6

Applications of Spatial Scan Statistics: A Review

Marcelo Azevedo Costa[1] **and Martin Kulldorff**[2]

[1]*Department of Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*
[2]*Department of Ambulatory Care and Prevention, Harvard University and Harvard Pilgrim Health Care, Boston, USA*

Abstract: In 1965, Joseph Naus published his now classical paper on spatial scan statistics, entitled 'Clustering of random points in two dimensions'. This paper set in motion an important statistical theory of spatial scan statistics and an avalanche of spatial scan statistics applications in a wide variety of fields, including archaeology, astronomy, brain imaging, criminology, demography, early detection of disease outbreaks, ecology, epidemiology, forestry, geology, history, psychology and veterinary medicine. In this chapter, we survey this wide variety of applications.

Keywords and phrases: Scan statistic, spatial, geography, applications

6.1 Introduction

Suppose we observe a number of points located within a geographical or spatial region. These points may, for example, reflect the locations of trees, ant nests, diseased individuals or post offices. The general aim of the spatial scan statistic is to detect and evaluate the statistical significance of a spatial cluster of events that cannot be explained by an underlying probability model defined by a null hypothesis of spatial randomness. There are spatial scan statistics for two, three and more dimensions. If the scanning is done over a three-dimensional area defined by both space and time, we have a space-time scan statistic, which is an important special case of the three-dimensional spatial scan statistic.

Since first presented by Naus in 1965, spatial scan statistics have been applied in many different fields such as infectious diseases, cancer, cardiology, pediatrics, rheumatology, auto-immune diseases, neurological diseases, liver diseases, diabetes, geriatrics, parasitology, alcohol and drugs, accidents, veterinary medicine, demography, forestry, toxicology, psychology, medical imaging, history, criminology, astronomy and geology. The aim of this chapter is to present

J. Glaz et al. (eds.), *Scan Statistics: Methods and Applications*, Statistics for Industry and Technology, DOI 10.1007/978-0-8176-4749-0_6,

a review of the areas in which the spatial scan statistic has been applied, providing a broad sense of how it is being used across the globe and across scientific disciplines. After a brief methodological review, we present examples of applications by field of study. A final discussion presents a brief summary of the main findings.

Although this chapter emphasizes the use of spatial scan statistics, there are many other important spatial statistical methods. From a user perspective, the spatial scan statistic is best viewed as one of several important tools for the successful analysis of geographical and spatial data. Other important methods include visualization techniques, descriptive statistics of rates and proportions, spatial smoothing methods, kriging, global clustering tests, regression for spatially correlated data and so on.

6.2 Brief Methodological Overview

Suppose we have a square region with a number of points. In its original form, studied by Naus (1965), the spatial scan statistic consists of a rectangular scanning window with a fixed size and shape. This window is continuously moved over the predefined square study region, covering all possible locations, and the definition of the spatial scan statistic is the maximum number of points in the scanning window at any given time. The next step is to find the probability of observing at least that many points within the window, under the null hypothesis of randomly located points, generated by a homogeneous Poisson process. In mathematical language, we want to know the probability of finding at least one rectangle with dimensions u and v with at least n out of N points uniformly distributed in the unit square. While simple to state, the complexity of this problem lies in the multiple testing inherent in the many window locations and the overlapping nature of those windows, resulting in the maximum being taken from a set of highly dependent observations. Using some very beautiful and powerful mathematics, Naus (1965) developed theoretical formulas to obtain upper and lower bounds for those probabilities, showing that the bounds converge to the true probability.

Following the pioneering paper by Naus, there have been a number of further methodological developments of spatial scan statistics, in order to handle different types of data. The spatial region to be scanned may be of different shapes; the scanning window may be of different sizes and shapes; the analysis may or may not be conditional on the total number of points observed; the observations may be generated by a homogeneous Poisson process, an inhomogeneous Poisson process, or by a Bernoulli, multinomial, normal or exponential distribution function; there may be a need to adjust for covariates or temporal trends; and so on. For each application, the scan statistic parameters and probabilistic models must be appropriately selected to fit the data and the scientific questions asked.

The study region is usually defined directly by the data and can be of a variety of shapes and sizes. The exact locations of the points may be known, so that we have a spatial point process. Alternatively, the data may be spatially aggregated, so that instead of points we have counts in a set of squares on a lattice or in a set of administrative geographical areas such as postal codes, census tracts or counties.

An important component of the spatial scan statistic is the shape and size of the scanning window. Naus (1965) used a rectangular window of any fixed shape and size, while Loader (1991) used a variable size rectangular window. Alm (1997, 1998) used circles, ellipses and triangles. Kulldorff (1997) defined a spatial scan statistic for any variably sized collection of windows, using a continuously variable size circle in his example. More recently, spatial scan statistics have also been defined using non-parametrically defined windows [Duczmal and Assunção (2004), Patil and Taillie (2003, 2004), Assunção *et al.* (2006), Tango and Takahashi (2005)], taking very irregular shapes. The shape of the window does not need to be the same as the shape of the study region.

Rather than defining the null hypothesis based on a homogeneous Poisson process, another assumption for the null hypothesis implies that intensity varies within the region, following an underlying known population defined by an inhomogeneous Poisson process [Turnbull *et al.* (1990)]. Areas with higher population are then expected to have more points under the null hypothesis, reflecting, for example, the fact that there are more cancer cases per geographical unit in urban compared to rural areas, simply because of the higher population density. Spatial scan statistics have also been developed for discrete 0/1 Bernoulli data [Chen and Glaz (1996), Kulldorff (1997)], as well as for multinomial [Jung, Kulldorff and Klassen (2007)], normal [Kulldorff, Huang and Konty (2008); Huang *et al.* (2009)] and survival type data [Huang, Kulldorff and Gregorio (2007); Cook, Gold and Li (2007)].

An important extension of the spatial scan statistic is to three or more dimensions [Alm (1998)]. The most common of these is the space-time scan statistic, where time is added as a third dimension [Kulldorff *et al.* (1998)]. The size and shape of the study region and scanning window can be defined as before for the purely spatial scan statistic, while time is added as a third dimension. Retrospective space-time scan statistics provide a mechanism to detect and evaluate past or present clusters that might have appeared anytime during the study period. Prospective space-time scan statistics only consider windows that touch the current date in order to only detect and evaluate the existence of clusters that are currently present. The latter method is used in early disease outbreak detection surveillance systems [Kulldorff (2001); Kulldorff *et al.* (2005)].

