Published in final edited form as:

Stat Med. 2014 July 10; 33(15): 2634–2644. doi:10.1002/sim.6143.

Relative risk estimates from spatial and space-time scan statistics: Are they biased?

Marcos O. Prates^{a,*}, Martin Kulldorff^b, and Renato M. Assunção^c

^aUniversidade Federal de Minas Gerais, Statistics Department, Av. Antônio Carlos, 6627, Belo Horizonte, MG, Brazil, CEP 30123-970

^bHarvard Medical School and Harvard Pilgrim Health Care Institute, Department of Population Medicine, 133 Brookline Avenue, 6th Floor Boston, MA 02215

^cUniversidade Federal de Minas Gerais, Computer Science Department, Av. Antônio Carlos, 6627, Belo Horizonte, MG, Brazil, CEP 30123-970

Abstract

The purely spatial and space-time scan statistics have been successfully used by many scientists to detect and evaluate geographical disease clusters. Although the scan statistic has high power in correctly identifying a cluster, no study has considered the estimates of the cluster relative risk in the detected cluster. In this paper we evaluate whether there is any bias on these estimated relative risks. Intuitively, one may expect that the estimated relative risks has upward bias, since the scan statistic cherry picks high rate areas to include in the cluster. We show that this intuition is correct for clusters with low statistical power, but with medium to high power the bias becomes negligible. The same behaviour is not observed for the prospective space-time scan statistic, where there is an increasing conservative downward bias of the relative risk as the power to detect the cluster increases.

Keywords

bias; relative risk; scan statistic; spatial statistics; space-time statistics

1. Introduction

The main goal of the scan statistic is to detect and evaluate the statistical significance of spatial or space-time clusters that cannot be explained by the assumption of spatial or space-time randomness. Suppose we observe a number of events located within a geographical region. Naus [1] introduced the notion of a spatial scan statistic with a rectangular window of fixed size, scanning over the space for an unexpectedly large cluster of observations. The restriction of the rectangular windows with fixed size was relaxed by Loader [2]. Turnbull et al.[3] used a fixed population size circular scan statistic on top of an inhomogeneous background population. Kulldorff and Nagarwalla[4] and Kulldorff [5] extended this work

to a variable size scanning window. More recently, scan statistics with arbitrarily shaped scanning windows have been proposed, permitting a variety of clusters shapes [6-10]. Spatial and space-time scan statistics has been widely applied in a variety of fields[11] including infectious diseases [12, 13], cancer epidemiology [14, 15], parasitology [16, 17], veterinary medicine [18], forestry[19], criminology[20], military science [21], history[22] and astronomy[23].

An important extension of the spatial scan statistic is the inclusion of time. The most commonly used space-time scan statistic defines the scanning window as a cylinder with a circular spatial base and a time interval height[24]. There are both prospective and the retrospective space-time scan statistics. With the prospective space-time permutation scan statistic analyses are done repeatedly at regular intervals and the scanning window is restricted so that the cylinder end point is fixed at the most current time, so it is only capable of detecting cluster that are still occurring. For this reason, the prospective scan statistic is used in early disease outbreak detection surveillance systems [25, 26]. The retrospective scan statistic relax the restriction on the cylinder end point searching for a cluster that occurred in anytime over the study period. It is used in applications where there is interest in both current and historical clusters [27-29].

For each cluster detected by the spatial and space-time scan statistics, inference is performed using Monte Carlo hypothesis testing[30], resulting in a p-value that is adjusted for the multiple testing inherent in the many cluster locations and sizes considered. That is, if the null hypothesis of no clusters is true, then the probability of detecting any statistically significant cluster anywhere on the map is 0.05 for an alpha level of 0.05. These p-values are unbiased respectively to the number of potential clusters evaluated, the edges of the study region and the nature of the population density. While the p-values are the main results of a scan statistic, researchers often look at the observed relative risks of the detected cluster to better understand and interpret the results. However, it is not clear how the observed relative risk works as an estimate of a true underlying relative risk of the area in the detected cluster. Intuitively, the estimates of the relative risk returned by a scan statistic may be upward biased, since it cherry picks areas with high rates and excludes those with low rates when determining the boundaries of the detected cluster. For example, when the null hypothesis is true, the observed relative risk for the most likely cluster is always greater than 1 and hence biased. This occurs because some areas will always have higher number of cases than expected, just by chance.

To date, no formal study on the behavior of the relative risk provided by these methods have been investigated. Using the purely spatial scan statistic [5] and the prospective space-time permutation scan statistic [31] has been conducted a study to investigate if there is any bias in the estimated relative risk, its direction and its magnitude. We use the benchmark datasets created by Kulldorff et al. [26, 32] for the purely spatial and the prospective space-time permutation scan statistics respectively. For the prospective space-time permutation scan statistic we introduce two novel relative risk definitions and describe how they can be estimated.

The results are that the relative risks observed for the spatial scan statistic indeed has an upward bias but the bias becomes negligible, tending to zero, when the true underlying relative risk increases and there is enough power to correctly detect the cluster. This asymptotic behavior is not observed for the prospective space-time permutation scan statistic, for which there is a conservative downward bias, so that the observed relative risks tend to be smaller than the true relative risks. This bias remain as the power of the test statistic approaches one. When interpreting the results of spatial and space-time scan statistics, it is important to be aware of these properties of the observed relative risk in the detected clusters.

