An Evaluation of the Extended Power of Cluster Detection Methods Under Sampling Strategies

J. F. Conley

West Virginia University
Department of Geology & Geography
98 Beechurst Ave.
330 Brooks Hall
Morgantown WV 26506
Email: Jamison.Conley@mail.wvu.edu

1. Introduction

This paper analyzes the power of spatial cluster detection methods to accurately detect a cluster's extent or shape under different sampling strategies. When detecting and analyzing disease clusters, it is important that the spatial extent of the detected cluster matches the spatial extent of the actual cluster, as shown by the development of cluster detection methods able to find arbitrarily-shaped clusters (Duczmal and Assunçáo 2004, Patil and Taillie 2004, Tango and Takahashi 2005, Assunçáo et al. 2006, Duczmal et al. 2007, 2008). It is not enough to simply report that the data contains clusters, or give a circular neighborhood that contains the cluster; the shape of the cluster, which may not be circular, needs to be accurately represented to efficiently analyze and respond to the cluster. When the fullest possible data of all known cases of the disease and the entire at-risk population is available, an accurate assessment of the spatial extent of the cluster is possible through the methods cited above. However, this fullest possible dataset may not be available, in which case, samples may be used. The smaller populations used in samples often weaken the standard power of statistics, even causing spatial variation in the power to detect clusters (Waller et al. 2006). Thus, the use of samples may distort the shape of the detected cluster by weakening the ability to determine whether or not locations are part of the cluster.

Recently, Takahashi and Tango (2006) have argued that when searching for clusters of arbitrary shape, the traditional power statistic (rejecting the null hypothesis of no clustering) is insufficient because the detected cluster may not match the shape of the actual cluster. To examine this mismatch, they proposed a bivariate power statistic comparing the number of locations in the detected cluster that are in the actual cluster with the total number of locations in the detected cluster. Further details are available in their paper (Takahashi and Tango 2006). This paper applies their power evaluation statistic to datasets generated with different sampling methods to quantify the resulting changes to the power of cluster detection methods to accurately identify the extent of the cluster.

2. Data

The data used in this experiment are disease simulations using the background population of Allegheny County, Pennsylvania, USA, which contains the city of Pittsburgh. The county's population of 1,281,666 is aggregated into 416 census tracts. There are seven