As we survey the application of spatial and space-time scan statistics, we will consider most of the above variants of the spatial scan statistic, as different versions are useful for different types of applications.

6.3 Applications in Medical Imaging

The spatial scan statistic has been applied to important problems in brain imaging. Naiman and Priebe (2001) have used it for positron emission tomography (PET) scan brain imagery data. Yoshida, Naya and Miyashita (2003) have applied it for neural response data in monkeys. Injections of retrograde tracers in a specific region (cases) and adjacent regions (control) in the brain generated maps with pixels associated to selective and non-selective neurons. Significant clusters of selective neurons were found, assuming a Bernoulli model.

Spatial scan statistics have also been used for breast cancer digital mammography data. The goal is the detection of clustered microcalcifications, which may be indicative of a cancerous tumor [Priebe, Olson and Healy (1997a); Naiman and Priebe (2001)]. Popescu and Lewitt (2006a, 2006b) mimic a cancer nodule detection system. A circular scanning window with fixed radius and variable center is used. The test statistic is the sum of the values of the pixels inside the window. The null distribution of the test statistic is generated by scanning background-only images.

6.4 Applications in Cancer Epidemiology

The incidence, prevalence or mortality rates of cancer may vary geographically for a number of reasons, including spatial variation of environmental or behavioral risk factors or the genetic make-up of the population. Spatial scan statistics have often been used to detect and/or evaluate the statistical significance of geographical cancer clusters, as cancer clusters will also occur simply by chance in some parts of the map. Leukemia was the first cancer that was observed using spatial scan statistics, with Turnbull *et al.* (1990) studying leukemia in upstate New York and Hjalmars *et al.* (1996) studying childhood leukemia in Sweden. Hjalmars *et al.* (1996) did not find any statistically significant clusters in their data, even though there had been one leukemia cluster alarm reported in the press a few years earlier. While the cluster was detected, it was not statistically significant and was not even among the three top clusters. In contrast, Viel *et al.* (2000) found a statistically significant cluster of soft-tissue sarcoma and non-Hodgkins lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels in France.

Under the null hypothesis, the incidence or mortality of cancer is assumed to follow a Poisson distribution, where the expected number of cases in a particular location is proportional to the covariate-adjusted population in that location. Age and other covariates are adjusted for by using indirect standardization. Let $b_{(i,k)}$ be the population in age group k in location i and let $B_{(k)}$ be the total population in age group k. Let $C_{(k)}$ be the total number of cases in age group k. The indirectly age standardized expected number of cases in location i is then

$$\mu_i = \sum_k b_{(i,k)} \cdot C_{(k)} / B_{(k)} \tag{6.1}$$

Age must always be adjusted for in cancer incidence and mortality studies. If not, there will be significant clusters in areas with a predominately older population, since older people are at higher risk of being diagnosed with and dying from cancer. It is often interesting to also adjust for other known risk factors, including socio-economic variables such as ethnicity, educational levels or urbanicity, as well as biological variables such as skin color for skin cancer or parity for breast cancer studies [Kulldorff *et al.* (1997); Hsu *et al.* (2004); Klassen, Kulldorff and Curriero (2005)]. A very interesting approach is to reduce the number of socio-economic variables by only taking a few components from principal component analysis [Sheehan *et al.* (2004), Sheehan and DeChello (2005), Fukuda *et al.* (2005)], usually two independent components. After an adjustment, cancer clusters will disappear if they can be explained by the covariates that were adjusted for. However, the number of clusters can also increase, as a true cluster can be hidden in an unadjusted analysis.

The spatial scan statistic is able to detect and evaluate the statistical significance of individual clusters, but it won't provide an estimate of incidence or mortality rates throughout the map. For that, other statistical methods are needed as a complement, such as the mapping of smoothed rates using conditional autoregressive models [Thomas and Carlin (2003) and Buntinx *et al.* (2003)].

For most cancer sites, there may be a long time between exposure and diagnosis and an even longer time between exposure and death. Han *et al.* (2004) presents a notable approach for breast cancer clustering analysis by using place of residency at the (i) time of birth, (ii) time of menarche and (iii) time of birth of the first child, as alternative geographical coordinates in separate spatial analyses. In this way, the study provides an opportunity to examine geographical clustering of breast cancer at various points during life. Significant clusters were found for the time of birth and time of menarche analyses, with similar results. There were fewer clusters when the data was analyzed using place of residence at time of diagnosis.

Spatial scan statistics can also be used to study the geographical variation of a particular subtype, in order to determine if there are geographical clusters of late stage cancer or cancer of a particular type or grade [Roche, Skinner and Weinstein (2002); Gregorio *et al.* (2002); Sheehan and DeChello (2005); Klassen, Kulldorff and Curriero (2005)]. The detection of a geographical cluster with a high proportion of late stage breast cancer cases may indicate a need to improve breast cancer mammography screening in that geographical area. In these analyses, no census population data are used. Rather, the total number of diagnosed cancer cases is the 'population' while the 'cases' are those cancer cases that are of a certain type, such as late stage. A Bernoulli probability model is suitable for this type of data. These types of spatial scan statistics have also been used to study the geographical variation in cancer treatments [Gregorio *et al.* (2001)]. When there are more than two different stages or grades, it is possible to use a spatial scan statistic for ordinal data, which Jung, Kulldorff and Klassen (2007) did for prostate cancer stage in Maryland, United States.

There may also be an interest in the geographical variation in the survival time after a cancer diagnosis, to determine if there are geographical areas with exceptionally poor survival. This is a continuous outcome. Such analyses must be able to handle censored data and adjust for differences in prognostic factors such as the age of the patient and the stage or grade of the cancer. Using a spatial scan statistic for exponentially distributed data with censoring, Huang, Kulldorff and Gregorio (2007) studied prostate cancer survival in Connecticut, United States.

6.5 Applications in Infectious Disease Epidemiology

In infectious disease surveillance, the spatial and space-time scan statistics are used for two different purposes. The first is retrospective in nature, where historical data are used to detect geographical areas with many cases of the disease. Such clusters can either be temporary in nature, due to an outbreak, or long lasting, if the area or population is especially prone to infection. Different aspects of the infectious disease will influence the proper choice of spatial scan statistic parameters. The incubation time of the disease, for example, is a very important feature to incorporate in the selection of the scanning time window length.