The rest of the article is organized as follows. In Section 2, we present the study for the purely spatial scan statistic. More specifically, the method for the purely spatial scan statistics is presented in Section 2.1. In Section 2.2 the definition of the relative risk is presented and in Section 2.4 the simulation analysis for the purely spatial is performed. In Section 3.1 the prospective space-time permutation scan method is presented. For the space-time scan statistic a new formulation of the relative risk is introduced in Section 3.2. Section 3.4 does the same as Section 2.4 for the prospective space-time permutation scan statistic. Section 4 concludes with a discussion.

2. Purely Spatial Scan Statistics

2.1. The Method

Kulldorff [5] defined a spatial scan statistic imposing a circular window and allowing the circle centroid to move across the map. For any given position of the centroid, the circle radius varies continuously taking value between zero and some pre-defined upper limit. Theoretically, the method uses an infinite number of distinct circles, each with a different location and size. For analysis purpose we set the circle to contain at most 50% of the total population.

Under the alternative hypothesis, there exists at least one circle with a higher relative risk compared to outside. For every circle, it is possible to calculate a likelihood that takes into account the observed number of cases inside and outside the circle. The circle that maximizes the likelihood function is called the most likely cluster or the candidate cluster. The most common disease mapping applications consider the Poisson or the Binomial distribution for the disease counts. We focus in the Poisson case and the model is defined as follows.

For the map A under study, let Z be the current circle of interest. Define N_Z as the observed number of cases in Z and let E_Z be the expected number of cases in circle Z under the null hypothesis, so that $E_A = N_A = N$. Under the alternative hypothesis that exists a cluster in circle Z . Define L_Z to be the likelihood, and define L_0 to be the likelihood under the null hypothesis. It can be showed that

$$L_Z = \left(\frac{N_Z}{E_Z} \right)^{N_Z} \left(\frac{N - N_Z}{N - E_Z} \right)^{N - N_Z} I(N_Z > E_Z) \quad (1)$$

and one otherwise [5]. The candidate cluster is the one in which the likelihood ratio is maximized over all possible circles. The test statistic is given by

$$S = \max_Z \frac{L_Z}{L_0}, \quad (2)$$

and $\hat{Z} = \arg \max L_Z/L_0$ is the candidate cluster. When derived as a likelihood ratio test, it is based on a set of alternative hypotheses, each with a single circular cluster of different size, location and relative risk. Its p-value is obtained through Monte Carlo hypothesis testing [30].

2.2. Relative Risk

The true relative risk (RR) is defined to be the risk λ_Z of region Z compared to the risk $\lambda_{A \setminus Z}$ in all other areas $A \setminus Z$. Let

$$\lambda_Z = \frac{E(Y_Z)}{E_Z}, \quad (3)$$

$$E_Z = N \frac{P_Z}{P_+}, \quad (4)$$

where Y_Z is the Poisson random variable of the counts in region Z , with expected value given by $E(Y_Z)$, P_Z is the population of region Z , P_+ is the total population at risk in map A and N is the observed total number of cases. Analogously we define $\lambda_{A \setminus Z}$. This way, the true relative risk is defined as

$$RR = \frac{\lambda_Z}{\lambda_{A \setminus Z}}.$$

If both Z and $A \setminus Z$ have the same $\lambda_Z = \lambda_{A \setminus Z} = \lambda$, the true relative risk is 1.

Suppose that Z is selected independently of the observed data. Then, the natural estimator of RR is given by

$$\hat{RR}^s = \frac{N_Z/E_Z}{(N - N_Z)/(E_A - E_Z)}, \quad (5)$$

where N is the total number of cases, N_Z is the number of cases in the cluster Z , E_A is the expected number of cases over the region under the null hypothesis and E_Z is the expected number of cases in the area Z over the null hypothesis. Moreover, this is an unbiased estimator when the region is chosen independently of the observed data. Thus, clearly if the estimated risk within region Z , N_Z/E_Z , is close to estimated risk outside the cluster, $(N - N_Z)/(E - E_Z)$, we have $\hat{RR}^s$ is close to 1 and a strong evidence that there are no cluster in map A . If N_Z/E_Z increases compared to $(N - N_Z)/(E - E_Z)$ the estimated relative risk $\hat{RR}^s$ increases indicating the existence of a cluster in regions Z .

However, we are only interested in estimating RR for the candidate cluster $\hat{Z}$. Because $\hat{Z}$ is cherry picked from all possible Z 's to maximize (1), it is strongly dependent of the realized random variables. Therefore, there is a suspicion that calculating (5) using $\hat{Z}$ will provide an upward biased estimator for RR .

2.3. Relative Risk Bias

After defining the estimated relative risk, we now propose the relative risk bias as

$$\hat{B} = \hat{RR}^s / RR^s, \quad (6)$$

where $\hat{RR}^s$ is the estimated relative risk and RR^s is the true relative risk. Thus, $\hat{B} = 1$ indicates no bias, while $\hat{B} > 1$ indicate positive bias and $\hat{B} < 1$ negative bias.

