J. R. Statist. Soc. A (2001)
164, *Part* 1, *pp.* 61–72

Prospective time periodic geographical disease surveillance using a scan statistic

Martin Kulldorff

University of Connecticut, Farmington, USA

[Received January 2000. Revised September 2000]

Summary. Most disease registries are updated at least yearly. If a geographically localized health hazard suddenly occurs, we would like to have a surveillance system in place that can pick up a new geographical disease cluster as quickly as possible, irrespective of its location and size. At the same time, we want to minimize the number of false alarms. By using a space–time scan statistic, we propose and illustrate a system for regular time periodic disease surveillance to detect any currently 'active' geographical clusters of disease and which tests the statistical significance of such clusters adjusting for the multitude of possible geographical locations and sizes, time intervals and time periodic analyses. The method is illustrated on thyroid cancer among men in New Mexico 1973–1992.

Keywords: Chronic disease surveillance; Geographical clusters; Infectious disease surveillance; New Mexico; Space–time clustering; Spatial statistics; Thyroid cancer

1. Introduction

A common form of geographical disease surveillance is to select a disease, a geographical region and a fixed time period for mapping disease rates. In addition to the most basic maps with the observed rates depicted for non-overlapping geographical areas, several statistical methods have been proposed both for disease mapping (Choynowski, 1959; Clayton and Kaldor, 1987; Besag *et al.*, 1991; Bernardinelli and Montomoli, 1992; Lawson and Williams, 1993; Martuzzi and Elliott, 1996; Kafadar, 1996; Rushton and Lolonis, 1996) and for testing whether an observed pattern is probably due to chance (Cliff and Ord, 1973; Turnbull *et al.*, 1990; Cuzick and Edwards, 1990; Besag and Newell, 1991; Diggle and Chetwynd, 1991; Walter, 1994; Tango, 1995; Martuzzi and Hills, 1995; Kulldorff, 1997; Bithell, 1999; Kulldorff, 1999; Tango, 2000). The former utilizes spatial smoothing techniques whereas the latter are referred to as tests for spatial randomness.

Many countries and regions maintain disease registries, most commonly for cancer and birth defects, and many national and local health departments continuously collect information about occurrences of different infectious diseases. Incident cases are typically added to a disease registry daily, monthly or yearly, with the timescale depending on the type of disease and the capacity of the registry system. To detect emerging geographical clusters due to some suddenly occurring risk factor quickly can be of great public health importance. It can be used to detect previously unknown diseases, previously unknown risk factors for known diseases or the previously unknown local existence of known risk factors.

Address for correspondence: Martin Kulldorff, Department of Community Medicine and Health Care, Division of Biostatistics, University of Connecticut School of Medicine, Farmington, CT 06030-6325, USA.
E-mail: MartinK@neuron.uchc.edu

 0964–1998/01/164061

For a well-defined geographical area, there are several temporal statistical techniques for monitoring and detecting a sudden temporal increase in the risk of disease (Bjerkedal and Bakketeig, 1975; Weatherall and Haskey, 1976; Chen, 1978, 1979; Chen *et al.*, 1982; Radaelli, 1996). If an outbreak is geographically localized though, a sudden and important increase in one area may be hidden as part of the aggregated data for the geographical region as a whole. One way to overcome this problem is to divide the larger geographical region into several smaller areas, and to monitor each of these areas individually. With many areas though, there is multiple testing, and we would be overwhelmed with false alarms if the nominal significance level is used. More importantly, if an outbreak occurs on the border between areas or in only a small part of an area, an important outbreak may be missed simply because it did not conform to the predefined geographical boundaries.

The one-dimensional scan statistic has long been used for purely temporal disease surveillance (Wallenstein, 1980; Weinstock, 1981), and a spatial scan statistic has been used for purely geographical surveillance (Kulldorff, 1997). In this paper, a time periodic geographical disease surveillance system is proposed, based on a space–time scan statistic. In this system, the statistical inference is adjusted for the multiple testing arising from many possible geographical locations and sizes of disease clusters. Except for the limits imposed by the aggregation level for which data are available, the method can detect spatial clusters irrespective of any predefined geographical boundaries, by combining any number of close locations into the same cluster. The method strives to detect only those clusters that are still 'alive', meaning that the excess risk is still present during the last time period for which data are available. It can detect clusters irrespective of the cluster's 'age', including long-existing clusters with moderately excess risk as well as recently emerged clusters with a high excess risk. The statistical inference adjusts for the many possible time lengths. Finally, as there is also multiple testing due to the repeated time periodic analyses, the surveillance system is constructed so that it adjusts for the multitude of analyses that have already been performed.

The method proposed in this paper should not be confused with traditional space–time disease clustering methods (Knox, 1964; Ederer *et al.*, 1964; Mantel, 1967; Wallenstein *et al.*, 1989; Jacquez, 1996; Kulldorff *et al.*, 1998a). These are retrospective in nature, designed to test whether a disease is randomly distributed over space and time for a predefined geographical region during a predetermined time period, whereas the currently proposed method is prospective in nature, with repeated time periodic analyses (Rogerson, 1997).

