.,

$$B_i = \left(\frac{N_i - p\xi_i}{\sqrt{\xi_i}}\right), \quad i = 1,\ldots, m \quad \text{and} \quad B = \begin{matrix} B_1 \\ \ldots \\ B_m \end{matrix} \qquad (5)$$

has a multinormal distribution $N_m(0, \sigma^2(I_m+\varepsilon W_m))$ with the mean of zero and covariance matrix of $\sigma^2(I_m+\varepsilon W_m)$. Here, the covariance structure has a similar model structure to *R*. When there is no spatial clustering, the ε term would be zero, leaving the usual σ^2 term. W_m is a spatial weight matrix with $w_{ij}= \exp(-d(ij)/\tau)$, and τ is a scale parameter.

Under the null hypothesis (H_0) of independent spatial distribution, ε should be statistically close to 0. The likelihood ratio test compares the observed pattern with the pattern under the null hypothesis (Casella, 1990). If the ε makes a difference between the observed and the one under the H_0, then the likelihood ratio statistic should be big enough with one degree of freedom. Hence, testing the existence of a cluster is equivalent to testing: $H_0 : \varepsilon=0$ vs. $H_1 : \varepsilon>0$. If the

largest log-likelihood ratio statistic is small, then the observed pattern is similar to the pattern of no spatial clustering. Given the density function of the multinormal distribution, we can construct a likelihood ratio test with $(Max_{Ho}f(X))/(Max_{H1}f(X))$ or its equivalent test of the 2loglikelihood=$2l(\varepsilon,\sigma,p)$. A measure of spatial clustering, therefore, can be obtained as the difference between $2max_{\varepsilon>0}$Log-likelihood and $2max_{\varepsilon=0}$ log-likelihood:

$$= -2m\ln\{\sqrt{(2\pi)}\sigma\} - \ln|I_m + \varepsilon W_m| - \frac{B^T(I_m + \varepsilon W_m)^{-1}B}{\sigma^2}$$

After working out some algebra, we have

$$T = \inf_{\varepsilon>0}[m\ln B^T(I_m + \varepsilon W_m)^{-1}B + \ln|I_m + \varepsilon W_m|] + m\ln(B^T B) \qquad (6)$$

Where, $\inf_{\varepsilon>0}$ refers to the minimization over all positive ε, hence its negative is a maximization process. $\hat{p}$ is the maximum likelihood estimator of p given by $\sum N_i/\sum\xi_i$. When the likelihood function achieves the maximum, we have $\hat{p}$, $\hat{\sigma}$ and the corresponding ε for the likelihood ratio test *T*. Like *R*, this likelihood ratio test asymptotically follows Chi-square distribution; unlike *R*, it explicitly consumes one degree of freedom, whereas the degrees of freedom in *R* are implicit within the approximation.

Similar to Rogerson *R*, the *LR-T* is designed to identify disease clusters or any other spatial clusters based on spatial incidence. A large *T* suggests that the observed distribution differs from the independent distribution with positively correlated ε term, and the H_0 is likely to be rejected. A small *T*, on the other hand, is likely to accept the H_0. Unlike general cases, where

spatial associations can be positive or negative, a possible negative spatial association in the case of Moran's *I* or the weak case in *G* statistics is not defined. Thus, if H_0 is rejected, it rejects of no clustering of hot or cool spots, but it does not reject other spatial associations.

In addition, there are some special features for the spatial likelihood ratio test. The test is conceptually easy to understand and mathematically easy to construct. Second, the test can be used in various ad hoc situations. Similar to Moran's I, the LR-T evaluates similarity of neighbor values. Unlike Moran's I, the LR-T is a simple test, rather than a statistic; it does not require the calculation of the expected value and variance, and other statistical properties, the exact distribution form for the observed pattern can be treated as an unknown. Moreover, like R, the LR-T treats spatial events in each area unit as an independent Poisson process under the H_0, if one of them is correlated with neighbors' processes, it is sufficient to reject H_0. In contrast, Moran's *I* and the Getis-Ord G treat the entire study area with multiple area units from a single parent distribution. Finally, since the likelihood ratio test is on the spatial term for one degree of freedom, potential covariate structure, such as age, sex and time-period, can be included as controls. In this regard, age and sex can be explicitly controlled rather than being standardized to an adjusted rate (Sun et al, 2000).