2.4. Estimation of the Relative Risk Bias

A simulation study is presented to assess the relative risk bias of the estimates of purely spatial scan statistic. The analysis was performed using the software SaTScan [33] which is freely available at <http://www.satscan.org>.

For the purely spatial scan statistic the Northeastern USA Benchmark Data set was used [32]. The benchmark datasets were created with the intention of unifying different analysis over spatial scan methods and are available at <http://www.satscan.org/datasets>.

The Northeastern USA Benchmark Data set consists of the female population in the 245 counties and county equivalents in the Northeastern United States, where each county is geographically represented by its centroid coordinate. The population is the number of women in each county according to the 1990 United States census.

Kulldorff et al. [32] constructed a rural, urban and mixed area hot-spot to represent three different sets of local clusters. The rural cluster was centered Grand Isle County in northern Vermont, on the Canadian border. This is the county with the smallest population in the Northeastern. The urban cluster is generated around Manhattan island in New York City, which is a highly populated region. The center of the mixed cluster is the city of Pittsburgh in Western Pennsylvania. Pittsburgh is a big city surrounded by rural areas. Within each of these three sets, different size clusters using 1, 2, 4, 8 and 16 counties were defined. Moreover, the counties within each cluster were given a higher risk than the remaining counties. For each of the 15 clusters, the relative risks and the expected number of cases under both the null and the alternative hypotheses are presented in Kulldorff et al. [32].

For the 3 different scenarios rural, mixed and urban and with cluster size of 1, 4 and 16, data sets with a variety of relative risks were generated. The chosen RR s were 1, 1.2, 1.3, 1.4, 1.5, 2, 3, 5, 10, 20, 50 and 100. For every possible scenario ten thousand data sets were generated and the relative risk were estimated by SaTScan. Using the estimated values of the relative risk for each data set, the relative bias B_l were estimated for $l = 1, \dots, 10,000$ and the mean relative bias $\bar{B}$ were used to study the bias of the estimated RR s.

Table 1 presents for the cluster sizes 1, 4 and 16 for the mixed scenario the estimation of $\overline{\hat{B}}$, the 95% confidence interval (CI) for the mean estimate, the standard deviation (STD) and the power of the purely spatial scan statistic for each RR . As the findings for all scenarios are very similar we present in Table 1 only the mixed scenario results. From Table 1 it is clear that, for all cluster sizes, as power increases, $\overline{\hat{B}}$ goes to 1 and CI becomes narrower.

Figure 1 presents a dot plot of all simulation scenarios showing that as the power increases the relative bias tend to approximately one. This is an overall empirical indication that there is very little bias for the spatial relative risk when we have enough power to estimate the cluster. The power is a function of the underlying relative risk and the population density. It increases in direct relationship with the relative risk and the population size.

Figure 2(a) shows the variation of the relative bias estimates and the power. We present only the results for the mixed scenario with size 4 cluster as it is representative of all others except for the rural with 1 area, which is presented in Figure 2(b). From Figure 2(a) we observe that for powers above 0.85 the relative bias is practically 1, indicating a good estimation of the true RR by equation (5).

From Figure 2(b) we do not observe the same pattern presented in Figure 2(a). This is because the rural scenario with only one area has a sparse density population at risk. In this situation, it is necessary to have a high RR for the method to be able to capture a signal. For RR from 1.2 up to 5 the power is negligible and only for extremely large RR s the power is above 0.80. Since the power is very low for $RR \leq 3$ most of the times $\hat{Z}$ is not significant, and hence the bias curves for these RR s are irrelevant. For RR between 5 and 10 the bias is large and positive as consequence of the cherry picking selection of $\hat{Z}$, as we anticipated. After that, increasing RR , the relative risk bias quickly converges to 1. Thus, when the population at risk is small, it is necessary a high RR to accurately detect the correct cluster.

Comparing Figure 2(a) and Figure 2(b) we can see that the bias curve in small clusters in rural areas behaves roughly. Increasing the number of datasets generated would not fix this problem. From our experience, smoothness would be achieved if we had a simulation study with more than 6,000 total cases. This way it would allow small regions (like the rural scenario) to get cases more often and the RR would be more precisely estimated. To exemplify, suppose that there is an area at risk with small population which lead to an initial probability of getting a case of 0.01. If the risk in that area is increased, say from $RR = 1$ to $RR = 5$, the probability of getting a case increases to 0.05, which is still very small. So, rarely cases will be observed in that area leading to small detection power. Increasing RR further up to 20, the probability that a new case comes from that area jumps to 0.20 and only then cases would be common in that region allowing for a more precise RR estimation and larger power.

As the power and RR are positively correlated, there is a confounding effect between these two possible explanatory variables for the variation in the relative bias. We prefer to explain the effect on the relative bias focusing on the power as an explanatory variable rather than the RR . The reason is that there exists a very simple summary in all spatial configurations.

When power is above 0.8 the relative bias is negligible. This simple conclusion cannot be directly obtained by the RR analysis.