Cousens *et al.* (2001) describe the spatial investigation of 84 cases of variant Creutzfeldt–Jakob disease (vCJD), a rare and fatal disease caused by the same transmissible agent as in bovine spongiform encephalopathy (mad cow disease) and therefore hypothetically associated with the consumption of beef products. With the spatial scan statistic, one statistically significant cluster with five cases was detected. A subsequent investigation revealed a local butcher shop as a likely common source of infection.

Fevre *et al.* (2001) used the spatial scan statistic to study sleeping sickness in Uganda. Sleeping sickness is caused by a parasite that is transmitted to humans by the *tsetse* fly, which picks up the infection from domestic cattle. A purely spatial analysis was performed using the number of cases diagnosed over a 32-day period, from the time of the first recorded case to the time when vector control measures started to be implemented. A case control study was designed, where each case was matched with one control by age, gender and month of admission. Consequently, the spatial analysis was carried out assuming that cases and controls followed a Bernoulli distribution. One significant cluster was found around an important regional cattle market.

Chaput, Meek and Heimer (2002) provide some useful insights into exploring the data through evaluating separate data streams from just-confirmed and confirmed plus probable cases of human granulocytic ehrlichiosis. A spatial analysis in a 12-town area for tick-borne infections is presented using cases during four years of surveillance. The cluster analyses were conducted using either confirmed or both confirmed and probable disease cases obtained from active and passive surveillance system reports. Both datasets provided similar results.

A purely spatial analysis of a variation of vCJD in France is presented by Huillard d'Aignaux *et al.* (2002). In addition to the use of the spatial scan statistic for cluster detection and evaluation, exploratory analyses are also provided, including maps and tests for global spatial clustering. Due to evidence that the incubation period for the disease can be longer, the cluster analyses were done for both place of residency and place of birth.

Listeriosis is a bacterial food-borne pathogen that may be present in 1 to 5 percent of common ready-to-eat food products and which can cause a rare severe invasive disease manifestation and even death in humans [Sauders *et al.* (2003)]. In particular, since the spread of the bacteria is associated with contaminated food, the source of exposure might come from either global food distribution or local sources. As a consequence, spatial-temporal clustering might detect large or small clusters. A cluster analysis using the spatial scan statistic was conducted using different molecular subtyping strategies (ribotype) from sterile sites. Clusters with the same subtyping may represent clusters with a common source of exposure, potentially increasing the ability to detect outbreaks.

When studying sexually transmitted diseases, Wylie, Cabral and Jolly (2005) also used the spatial scan statistic by differentiating the cases by genotype. According to the authors, the underlying assumption behind genotyping is that two individuals infected by the same strain of an infectious agent are more likely to have an epidemiological link to each other than two individuals infected by a different strain.

Pearl *et al.* (2006) used the scan statistic to detect outbreaks of *Escherichia coli* O157. The study used a sequential cluster detection procedure, which starts with a purely temporal analysis followed by a purely spatial analysis for each year and finally by a spatio-temporal analysis.

The second purpose for using scan statistics for infectious disease data is prospective in nature, when continuously collected data are analyzed in real or near real time in order to quickly detect an emerging infectious disease outbreak. In most cases, a space-time scan statistic is then used. As soon as a new cluster is detected, specific actions to contain and eradicate the contaminant source of the disease or to stop the disease dynamics would be taken.

Mostashari *et al.* (2003) have proposed a surveillance system for West Nile virus through the daily reporting of dead birds by the public. The county-level density of dead birds and crows was strongly correlated with levels of West Nile virus activity in 2000, suggesting that dead bird surveillance could detect subsequent outbreaks in. Multiple dead bird reports for the same location on the same day were counted as one. Results show that in most cases, dead bird clusters not only preceded the time of collection of mosquitoes and birds that were tested positive for West Nile virus but also the reports of human cases near the cluster area.

Space-time scan statistics have also been used for syndromic surveillance, where a daily feed of automated medical health records is used for the early detection of infectious disease outbreaks [Kulldorff *et al.* (2005)].

6.6 Applications in Parasitology

Enemark *et al.* (2002), Washington *et al.* (2004), Odoi *et al.* (2004) and Reperant and Deplazes (2005) have all used spatial statistics in parasitology. A very nice subtype clustering analysis is presented by Enemark *et al.* (2002) for *Crypstosporidium parvum*, a protozoan parasite that infects the gastrointestinal tract and is recognized as a major cause of diarrhea. Washington *et al.* (2004) performed clustering analysis in sentinel sites before and after a public intervention program for the elimination of lymphatic filariasis in Haiti. After the intervention occurred, the most significant cluster was found in an area where drug coverage was low. Odoi *et al.* (2004) used the spatial scan statistic to study giardiasis in Canada and Reperant and Deplazes (2005) used it to study *Capillaria hepatica* infection in Switzerland.

6.7 Other Medical Applications

Hypoplastic left heart malformation is a congenital cardiovascular malformation. Parental exposure to various categories of solvents is correlated to the occurrence of cases in newborn children. Kuehl and Loffredo (2006) used the spatial scan statistic to search for disease clusters and evidence of industrial release of solvents in Baltimore, Maryland, United States. After geographical clusters were detected, the results were used to fit different multiple logistic regression models stratified by residence within or outside the clusters at the time of conception.

Several papers [Sankoh *et al.* (2001), George *et al.* (2001), Forand *et al.* (2002), Andrade *et al.* (2004), Ozdenerol *et al.* (2005), Ali *et al.* (2005)] have used spatial and space-time scan statistics for pediatric data. Sankoh *et al.* (2001) analyzed childhood mortality in northwest Burkina Faso (West Africa) in the 1993–1998 period. A purely spatial analysis was performed for each year of data, providing time-independent analyses that detect clusters for specific years. Their results show that a particular village was found as the most likely cluster in both the purely spatial and space-time analyses. When this village was omitted from the analysis, a new analysis was conducted which identified the previous secondary cluster as the most likely. Data exclusion is one way to focus spatial clustering away from an evident area.

Sabel *et al.* (2003) used the spatial scan statistic to detect and evaluate geographical clusters of amyotrophic lateral sclerosis in Finland. Separate analyses were done using place of birth and place of death as the geographical coordinates. The cluster found using the place of birth overlapped with the most significant cluster found using the place of death.

Using the spatial scan statistic, Ala *et al.* (2006) showed that the prevalence of primary biliary cirrhosis patients listed for transplantation was higher near a New York City superfund toxic waste site. In this particular analysis, a focused cluster analysis was also done by including the longitude and latitude of each New York City superfund site. This approach changed the center of the most significant cluster to a new location.

