Epidemiol. Infect. (2005), **133**, 409–419. © 2005 Cambridge University Press
doi:10.1017/S0950268804003528 Printed in the United Kingdom

A model-adjusted space–time scan statistic with an application to syndromic surveillance

K. P. KLEINMAN[1*], A. M. ABRAMS[1,2], M. KULLDORFF[1,3] AND R. PLATT[1,4]

[1] *Department of Ambulatory Care and Prevention, Harvard Medical School, Harvard Pilgrim Health Care, and CDC Eastern Massachusetts Prevention Epicenter and HMO Research Network Center for Education and Research in Therapeutics, Boston, MA, USA*
[2] *University of Minnesota School of Public Health, Minneapolis, MN, USA*
[3] *University of Connecticut Health Center, Farmington, CT, USA*
[4] *Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

(*Accepted 19 November 2004*)

SUMMARY

The space–time scan statistic is often used to identify incident disease clusters. We introduce a method to adjust for naturally occurring temporal trends or geographical patterns in illness. The space–time scan statistic was applied to reports of lower respiratory complaints in a large group practice. We compared its performance with unadjusted populations from: (1) the census, (2) group-practice membership counts, and on adjustments incorporating (3) day of week, month, and holidays; and (4) additionally, local history of illness. Using a nominal false detection rate of 5%, incident clusters during 1 year were identified on 26, 22, 4 and 2% of days for the four populations respectively. We show that it is important to account for naturally occurring temporal and geographic trends when using the space–time scan statistic for surveillance. The large number of days with clusters renders the census and membership approaches impractical for public health surveillance. The proposed adjustment allows practical surveillance.

INTRODUCTION

The public health community's interest in the prompt detection of clusters of illnesses has increased recently because of the need for the earliest possible detection of bioterrorism attacks and the perceived utility of enhanced detection of both endemic and new diseases, the earliest manifestations of which may consist of non-specific signs and symptoms [1, 2]. Useful cluster detection systems should be able to identify clusters of various size, shape, and duration, centred anywhere within the surveillance region. In order to identify clusters, it is necessary to take into account the size and location of the population subject to surveillance. Information about the population's usual patterns of medical care utilization for specific types of illness and historical information about temporal variation (for instance, by month of year and day of week) may also be useful.

Kulldorff has proposed a statistical approach to surveillance based on the space–time scan statistic that meets the primary criteria noted above in being able to identify clusters of varying size and duration, centred anywhere in the surveillance area [3]. A software package (SaTScan) to apply the approach is freely available [4].

We propose here a model-based method for adjusting the space–time scan statistic in a way that can accommodate both temporal and geographical variation in syndromic event rates, and examine its

* Author for correspondence: K. P. Kleinman, Sc.D., Department of Ambulatory Care and Prevention, 133 Brookine Ave, 6th Floor, Boston, MA 02215, USA.
(Email: ken_kleinman@harvardpilgrim.org)

impact on syndromic surveillance. In contrast, previous work has mainly discussed adjusting the spatial and space–time scan statistics for individual and area level covariates such as age, gender, urbanicity and socio-economic status, or by using historical counts as controls [5–8]. SaTScan offers a simple adjustment for purely temporal trends, but this approach cannot adapt to regular periods of increased and decreased risk as may sometimes be necessary.

We consider four possible space–time scan statistic analyses. First, we consider using the census to define the background population at risk for each census tract, even though not everyone in that population would be 'caught' by our system when they go to the doctor. Second, we consider using the number of individuals in each census tract who are eligible to appear in the surveillance system. Third, we adjust the eligible population for seasonal and other temporal patterns. Finally, we also adjust for different baseline risks of illness in each census tract, accounting for the possibility that different individuals may have a different propensity to appear as cases (e.g. for going to the doctor) depending on which census tract they live in.

METHODS

Dataset

We have previously described syndromic surveillance based on an automated electronic medical record system [9]. The dataset represents the ambulatory medical encounters of insured individuals. For each encounter, a clinician enters diagnoses, to which diagnostic codes are attached. Diagnoses are then grouped into syndromes. Geocoding patients' addresses provides the census tract of residence. Although the accuracy and utility of commercial geocoding has been questioned [10], the addresses are used for billing and other business purposes, so they are likely to be accurate.

We use syndromic surveillance of lower respiratory infection (LRI) as our example. A case of anthrax in the initial phase would include symptoms that would probably cause it to be classified into the LRI syndrome [9] so our example reflects on bioterrorism surveillance. As would be expected, incidence rates of LRI are much higher in the winter than summer [1]. Local baseline risk of LRI is also thought to vary due to local variability in the proportion of individuals with weaker immune systems or cultural propensity to seek health care.