families of datasets based on the cluster. One family contains no clusters, and the other six families contain one cluster in one of the six sets of fifteen tracts shown in fig. 1. These six clusters are chosen because they have different shapes and sizes. The simulated disease has a background rate of 100 per 100,000 residents. To determine the number of cases in each tract, each person in that tract is given a 0.001 probability of having the simulated disease if the tract is not part of the cluster. Whether or not a person has the disease is calculated independently for each individual, and these individuals are aggregated to the tract level, giving a total number of cases in each tract. For tracts within clusters, the rate is doubled to 200 per 100,000 or a 0.002 probability for each person. To facilitate stratified random sampling and case-control sampling, each person also has an age, race, and gender value which matches data from the 2000 United States Census. These demographic variables are only used for sampling; the age, race, or gender of a person does not impact the probability that that person has the simulated disease.
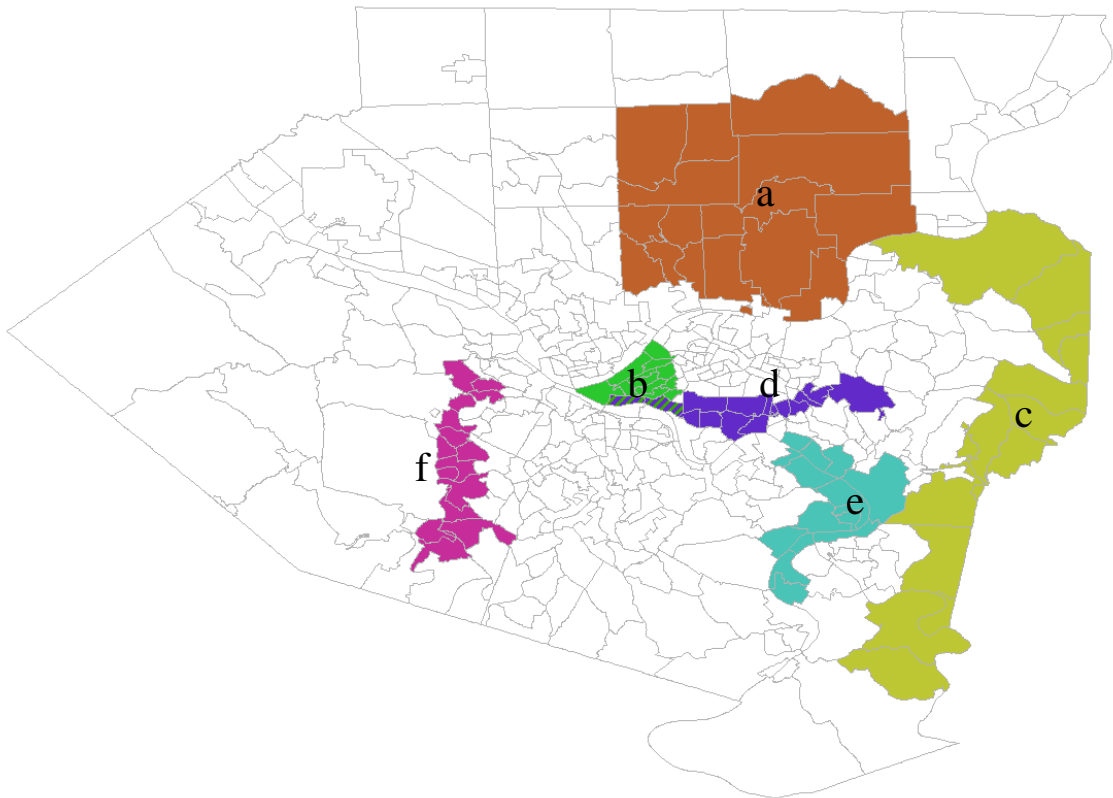

Figure 1. Cluster locations in the simulated datasets. a) compact suburban cluster. b) compact downtown cluster. c) cluster along edge of county. d) cluster along part of I-376. e) cluster along part of Monongahela River. f) elongated suburban cluster.

Additionally, three sampling methods are applied to the complete datasets: (1) random sampling, which uses 1 percent of the people in the complete dataset; (2) stratified

random sampling, which uses a random subset of approximately 5,000 of the people in the following race/gender categories: white males, white females, black males, black females, Hispanic males, Hispanic females, Asian males, Asian females, mixed-race males, and mixed-race females; and (3) case-control sampling, in which one control person is found for each case who shares that case's age, race, and gender values. The control person's location is not restricted to the case person's tract.

3. Methodology

One thousand simulations of each cluster/sampling combination (e.g., compact downtown cluster with stratified random sampling) were generated, giving a total of 28,000 datasets. Three cluster detection statistics were applied each of the 28,000 datasets: the genetic algorithm approach from PROCLUDE (Conley et al. 2005), the circular spatial scan statistic (Kulldorff 1997, Kulldorff and Information Management Services 2009), and the flexible scan statistic (Tango and Takahashi 2005). These three methods each provide a different level of detail available for the cluster shape. SaTScan is one of the most commonly used methods for spatial cluster detection, but under common configurations, only gives circular clusters. The genetic algorithm from PROCLUDE is a fast method for finding elliptical clusters, but cannot generalize the shape beyond ellipses. The flexible scan statistic can find clusters of any shape, although its time complexity limits the size of the detected clusters by requiring longer running times for larger clusters. Due to space constraints, the parameterizations of the algorithms are not presented here, but are available from the author. The results were then analyzed using both the traditional statistical power calculation and the bivariate power statistics proposed by Takahashi and Tango (2006).