The spatial scan statistic has also been used for systemic sclerosis in the United States [Walsh and Fenster (1997)], lupus in the United States [Walsh and DeChello (2001)], diabetes in Canada [Green *et al.* (2003)], multiple sclerosis in Scotland [Donnan *et al.* (2005)] and asthma in the United States [Cook, Gold and Li (2007)], among many other diseases and locations. It has also been used to study the geography of alcohol and drug use [Hanson and Wieczorek (2002)] and pesticide exposure [Sudakin, Horowitz and Giffin (2002)].

6.8 Applications in Veterinary Medicine

In veterinary medicine, spatial scan statistics have been used for domestic animals as well as wildlife. Many different domestic animals have been studied, including cattle [Norström, Pfeiffer and Jarp (2000)], horses [USDA (2001)], sheep [Ward (2001); Falconi, Ochs and Deplazes (2002)], pigs [Berke and Grosse (2003)], chickens and turkeys [Guerin *et al.* (2005)], farmed salmon [Knuesel, Segner and Wahli (2003)] and dogs [Ward (2002)]. The spatial scan statistic has been especially popular for epidemiological investigations of bovine spongiform encephalopathy (mad cow disease), with studies in Switzerland [e.g. Schwermer *et al.* (2002)], France [Abrial *et al.* (2003)], Ireland [Sheridan *et al.* (2005)], Spain [Allepuz *et al.* (2007)] and the Netherlands [Heres, Brus and Hagenaars (2008)].

For wildlife data, the spatial scan statistic has been used to study various diseases among foxes in Germany [Berke *et al.* (2002)], sea otters in California [Miller *et al.* (2002)], coyotes in California [Hoar *et al.* 2003], deer in Wisconsin [Joly *et al.* (2003)] and badgers in Ireland [Olea-Popelka *et al.* (2003)]. When evaluating spatial clusters for wildlife data, a main challenge is the nonstationarity of many animals, and their ability to travel a long distance before being sampled [Hoar *et al.* 2003]. Miller *et al.* (2002) tried to detect spatial clusters of parasites in sea otters, but possibly due to high mobility, the spatial analysis did not detect any statistically significant clusters. An alternative is to sample static sources of isolates such as animal carcasses [Smith *et al.* (2000)].

In the geographical analysis of disease, it is often useful to use multiple spatial statistical methods to investigate different aspects or features of the spatial pattern. For example, in their study of acute respiratory disease in Norwegian cattle, Norström, Pfeiffer and Jarp (2000) also used the Knox test (1964) and Jacquez's k-nearest neighbor test (1996) to look at space-time interaction and a kernel-density interpolation for exploratory analysis. Sheridan *et al.* (2005) used the coordinates of major cattle feed suppliers to evaluate clusters around such prespecified locations by using a focused cluster test. Results provided evidence of association between significant clusters and feed sources.

6.9 Applications in Forestry

Coulston and Riitters (2003) and Riitters and Coulston (2005) have used the spatial scan statistic for forest data from the eastern United States. In a purely spatial analysis, the population size is the number of 0.009-ha units of forest land in a county and the number of cases are the number of units with perforated forest, which is forest located near holes in an otherwise intact and continuous forest cover. So, counties with a higher population mean more forest land, and counties with a high ratio of cases to population mean a high proportion of perforated forest. In a posterior spatial analysis, they take a previously detected primary cluster as the new study region and apply the spatial scan statistic a second time to see if there are any new smaller clusters within the old larger cluster. In this way, they have found several small clusters arranged in a linear fashion along the I-95 highway. This result shows that the primary cluster had an irregular spatial component. In another analysis, using the space-time scan statistic and 10 years of data, they defined cases as the number of units with insects or pathogens.

Tuia *et al.* (2008) used the spatial scan statistic to detect and evaluate space-time clusters of forest fires. They conclude that the 'evaluation of the presence of spatial and temporal patterns in fire occurrence and their significance could have a great impact in forthcoming studies on fire occurrences prediction'.

6.10 Applications in Geology

Conover, Bement and Iman (1979) applied the spatial scan statistic to geology data, where the aim was to detect uranium deposits by using radiation measurement taken from an airplane. As the measurements contain a fair amount of random background noise, the goal was to detect clusters of high radiation readings.

6.11 Applications in Astronomy

Astronomy would seem like a natural area of application for the three-dimensional scan statistic, but we are not aware of any such application. However, the two-dimensional scan statistic has been applied in astronomy. In a study on star formation, Marcos and Marcos (2008) used the two-dimensional scan statistic to study the spatial clustering of 'open star clusters', which are physically related groups of stars held together by mutual gravitational attraction. The 'spatial' study regions were defined by galactic longitude as the first dimension and either radial velocity, proper motion or inclination as the second dimension, in three different analyses. A number of statistically significant clusters were found.

6.12 Applications in Psychology

Margai and Henry (2003) used the spatial scan statistic to detect geographical clusters of high prevalence of learning disabilities among children in Binghamton, New York, United States. They found a statistically significant cluster in the northwestern part of the city. As a complement to the spatial scan statistic, they used Moran's I to evaluate whether there was general evidence of global spatial clustering throughout the city. They also explored a set of socio-economic variables potentially correlated to the spatial occurrence of individuals with learning disabilities. They compared the means of these variables inside and outside the detected spatial cluster through t-tests. They also applied discriminant analysis using the cluster status as the dependent variable and significant variables obtained from previous t-test analyses. This last approach represents an alternative and indirect method to associate detected geographical clusters to a set of socio-economic variables.

6.13 Applications to Accidents

Nkhoma *et al.* (2004) applied spatial scan statistics for accidental poisoning mortality data. Cases were divided according to specific toxic agents. Both spatial and space-time scan statistics were used to evaluate the data with and without the influence of a time trend. Yiannakoulias *et al.* (2003) used the spatial scan statistic to study the geography of fall injuries in the elderly.

6.14 Applications in Criminology and Warfare

Beato *et al.* (2001) used both the spatial scan statistic and Bayesian smoothing techniques to study the geographical distribution of homicides in Belo Horizonte, Brazil. Statistically significant clusters were found in areas known for drug trafficking activities. Ceccato and Haining (2004) used the spatial scan statistic to compare the location of crime events during two distinct periods in Malmö, Sweden, before and after the building of the new Öresund bridge connecting Malmö with Copenhagen, Denmark. No significant clusters were found close to the vicinity of the bridge, but there were notable shifts in the geographical locations of some clusters as well as new clusters for some of the crimes.