The dataset used for the example incorporates all LRI encounters between 1 January 1996 and 31 October 1999. This represents the 80 643 encounters of approximately 240 000 individuals, about 10% of the population in the 566 census tracts with centroids in the greater Boston area between west longitudes 70·85 and 71·40 and north latitudes 42·15 and 42·67.

Spatial surveillance scan statistic

The space–time scan statistic can be used to detect clusters of disease in time and space [3]. Cylindrical risk regions are used to detect clusters. The radius represents an area of the map, while the height represents time. The radius varies from zero to a chosen upper bound that restricts the percentage of the population included in the cylinder, in our case 25% of the total population. The maximum number of days included in the cluster can also be limited [3, 4]. We set this maximum to 1 day. These limits are somewhat arbitrary and we chose them as relevant to the example of surveillance for anthrax.

We simplify the space–time scan statistic by making an initial summary within census tracts. This means that we treat all individuals as if they live at the centroid of their census tract. The entire census tract population will be included only if the centroid is included in the cylinder. There are two reasons we used the census tract summed counts rather than the exact locations. First, our privacy agreement with the health plan that gave us the data requires we use the exact locations only to get the grouped data. Second, the computational time required to do the exact-location analysis on approximately 200 000 addresses would be prohibitive, compared with the 529 populated census tract locations. Although possibly not a reason to summarize spatially in itself, we note that the type of adjustment we outline below would be difficult to apply to individual addresses.

The space–time scan procedure first calculates a Poisson-based likelihood ratio. This is proportional to $(n/\mu)^n[(N-n)/(N-\mu)]^{N-n}$, where N is the total number of visits, n is the number of visits among those people who live within the cylinder, and μ is the expected number of visits within the cylinder. The expected number of visits is simply N times the proportion of the population across time that lives within the cylinder. Note the central role of μ in the likelihood ratio and that μ depends on the denominator from which the cases develop; thus, we can alter the

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

expected number of cases μ by modifying the denominator.

The value of the likelihood ratio is calculated for every possible cylinder, and the cylinder associated with the maximum value is the most likely cluster [11]. This maximum-likelihood ratio is the space–time scan statistic. A P value is assigned to the cluster using Monte Carlo hypothesis testing [12]. The number of cases actually observed (N) is randomly re-distributed over time and space under the null hypothesis (that the cases are distributed proportionately to population over time and space). This is done 999 times. In each Monte Carlo replicate, a most likely cluster is found and its likelihood ratio (the scan statistic) is recorded. The proportion of the scan statistics from all the $999+1$ datasets that are larger than the scan statistic from the one observed dataset is the P value. In this way, the method adjusts for the multiple testing inherent in the many cylinder sizes and locations evaluated.

In prospective surveillance applications, the space–time scan statistic will be calculated repeatedly every day, resulting in an additional multiple testing problem. Kulldorff provides a *spatio-temporal surveillance scan* routine [3]. First, in the observed data, the space–time scan statistic among clusters that include the study end date is found. Then, in each Monte Carlo replication the space–time scan statistic is found with no restriction that the clusters include the study end date. The 'surveillance-corrected' P value is the proportion of these unrestricted scan statistics that are larger than the restricted scan statistic based on the observed data. Instead of expecting ten P values smaller than 0·05 in every 200 days of surveillance under the null (as would be the case without correction), the probability is now 0·05 of observing one or more significant P value during this 200-day period. In the remainder of this article, all referenced P values are of this sort.

All analyses were performed using SaTScan software [4].

Populations and denominators

We use the terms 'population' and 'denominator' to refer to the census and eligible populations and the adjustments described below. The four denominators reflect increasing amounts of information about the population subject to surveillance that may be available. The first of these is the population residing in each census tract as determined by the most recent decennial U.S. census. This population is the minimum available data, and might be considered if the eligible population were all residents of an area, although it is somewhat implausible in the example. We call this the census population. In Figure 1, we show the 1990 census population of our area.