3. Prospective Space–Time Permutation Scan Statistic for Outbreak Detection

3.1. The Method

Analogously to the purely spatial scan statistic (Section 2), the prospective space-time permutation scan statistic [31] imposes a cylindrical window over the study region and allows the centroid of the base to move across the map, so that, at different positions, the window includes different sets of neighboring areas. As in the spatial scan statistic, for each centroid, the radius of the base circle varies continuously until it includes at most 50% of the total population. Thus, the window remains flexible, both in location and size. The cylinder height contain one or more days up to an upper limit. In contrast to the retrospective scan statistic, the prospective space-time permutation scan statistic considers only those cylinders that reach all the way to the end of the study period are considered. Let $[T_1, T_2]$ represents the time interval of the study, and let s and e represent the start and end times of the cylinder, respectively. The cylinders considered by the prospective space-time permutation scan statistics have the form

$$T_1 \leq s < e = T_2.$$

Hence, the window can cover a geographically small outbreak in a single area having lasted multiple days (a long and narrow cylinder), a geographically large outbreak affecting the entire map but present only during the last day (a short and fat cylinder), or any combination of geographic size and temporal height. In theory, the method creates a set with a very large number of cylinders constituted by different neighboring areas and times within it, and each cylinder is a possible candidate to contain a disease outbreak. Different parameter options can be chosen in terms of the maximum geographic and temporal cluster size being considered. For our analysis purpose we set a period of 3 days as the upper limit on the temporal cluster size.

Similarly to the purely spatial scan statistic, let A represent the map under analysis and let N_Z represent the number of cases in cylinder Z . Define E_Z as the expected number of counts under the null hypothesis, so that $E_A = N_A = N$. Under the alternative hypothesis that exists a cluster in cylinder Z , define L_Z to be the likelihood, and define L_0 to be the likelihood under the null hypothesis. Then it is possible to show that, under the assumption that the observed number of counts is Poisson distributed, the likelihood ratio is the same as equation (1). The definition of the space-time scan statistic S is equivalent to the one presented in equation (2). Finally, a p-value is returned through Monte Carlo hypothesis testing [30].

3.2. Relative Risk

Unlike the purely spatial scan statistic where the estimated relative risk can be calculated by simply dividing the observed by the expected counts, for the prospective space-time permutation scan statistic this is not valid. In the prospective space-time permutation

method, the expected counts depends simultaneously on the specific spatial and temporal location of the observed cases and not only on the spatial and temporal marginal distribution of the population at risk. Thus, the space-time scan statistic allows for purely spatial heterogeneity as well as purely temporal heterogeneity. The three dimensional space is scanned looking for the cylinder that has more events than what we expect based on the marginal spatial and temporal patterns. This fact motivates two definitions of the relative risk. One takes into account the spatial heterogeneity, while the other takes into account the temporal heterogeneity.

The spatially oriented estimated relative risk is given by

$$\hat{RR}_S^{st} = \frac{N_Z/N_T}{(N_B - N_Z)/(N - N_T)}, \quad (7)$$

where N is the total number of cases, N_Z is the number of cases in the cylinder Z with spatial base B and time period $[s, T_2]$, N_T is the total number of cases in the time interval and N_B is the total number of cases in the area B over all times (T_1, T_2) .

The temporally oriented relative risk is defined as

$$\hat{RR}_T^{st} = \frac{N_Z/N_B}{(N_T - N_Z)/(N - N_B)}. \quad (8)$$

Note that both estimated relative risk uses the same statistics N , N_B , N_T and N_Z .

In Table 2 we present two simple scenarios to exemplify the difference between $\hat{RR}_S^{st}$ and $\hat{RR}_T^{st}$. Suppose that in our map A we have a total number of cases fixed as $N = 1000$. In scenario 1, $N_B = 150$ and $N_T = 200$ are fixed, while in scenario 2, N_B and N_T are fixed as 200 and 150 respectively. This way the expect number of cases in the cylinder Z constructed by region B and height $[s, T_2]$, under the null hypothesis, is always $E = 30$ for any setup. Table 2 shows that under the null hypothesis where $N_T \times N_B = N_Z$ the relative risks are estimated to be the same. However, in scenario 1, as the number of observed cases in the cylinder N_Z increases, $\hat{RR}_S^{st}$ increases faster than $\hat{RR}_T^{st}$. This occurs because $N_T < N_B$ and so as N_Z increases, proportionally $N_T - N_Z$ decrease faster than $N_B - N_T$. Clearly, in scenario 2 which $N_B < N_T$, we have that $\hat{RR}_T^{st}$ increase faster than $\hat{RR}_S^{st}$ presenting a symmetrical behavior depending of the number of cases in the region B and time $[s, T_2]$.

In areas where the chance of occurrence of events are rare it may occur that $N_{ZS} = N_S$ or $N_{ZS} = N_Z$ when the true relatives risk is very high in the outbreak period. If that happens we have $\hat{RR}_S^{st} = +\infty$ and $\hat{RR}_T^{st} = +\infty$ respectively. This remarkably fact breaks the relative risk estimation making it biased by definition. We proceed with our simulation analysis selecting $\hat{RR}_S^{st}$ as our standard estimated space-time relative risk.