Priebe, Olson and Healy (1997b) have used the spatial scan statistic for minefield detection using remote sensing data.

6.15 Applications in Demography

Callado Chavez (2003) has used the spatial scan statistic to evaluate the geography of fecundity, the potential for reproduction, in Costa Rica.

6.16 Applications in the Humanities

Spatial scan statistics are not widely used in the humanities, but there are some examples from anthropology, archaeology and history. In a very interesting study, Witham and Oppenheimer (2004) used the spatial scan statistic to study the geographical distribution of excess deaths in England due to the 1783 Laki Craters volcanic eruption in Iceland, which fumigated many parts of Europe with volcanic gases and particles. They found that the eastern part of England was the most affected region. In anthropology, Usher and Allen (2005) used the scan statistic for spatial genetic analysis to evaluate kinship clusters in cemeteries. Waller (2006) used the spatial scan statistic as well as many other spatial statistical techniques to compare the geographical distribution of early versus late period archaeological sites from the Anasazi culture in Black Mesa, Arizona.

6.17 Scan Statistic Software

Different versions of the spatial scan statistic have been included in a couple of statistical software packages. The freely available SaTScan$^{\mathrm{TM}}$ software (www.satscan.org) can be used to run the purely spatial and space-time scan statistics for Poisson, Bernoulli, multinomial, normal and exponentially distributed data. ClusterSeer (www.terraseer.com) is a commercial software that includes the purely spatial and space-time scan statistics together with a number of other spatial statistical methods.

6.18 Discussion

Different types of data require different forms of the spatial scan statistic, but the underlying principle is the same as in the pioneering paper by Naus in 1965. In this chapter, we have presented a partial sample of the applications for which the spatial scan statistic has been used. As can be seen from the literature review, the spatial scan statistic has been applied in a remarkable number of different subject areas, from the small spatial scale of medical imaging to the large spatial scale of astronomy. The method is most commonly used in cancer, infectious disease and veterinary epidemiology. These are areas with a long and strong interest in epidemiology in general. They are also areas with a long tradition of disease cluster and outbreak investigations, for which the spatial scan statistic is ideally suited.

The spatial scan statistic is increasingly being used for other diseases as well. The number of applications in non-medical areas is more limited, but we think that may change with time. With the increasing use of geographical information systems in many different disciplines, there will be an increase in the use of formal methods of statistical inference to complement the beautiful maps that are created. Areas for which we think that the spatial scan statistic will play an especially important role include archaeology, astronomy, criminology, demography, ecology, geography and medical imaging.

The spatial scan statistic is also used in ways that do not lead to publications in scientific journals. For example, many public health officials use it for routine disease surveillance on a daily, weekly or yearly basis to monitor the geographical distribution of disease. Likewise, spatial scan statistics are used by law enforcement agencies for the routine monitoring of crime activities.

Acknowledgment

This research was funded by grant #RO1CA095979 from the National Cancer Institute.

References

1. Abrial, D., Calavas, D., Lauvergne, N., Morignat, E. and Ducrot, C. (2003). Descriptive spatial analysis of BSE in western France, *Veterinary Research*, **34**, 749–760.

2. Ala, A., Stanca, C.M., Bu-Ghanim, M., Ahmado, I., Branch, A.D., Schiano, T.D., Odin, J.A. and Bach, N. (2006). Increased prevalence of primary biliary cirrhosis near superfund toxic waste sites, *Hepatology*, **43**, 525–531.

3. Ali, M., Asefaw, T., Byass, P., Beyene, H. and Karup Pedersen, F. (2005). Helping northern Ethiopian communities reduce childhood mortality: population-based intervention trial, *Bulletin of the World Health Organization*, **83**, 27–33.

4. Allepuz, A., López-Quílez, A., Forte, A., Fernández, G. and Casal, J. (2007). Spatial analysis of bovine spongiform encephalopathy in Galicia, Spain (2000-2005), *Preventive Veterinary Medicine*, **79**, 174–185.

5. Alm, S.E. (1997). On the distributions of scan statistics of a two dimensional Poisson process, *Adv. in Appl. Probab.*, **29**, 1–18.

6. Alm, S.E. (1998). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions, *Extremes*, **1**, 111–126.

7. Andrade, A.L., Silva, S.A., Martelli, C.M., Oliveira, R.M., Morais Neto, O.L., Siqueira Junior, J.B., Melo, L.K. and Di Fabio, J.L. (2004). Population-based surveillance of pediatric pneumonia: use of spatial analysis in an urban area of central Brazil, *Cadernos de Saúde Pública*, **20**, 411–421.

8. Assunção, R., Costa, M.A., Tavares, A. and Ferreira, S. (2006). Fast detection of arbitrarily shaped disease clusters, *Statistics in Medicine*, **25**:5, 723–742.

9. Beato Filho, C.C., Assunção, R.M., Silva, B.F., Marinho, F.C., Reis, I.A. and Almeida, M.C. (2001). Homicide clusters and drug traffic in Belo Horizonte, Minas Gerais, Brazil from 1995 to 1999, *Cadernos de Saúde Pública*, **17**, 1163–1171.

10. Berke, O. and Grosse Beilage, E. (2003). Spatial relative risk mapping of pseudorabies-seropositive pig herds in an animal-dense region, *Journal of Veterinary Medicine*, **B50**:4, 322–325.

11. Berke, O., von Keyserlingk, M., Broll, S. and Kreienbrock, L. (2002). On the distribution of *Echinococcus multilocularis* in red foxes in Lower Saxony: identification of a high risk area by spatial epidemiological cluster analysis. *Berliner und Munchener Tierarztliche Wochenschrift*, **115**, 428–434.

12. Buntinx, F., Geys, H., Lousbergh, D., Broeders, G., Cloes, E., Dhollander, D., Op De Beeck, L., Vanden Brande, J., Van Waes, A. and Molenberghs, G. (2003). Geographical differences in cancer incidence in the Belgian province of Limburg, *European Journal of Cancer*, **39**, 2058–2072.

13. Callado Chavez, A. (2003). Fecundidad adolescente en el gran área metropolitana de Costa Rica, *Población y Salud en Mesoamérica*, **1**, 4.

14. Ceccato, V. and Haining, R. (2004). Crime in border regions: The Scandinavian case of Öresund, 1998-2001, *Annals of the Association of American Geographers*, **94**, 807–826.