3. COMPARING *R*, *G* AND *LR-T* TESTS

3.1. Simulations on a lattice

To see how these tests perform, we shall first compare test results when the pattern is completely random using simulated data. As mentioned in the previous section, the development of *R* is based on the properties of Poisson distribution implicitly for rare disease clustering. The

G-test, on the other hand, is based on the properties of normal distribution. While the *LR*-*T* is based on Poisson assumptions, it is less restrictive to the underlying variable distribution due to the nature of the test. To better understand of disease risk along the line of rare to common diseases, a range of risk levels are generated corresponding to mean risks of 0.3, 0.5, 1.5, 3, 4, 5, 6, 7 percent. First, populations at risk were randomly generated around 1000 for each cell in a 10 by 10 lattice. Then sample events based on Poisson distribution are randomly generated by S-Plus[1] with the following sample mean intervals: 1 to 5, 1 to10, 10 to 20, 30 to 50, 20 to 60, 30 to 70, 40 to 80, and 50 to 90. Note that at the lower end of disease risks, the *G*-test is deemed not stable due to purer Poisson distribution. However, when the sample event becomes more frequent, the Poisson random number generator is gradually approximated by normal distribution, where *R* could equally unstable. Note also that the reason for having a finer mean interval for small sample events is to fine tune the threshold for a rare disease/event, where a particular test may begin to change the statistical power to the acceptable level of 5% due to different assumptions about sample distributions. The weight matrix was based on the exponential distance function identical to *R* with λ =1. Since the binary (0,1) weight was used for G test, two W matrices were generated based on first neighbor and second-order neighbor respectively. The first neighbor weight matrix includes all the regions immediately adjacent to the reference region; while the second-order neighbor weight matrix includes regions adjacent to the first neighbor regions. All the simulations were repeated 1,000 times.

Table 1: Power tests based on 1,000 Simulation for each disease prevalence rate

Sample Mean Intervals*	R test	LR-Test	G test 1st neighbors	G test 2nd neighbors|
1 to 5 (0.3%)	4.5%	0.1%	4.7%	4.3%
1 to 10 (0.5%)	5.3%	0.0%	5.5%	5.3%
10 to 20 (1.5%)	5.9%	0.2%	5.9%	5.4%
10 to 50 (3%)	6.1%	0.1%	5.1%	3.9%
20 to 60 (4%)	7.2%	0.3%	4.9%	4.5%
30 to 70 (5%)	6.6%	0.1%	4.1%	4.2%
40 to 80 (6%)	9.7%	0.3%	3.8%	3.4%
50 to 90 (7%)	10.2%	0.0%	3.2%	3.4%

* numbers in parentheses are average disease rates for each mean interval. For R and T

Table 1 lists the simulation results for the power tests. It appears that the power for *R* is increasing as disease risks move from common (5-9%) to rare (0.3%). On a regular lattice, the *R* test tends to reject a random pattern more than 5% when the disease rate is above 1%. When the means are less than 5, which correspond to λ =1, 2, 3, 4 or 5, the power for R improves to 4.5%. Hence, on a regular lattice, a type I error for *R* is likely to be avoided if the disease rate is less than 0.5% for an average population exposure of 1,000. For the Likelihood ratio test (*LR-T*), the power score is better than both R and G tests. In all simulations; it accepts the H_0 of no cluster consistently less than 0.3%, which corresponds to the 0.003 significance level. The results for G show that the test based on first-order neighbors tends to reject more randomly generated patterns than the one that includes second-order neighbors. In addition, it seems that when sample means decrease, the power of *G* statistic gets weaker until the sample mean reaches 5. When the sample size is between 5 and 20, or 0.5% and 2% risks in the simulations, the rejection rates for a random pattern reaches a level above 5%. When the sample mean is below 5, which corresponds to an average risk of 0.3%, the rejection rate improves to 4.5%.