Thyroid cancer incidence data for men in New Mexico are used to illustrate the time periodic geographical disease surveillance system. All analyses are adjusted for age by using indirect standardization. *p*-values are based on Monte Carlo hypothesis testing (Dwass, 1957) using 999 random replications of the data under the null hypothesis of spatial randomness. All analyses were done using the publicly and freely available software SaTScan (Kulldorff *et al.*, 1998b), developed at the National Cancer Institute.

2. Illustrative data set: male thyroid cancer in New Mexico

For illustration, we use male thyroid incidence data from New Mexico during 1973–1992, collected by the New Mexico Tumor Registry for the National Cancer Institute's 'Surveillance, epidemiology and end results program'. The data are geographically aggregated to the 32 counties that existed in 1973. A total of 333 cases were reported. The incidence rate was 2.4 per 100 000. As the denominator we use the official age-specific population estimates for each individual year and for each county. These are based on the decennial US census

counts in conjunction with estimates of births, deaths and migration. The total male population increased from 546 000 in 1973 to 779 000 in 1992.

Thyroid cancer is rare. The annual incidence in the USA is estimated at 5.2 cases per 100 000 people. The incidence is lower for men than for women, with 2.9 and 7.4 cases per 100 000 respectively. The lifetime risk of being diagnosed with thyroid cancer is 0.26% for men and 0.72% for women (Ries *et al.*, 1999). In an international comparison of places with available data, the age-adjusted rates were highest in Hawaii and Iceland, and lowest in Bombay, England and Wales (Parkin *et al.*, 1992). Known risk factors include the exposure to ionizing radiation during childhood (Duffy and Fitzgerald, 1950; Ron *et al.*, 1989), radiation treatment for benign head and neck conditions (Shore *et al.*, 1985), radioactivity from nuclear explosions (Robbins and Adams, 1989; Prentice *et al.*, 1982; Kazakov *et al.*, 1992) and work as a radiology technician (Boice *et al.*, 1992) or X-ray operator (Carstensen *et al.*, 1990).

With a 5-year relative survival rate of 95% thyroid cancer is one of the least fatal forms of cancer. The annual mortality rates in the USA are about the same for men and women, with 0.3 deaths per 100 000 during 1992–1996. Among the 50 states and the District of Colombia, and for the years 1992–1996, New Mexico had the 21st highest age-adjusted male mortality rate with 0.3 deaths per 100 000 (Ries *et al.*, 1999).

3. Purely geographical surveillance

Subsequent theory builds on the purely spatial scan statistic (Kulldorff, 1997), and we shall first briefly describe this method and apply it to the New Mexico thyroid cancer data for different time periods to facilitate comparisons.

3.1. The spatial scan statistic

The spatial scan statistic has been developed to test for geographical clusters and to identify their approximate location (Kulldorff, 1997). The number of events, e.g. incident cases, cases with a late stage breast cancer diagnosis or deaths, may be assumed to be either Poisson or Bernoulli distributed. The Poisson model should be used if for example the denominator reflects person-years lived during a 5-year period, and we are interested in the geographical distribution of incident breast cancer cases during these 5 years, adjusting for the population at risk. The Bernoulli model should be used if the denominator consists of specific individuals, such as all incident breast cancer cases, and we are interested in the geographical distribution of late stage incident cases adjusting for the geographical distribution of all cases. Depending on the availability of data, the spatial scan statistic can be used for either aggregated data such as census areas or for the special case of precise geographical coordinates, where each 'census area' contains only one person at risk.

The spatial scan statistic imposes a circular window on the map and lets the centre of the circle move over the area so that at different positions the window includes different sets of neighbouring census areas. If the window contains the centroid of a census area, then that whole area is included in the window. For each circle centroid, the radius of the circular window is varied continuously from 0 up to a maximum radius so that the window never includes more than 50% of the total population at risk. In this way, the circular window is flexible both in location and in size. In total the method creates a very large number of distinct circular windows, each with a different set of neighbouring counties within it, and each a possible candidate for containing a cluster of events.

Conditioning on the observed total number of cases, N, the definition of the spatial scan statistic S is the maximum likelihood ratio over all possible circles Z,

$$S = \frac{\max_Z\{L(Z)\}}{L_0} = \max_Z \left\{ \frac{L(Z)}{L_0} \right\}, \tag{1}$$

where $L(Z)$ is the maximum likelihood for circle Z, expressing how likely the observed data are given a differential rate of events within and outside the zone, and where L_0 is the likelihood function under the null hypothesis.