15. Chaput, E.K., Meek, J.I. and Heimer, R. (2002). Spatial analysis of human granulocytic ehrlichiosis near Lyme, Connecticut, *Emerging Infectious Diseases*, **8**, 943–948.

16. Chen, J. and Glaz, J (1996). Two-dimensional discrete scan statistics, *Statist. Probab. Lett.*, **23**, 751–771.

17. Conover, W.J., Bement, T.R. and Iman, R.L. (1979). On a method for detecting clusters of possible uranium deposits, *Technometrics*, **21**, 277–282.

18. Cook, A.J., Gold, D.R. and Li, Y. (2007). Spatial cluster detection for censored outcome data, *Biometrics*, **63**, 540–549.

19. Coulston, J.W. and Riitters, K.H. (2003). Geographic analysis of forest health indicators using spatial scan statistics, *Environmental Management*, **31**, 764–773.

20. Cousens, S., Smith, P.G., Ward, H., Everington, D., Knight, R.S.G., Zeidler, M., Stewart, G., Smith-Bathgate, E.A.B., Macleod, M.A., Mackenzie, J. and Will, R.G. (2001). Geographical distribution of variant Creutzfeldt-Jakob disease in Great Britain, *The Lancet*, **357**, 1002–1007.

21. Donnan, P.T., Parratt, J.D.E., Wilson, S.V., Forbes, R.B., O'Riordan, J.I. and Swingler, R.J. (2005). Multiple sclerosis in Tayside, Scotland: detection of clusters using a spatial scan statistic, *Multiple Sclerosis*, **11**, 403–408.

22. Duczmal, L. and Assunção, R.A. (2004). Simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics and Data Analysis*, **45**, 269–286.

23. Enemark, H.L., Ahrens, P., Juel, C.D., Petersen, E., Petersen, R.F., Andersen, J.S., Lind, P. and Thamsborg, S.M. (2002). Molecular characterization of Danish *Cryptosporidium parvum* isolates, *Parasitology*, **125**, 331–341.

24. Falconi, F., Ochs, H. and Deplazes, P. (2002). Serological cross-sectional survey of psoroptic sheep scab in Switzerland, *Veterinary Parasitology*, **109**, 119–127.

25. Fevre, E.M., Coleman, P.G., Odiit, M., Magona, J.W., Welburn, S.C. and Woolhouse, M.E.J. (2001). The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda, *The Lancet*, **358**, 625–628.

26. Forand, S.P., Talbot, T.O., Druschel, C. and Cross, P.K. (2002). Data quality and the spatial analysis of disease rates: congenital malformations in New York State, *Health and Place*, **8**, 191–199.

27. Fukuda, Y., Umezaki, M., Nakamura, K. and Takano, T. (2005). Variations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan, *International Journal of Health Geographics*, **4**, 16.

28. George, M., Wiklund, L., Aastrup, M., Pousette, J., Thunholm, B., Saldeen, T., Wernroth, L., Zaren, B. and Holmberg, L. (2001). Incidence and geographical distribution of sudden infant death syndrome in relation to content of nitrate in drinking water and groundwater levels, *European Journal of Clinical Investigation*, **31**, 1083–1094.

29. Green, C., Hoppa, R.D., Young, T.K. and Blanchard, J.F. (2003). Geographic analysis of diabetes prevalence in an urban area, *Social Science and Medicine*, **57**, 551–560.

30. Gregorio, D.I., Kulldorff, M., Barry, L., Samociuk, H. and Zarfos, K. (2001). Geographic differences in primary therapy for early stage breast cancer. *Annals of Surgical Oncology*, **8**, 844–849.

31. Gregorio, D.I., Kulldorff, M., Barry, L. and Samociuk, H. (2002). Geographic differences in invasive and in situ breast cancer incidence according to precise geographic coordinates, Connecticut, 1991–1995. *International Journal of Cancer*, **100**, 194–198.

32. Guerin, M.T., Martin, S.W., Darlington, G.A. and Rajic, A. (2005). A temporal study of *Salmonella* serovars in animals in Alberta between 1990 and 2001, *Canadian Journal of Veterinary Research*, **69**, 88–89.

33. Han, D.W., Rogerson, P.A., Nie, J., Bonner, M.R., Vena, J.E., Vito, D., Muti, P., Trevisan, M., Edge, S.B. and Freudenheim, J.L. (2004). Geographic clustering of residence in early life and subsequent risk of breast cancer (United States), *Cancer Causes and Control*, **15**, 921–929.

34. Hanson, C.E. and Wieczorek, W.F. (2002). Alcohol mortality: a comparison of spatial clustering methods, *Social Science and Medicine*, **55**, 791–802.

35. Heres, L., Brus, D.J. and Hagenaars, T.J. (2008). Spatial analysis of BSE cases in the Netherlands, *BMC Veterinary Research*, **4**:21.

36. Hjalmars, U., Kulldorff, M., Gustafsson, G. and Nagarwalla, N. (1996). Childhood leukemia in Sweden: using GIS and a spatial scan statistic for cluster detection, *Statistics in Medicine*, **15**, 707–715.

37. Hoar, B.R., Chomel, B.B., Rolfe, D.L., Chang, C.C., Fritz, C.L., Sacks, B.N. and Carpenter, T.E. (2003). Spatial analysis of *Yersinia pestis* and *Bartonella vinsonii* subsp. *berkhoffii* seroprevalence in California coyotes (*Canis latrans*), *Preventive Veterinary Medicine*, **56**, 299–311.

38. Hsu, C.E., Jacobson, H.E. and Soto Mas, F. (2004). Evaluating the disparity of female breast cancer mortality among racial groups - a spatiotemporal analysis, *International Journal of Health Geographics*, **3**:4.

39. Huang, L., Kulldorff, M. and Gregorio, D. (2007). A spatial scan statistic for survival data, *Biometrics*, **63**, 109–118.

40. Huang, L., Tiwari, R., Zuo, J., Kulldorff, M. and Feuer, E. (2009). Weighted normal spatial scan statistic for heterogenous population data, *Journal of the American Statistical Association*, in press.

41. Huillard d'Aignaux, J., Cousens, S.N., Delasnerie-Laupretre, N., Brandel, J.P., Salomon, D., Laplanche, J.L., Hauw, J.J. and Alperovitch, A. (2002). Analysis of the geographical distribution of sporadic Creutzfeldt-Jakob disease in France between 1992 and 1998, *International Journal of Epidemiology*, **31**, 490–495.