4. Results

Tables 1 through 3 show the standard statistical power ($P(+,+)$ in the notation of Takahashi and Tango (2006)) for each of the three methods. For example, the power for SaTScan is 1.000 for all of the cluster families with the complete data, so SaTScan rejected the null hypothesis of no clustering in every single dataset without sampling, even the datasets in the no clustering family, where the null hypothesis should not have been rejected. However, when sampling is used, SaTScan never rejects the null hypothesis, giving a power of 0.000.

	downtown	suburb	edge	stretch	river	interstate	none
complete	1.000	1.000	1.000	1.000	1.000	1.000	1.000
random	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stratified	0.000	0.000	0.000	0.000	0.000	0.000	0.000
caseControl	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 1. SaTScan traditional power

	downtown	suburb	edge	stretch	river	interstate	none
complete	0.988	1.000	1.000	0.998	0.995	0.988	0.991
random	0.641	0.636	0.598	0.602	0.621	0.631	0.604
stratified	0.968	0.989	0.975	0.987	0.992	0.979	0.986
caseControl	0.653	0.647	0.690	0.666	0.642	0.639	0.674

Table 2. PROCLUDE traditional power

	downtown	suburb	edge	stretch	river	interstate	none
complete	1.000	1.000	1.000	1.000	1.000	1.000	1.000
random	1.000	0.999	0.996	0.997	0.997	0.997	0.995
stratified	1.000	1.000	1.000	1.000	1.000	1.000	1.000
caseControl	0.689	0.727	0.744	0.716	0.717	0.709	0.719

Table 3. FleXScan traditional power

Tables 4 through 6 show the power of the methods to include at least one correct region in the cluster ($P(+, >0)$). In the first set of tables, the traditional power is not accounting for the location of the detected cluster. Even though the null hypothesis may be correctly rejected, the results are still problematic if the detected cluster is incorrectly located far away from the actual cluster. This set of tables gives the power of each method to detect a cluster that overlaps the actual cluster by at least one region. Since the cluster detection methods can detect multiple clusters, these tables are broken down into whether or not the most likely of the detected clusters ("BEST CLUSTER") overlaps with the actual cluster, and whether or not any of the detected clusters ("ALL CLUSTERS") overlaps with the actual cluster. For example, using PROCLUDE with the compact downtown cluster from fig. 1, the best cluster overlapped the actual cluster in 108 of the 1000 simulations using the complete dataset, in 350 of the 1000 simulations when random sampling is used, 169 of the 1000 simulations using stratified random sampling, and only 5 of the 1000 simulations with case control sampling. When all clusters are examined, rather than just the cluster reported by the algorithm as most likely, the number of simulations where the actual cluster overlaps one of the detected clusters increases to 679 for the complete dataset, 375 with random sampling, 441 with stratified random sampling, and 10 for case control sampling.

ALL CLUSTERS	downtown	suburb	edge	stretch	river	interstate
complete	0.999	1.000	0.999	1.000	0.995	0.996
random	0.000	0.000	0.000	0.000	0.000	0.000
stratified	0.000	0.000	0.000	0.000	0.000	0.000
caseControl	0.000	0.000	0.000	0.000	0.000	0.000

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.948	0.993	0.979	0.974	0.914	0.889
random	0.000	0.000	0.000	0.000	0.000	0.000
stratified	0.000	0.000	0.000	0.000	0.000	0.000
caseControl	0.000	0.000	0.000	0.000	0.000	0.000

Table 4. SaTScan $P(+, >0)$

ALL CLUSTERS	downtown	suburb	edge	stretch	river	interstate
complete	0.679	1.000	1.000	0.954	0.930	0.699
random	0.375	0.336	0.156	0.259	0.305	0.445
stratified	0.441	0.703	0.622	0.572	0.678	0.608
caseControl	0.010	0.263	0.328	0.191	0.110	0.024