Let n_Z be the number of cases in circle Z. For the Poisson model, let $\mu(Z)$ be the expected number under the null hypothesis, so that $\mu(A) = N$ for A, the total region under study. It can then be shown that

$$\frac{L(Z)}{L_0} = \left\{ \frac{n_Z}{\mu(Z)} \right\}^{n_Z} \left\{ \frac{N - n_Z}{N - \mu(Z)} \right\}^{N - n_Z} \tag{2}$$

if $n_Z > \mu(Z)$ and $L(Z)/L_0 = 1$ otherwise. Details, including derivations as likelihood ratio tests, have been given elsewhere (Kulldorff, 1997).

As this likelihood ratio is maximized over all the circles, it identifies the circle that constitutes the most likely cluster. Its p-value is obtained through Monte Carlo hypothesis testing (Dwass, 1957).

The spatial scan statistic has the following features, which make it particularly suitable as a tool for surveillance purposes:

(a) it adjusts both for the inhomogeneous population density and for any number of confounding variables;
(b) by searching for clusters without specifying their size or location the method ameliorates the problem of preselection bias;
(c) the likelihood-ratio-based test statistic takes multiple testing into account and delivers a single p-value for the test of the null hypothesis;
(d) if the null hypothesis is rejected, we can specify the approximate location of the cluster that caused the rejection.

In addition to the most likely cluster, the method identifies secondary clusters in the data set and can order them according to their likelihood ratio. There will always be sets of census areas that overlap in part with the most likely cluster and that have almost as high a likelihood, since adding or subtracting a few areas does not normally change the likelihood greatly. We do not report on all clusters of this type since most of them provide little additional information, but their existence means that, although we can pin-point the general location of a cluster, its exact boundaries must remain uncertain. Thus we always refer to a cluster's 'approximate location'.

The spatial scan statistic has been used for different health events such as the incidence of leukaemia (Hjalmars *et al.*, 1996), systemic sclerosis mortality (Walsh and Fenster, 1997), late stage breast cancer (Sheehan *et al.*, 2000) and the incidence of soft tissue sarcoma (Viel *et al.*, 2000).

3.2. Purely spatial analysis of the New Mexico data

Suppose that we travel back in time, and after collecting the 1978 data we do a purely spatial analysis of the 1973–1978 data using the spatial scan statistic. During this time period 99 male

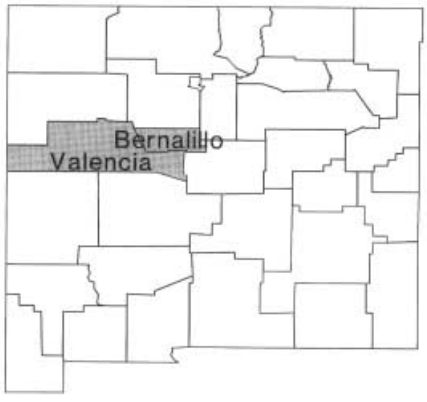

Fig. 1. Most likely cluster of male thyroid cancer in New Mexico for the time period 1973–1978, using a purely spatial age-adjusted analysis: the cluster consists of Bernalillo and Valencia counties ($p = 0.44$)

thyroid cancer cases were detected in New Mexico. The most likely cluster consists of Bernalillo and Valencia counties in the upper central part of the state, including Albuquerque, the largest city in New Mexico (Fig. 1). In this area 46 cases were observed whereas 34.6 were expected, with $p = 0.441$ (Table 1). Our conclusion is hence that, for this time period, there is no evidence for a geographically differential risk of thyroid cancer in New Mexico.

However, in a cancer registry, new data are collected and added to the data files yearly. It would be simple to repeat the purely spatial analysis each year with the additional data added together with the corresponding updated population numbers. The results of such yearly analyses are given in Table 1.

Table 1. Purely spatial analyses of male thyroid cancer in New Mexico, by using the purely spatial scan statistic†

Data years	*Most likely cluster‡*	*Cluster years*	*Cases*	*Expected*	*Relative risk*	*p*
1973–1978	Bernalillo, Valencia	1973–1978	46	35	1.3	0.44
1973–1979	Bernalillo, Valencia	1973–1979	52	40	1.3	0.48
1973–1980	Bernalillo, Valencia	1973–1980	58	45	1.3	0.39
1973–1981	Bernalillo, Valencia	1973–1981	68	51	1.3	0.10
1973–1982	Bernalillo, Valencia	1973–1982	79	58	1.4	0.04
1973–1983	Bernalillo, Valencia	1973–1983	84	62	1.4	0.04
1973–1984	North-central counties	1973–1984	113	90	1.3	0.05
1973–1985	North-central counties − San Miguel	1973–1985	115	95	1.2	0.18
1973–1986	North-central counties + Colfax, Harding	1973–1986	129	108	1.2	0.16
1973–1987	North-central counties + Colfax, Harding	1973–1987	142	117	1.2	0.05
1973–1988	North-central counties − San Miguel	1973–1988	143	115	1.2	0.02
1973–1989	North-central counties + Colfax, Harding	1973–1989	165	134	1.2	0.01
1973–1990	North-central counties + Torrance	1973–1990	174	144	1.2	0.02
1973–1991	North-central counties + Colfax, Harding	1973–1991	186	155	1.2	0.02
1973–1992	North-central counties + Torrance	1973–1992	199	164	1.2	0.01

†The *p*-values are not adjusted for the repeated time periodic analyses conducted since 1978.
‡The north-central counties are Bernalillo, Los Alamos, Mora, Rio Arriba, Sandoval, San Miguel, Santa Fe and Taos.