42. Jacquez, G.M. (1996). A *k*-nearest neighbour test for space-time interaction, *Statistics in Medicine*, **15**:18, 1935–1949.

43. Joly, D.O., Ribic, C.A., Langenberg, J.A., Beheler, K., Batha, C.A., Dhuey, B.J., Rolley, R.E., Bartelt, G., Van Deelen, T.R. and Samual, M.D. (2003). Chronic wasting disease in free-ranging Wisconsin white-tailed deer, *Emerging Infectious Disease*, **9**, 599–601.

44. Jung, I., Kulldorff, M. and Klassen, A. (2007). A spatial scan statistic for ordinal data, *Statistics in Medicine*, **26**, 1594–1607.

45. Klassen, A., Kulldorff, M. and Curriero, F. (2005). Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors, *International Journal of Health Geographics*, **4**, 1.

46. Kleinman K., Abrams A., Kulldorff M. and Platt R. (2005). A model-adjusted space-time scan statistic with an application to syndromic surveillance, *Epidemiology and Infection*, **133**, 409–419.

47. Knox, E.G. (1964). The detection of spacetime interactions, *J. Appl. Stat.*, **13**, 24–30.

48. Knuesel, R., Segner, H. and Wahli, T. (2003). A survey of viral diseases in farmed and feral salmonids in Switzerland., *Journal of Fish Diseases*, **26**:4, 167–182.

49. Kuehl, K.S. and Loffredo, C.A. (2006). A cluster of hypoplastic left heart malformation in Baltimore, Maryland, *Pediatric Cardiology*, **27**, 25–31.

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

51. Kulldorff, M., Feuer, E.J., Miller, B.A. and Freedman, L.S. (1997). Breast cancer in northeastern United States: a geographical analysis, *American Journal of Epidemiology*, **146**, 161–170.

52. Kulldorff M., Athas W., Feuer E., Miller B. and Key C. (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, *American Journal of Public Health*, **88**, 1377–1380.

53. Kulldorff M. (2001). Prospective time-periodic geographical disease surveillance using a scan statistic, *Journal of the Royal Statistical Society*, **A164**, 61–72.

54. Kulldorff M., Heffernan R., Hartman J., Assunção R.M. and Mostashari F. (2005). A space-time permutation scan statistic for the early detection of disease outbreaks, *PLoS Medicine*, **2**, 216–224.

55. Kulldorff. M., Huang, L. and Konty, K. (2008). A spatial scan statistic for normally distributed data, *Manuscript*.

56. Loader, C.R. (1991). Large-deviation approximations to the distribution of scan statistics, *Adv. in Appl. Probab.*, **23**, 751–771.

57. Marcos, R.D.L.F. and Marcos, C.D.L.F. (2008). From star complexes to the field: open cluster families, *Astrophysical Journal*, **672**, 342–351.

58. Margai, F. and Henry, N. (2003). A community-based assessment of learning disabilities using environmental and contextual risk factors, *Social Science and Medicine*, **56**, 1073–1085.

59. Miller, M.A., Gardner, I.A., Kreuder, C., Paradies, D.M., Worcester, K.R., Jessup, D.A., Dodd, E., Harris, M.D., Ames, J.A., Packham, A.E. and Conrad, P.A. (2002). Coastal freshwater runoff is a risk factor for *Toxoplasma gondii* infection of southern sea otters (*Enhydra lutris nereis*), *International Journal for Parasitology*, **32**, 997–1006.

60. Mostashari, F., Kulldorff, M., Hartman, J.J., Miller, J.R. and Kulasekera, V. (2003). Dead bird clustering: a potential early warning system for West Nile virus activity, *Emerging Infectious Diseases*, **9**, 641–646.

61. Naiman, D.Q. and Priebe, C.E. (2001). Computing scan statistic p-values using importance sampling, with applications to genetics and medical image analysis, *Journal of Computational & Graphical Statistics*, **10**, 296–328.

62. Naus, J. I. (1965). Clustering of random points in two dimensions, *Biometrika*, **52**, 263–267.

63. Norström, M., Pfeiffer, D.U. and Jarp, J. (2000). A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds, *Preventive Veterinary Medicine*, **47**, 107–119.

64. Nkhoma, E.T., Hsu, C.E., Hunt, V.I. and Harris A.M. (2004). Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980 - 2001, *International Journal of Health Geographics*, **3**:25.

65. Odoi, A., Martin, S.W., Michel, P., Middleton, D., Holt, J. and Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial scan statistic, *International Journal of Health Geographics*, **3**:11.

66. Olea-Popelka, F.J., Griffin, J.M., Collins, J.D., McGrath, G. and Martin, S.W. (2003). Bovine tuberculosis in badgers in four areas in Ireland: does tuberculosis cluster? *Preventive Veterinary Medicine*, **59**, 103–111.

67. Ozdenerol, E., Williams, B.L., Kang, S.Y. and Magsumbol, M.S. (2005). Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters, *International Journal of Health Geographics*, **4**:19.

68. Patil, G.P. and Taillie, C. (2003). Geographic and network surveillance via scan statistics for critical area detection, *Statistical Science*, **18**:4, 457–465.

69. Patil, G.P. and Taillie, C. (2004). Upper level set scan statistic for detecting arbitrarily shaped hotspots, *Environmental and Ecological Statistics*, **11**, 183–197.

70. Pearl, D.L., Louie, M., Chui, L., Dore, K., Grimsrud, K.M., Leedell, D., Martin, S.W., Michel, P., Svenson, L.W. and McEwen, S.A. (2006). The use of outbreak information in the interpretation of clustering of reported cases of *Escherichia coli* O157 in space and time in Alberta, Canada, 2000-2002, *Epidemiology and Infection*, **134**, 699–711.

71. Popescu, L.M. and Lewitt, R.M. (2006a). Comparison between TOF and non-TOF PET using a scan statistic numerical observer, In *2006 IEEE Nuclear Science Symposium Conference Record*, **3**, 1774–1780.

72. Popescu, L.M. and Lewitt, R.M. (2006b). Small nodule detectability evaluation using a generalized scan statistic model, *Physics in Medicine and Biology*, **51**, 6225–6244.

73. Priebe, C. E., Olson, T. and Healy D.M. Jr (1997a). A spatial scan statistic for stochastic scan partitions, *Journal of the American Statistical Association*, **92**, 1476–1484.