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.108	0.704	0.773	0.456	0.437	0.158
random	0.350	0.314	0.135	0.231	0.280	0.423
stratified	0.169	0.402	0.334	0.275	0.334	0.253
caseControl	0.005	0.170	0.211	0.112	0.060	0.012

Table 5. PROCLUDE $P(+, >0)$

ALL CLUSTERS	downtown	suburb	edge	stretch	river	interstate
complete	1.000	1.000	1.000	1.000	0.999	0.999
random	0.455	0.444	0.442	0.395	0.468	0.518
stratified	0.740	0.632	0.578	0.550	0.700	0.809
caseControl	0.374	0.573	0.548	0.433	0.450	0.461

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.934	0.995	0.983	0.959	0.943	0.910
random	0.151	0.151	0.143	0.124	0.175	0.196
stratified	0.307	0.221	0.185	0.156	0.307	0.393
caseControl	0.302	0.523	0.468	0.352	0.380	0.377

Table 6. FleXScan $P(+, >0)$

Takahashi and Tango provide a weighted power that incorporates penalties for both false positives (w^+), which are tracts that are not part of the actual cluster, but are detected as being in the cluster, and false negatives (w^-), which are tracts that are part of the actual cluster, but are not included in the detected cluster. This weighted power provides an overall view of the effectiveness of the methods at capturing the shape accurately. It is 1.000 when the actual clusters are always detected perfectly (no false positives or false negatives) and decreases for each false positive or false negative, reaching zero when the detected clusters have so many false positives and negatives that the penalties add up to 1.000 or greater in every one of the 1000 simulations. Tables 7 through 9 show the weighted power using the value of 1/15 for both the w^- and w^+ penalties ($I(1/15, 1/15)$ in Takahashi and Tango's notation). This value for the penalties was chosen because each of the clusters contains 15 tracts.

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.3678	0.4931	0.3904	0.5956	0.2421	0.1512
random	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
stratified	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
caseControl	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 7. SaTScan weighted power

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.0443	0.1976	0.2301	0.1543	0.1456	0.0485
random	0.0000	0.0073	0.0032	0.0013	0.0004	0.0001
stratified	0.0218	0.1403	0.0973	0.0556	0.0495	0.0280
caseControl	0.0017	0.0449	0.0602	0.0328	0.0168	0.0032

Table 8. PROCLUDE weighted power

BEST CLUSTER	downtown	suburb	edge	stretch	river	interstate
complete	0.5459	0.5648	0.6436	0.7048	0.5227	0.3136
random	0.0567	0.05302	0.0510	0.0434	0.0596	0.0509
stratified	0.1324	0.0742	0.0720	0.0654	0.1154	0.1025
caseControl	0.1490	0.2473	0.2517	0.2100	0.1681	0.1060

Table 9. FleXScan weighted power

Additional tables for the power to include all of the regions in the true cluster ($P(+, 15)$), the power to avoid false positives ($\sum_{i=1}^{15} P(i, i)$), and the power to detect the cluster perfectly ($P(15, 15)$) were also calculated, but are not given to conserve space.

5. Discussion
As expected, tables 1-3 demonstrate that the traditional statistical power decreased for all methods when switching from the complete dataset to a sample, regardless of the sampling technique. The complete decrease in power for SaTScan, dropping from 1.000

to 0, is surprisingly large. FleXScan maintained its power the best of all the methods. Also note that the traditional power is still high for all three methods when there is no cluster, indicating that they reject the null hypothesis even when they should not. This is largely an artifact of some tracts with small populations having high rates; this analysis examines the most likely cluster that is returned, regardless of its p-value because one of the three methods (PROCLUDE) does not readily return p-values. When analysis using SaTScan and FleXScan is limited to clusters with p-values less than 0.1, the power decreases for all 28 cluster/sampling combinations. This decrease is especially large for the no cluster situation, as it should be, and the decrease is also much greater for the datasets using sampling than for the complete dataset. This, too, is not surprising because the samples have smaller populations than the complete dataset.