3.3. Limitations of repeated purely spatial analyses

There are two problems with frequently repeating the purely spatial analyses as part of a time periodic surveillance system. Firstly, we have low power to detect emerging clusters quickly. If a true excess risk is only present during the last few years, then the fact that we are doing a purely spatial analysis for the whole time period dilutes the strength of the cluster by including random fluctuations of the rate for earlier years when the risk was not high. Secondly, while we are adjusting for the multiple testing stemming from many possible cluster locations and cluster sizes, we are not adjusting for the multiple testing due to repeated analyses every year. For example, of the 15 analyses presented in Table 1 some are statistically significant whereas others are not, making it difficult to interpret the results.

We propose a solution to the first problem in Section 4, and then a simultaneous solution to both problems in Section 5.

4. Detecting emerging clusters

If we do a purely spatial analysis for an extensive time period, we have low power to detect recently emerging clusters. One way to resolve this is to do a purely spatial analysis including only the last few years. The suitable number of years is typically unknown though. If we include too few years, we might not have enough power to detect a low to moderate excess risk that has been present for a considerable time, and, if we include too many years, we might not have enough power to detect a very recent high excess risk cluster. A solution is to use a space–time scan statistic.

Instead of a circular window in two dimensions, the space–time scan statistic uses a cylindrical window in three dimensions. The base of the cylinder represents space, exactly as with the purely spatial scan statistic, whereas height represents time. The cylinder is flexible in its circular geographical base as well as in its starting date, independently of each other. This means that for each possible circle location and size we consider each possible starting date for the cluster and vice versa. We consider only those cylinders that reach all the way to the end of the study period. Hence, we consider only 'alive' clusters, those that are still active or present, while ignoring those that may have existed historically but which are no longer a public health problem. In mathematical notation, let $[Y_1, Y_2]$ be the time interval for which data exist, and let s and t be the start and end dates of the cylinder respectively. We then consider all cylinders for which $Y_1 \leqslant s \leqslant t = Y_2$.

The likelihood ratio test statistic is constructed in the same way as for the purely spatial scan statistic, using equations (1) and (2). For the random data sets, cases are generated so that both spatial areas and temporal intervals have a random number of cases, independently of each other. The computational algorithm for calculating the likelihood for each window has to work in three rather than two dimensions though, and it is hence more computer intensive.

Analysing the years 1973–1978, we find that the most likely cluster consists of Bernalillo, McKinley, Valencia, San Juan, Catron, Sandoval, Los Alamos and Rio Arriba counties, during the time period 1975–1978 (Table 2). Of the six years, only the last four are included in the cluster. With 48 cases observed, and 36 expected, $p = 0.60$. Hence, even when allowing for newly emerging clusters, there is no indication that there are any areas in New Mexico that by 1978 had a geographical excess risk of male thyroid cancer.

As with the purely spatial analysis, it is possible to repeat this procedure every year as new data are obtained. The results are presented in Table 2. In 1985 there suddenly appears a very

Table 2. Detection of recently emerging clusters of male thyroid cancer in New Mexico, by using the space–time scan statistic†

Data years	*Most likely cluster*‡	*Cluster years*	*Cases*	*Expected*	*Relative risk*	*p*
1973–1978	Bernalillo + 7 counties west	1975–1978	48	36	1.4	0.60
1973–1979	Los Alamos, Rio Arriba	1975–1979	9	3.3	2.7	0.58
1973–1980	Los Alamos, Rio Arriba	1975–1980	10	3.8	2.6	0.54
1973–1981	North-central counties − San Miguel	1975–1981	72	53	1.4	0.19
1973–1982	North-central counties − San Miguel	1975–1982	85	62	1.4	0.08
1973–1983	Bernalillo, Valencia	1975–1983	84	62	1.4	0.13
1973–1984	North-central counties	1973–1984	113	90	1.3	0.14
1973–1985	Lincoln	1985	3	0.2	13.8	0.23
1973–1986	North-central counties + Colfax, Harding	1973–1986	129	108	1.2	0.49
1973–1987	North-central counties + Colfax, Harding	1973–1987	142	117	1.2	0.21
1973–1988	North-central counties − San Miguel	1973–1988	143	115	1.2	0.08
1973–1989	North-central counties + Colfax, Harding	1973–1989	165	134	1.2	0.06
1973–1990	Los Alamos, Rio Arriba, Santa Fe, Taos	1979–1990	41	22	1.8	0.06
1973–1991	Los Alamos	1989–1991	7	0.9	7.6	0.02
1973–1992	Los Alamos	1989–1992	9	1.2	7.6	0.002