To test some other patterns, three distinct spatial patterns were generated along with a random pattern on the same lattice using the same sampling framework from the previous simulation (Table 2). First, sample events (λ=5, 10, 30, 50) on the grid were generated under the condition that all random samples must be accepted by all three tests for the H null of no clustering. This step guarantees that all the initial samples are randomly distributed on the lattice for the three tests. Second, an elevated number of cases by 60%- 100% for each cell around a central 3 by 3 grid (i.e., the 9 cells must have at least one grid away from the margin to avoid edge effects). For example, if the λ=10, then an elevated risk would correspond to λ=16 to 20 for a hot spot. Third, values in this 3 by 3 grid were replaced by 9 low values for a cool spot with a reduced rate by 60-90%. Finally, this cool spot is replaced with a chessboard pattern juxtaposing 5 high values and 4 low values with the identical high and low value ranges from the previous two steps[2].

The four columns in Table 2 represent four patterns set out above, and the four panels correspond to four levels of disease risks. According to the condition for a random sample (A), all the tests accept the H_0 of no clustering for the pattern in column one. Moving to the second column for a hot-spot, we find that all tests consistently reject the H_0 for an elevated risk of 80% more on average than the sample A. For the cool spot cases in the next column, *R* and the likelihood ratio tests reject the H_0 for all cases, while results from the *G* test are not so sensitive. Except the one for λ=30, all the results from *G* tests are not significant. When λ=30, the average values of 9 inserted cells for the cool area is 4.2, which is 14% of original risk of 29.9, or the reduction of 86% of the original risk. In this case, however, the *Z(g)* is positively significant suggesting an existence of a hot spot instead.

Table 2: Tests of significances based on four spatial patterns

Sample Mean	Random sample **A**	Add a hot spot to **A** HHH HHH HHH	Add a cool Spot to **A** LLL LLL LLL	Add hot-cool spot to **A** HLH LHL HLH
Sample mean 5 (0.5%)				
L test	0.577	0.035	0.026	1
R test	0.635	0	0.001	0.093
G test 1st neighbor	0.511	3.142	1.070	0.600
G test 2nd neighbor	0.7180	3.875	1.874	-1.159
Sample mean 10 (1%)				
L test	7.367	0.006	0.012	1
R test	0.354	0	0	0.031
G test 1st neighbor	-1.400	4.288	1.624	0.488
G test 2nd neighbor	-1.082	4.933	2.271	-1.957
Sample mean 30 (3%)				
L test	1	0.009	0.001	1
R test	0.356	0	0	0
G test 1st neighbor	1.434	3.952	2.077	0.290
G test 2nd neighbor	1.289	4.844	3.424	0.126
Sample mean 50 (5%)				
L test	1	0.001	0.002	1
R test	0.251	0	0	0
G test 1st neighbor	-0.510	3.391	-0363	-0.534
G test 2nd neighbor	-0.386	4.217	1.193	-3.934

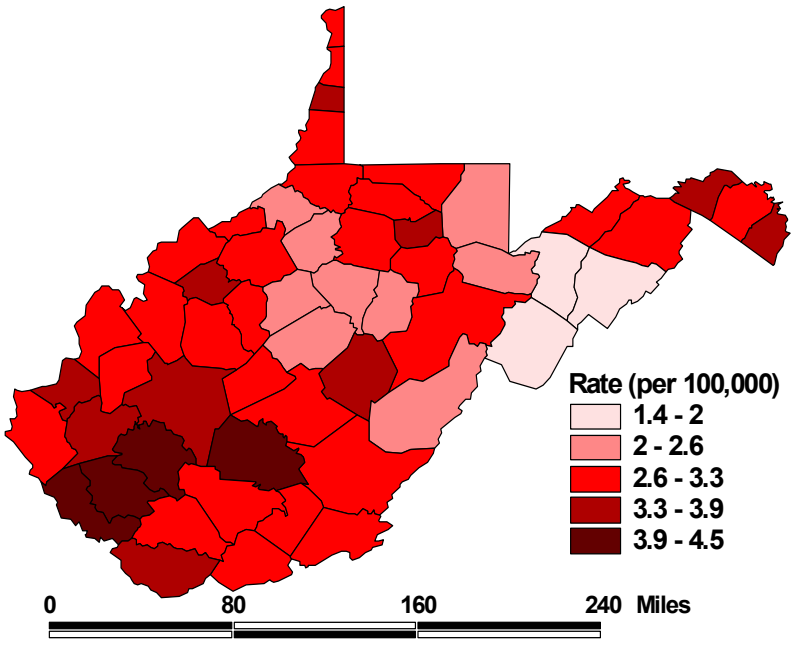
Moving to the last column, *R* test rejects H_0 all chessboard inserts for various sample means, suggesting an existence of a clustering. For the likelihood ratio test, all the patterns are accepted as no clustering. For G test, two cases accepted the H null of no clusters are mean=5 and 30. *Z(g)*s for means =10 and 50 are considered significantly different from the H null with a negative sign, and it suggests an existence of either a negative spatial association similar to Moran's I, or a cool spot. It is hard to tell though which one is at work when people do not have a firm handle of the observed pattern.