3.3. Relative Risk Bias

After defining the possible relative risks estimates for the space-time scan statistic, we now propose the relative risk bias estimator as

$$\hat{B} = \hat{RR}_i^{st} / RR_i^{st}, \quad i = S, T, \quad (9)$$

where $\hat{RR}_i^{st}$ is the estimated relative risk and RR_i^{st} is the true relative risk. Similar to the conclusion presented in Section 2.3, then $\hat{B} = 1$ indicates no bias, while $\hat{B} > 1$ indicate positive bias and $\hat{B} < 1$ negative bias.

3.4. Estimation of the Relative Risk Bias

To perform the study of the relative risk bias of the estimates of the prospective permutation space-time scan statistic we used the New York City (NYC) Benchmark Data [26] (available at <http://www.satscan.org/datasets>).

The NYC Benchmark Data consists in the geographic coordinates, representing the approximate center of each zip code, and population numbers for 176 NYC zip codes. Kulldorff et al. [26] generates a total of 134,977 data sets where a hypothetical disease or syndrome occurred, these data sets were structured either under the null model or one of 35 alternative models. The alternative models included a citywide outbreak as well as 34 geographically localized outbreaks. Under each of the proposed models three different sets of data sets were generated with 31, 32, and 33 days, respectively. For more information over the clusters structures see the paper by Kulldorff et al. [26]. All analysis was performed using the software SaTScan [33].

To assess the bias of the prospective space-time permutation scan statistic, we propose a simulation study under the NYC benchmark. Kulldorff et al. [26] proposes different scenarios for the true cluster. The borough scenario contain as a cluster 5 different borough in NYC. Within each borough 2 other scenarios were created (one zip and neighboring zip codes as cluster within the borough). The RRs were chosen to be the same as presented in Kulldorff et al. [26]. For each one of the defined spatial setup, one thousand data sets were generated, with the cluster starting in day 31 and ending in day 31, 32 and 33 when applicable. For all scenarios the relative risk $\hat{RR}_S^{st}$ were estimated using the output provided by SaTScan. We defined to use $\hat{RR}_S^{st}$ to continue with our analysis, since this was the true relative risk used to generate the data in Kulldorff et al. [26]. We use the relative bias $\bar{B}$ to study the overall bias of the method. Under the null hypothesis of no cluster $\bar{B}$ was estimated to be 5.79, 5.37 and 4.69 for the scenarios with 31, 32 and 33 days, respectively. This observation indicates under the null hypothesis the significant detected clusters are those by chance a much higher relative risk.

Table 3 presents the simulation analysis for the prospective space-time permutation scan statistic. When the true relative risk is medium, for all scenarios it is clear that both power and relative risk estimates improves as we have an outbreak with longer duration. However, when the underlying relative risk is originally high, the method estimates very well the one

day outbreak but tends to underestimate the relative risk for longer duration clusters. For some of the simulations all the cases in the One Zip Manhattan cluster (Roosevelt Island) occurs during the outbreak. This clearly biases our relative risk estimation because $\bar{B} = +\infty$. Those simulations were removed from our bias analysis. Besides the bias detection that may occur with very small population at risk areas, from Table 3 there is no clear relationship between population at risk and estimated relative risk. This indicates that the method controls for the population at risk uniformly in the different scenarios.

Figure 3 present the overall trend for the $\bar{B}$ estimates of the space-time scan statistic. As we can visualize the relative bias reduces as the power increases. However, differently from the purely spatial case, the relative risk estimate for high RR tend to underestimate the true RR since for clusters with 32 and 33 days we have $\bar{B}$ smaller than 1.

These results seem to be non-intuitive. However this is not the case. We know that for the spatial scan statistic the $\bar{B}$ estimates converges to 1 when we increase the power, so let us focus on the temporal aspect of the relative risk estimate. If the estimated cluster have more days than expected, the days added do not belong to the true cluster, thus it reduces true RR inside the detected cluster and therefore the estimated RR for the estimated cluster. On the other hand, if the estimated cluster have less days than expected, the true RR inside the detected cluster should be correct, but there is an overestimation of the RR of the other time period since one or more days with higher RR are included in the period outside the estimated cluster. From this point of view and since equation (7) depends on the ratio of the estimated RR inside and outside of the estimated cluster, in both cases the estimate of $\hat{RR}_s^{st}$ reduces providing a biased estimator.

Although it would be preferable for an unbiased estimator of the true RR , for clusters with long duration and high RR the method guarantees detection. However, it provides a RR estimate that is conservative. In this specific situation, one should know that the true RR tends to be larger than the one reported by the method.

4. Discussion

In this paper we formally introduced the mathematical formulation for how to estimate relative risks of the two proposed true relative risk that can occur in the prospective space-time permutation scan statistic. Further, using the benchmark datasets we successfully proposed a simulation study to analyze the existence or nonexistence of bias in the relative risk estimates provided by the purely spatial scan statistic and the prospective space-time permutation scan statistic.

As expected, the estimates of the bias provided by the scan statistics tend to overestimate the true relative risks for scenarios with low power. Since the scan method selects regions to be the candidate cluster where the underlying relative risk were abnormally higher than the remaining areas, but as the power increases, the spatial scan statistic is capable of providing an unbiased estimate of the relative risk of the true cluster. In fact, most of the time the true border of the clusters is not identified, but the RR estimate is still good.