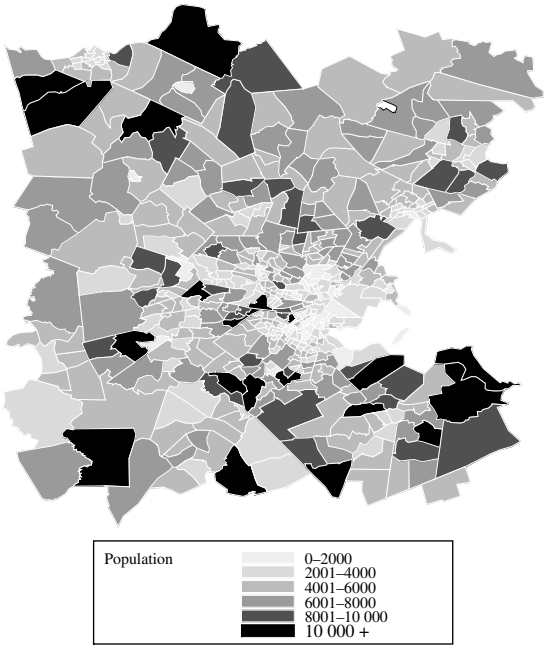

Fig. 1. 1990 census population, by 1990 census tract, Eastern Massachusetts.

The second denominator comprises subjects eligible for surveillance. This population would be relevant if it were possible to enumerate the population under surveillance, as in the example. We call this the eligible population. In Figure 2, we show the distribution of the approximately 240 000 eligible subjects in our dataset.

In contrast to the two measured populations, we also used two model-based adjustments. The first adjustment accounts for day of week, month of the year, holiday status, and secular time. To calculate the adjustment in this case, we first generated $\hat{P}_t$, the estimated probability of a visit on day t based on these factors. The method used for generating $\hat{P}_t$ is discussed in the Appendix. Next we multiplied this by the eligible population, $\text{pop}_{\text{elig}(i)}$ in census tract i. We then standardized these adjusted denominators by multiplying by a constant so that the apparent population after adjustment and the eligible population are equal over the course of a year. So, if the average daily eligible population was 240 000 over the course of a year, the average apparent adjusted population was also 240 000. Thus, the first adjustment results

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

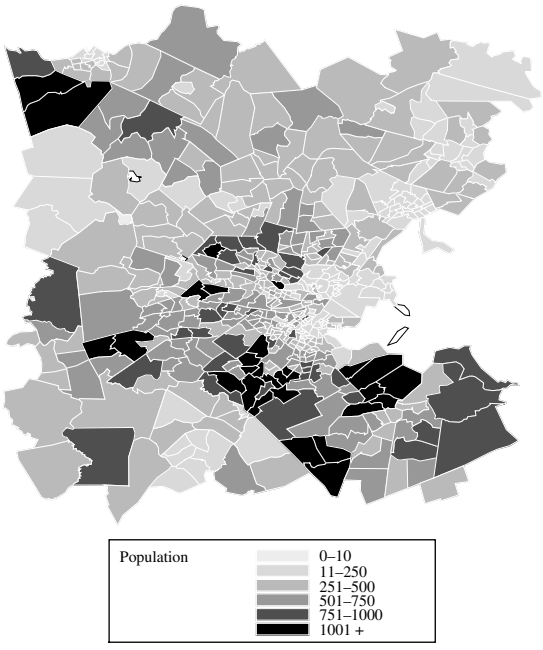

Fig. 2. Population eligible for surveillance on Monday, 11 January 1999, by 1990 census tract, Eastern Massachusetts.

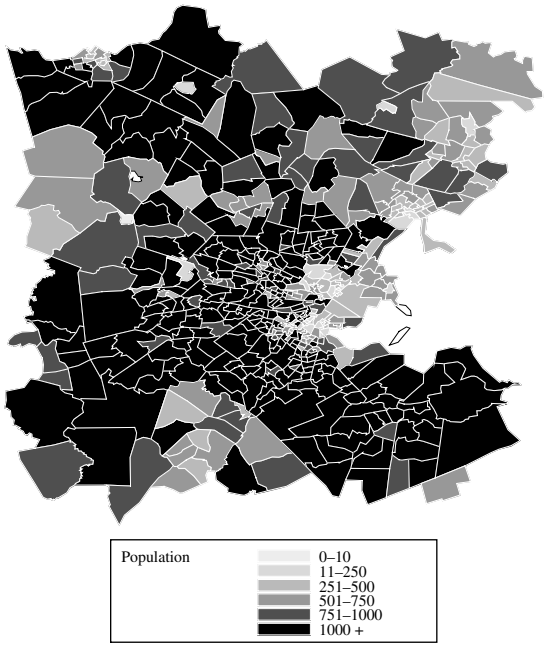

Fig. 3. Apparent population under date adjustment on Monday, 11 January 1999, by 1990 census tract, Eastern Massachusetts.

in denominators in tract i on day t that are $c_1 \times \hat{P}_t \times \text{pop}_{\text{elig}(i)}$, where c_1 is the constant used to standardize.