3.2. A case of West Virginia lung cancer

West Virginia is known for its coal industry and high concentration of smokers, where Lung Cancer related both to smoking and occupation health, is a major concern. To have a general assessment of spatial risks of Lung Cancer in the state, it is necessary to evaluate both general and focused tests of spatial clustering. Here, I restrict the assessment to the three global tests using age-standardized annual Lung-Cancer rate (per 100,000) for males based on 1970-94 data from US. National Cancer Institute. The annual Lung Cancer rate is displayed in Figure 1 in 5 equal intervals. It appears that several counties in the Southwestern West Virginia, where most coal mines are located, have the highest prevalence. The four counties, Logan, Mingo, Fayette, and Boone, had a rate above 4.0 per 100,000, while the rest of counties in the state are below 3.9. If the natural break were used to divide county groups in the figure, the four counties would still form a unique group with the highest Lung Cancer risk.

To calculate the *G*-test, the standardized rates can be used directly. For *R* and *LR*-*T*, the population at risk was 100,000 standardized by 1970 county populations, and the standardized rates are the sample events. For *G*, both first-order neighbor and d=40 are used to separately generate the spatial weight matrices. For *R* and *LR*-*T*, a power function $d^{-0.9}$ is used. The results [G(Z)s] for both first-order neighbor and 40 mile distance range are not significant[3]. The P-value for *R* is 0.0550, a very marginally non-significant result, while the *LR*-*T* yields a P-value of 0.0233 suggesting an existence of a cluster. Although these tests are used in ad hoc fashion, and the results may be preliminary, they nevertheless provide a range of utility along a continuum of rare to common spatial events. In this particular case, where the average Lung Cancer rate is 3.01 per 100,000, the *LR*-*T* may complement other tests to confirm an existence of a cluster.

Figure 1: Average Annual Lung Cancer Rate in West Virginia



Data source: National Cancer Institute County Data Sheet 1970-94.

SUMMARY

In this research note, I have compared three cluster tests in terms of their intent, and the range of utility for rare to common spatial events. Compared to Rogerson's *R*, the *LR-T* is less restrictive to the assumption of sample distribution, as the method primarily models the neighboring covariance, thus requiring no calculation of expected value and variance. Simulations on a regular lattice showed that when the events are relative common ranging from 1

to 15 in Poisson sampling, the *LR-T* is more robust than both *R*, and *G* in detecting hot or cool spots. In addition, the *LR-T* is more sensitive to local variation, and it tends to treat a low-and-high juxtaposition as no clustering, while other tests may find an existence of a cluster. For a rare event, the less robust results for the G-test is expected, and when the event becomes more common, the results are fairly stable and consistent with the expectations. As for R, some inconsistencies are due to different cluster definitions essentially from the concept of Chi-square, some may be due to approximation procedure used to adjust for the degrees of freedom based on the cumulative deviations under a slightly relaxed distribution assumption (i.e., from Poisson toward Normal distributions). It seems that very small variances resulting from the formulation make a huge difference (see Tiefelsdorf and Boots, 1997 for a discussion of the variance in a similar context). However, when the null distribution is Poisson with less than 0.3% disease rate in the simulation, R approach to robust and expected results.