†The *p*-values are not adjusted for the repeated time periodic analyses conducted since 1978.
‡The north-central counties are Bernalillo, Los Alamos, Mora, Rio Arriba, Sandoval, San Miguel, Santa Fe and Taos; the counties west are McKinley, Valencia, San Juan, Catron, Sandoval, Los Alamos and Rio Arriba.

localized cluster in Lincoln County with three cases and a relative risk of 13.8. Although not statistically significant that year, it can be viewed as a potential cluster awaiting confirmation or rejection after one more year of data has been collected. With no cases in Lincoln County in 1986, it was no longer the most likely cluster that year, and most probably a chance occurrence.

In 1991, we suddenly have a statistically significant cluster emerge, consisting of Los Alamos County and spanning the years 1989–1991 with seven cases when 0.9 were expected ($p = 0.02$) (Fig. 2). The subsequent years' analysis identifies a cluster in the same location for the time period 1989–1992, with two additional cases, for a total of nine cases when 1.2 were expected ($p = 0.002$). It is important to note though, that although these *p*-values are adjusted for the many possible geographical cluster locations and sizes, as well as for the many possible temporal lengths of the clusters, they reflect the *p*-value for a single analysis

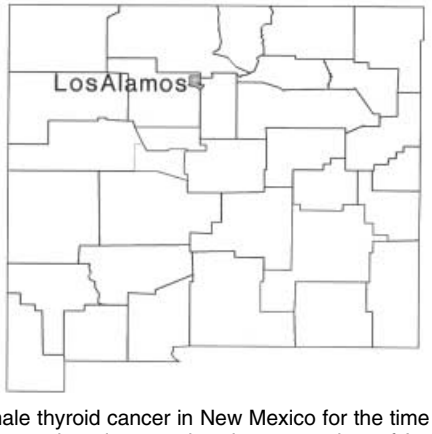

Fig. 2. Most likely cluster of male thyroid cancer in New Mexico for the time period 1973–1992, using the time periodic surveillance to detect emerging clusters: the cluster consists of Los Alamos county ($p = 0.016$); the analysis was adjusted for age and the multiple time period analyses

and do not adjust for the multiple time periodic analyses conducted over the years. That will be resolved in the next section.

5. Time periodic surveillance

New cases are added to disease registries regularly as they occur, and for surveillance purposes it is important to reanalyse the data in periodic intervals. One should then adjust for the multiple analyses conducted over time. This can be done by not only adjusting for the many possible geographical and temporal cluster locations and sizes but also for all the previous analyses that have already been conducted. It is accomplished through a slight modification to the space–time scan statistic.

Since we are interested in 'alive' clusters, the likelihood for the real data set is taken as the maximum over all cylinders reaching the end of the study period, just as before, i.e. we consider those cylinders for which $Y_1 \leqslant s \leqslant t = Y_2$. For the random data sets though, the likelihood is maximized over all cylinders used in previous analyses in addition to the current cylinders, to adjust for the multiple analyses conducted, i.e. those cylinders for which $Y_1 \leqslant s \leqslant t \leqslant Y_2$ and $t \geqslant Y_m$, where Y_m is the year in which the time periodic surveillance began. At any given moment, an observed cluster is then statistically significant at the α-level if the probability of having detected a cluster with higher likelihood during any of the previous analyses or the present analysis is at most α. The most likely clusters will always be the same as with the analysis from the previous section, but their p-values will be higher, reflecting the time period adjusted analyses.

Using this procedure, the results for the thyroid cancer data are given in Table 3. As can be seen for the 1973–1991 data, with $p = 0.13$ the Los Alamos cluster is not statistically significant when adjusting for the time periodic surveillance. For the 1973–1992 data, however, the Los Alamos cluster is statistically significant, with $p = 0.016$. We have not shown all yearly analyses in Table 3 as the entries are identical with those in Table 2 except for the p-values, which are all larger and hence non-significant for the earlier years.

Fig. 3 depicts the cumulative observed and expected number of male thyroid cases in Los Alamos County from 1973 to 1992. Although four cases occurred in 1989, that was not enough to declare Los Alamos a statistically significant cluster. Not until 1992 was it clear that there was a local excess which was not explainable by chance.

If a surveillance system is in place for a long time period, it might be too conservative to adjust the inference for all previous analyses. In such a setting, we could instead specify that we want at most a probability α of falsely rejecting the null hypothesis during any specific length of time, say 20 years for a chronic disease. The method proposed can easily be modified to accomplish this, by only including those cylinders analysed during the preceding

Table 3. Detection of recently emerging clusters of male thyroid cancer in New Mexico, by using the space–time scan statistic†

Data years	*Most likely cluster*	*Cluster years*	*Cases*	*Expected*	*Relative risk*	*p*
1973–1991	Los Alamos	1989–1991	7	0.9	7.6	0.13
1973–1992	Los Alamos	1989–1992	9	1.2	7.6	0.016

†The p-values are adjusted for the repeated time periodic analyses conducted since 1978. Entries in this table are identical with those of Table 2 except for the p-values, which are larger.