The adjustment makes the population bigger on days when the model suggests that more visits are likely and smaller on days when fewer visits are likely. Since the SaTScan expected cases are based on the proportional population, this adjustment to the population input to SaTScan results in an adjustment to the expected number of cases. The effect is that individual visits are less likely to signify interesting events on busy days, such as Mondays in flu season: the proportional population is bigger on those days so the expected number of cases is larger. Similarly, the expected number of cases on days like Sundays in June is smaller, and actual visits are more interesting events. We call this the 'date adjustment'; in Figure 3 we show the apparent population after date adjustment on Monday, 11 January 1999.

The second adjustment includes all the effects of the date adjustment, but also takes into account the differing baseline risk in each census tract. To do this, we find $\hat{P}_{it}$, the estimated probability that a person in *census tract i* seeks health care on day t. The method for obtaining $\hat{P}_{it}$ is discussed in Kleinman et al. [13]. Briefly, we fit the generalized linear mixed model version of logistic regression to the repeated counts for each tract. The model included a random intercept for each tract. This is similar to including an indicator variable or dummy code for each tract, but incorporates plausible assumptions about how differing amounts of information from each tract should be incorporated. The net meaning of the model, however, is that each tract has a different intercept in the logit scale, while all tracts share common day of week, month, holiday, and secular time effects.

We multiplied this probability by the eligible population in the census tract and then standardized so that the average apparent population after adjustment was the same as the eligible population over the course of a year. This gave us a second adjustment where the SaTScan denominator is $c_2 \times \hat{P}_{it} \times \text{pop}_{\text{elig}(i)}$, where c_2 is the constant used to standardize. We call this the 'date-and-tract adjustment' and display the apparent population in Figures 4–6 for Sunday 10 January 1999, Monday 11 January 1999 and Monday 12 July 1999 respectively.

The intuition here is that in addition to inflating the apparent population on days when many visits are predicted by the model, we also make it bigger in tracts that have a larger baseline risk. Visits in those tracts are generally more common than visits in tracts with reduced baseline risk. The difference between the two adjustments is that in the date adjustment, the probability of a visit ($\hat{P}_t$) is the same in each census tract on day t, whereas in the date-and-tract adjustment, the probability ($\hat{P}_{it}$) is different in each tract i.

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

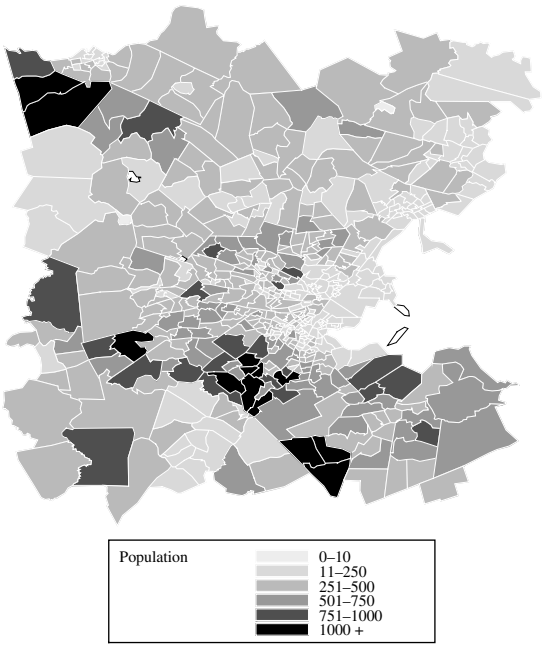

Fig. 4. Apparent population under date-and-tract adjustment on Sunday, 10 January 1999, by 1990 census tract, Eastern Massachusetts.

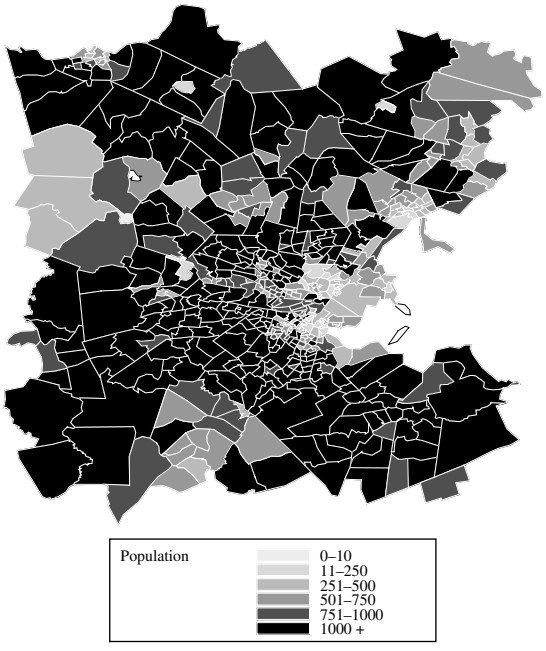

Fig. 5. Apparent population under date-and-tract adjustment on Monday, 11 January 1999, by 1990 census tract, Eastern Massachusetts.