From our analysis, we concluded that the purely spatial scan statistic has no major estimation bias as the relative risk increases and the method have enough power to detect the correct cluster. For the prospective space-time permutation scan statistic an interesting discovery was made. Even though longer duration cluster increases the power of detection, the relative risk estimates seems to be downward biased giving a conservative estimation of the true relative risk. It is also relevant to recall that not all of the real clusters in each studied scenario were circular for the prospective space-time permutation scan statistic and thus it was impossible for the circular scan statistic, investigated in this manuscript, to correctly find the underlying cluster. However, even in scenarios where the underlying clusters were not circular the estimated relative risk bias presented to be limited specially in simulations with higher power.

In this paper we studied the behavior of the relative risk estimates returned by scan statistics methods and we only considered the purely spatial Poisson and the prospective space-time permutation scan statistics which may not be enough to fully understand the real properties of the relative risk estimates. Further investigation for the Bernoulli [4], ordinal [34], multinomial [35], normal [36], and survival data [37] as well as retrospective analysis [24] are examples of different methods that can be used to confirm the results presented. Moreover, different types of scanning windows can have an impact over the estimation of the relative risk and thus of its bias [6-10]. Finally, we think that the presented study provides an initial insight in better understanding the reliability and the behavior of the estimated relative risks returned by the scan statistics methods.

Acknowledgments

This research was supported by cooperative agreement #U01GM076672 from the National Institute of General Medical Sciences under the Models of Infectious Disease Agent Study (MIDAS) program (Marcos Prates and Martin Kulldorff); by grant #R01CA165057 from the National Cancer Institute (Martin Kulldorff); and by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Marcos Prates and Renato Assunção).

References

1. Naus JI. Clustering of random points in two dimensions. *Biometrika*. 1965; 52:263–267.
2. Loader CR. Large-deviation approximations to the distribution of scan statistics. *Advances in Applied Probability*. 1991; 23:751–771.
3. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: Application to leukemia incidence in upstate New York. *American Journal of Epidemiology*. 1990; 132(suppl):136–143. URL <http://aje.oxfordjournals.org/content/132/suppl1/136.abstract>, <http://aje.oxfordjournals.org/content/132/suppl1/136.full.pdf+html>. [PubMed: 2356805]
4. Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. *Statistics in Medicine*. 1995; 14:799–810. [PubMed: 7644860]
5. Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*. 1997; 26:1481–1496.
6. Patil GP, Taillie C. Geographic and network surveillance via scan statistics for critical area detection. *Statistical Science*. 2003; 18:457–465.
7. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*. 2004; 11:183–197.

8. Duczmal L, Assunção RM. Simulated annealing strategy for detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis*. 2004; 45:269–286.
9. Tango T, Takahashi K. A flexible shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*. 2005; 4:11. [PubMed: 15904524]
10. Assunção R, Costa MA, Tavares A, Ferreira S. Fast detection of arbitrarily shaped disease cluster. *Statistics in Medicine*. 2006; 25:723–742. [PubMed: 16453376]
11. Costa, MA.; Kulldorff, M. chap Applications of Spatial Scan Statistics: A Review. Birkhauser; Boston: 2009. *Scan Statistics: Methods and Applications (Statistics for Industry and Technology)*; p. 129-152.
12. Chaput EK, Meek JI, Heimer R. Spatial analysis of human granulocytic ehrlichiosis near Lyme, Connecticut. *Emerging Infectious Diseases*. 2002; 8:943–948. [PubMed: 12194771]
13. Wylie JL, Cabral T, Jolly AM. Identifications of networks of sexually transmitted infection: A molecular, geographic, and social network analysis. *Journal of Infectious Diseases*. 2005; 191:899–906. [PubMed: 15717265]
14. Hjalmar U, Kulldorff M, Gustafsson G, Nargarwalla N. Childhood leukemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine*. 1996; 15:707–715. [PubMed: 9132898]
15. Thomas AJ, Carlin BP. Late detection of breast and colorectal cancer in Minnesota counties: An application of spatial smoothing and clustering. *Statistics in Medicine*. 2003; 22:113–127. [PubMed: 12486754]
16. Odoi A, Martin SW, Michel P, Middleton P, Holt J, Wilson J. Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International Journal of Health Geographics*. 2004; 3:11. [PubMed: 15176979]
17. Riitters KH, Coulston JW. Cluster of *Cappilaria hepatica* infections in non-commensal rodents from the canton of Geneva, Switzerland. *Parasitology Research*. 2005; 96:340–342. [PubMed: 15924224]
18. Heres L, Brus D, Hagenaars T. Spatial analysis of BSE cases in the Netherlands. *BMC Veterinary Research*. 2008; 4:21. URL <http://www.biomedcentral.com/1746-6148/4/21>. [PubMed: 18559085]
19. Fei S. Applying hotspot detection methods in forestry: A case study of chestnut oak regeneration. *International Journal of Forestry Research*. 2010; 2010:1–8. Article ID 815292.
20. Minamisava R, Nouer SS, de Moraes Neto OL, Melo LK, Andrade ALSS. Spatial clusters of violent deaths in a newly urbanized region of Brazil: Highlighting the social disparities. *International Journal of Health Geographics*. 2009; 8:66. [PubMed: 19943931]
21. O'Loughlin J, Witmer F, Linke A. The Afghanistan-Pakistan wars, 20082009: Microgeographies, conflict diffusion, and clusters of violence. *Eurasian Geography and Economics*. 2010; 51:437–471.
22. Wang F, Hartmann J, Luo W, Huang P. GIS-based spatial analysis of Tai place names in southern China: An exploratory study of methodology. *Annals of GIS*. 2006; 12:1–9.
23. Moni Bidin C, de la Fuente Marcos R, de la Fuente Marcos C, Carraro G. Not an open cluster after all: the ngc 6863 asterism in Aquila*. *Astronomy & Astrophysics*. 2010; 510:A44. URL <http://dx.doi.org/10.1051/0004-6361/200912761>.
24. Kulldorff M, Athas W, Feuer E, Miller B, K C. Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *American Journal of Public Health*. 1998; 88:1377–1380. [PubMed: 9736881]
25. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A*. 2001; 164:61–72.
26. Kulldorff M, Zhang Z, Hartman J, Heffernan R, Huang L, Mostashari F. Benchmark data and power calculations for evaluating disease outbreak detection methods. *Morbidity and Mortality Weekly Report*. 2004; 53:144–151. [PubMed: 15714644]
27. Norström M, Pfeiffer DU, Jarp J. A space-time cluster investigation of an outbreak of acute respiratory disease in Norowegian cattle herds. *Preventive Veterinary Medicine*. 2000; 47:107–119. [PubMed: 11018738]