Even though the *LR-T* test is computationally more intensive and time consuming than *R* and *G,* because it needs to search the maximum of the log-likelihood, it conceptually simple and easy to formulate mathematically. Besides its' less restrictive distribution assumptions, *LR-T* ignores juxtapositions of high and low values regions as a cluster. If regular and irregular high and low patterns are considered as clusters, then a distinct sign, or value range should be associated with this type of pattern. Otherwise, a significant result might be misinterpreted as a cluster of disease when in reality, both high and low values coexisted in the same clustered area. In both *R* and *G*, there is no distinct sign or value range to treat this situation. Moran's *I* has a distinct negative sign associated with it, but the statistic is for spatial autocorrelation rather than cluster detection. It seems that that cluster detection methods must be related to type of data and

the research questions that the analyst want to answers.

This research note is restricted to the comparison of general test of spatial cluster. However, both *G* and *R* have a corresponding local version of the tests. It could be argued that the local test is more important regardless of the result from the global test (however, see Sokal et al., 1998 for a counter argument). Although it is possible to decompose *LR*-*T* the sum of local contributions to the global statistic by partitioning the *W* matrix into *m* local weight matrices, the local version of test cannot be directly compared with the other tests for the existence of multiple clusters. Our next step is to explore various ways of measuring local clusters under the Poisson distribution and comparing their statistic powers among themselves.

Endnotes:

[1] S-Plus is statistic software registered by MathSoft, Inc. All the statistic results in this paper are generated using S-Plus. Any data and program codes used in this study are available upon request.

[2] In the real simulation, the juxtapositions of high or low values are randomized; the chance of 4 high values with 5 low values is the same as the one reported (5 high value and 4 low values). In the preliminary analysis, I did a dozen runs, the results for 4 high with 5 low values are identical to the reported in the table in the majority cases.

[3] Local G_i*s are positive and significant around the central two of the four counties; G_i*s for Logan and Boone counties are 0.136 (Z=3.48) and 0.173 (Z=3.23) respectively, suggesting a cluster around these two counties with an excessive Lung Cancer rate. This result consistent with the notion that G_i* can be used independently to confirm the existence of a cluster, although extra caution needs to be taken when the global G is not significant.

References:

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographic Analysis* 27:93-115.

Bao, S. and M. S. Henry (1996). Heterogeneity issues in local measurements of spatial association. *Geographic Systems* 3:1-13.

Basag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistic Society* B 26: 192-225.

Basag, J. and J. Newell (1991). The detection of clusters in rare diseases. *Journal of Royal Statistic Society* A 154 part1:143-55.

Casella, G. (1990). *Statistical inference*. Pacific Grove, CA.

Cliff, A. D. and K. J. Ord (1971). Evaluating the percentage points of a spatial autocorrection coefficient. *Geographical Analysis* 3:51-62.

Cliff, A. D. and K. J. Ord (1981). *Spatial processes: Models and applications.* Pion, Ltd., London.

Dutilleual, P. (1993). Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49, 305-314.

Graglia, M, R. Haining, and P. Wiles (2000). A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies* 37: 711-29.

Getis, A. and J. Ord (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189-206.

Haining, Robert (1990). *Spatial data analysis in the social and environmental scienc*es. Cambridge University Press, Cambridge, UK.

Hubert, L. J. (1977). Generalized proximity function comparisons." *British Journal of Mathematical and Statistical Psychology* 31:179-82.

Johnson, A. Richard (1982). *Applied multivariate statistical analysis.* Englewood Cliffs, N.J.: Prentice-Hall.

Lin, G. and T. Zeng (1999). Spatial cluster of diseases: Remodeling the concept. *Geographic Information Sciences* 5(2): 175-1180.

Oden, N. (1995). Adjusting Moran's I for population density. *Statistics in Medicine* 14:17-26.

Rogerson, P.A. (1999). The detection of clusters using a spatial version of the Chi-square

goodness-of-fit statistics. *Geographical Analysis* 31:130-147.

Sokal, R., N. L. Oden, and B. A. Thomson (1998). Local spatial autocorrelation in a biological model. *Geographic Analysis* 30:331-351.

Sun, D., R. K. Tsutakawa, H. Him, and Z. He (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 19:2015-35.

Tiefelsdorf, M. and B. Boots (1997). A note on the extremities of local Moran's I's and their impact on global Moran's I. *Geographical Analysis* 29, 249-257.

Tango, T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine* 14: 2323-2334.

Wartenberg, D. and M. Greenberg (1990). Detecting disease clusters: The importance of statistic power. *American Journal of Epidemiology* 132 s156-166 (supplement).