Note that the scales are the same in Figures 2–6, with darker regions indicating larger effective populations. The figures show the effect of the adjustments, especially for day of week and month. The date-adjustment and date-and-tract-adjustment maps

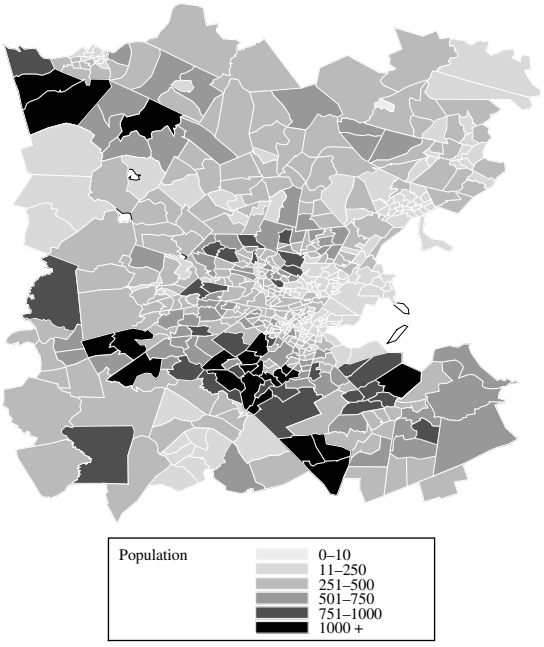

Fig. 6. Apparent population under date-and-tract adjustment on Monday, 12 July 1999, by 1990 census tract, Eastern Massachusetts.

appear similar. Comparing the light map on Sunday to the darker map for Monday indicates the increased probability, under the model, of a visit to a doctor on a Monday compared to Sunday.

Analyses

We retrospectively conducted 1 year of surveillance. We treated each day in the year beginning 1 November 1998 as if it were the last day in the dataset, the day for which surveillance results were required. For each day, we used SaTScan to find the most likely cluster for each denominator and its *P* value.

RESULTS

Table 1 shows the results, reporting the number of days when the space–time scan statistic produced a *P* value smaller than 0·05, by month and denominator. There were no significant clusters on any weekend day, so we report the number of weekdays in each month as well. From October to the end of February, the census and eligible denominators resulted in significant clusters on a majority of weekdays. In contrast, the date-adjusted SaTScan identifies clusters on 13 days and the date-and-tract-adjusted SaTScan on 7 days.

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

Table 1. *Number (%) of weekdays per month between 1 November 1998 and 31 October 1999 for which SaTScan produced surveillance P values less than 0·05, under each of four denominators*

	No. of weekdays	Census	Eligible	Date adjustment	Date-and-tract adjustment
1998					
Nov.	21	19 (90%)	18 (86%)	2 (10%)	1 (5%)
Dec.	23	20 (87%)	14 (61%)	0	0
1999					
Jan.	21	17 (81%)	17 (81%)	4 (19%)	1 (5%)
Feb.	20	18 (90%)	16 (80%)	6 (30%)	4 (20%)
Mar.	22	1 (5%)	0	0	0
Apr.	21	0	0	0	0
May	22	1 (5%)	0	0	0
June	22	0	0	1 (5%)	1 (5%)
July	22	0	0	0	0
Aug.	22	0	0	0	0
Sept.	22	4 (18%)	2 (9%)	0	0
Oct.	21	16 (76%)	15 (71%)	0	0
Total (weekdays)	259	96 (37%)	82 (32%)	13 (5%)	7 (3%)
Total (365 days)		96 (26%)	82 (22%)	13 (4%)	7 (2%)

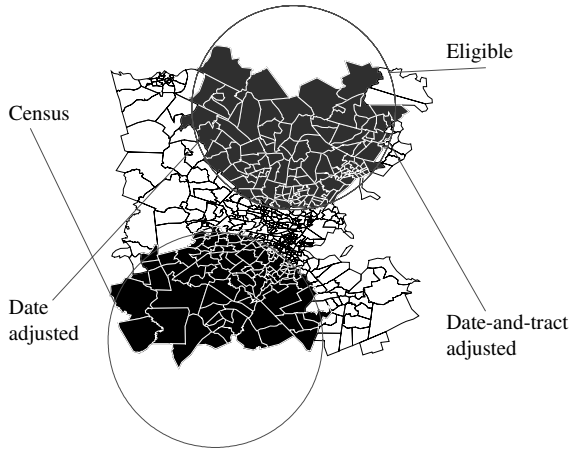

Fig. 7. Clusters identified by SaTScan on Thursday, 30 November 1998 using two populations and two model adjustments.