Tables 4-6 show that the power to detect a cluster that overlaps the actual cluster ($P(+, >0)$) declines more than the traditional power when switching from the complete dataset to samples. Even though the methods are detecting clusters, they are not necessarily located where the actual clusters are, especially with PROCLUDE's genetic GAM algorithm. For example, comparing the interstate cluster with the complete dataset, while PROCLUDE finds a cluster 98.8% of the time, the most likely detected cluster overlaps the actual cluster only 15.8% of the time, and any cluster detected by PROCLUDE overlaps the actual cluster only 69.9% of the time. Also, this decline in power is consistent for all cluster shapes. The shapes of the compact clusters ("downtown" and "suburb") are not consistently detected better or worse than the other, more elongated clusters.

Tables 7-9 continue this pattern of the power being weaker for the samples than for the complete dataset. This is especially true for FleXScan, which has the highest power using the complete data, yet the decrease when sampling is applied is very dramatic, at times showing powers only 10% of the complete dataset power, even when the traditional power remains at or near 1.000, as with the random sample dataset. This is likely the case because in the random and the stratified sampling, which showed the greatest drops in power, one or more tracts in the cluster may end up containing no cases after the sample is drawn. This greatly reduces the likelihood that FleXScan will include that tract in the detected cluster, even if it was originally part of the actual cluster. Therefore, even though a cluster is almost always detected, as shown in table 3, and the detected cluster often overlaps the actual cluster, as shown in table 6, there are still enough false positives and false negatives to decrease the weighted power considerably. As with the power to detect a cluster that overlaps the actual cluster, this weighted power decreases more for all the sampling strategies than for the complete dataset.

This paper demonstrates that the ability of cluster detection methods to accurately detect the shape of a cluster decreases more than the standard power under sampling conditions. Of these three methods, only FleXScan is specifically designed to detect clusters of arbitrary shape, and for the majority of cluster/sampling combinations, it performed the best. However, once sampling is applied, its ability to accurately delineate the shape of the cluster was still poor, with weighted powers around 0.05 under random sampling. This difficulty in finding the correct extent and location of the cluster should be kept in mind when analyzing data drawn from samples of the population, as is often the case in disease clustering situations. Additionally, work should continue to find ways of minimizing this drop in power for samples.

6. Acknowledgements

7. References

Assunçáo R, Costa M, Tavares A, and Ferreira M, 2006, Fast detection of arbitrarily shaped disease clusters. *Statistics in Medicine*, 25:723-742.

Conley J, Gahegan M and Macgill J, 2005, A genetic approach to detecting clusters in point datasets. *Geographical Analysis*, 37:286-314.

Duczmal L and Assunçáo R, 2004, A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics & Data Analysis*, 45:269-286.

Duczmal L, Cançado ALF and Takahashi RHC, 2008, Delineation of irregularly shaped disease clusters through multiobjective optimization. *Journal of Computational and Graphical Statistics*, 17(1):243-262.

Duczmal L, Cançado ALF Takahashi RHC and Bessegato L, 2007, A genetic algorithm for irregularly shaped spatial scan statistics. *Computational Statistics & Data Analysis*, 52:43-52.

Kulldorff M, 1997, A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26(6):1481-1496

Kulldorff M and Information Management Services, Inc., 2009, SaTScan™ v8.0: Software for the spatial and space-time scan statistics. Downloaded from http://www.satscan.org.

Patil GP and Taillie C, 2004, Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*, 11:183-197.

Takahashi K and Tango T, 2006, An extended power of cluster detection tests. *Statistics in Medicine*, 25:841-852.

Tango T and Takahashi K, 2005, A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4:11.

Waller LA, Hill EG and Rudd RA, 2006, The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statistics in Medicine*, 25:853-865.