28. Nkhoma ET, Hsu CE, Hunt VI, Harris AM. Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980-2001. *International Journal of Health Geographics*. 2004; 3:25. [PubMed: 15509301]
29. Sheehan TJ, DeChello LM. A space-time analysis of the proportion of late stage breast cancer in Massachusetts, 1988 to 1997. *International Journal of Health Geographics*. 2005; 4:15. [PubMed: 15943865]
30. Dwass M. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*. 1957; 28:181–187.
31. Kulldorff M, Heffernan R, Hartman J, Assuno R, Mostashari F. A spacetime permutation scan statistic for disease outbreak detection. *PLoS Med*. 2005; 2(3):e59. URL <http://dx.doi.org/10.1371%2Fjournal.pmed.0020059>. [PubMed: 15719066]
32. Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Computational Statistics & Data Analysis*. 2003; 42:665–684.
33. Kulldorff, M. Information management services, inc. satscan version 9.1: Software for the spatial and space-time scan statistics. 2010. URL <http://www.satscan.org>
34. Jung I, Kulldorff M, Klassen A. A spatial scan statistic for ordinal data. *Statistics in Medicine*. 2007; 26:1594–1607. [PubMed: 16795130]
35. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Statistics in Medicine*. 2010; 29:1910–1918. [PubMed: 20680984]
36. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics*. 2009:8–58. [PubMed: 19208225]
37. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. *Biometrics*. 2007; 63:109–118. [PubMed: 17447935]

Figure 1. Dot plot of the estimated relative risk bias $\overline{\hat{B}}$ versus power for all purely spatial scenarios combined

Figure 2.

Relative bias estimates (black continuous line) and power (grey dashed line) versus RR for the mixed with 4 areas scenarios (left panel) and rural with 1 area scenario (right panel). The left vertical axis represents the estimated bias scale and the right vertical axis represents the power scale.

Figure 3.
Dot plot of the relative risk estimates versus power for all scenarios combined for the prospective space-time permutation scan statistics.

For the purely spatial scan statistic with 3 different clusters sizes (1, 4 and 16) it is presented the relative bias estimate $\widehat{\bar{B}}$, its confidence interval (CI), its standard deviation (STD) and the power for the mixed scenario with a 0.05 significance level under different RR s.

Table 1

Size	RR	$\widehat{\bar{B}}$	CI	STD	Power
H_0	1.0	1.6532	(1.6497,1.6566)	0.938	0.05
	1.2	1.1718	(1.1713,1.1722)	0.345	0.14
	1.3	1.0614	(1.0613,1.0615)	0.238	0.41
	1.4	1.0135	(1.0134,1.0135)	0.102	0.73
	1.5	0.9939	(0.9939,0.9939)	0.079	0.90
	2.0	0.9974	(0.9974,0.9975)	0.063	0.94
	5.0	0.9997	(0.9997,0.9997)	0.042	0.95
1	20.0	1.0004	(1.0004,1.0004)	0.028	0.93
	1.2	1.1194	(1.1192,1.1196)	0.302	0.29
	1.3	1.0370	(1.0370,1.0371)	0.098	0.70
	1.4	1.0070	(1.0070,1.0070)	0.064	0.91
	1.5	1.0001	(1.0001,1.0001)	0.061	0.94
	2.0	0.9981	(0.9981,0.9981)	0.050	0.95
	5.0	1.0003	(1.0002,1.0003)	0.035	0.94
4	20.0	1.0006	(1.0005,1.0006)	0.026	0.93
	1.2	1.0929	(1.0928,1.0930)	0.286	0.54
	1.3	1.0292	(1.0292,1.0292)	0.080	0.90
	1.4	1.0158	(1.0157,1.0158)	0.094	0.95
	1.5	1.0093	(1.0093,1.0093)	0.048	0.95
	2.0	1.0010	(1.0009,1.0010)	0.042	0.95
	5.0	1.0003	(1.0002,1.0003)	0.031	0.95
16	20.0	1.0002	(1.0002,1.0002)	0.026	0.94