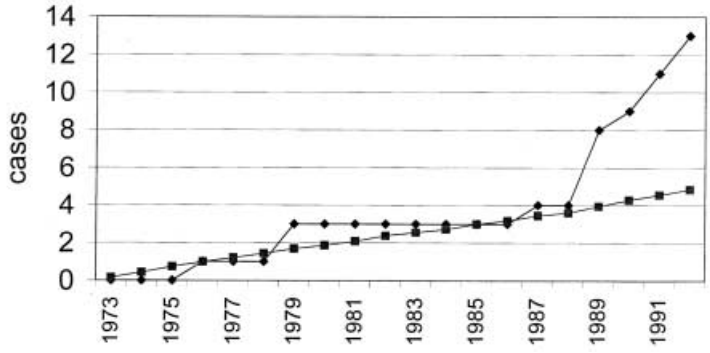

Fig. 3. Cumulative observed (◆) and expected (■) numbers of male thyroid cases in Los Alamos County from 1973 to 1992

20 years when calculating the likelihood for the most likely cluster in the random data sets, i.e. those cylinders for which $Y_1 \leqslant s \leqslant t$ and $Y_2 - 20 < t \leqslant Y_2$.

6. Discussion

As with chronic diseases, it is of interest to detect localized infectious disease outbreaks as early as possible without having to specify either the geographical location and size or the temporal length of the cluster *a priori*. In infectious disease surveillance though, we are not always interested in clusters present during a very long time period, as these would be reflective of populations that are prone to infection rather than of an important emerging health hazard. Although both types of cluster may be of interest, time periodic infectious disease surveillance is typically geared towards the latter. For such analyses, one should not consider all cylinders of all possible temporal lengths but only those up to some prespecified maximum. Depending on the particular infectious disease, that maximum could be a couple of days or a few years.

Within a surveillance system, it is not advisable to maintain a strict cut-off for the p-value to determine whether a detected cluster should be investigated or not. Rather, the p-value should serve as an indicator concerning the evidence for a real cluster, and the amount of effort devoted to an initial investigation should depend on how strong this evidence is. When the evidence is strong, a detailed epidemiological study with individual case histories should be conducted. In looking for a local explanation for the observed excess, it is then important to realize that some cases will be unrelated to that explanation and would have occurred anyway, constituting what may be called the background rate.

The proposed prospective surveillance system cannot detect historic clusters that both began and ended before the time periodic surveillance began. Even though it is too late to do anything about such clusters, they may nevertheless be of great interest for aetiological reasons. To detect such clusters we should instead do a single analysis using a version of the space–time scan statistic where both the start and the end date of the cylinders vary (Kulldorff *et al.*, 1998a), or some other retrospective surveillance method.

Other prospective time periodic geographical surveillance systems have been proposed by Rogerson (1997, 2001) and Järpe (1999). These are geared towards detecting the sudden appearance of clustering as a global phenomenon throughout the geographical area under

study, as opposed to the currently proposed method, which is used to detect the specific geographical location of emerging clusters and to evaluate their significance. The two types of method complement each other, being suitable for different types of inquiries.

A statistically significant cluster of male thyroid cancer was found in Los Alamos, emerging in 1989 and detectable in 1992. The observed cluster could be due to errors in the data, either in the numerator, denominator or geographical co-ordinates, and errors in data should always be the first suspect when evaluating a statistically significant cluster. No such errors were found. The New Mexico Department of Health have investigated the individual nature of each of 17 male cases reported during 1970–1995, including the nine cases diagnosed in 1989–1992 as well as three cases diagnosed in 1993–1995, confirming all cases (Athas, 1996). The 'stage-of-disease' distribution was similar to that for the rest of New Mexico, excluding better screening programmes as a likely explanation for the Los Alamos cluster. Three of the 17 cases had a history of therapeutic ionizing radiation treatment to the head and neck during the 1940s and 1950s, which is an established risk factor. Two of these were diagnosed after 1988. At least eight of the 17 cases had been regularly monitored for exposure to radiation due to their particular work tasks at the Los Alamos National Laboratory. Two of these, both diagnosed after 1988, had significant workplace-related exposure to ionizing radiation from atmospheric weapons testing fieldwork, another established risk factor. It is hence safe to say that the observed excess is at least in part due to known risk factors of a local nature that is unique to the population of Los Alamos.

When interpreting the results of prospective surveillance, it is important to consider the latency period of the disease. For example, in our example there were several subjects whose exposure to known risk factors had occurred many years before the diagnosis. If the exposure is in the local environment, long latency reduces the power of the analysis as some exposed individuals will have moved away between exposure and diagnosis, while unexposed individuals have moved in. Moreover, when a cluster is detected, it is important not only to look at recent exposures when searching for the cause. If the latency is long then new cases will continue to occur even if the exposure is removed, and that is important to keep in mind when doing follow-up monitoring and evaluation.