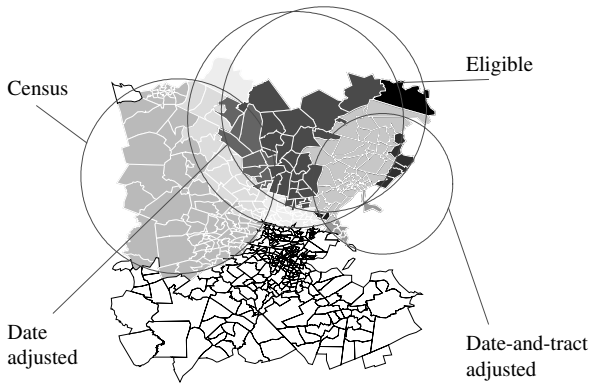

Fig. 8. Clusters identified by SaTScan on Thursday, 21 January 1999 using two populations and two model adjustments.

Note that the lack of a weekend cluster for the census and eligible denominators is not surprising: the small proportion of visits by population in these unadjusted denominators merely suggests a day with little illness. Under either adjustment, a smaller number of visits is anticipated via the model and adjusted for.

To assess the probability that the rate of clusters on weekends differed from the rate on weekdays, we used Fisher's exact test for the table of weekend/weekday by cluster/no cluster. The null hypothesis is that there are equal probabilities of significant clusters on weekends and weekdays. The *P* values for the four denominators were 0·20 (date-and-tract adjustment) 0·02 (date adjustment) and $<0·0001$ (eligible and census). There is little evidence that the date-and-tract adjustment reduced the probability of identifying clusters on weekends. Although the null hypothesis would be rejected under a strict 5% error-level test, the *P* value under the date adjustment is large enough to leave some ambiguity as to whether there is truly a reduced probability of identifying clusters on weekends using this adjustment, especially since this is a *post-hoc* test of an unplanned hypothesis. In contrast, there is little doubt that there is a reduced probability of a cluster when using the census and eligible populations.

The nature of the clusters identified also differs. In Figures 7–13, we display the most likely clusters

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

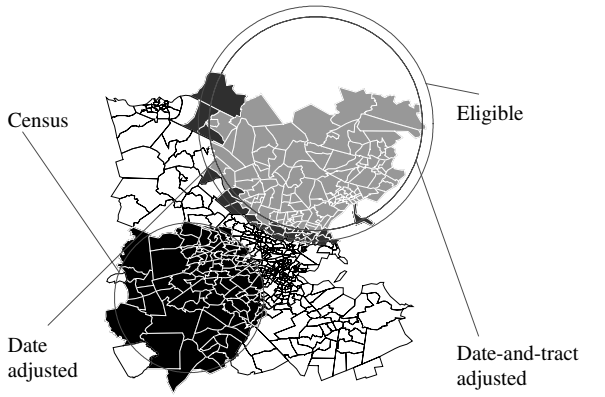

Fig. 9. Clusters identified by SaTScan on Friday, 5 February 1999 using two populations and two model adjustments.

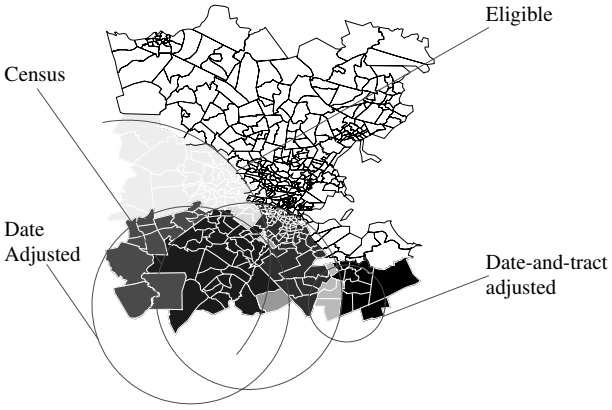

Fig. 10. Clusters identified by SaTScan on Wednesday, 10 February 1999 using two populations and two model adjustments.