74. Priebe, C. E., Olson, T. and Healy D.M. Jr (1997b). Exploiting stochastic partitions for minefield detection. *Proceedings of SPIE, the International Society for Optical Engineering*, **3079**, 508.

75. Reperant, L.A. and Deplazes, P. (2005). Cluster of *Capillaria hepatica* infections in non-commensal rodents from the canton of Geneva, Switzerland, *Parasitology Research*, **96**, 340–342.

76. Riitters, K.H. and Coulston, J.W. (2005). Hot spots of perforated forest in the eastern United States, *Environmental Management*, **35**, 483–492.

77. Roche, L.M., Skinner, R. and Weinstein, R.B. (2002). Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer, *Journal of Public Health Management and Practice*, **8**, 26–32.

78. Sabel, C.E., Boyle, P.J., Lytnen, M., Gatrell, A.C., Jokelainen, M., Flowerdew, R. and Maasilta P. (2003). Spatial clustering of amyotrophic lateral sclerosis in Finland at place of birth and place of death, *American Journal of Epidemiology*, **157**, 898–905.

79. Sankoh, O.A., Ye, Y., Sauerborn, R., Muller, O. and Becher, H. (2001). Clustering of childhood mortality in rural Burkina Faso, *International Journal of Epidemiology*, **30**, 485–492.

80. Sauders, B.D., Fortes, E.D., Morse, D.L., Dumas, N., Kiehlbauch, J.A., Schukken, Y., Hibbs, J.R. and Wiedmann, M. (2003). Molecular subtyping to detect human listeriosis clusters, *Emerging Infectious Diseases*, **9**, 672–680.

81. Schwermer, H., Rufenacht, J., Doherr, M.G. and Heim, D. (2002). Geographic distribution of BSE in Switzerland, *Schweizer Archiv fur Tierheilkunde*, **144**, 701–708.

82. Sheehan, T.J., DeChello, L.M., Kulldorff, M., Gregorio, D.I., Gershman, S. and Mroszczyk, M. (2004). The geographic distribution of breast cancer incidence in Massachusetts 1988-1997, adjusted for covariates, *International Journal of Health Geographics*, **3**, 17.

83. Sheehan, T.J. and DeChello, L.M. (2005). A space-time analysis of the proportion of late stage breast cancer in Massachusetts, 1988 to 1997, *International Journal of Health Geographics*, **4**, 15.

84. Sheridan, H.A., McGrath, G., White, P., Fallon, R., Shoukri, M.M. and Martin, S.W. (2005). A temporal-spatial analysis of bovine spongiform encephalopathy in Irish cattle herds, from 1996 to 2000, *Canadian Journal of Veterinary Research*, **69**, 19–25.

85. Smith, K.L., DeVos, V., Bryden, H., Price, L.B., Hugh-Jones, M.E. and Keim, P. (2000). *Bacillus anthracis* diversity in Kruger National Park, *Journal of Clinical Microbiology*, **38**, 3780–3784.

86. Sudakin, D.L., Horowitz, Z. and Giffin, S. (2002). Regional variation in the incidence of symptomatic pesticide exposures: applications of geographic information systems, *Journal of Toxicology - Clinical Toxicology*, **40**, 767–773.

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

88. Thomas, A.J. and Carlin, B.P. (2003). Late detection of breast and colorectal cancer in Minnesota counties: an application of spatial smoothing and clustering, *Statistics in Medicine*, **22**, 113–127.

89. Tuia, D., Ratle, F., Lasaponara, R., Telesca, L. and Kanevski, M. (2008). Scan statistics analysis of forest fire clusters. *Communications in Nonlinear Science and Numerical Simulation*, **13**, 1689–1694.

90. Turnbull, B., Iwano, E.J., Burnett, W.S., Howe, H.L. and Clark, L.C. (1990). Monitoring for clusters of disease: application to leukemia incidence in upstate New York, *Amer. J. Epidemiology*, **132**, 136–143.

91. United States Department of Agriculture. (2001). West Nile virus in equids in the Northeastern United States in 2000. USDA, APHIS, Veterinary Services.

92. Usher, B.M. and Allen, K.L. (2005). Identifying kinship clusters: SatScan for genetic spatial analysis, *American Journal of Physical Anthropology, Supplement*, **126**, S40, 210.

93. Viel, J.F., Arveux, P., Baverel, J. and Cahn, J.Y. (2000). Soft-tissue sarcoma and non-Hodgkins lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels, *American Journal of Epidemiology*, **152**, 13–19.

94. Waller, L.A. (2006). Detection of Clustering in Spatial Data. Emory University, Department of Biostatistics, Technical Report **06-12**.

95. Walsh, S.J. and Fenster, J.R. (1997). Geographical clustering of mortality from systemic sclerosis in the Southeastern United States, *Journal of Rheumatology*, **24**, 2348–2352.

96. Walsh, S.J. and DeChello, L.M. (2001). Geographical variation in mortality from systemic lupus erythematosus in the United States, *Lupus*, **10**, 637–646.

97. Ward, M.P. (2001). Blowfly strike in sheep flocks as an example of the use of a time-space scan statistic to control confounding, *Preventive Veterinary Medicine*, **49**, 61–69.

98. Ward, M.P. (2002). Clustering of reported cases of leptospirosis among dogs in the United States and Canada, *Preventive Veterinary Medicine*, **56**, 215–226.

99. Washington, C.H., Radday, J., Streit, T.G., Boyd, H.A., Beach, M.J., Addiss, D.G., Lovince, R., Lovegrove, M.C., Lafontant, J.G., Lammie, P.J. and Hightower, A.W. (2004). Spatial clustering of filarial transmission before and after a Mass Drug Administration in a setting of low infection prevalence, *Filaria Journal*, **3**, 3.

100. Witham, C.S. and Oppenheimer, C. (2004). Mortality in England during the 1783-4 Laki Craters eruption, *Bulletin of Volcanology*, **67**, 15–25.

101. Wylie, J.L., Cabral T. and Jolly, A.M. (2005). Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis, *Journal of Infectious Diseases*, **191**, 899–906.

102. Yiannakoulias, N., Rowe, B.H., Svenson, L.W., Schopflocher, D.P., Kelly, K. and Voaklander, D.C. (2003). Zones of prevention: the geography of fall injuries in the elderly, *Social Science and Medicine*, **57**, 2065–2073.

103. Yoshida, M., Naya, Y. and Miyashita, Y. (2003). Anatomical organization of forward fiber projections from area TE to perirhinal neurons representing visual long-term memory in monkeys, *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4257–4262.