Example of the relative risk estimates: $\hat{RR}_S^{st}$, with spatial heterogeneity and $\hat{RR}_T^{st}$, with temporal heterogeneity for a fixed space-time scenario. Let N_B be number of cases in area B for all times; N_T the number of cases in the time interval (T_1, T_2) ; N_Z the number of cases in the cylinder Z , and E the expected number of cases in the cylinder Z .

Table 2

Scenario 1						Scenario 2					
N_B	N_T	N_Z	N_Z/E	$\hat{RR}_S^{st}$	$\hat{RR}_T^{st}$	N_B	N_T	N_Z	N_Z/E	$\hat{RR}_S^{st}$	$\hat{RR}_T^{st}$
150	200	10	0.33	0.29	0.30	200	150	10	0.33	0.30	0.29
150	200	20	0.67	0.62	0.63	200	150	20	0.67	0.63	0.62
150	200	30	1.00	1.00	1.00	200	150	30	1.00	1.00	1.00
150	200	40	1.33	1.45	1.42	200	150	40	1.33	1.42	1.45
150	200	50	1.67	2.00	1.89	200	150	50	1.67	2.00	1.89
150	200	60	2.00	2.67	2.43	200	150	60	2.00	2.43	2.67
150	200	70	2.33	3.50	3.05	200	150	70	2.33	3.05	3.50
150	200	80	2.67	4.57	3.78	200	150	80	2.67	3.78	4.57
150	200	90	3.00	6.00	4.64	200	150	90	3.00	4.64	6.00
150	200	100	3.33	8.00	5.67	200	150	100	3.33	5.76	8.00

Table 3

Prospective space-time permutation scan statistic using the NYC borough benchmark datasets. Presented are the estimated relative risk bias $\widehat{\bar{B}}$ and power for all borough scenarios with a 0.05 significance level.

Region	Cluster Type	Relative risk												
		Medium						High						
		$\widehat{B}$						$\widehat{B}$						
		Power						Power						
31	32	33	31	32	33	31	32	33	31	32	33	31	32	33
<hr/>														
H_0		5.79	5.37	4.69	0.05	0.06	0.05	5.79	5.37	4.69	0.05	0.06	0.05	0.05
<hr/>														
Brooklyn	One Zip	1.35	1.11	1.04	0.30	0.64	0.82	0.97	0.91	0.88	0.84	0.99	1.00	
	Neighbors Zip	1.49	1.11	1.03	0.33	0.70	0.85	1.01	0.89	0.85	0.79	0.99	1.00	
	Borough	1.44	1.05	0.97	0.31	0.66	0.86	0.97	0.79	0.74	0.73	0.98	1.00	
<hr/>														
Manhattan	One Zip *	1.92	1.72	1.45	0.32	0.57	0.79	1.25	1.02	0.97	0.88	1.00	1.00	
	Neighbors Zip	1.36	1.12	1.01	0.34	0.70	0.88	1.01	0.88	0.86	0.80	0.99	1.00	
	Borough *	1.90	1.27	1.08	0.29	0.59	0.77	1.14	0.90	0.83	0.65	0.95	0.99	
<hr/>														
Staten Island	One Zip	1.34	1.14	1.03	0.65	0.83	0.85	0.98	0.90	0.88	0.83	0.99	1.00	
	Neighbors Zip	1.45	1.13	1.04	0.32	0.84	0.68	0.97	0.87	0.82	0.77	0.98	1.00	
	Borough	1.67	1.22	1.08	0.37	0.71	0.90	1.13	0.94	0.89	0.79	0.99	1.00	
<hr/>														
Queens	One Zip	1.56	1.25	1.15	0.61	0.81	1.06	0.96	0.94	0.84	0.84	0.99	1.00	
	Neighbors Zip	1.41	1.11	1.00	0.33	0.68	0.87	1.00	0.87	0.84	0.77	0.99	1.00	
	Borough	1.71	1.19	1.07	0.32	0.66	0.86	1.06	0.84	0.77	0.71	0.97	1.00	
<hr/>														
Bronx	One Zip	1.33	1.18	1.07	0.31	0.65	0.83	1.02	0.93	0.91	0.82	1.00	1.00	
	Neighbors Zip	1.35	1.13	1.01	0.34	0.67	0.88	1.00	0.89	0.86	0.80	0.98	1.00	
	Borough	1.45	1.15	1.01	0.37	0.72	0.90	1.05	0.85	0.81	0.74	0.98	1.00	

* removed cases where $\widehat{\bar{B}}$ was estimated to be infinity.