The choice of 50% of the population at risk as the maximum circle size is natural as it allows both small and large clusters to be detected, while ignoring those 'clusters' that contain more than half the population at risk, as these would be more suitably interpreted as a 'negative cluster' of lower than expected risk outside the circle. Other choices of the maximum are also possible. In this paper we were only interested in detecting areas with an excess risk for disease, but the space–time scan statistic can also be used to detect areas with exceptionally low rates (Kulldorff *et al.*, 1998b).

In conclusion, the space–time scan statistic can serve as an important tool for systematic time periodic geographical disease surveillance. It is possible to detect emerging clusters, and we can adjust for the multiple tests performed over time. No *a priori* hypotheses about cluster location, size or length need to be made. The method can be used at different levels of geographical and temporal aggregation, and for different types of disease. Although it is computer intensive, the method is not overly complex. The analyses in this paper took between 5 and 75 s to run on a 400 MHz Pentium Pro computer.

Acknowledgements

Thoughtful comments from Dr John Bithell, Dr Helen Dolk and two reviewers are gratefully acknowledged.

References

Athas, W. F. (1996) *Investigation of Excess Thyroid Cancer Incidence in Los Alamos County*. Santa Fe: New Mexico Department of Health.

Bernardinelli, L. and Montomoli, C. (1992) Empirical Bayes versus fully Baysian analysis of geographical variation in disease risk. *Statist. Med.*, **11**, 983–1007.

Besag, J. and Newell, J. (1991) The detection of clusters in rare diseases. *J. R. Statist. Soc.* A, **154**, 143–155.

Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1–59.

Bithell, J. F. (1999) Disease mapping using the relative risk function estimated from areal data. In *Disease Mapping and Risk Assessment for Public Health* (eds A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J. F. Viel and R. Bertollini). Chichester: Wiley.

Bjerkedal, T. and Bakketeig, L. S. (1975) Surveillance of congenital malformations and other conditions of the newborn. *Int. J. Epidem.*, **4**, 31–36.

Boice, J. D., Mandel, J. S., Doody, M. M., Yoder, R. L. and McGrowan, R. (1992) A health survey of radiologic technologists. *Cancer*, **69**, 686–698.

Carstensen, J. M., Wingren, G., Hatschek, T., Fredriksson, M., Noorlinde-Brage, H. and Axelson, O. (1990) Occupational risks of thyroid cancer: data from the Swedish Cancer Environment Register, 1961-1979. *Am. J. Indust. Med.*, **18**, 535–540.

Chen, R. (1978) A surveillance system for congenital malformations. *J. Am. Statist. Ass.*, **73**, 323–327.

———(1979) Statistical techniques in birth defects surveillance systems. *Contrib. Epidem. Biostatist.*, **1**, 184–189.

Chen, R., Mantel, N., Connelly, R. R. and Isacson, P. (1982) A monitoring system for chronic diseases. *Meth. Inform. Med.*, **21**, 86–90.

Choynowski, M. (1959) Maps based on probabilities. *J. Am. Statist. Ass.*, **54**, 385–388.

Clayton, D. and Kaldor, J. (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–681.

Cliff, A. D. and Ord, J. K. (1973) *Spatial Autocorrelation*. London: Pion.

Cuzick, J. and Edwards, R. (1990) Spatial clustering for inhomogeneous populations (with discussion). *J. R. Statist. Soc.* B, **52**, 73–104.

Diggle, P. J. and Chetwynd, A. D. (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, **47**, 1155–1163.

Duffy, B. J. and Fitzgerald, P. J. (1950) Cancer of the thyroid in children: a report of 28 cases. *J. Clin. Endocrin. Metablm*, **10**, 1296–1308.

Dwass, M. (1957) Modified randomization tests for nonparametric hypotheses. *Ann. Math. Statist.*, **28**, 181–187.

Ederer, F., Myers, M. H. and Mantel, N. (1964) A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics*, **20**, 626–638.

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

Jacquez, G. M. (1996) A k nearest neighbor test for space-time interaction. *Statist. Med.*, **15**, 1935–1949.

Järpe, E. (1999) Surveillance of the interaction parameter of the Ising model. *Communs Statist. Theory Meth.*, **28**, 3009–3027.

Kafadar, K. (1996) Smoothing geographical data, particularly rates of disease. *Statist. Med.*, **15**, 2539–2560.

Kazakov, V. S., Demidchik, E. P. and Astakhova, L. N. (1992) Thyroid cancer after Chernobyl. *Nature*, **359**, 21.

Knox, E. G. (1964) Detection of space-time interactions. *Appl. Statist.*, **13**, 25–29.