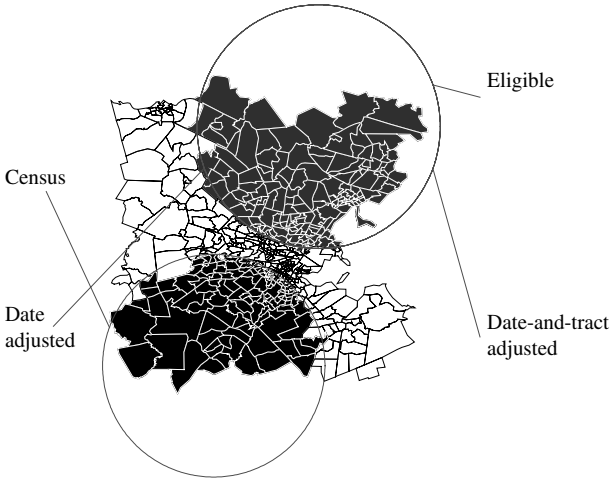

Fig. 11. Clusters identified by SaTScan on Thursday, 11 February 1999 using two populations and two model adjustments.

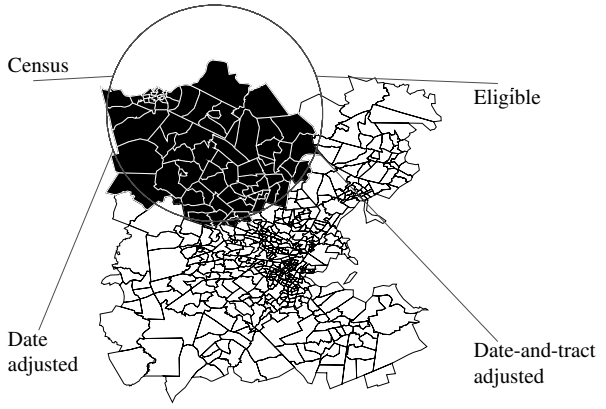

Fig. 12. Clusters identified by SaTScan on Tuesday, 16 February 1999 using two populations and two model adjustments.

Fig. 13. Clusters identified by SaTScan on Tuesday, 1 June 1999 using two populations and two model adjustments.

found for the 7 days when the date-and-tract adjustment resulted in scan-statistic P values below 0·05. The other denominators also generated P values below 0·05, except for 1 June, when only the date-and-tract and date adjustments had P values below 0·05. In that year, 1 June was the day after Memorial Day; there tend to be increased visits on the day after a holiday, a fact not accounted for in either adjustment model. In general, the census and eligible clusters contain close to 25% of the population, the maximum cluster size we allowed. This is also true for the days when only these denominators result in P values below 0·05 (results not shown). In contrast, the size of clusters identified when using the adjustments varies widely.

Table 2 provides further information about the clusters on those 7 days. For 16 February and 1 June, the denominators result in identical clusters. For 16 February, the expected number of cases is 50–80% higher for the adjustments, resulting in less significant P values. In contrast, the expected number of cases is

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

Table 2. *P value, number of cases, and expected number of cases under four denominators for the 7 days between 1 November 1998 and 31 October 1999 on which the date-and-tract adjusted denominator resulted in a P value below 0·05*

Date	Population or adjustment	*P* value for cluster*	No. of cases in cluster	Expected no. of cases in cluster area
Monday, 30 Nov. 1998	Census	0·001	54	10·25
	Eligible	0·001	53	9·22
	Date adjustment	0·002	53	22·01
	Date-and-tract adjustment	0·009	53	23·76
Thursday, 21 Jan. 1999	Census	0·001	50	14·68
	Eligible	0·001	56	14·95
	Date adjustment	0·001	37	13·14
	Date-and-tract adjustment	0·010	26	8·35
Friday, 5 Feb. 1999	Census	0·001	67	14·56
	Eligible	0·001	65	14·41
	Date adjustment	0·001	55	17·36
	Date-and-tract adjustment	0·002	55	19·13
Wednesday, 10 Feb. 1999	Census	0·001	76	15·77
	Eligible	0·001	64	17·52
	Date adjustment	0·016	46	17·52
	Date-and-tract adjustment	0·025	21	4·60
Thursday, 11 Feb. 1999	Census	0·001	67	17·66
	Eligible	0·001	69	17·05
	Date adjustment	0·001	69	27·98
	Date-and-tract adjustment	0·003	69	29·62
Tuesday, 16 Feb. 1999	Census	0·001	53	12·86
	Eligible	0·001	53	10·52
	Date adjustment	0·001	53	17·74
	Date-and-tract adjustment	0·003	53	18·64
Tuesday, 1 June 1999	Census	1·0	32	15·55
	Eligible	1·0	32	14·54
	Date adjustment	0·018	32	9·63
	Date-and-tract adjustment	0·027	32	9·77

* *P* values calculated based on 999 simulations, therefore 0·001 is the smallest possible value.