Kulldorff, M. (1997) A spatial scan statistic. *Communs Statist. Theory Meth.*, **26**, 1481–1496.

———(1999) An isotonic spatial scan statistic for geographical disease surveillance. *J. Natn. Inst. Publ. Hlth*, **48**, 94–101.

Kulldorff, M., Athas, W., Feuer, E., Miller, B. and Key, C. (1998a) Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos. *Am. J. Publ. Hlth*, **88**, 1377–1380.

Kulldorff, M., Rand, K., Gherman, G., Williams, G. and DeFrancesco, D. (1998b) *SaTScan v 2.1: Software for the Spatial and Space-time Scan Statistics*. Bethesda: National Cancer Institute. (Available from `http://dcp.nci.nih.gov/BB/SaTScan.html`.)

Lawson, A. B. and Williams, F. L. R. (1993) Applications of extraction mapping in environmental epidemiology. *Statist. Med.*, **12**, 1249–1258.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.*, **27**, 209–220.

Martuzzi, M. and Elliott, P. (1996) Empirical Bayes estimation of small area prevalence of non-rare conditions. *Statist. Med.*, **15**, 1867–1873.

Martuzzi, M. and Hills, M. (1995) Estimating the degree of heterogeneity between event rates using likelihood. *Am. J. Epidem.*, **141**, 369–374.

Parkin, D. M., Muir, C. S., Whelan, S. L., Gao, Y. T., Ferlay, I. and Powell, J. (1992) *Cancer Incidence in Five Continents*, vol. VI. Lyon: International Agency for Research on Cancer.

Prentice, R. L., Kato, H., Yoshimoto, K. and Mason, M. (1982) Radiation exposure and thyroid cancer incidence among Hiroshima and Nagasaki residents. *Natn. Cancer Inst. Monogr.*, **62**, 207–212.

Radaelli, G. (1996) Detection of an unknown increase in the rate of a rare event. *J. Appl. Statist.*, **23**, 105–113.
Ries, L. A. G., Kosary, C. L., Hankey, B. F., Miller, B. A., Clegg, L. and Edwards, B. K. (1999) *SEER Cancer Statistics Review, 1973-1996*. Bethesda: National Cancer Institute.
Robbins, J. and Adams, W. H. (1989) Radiation effects in the Marshall Islands. In *Radiation and the Thyroid* (ed. S. Nagataki). Amsterdam: Excerpta Medica.
Rogerson, P. A. (1997) Surveillance systems for monitoring the development of spatial patterns. *Statist. Med.*, **16**, 2081–2093.
———(2001) Monitoring point patterns for the development of space–time clusters *J. R. Statist. Soc.* A, **164**, 87–96.
Ron, E., Modan, B., Preston, D., *et al.* (1989) Thyroid neoplasia following low-dose radiation in children. *Radian Res.*, **120**, 516–531.
Rushton, G. and Lolonis, P. (1996) Exploratory spatial analysis of birth defect rates in an urban population. *Statist. Med.*, **15**, 717–726.
Sheehan, T. J., Gershman, S. T., MacDougal, L., Danley, R., Mrosszczyk, M., Sorensen, A. M. and Kulldorff, M. (2000) Geographical surveillance of breast cancer screening by tracts, towns and zip codes. *J. Publ. Hlth Mangmnt Pract.*, **6**, 48–57.
Shore, R. E., Woodward, E., Hildreth, N., Dvoretsky, P., Hempelmann, L. and Pasternack, B. (1985) Thyroid tumors following thymus irradiation. *J. Natn Cancer Inst.*, **74**, 1177–1184.
Tango, T. (1995) A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statist. Med.*, **14**, 2323–2334.
———(2000) A test for spatial disease clustering adjusted for multiple testing. *Statist. Med.*, **19**, 191–204.
Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L. and Clark, L. L. (1990) Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *Am. J. Epidem.*, **132**, S136–S143.
Viel, J. F., Arveux, P., Baverel, J. and Cahn, J. Y. (2000) Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *Am. J. Epidem.*, **151**, 13–19.
Wallenstein, S. (1980) A test for detection of clustering over time. *Am. J. Epidem.*, **111**, 367–372.
Wallenstein, S., Gould, M. and Kleinman, M. (1989) Use of the scan statistic to detect time-space clustering. *Am. J. Epidem.*, **130**, 1057–1064.
Walsh, S. J. and Fenster, J. R. (1997) Geographical clustering of mortality from systemic sclerosis in the Southeastern United States, 1981-90. *J. Rheumatol.*, **24**, 2348–2352.
Walter, S. D. (1994) A simple test for spatial pattern in regional health data. *Statist. Med.*, **13**, 1037–1044.
Weatherall, J. A. C. and Haskey, J. C. (1976) Surveillance of malformations. *Br. Med. Bull.*, **32**, 39–44.
Weinstock, M. A. (1981) A generalized scan statistic test for the detection of clusters. *Int. J. Epidem.*, **10**, 289–293.