50% smaller in June, resulting in much smaller *P* values. 30 November, 5 and 11 February all show the census cluster in a completely distinct location from the others. Thus, it may be most helpful to compare the adjustments with the eligible population in these cases. These cases demonstrate the impact of the seasonal adjustment and the smaller but noticeable effect of date-and-tract adjustment compared to the date adjustment.

DISCUSSION

These results indicate that some type of temporal and spatial adjustment is needed when using the space–time scan statistic for syndromic surveillance. During the influenza season, the space–time scan statistic will detect large clusters almost every weekday without adjustment; corresponding results should be expected for other diseases or syndromes with seasonal patterns. This would make it unfeasible to investigate all 'unusual' events, meaningfully diminishing the value of the surveillance.

It is also worthwhile to compare the two adjustments. While each represents a marked difference from the unadjusted approaches, the simpler date adjustment identified twice as many days as the date-and-tract adjustment. Heuristically, the date adjustment will suggest clusters too often around tracts that have a higher baseline risk; they get more visits than expected under this model. In addition, it will be less sensitive to clusters in areas with a lower baseline risk.

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press

Again heuristically, we might consider the question of how the adjustments affect the identified clusters: are they smaller or larger? Do they have more or fewer expected cases? Do they mostly affect location? It may be more effective to consider the question from the other side: How does failing to adjust affect these features of clusters? The clusters found with the census and eligible population generally tend to be spatially larger and to have smaller excess risk than clusters found with either adjustment (results not shown). Their location is not notably different from either adjustment method. We speculate that the effect of not adjusting is generally to make clusters appear to be larger and with smaller excess risk than with appropriate adjustment.

Some mention might profitably be made about edge effects. In spatial analysis, there is often a question about how the edge of the map is handled. Is estimation as good at the edge of the map? Can clusters include areas not appearing on the map? The spatio-temporal scan statistic as implemented in SaTScan treats areas off of the map as if they were water. That is, since they are not included in the analysis, they are treated as if no one lives there, and there is no suggestion that the risk extends beyond the map. Similarly, the ability to detect clusters is not necessarily diminished at the edge.

The generalizability of these results depends on many things. Among these are the region of the country, the population eligible for coverage, the particular year examined, the use of a census population several years removed from the observed data, and the suitability of the adjustments. Of these, the region of the country may attenuate differences between months. However, weekday variability seems likely to generalize. The eligible population considered here is generally more prosperous and better educated than the average in the area. These differences seem unlikely to affect when people get sick. The fact that the census results and the eligible population results were similar suggests that the main problem with unadjusted results is the lack of response to season, rather than the particular constant population used. In addition, if a population shift were the source of alarms, we would expect those alarms to appear mainly in one area or another. This is not the case. Thus, the temporal distance between the census data and the example surveillance is not a likely source of any observed alarms when using the census population.

The year we surveyed could have had an effect on the observed results. For example, the winter of 1998–1999 was atypical in that winter respiratory illness was particularly prevalent in late January and February of that year. This may explain why the adjustments produced about half of their significant clusters in February. In a year with low prevalence of winter respiratory illness, the differences between the adjusted and unadjusted methods might be less striking.

Finally, the model used for adjustment may affect the results. Adjustments from less appropriate models would diminish the advantages of adjusting. Similarly, better models would improve the advantages. The results point to a deficiency in the models: they do not allow for an increased probability of a visit on the day after a holiday. This is demonstrated in the results for 1 June, which was the day after Memorial Day. Improved models would remove the significant cluster on the day after Memorial Day for both adjustments. This change would be simple to make, and in practice we include an indicator variable for the day after a holiday in our predictive model. We omitted it here for simplicity of presentation and to help display the limitations of any model-based approach.

One goal of surveillance such as that used in the example is to discover instances of bioterrorist attack. For this perspective, a failure to use an adjustment such as those proposed here would result in so many signals that any attack would be masked by the seasonally unsurprising signals or alternatively that an attack in the summer would go undetected through a failure to recognize that the count should be smaller in the summer.

We note here that surveillance for biological terrorism, and indeed perhaps any type of surveillance, implies the need for a decision to be made: whether or not to send a team of field investigators to assess a possible cluster of disease, for example. For this purpose, it is convenient to have a simple metric upon which to base the decision. We have followed this observation in this article and used P values to assess the unusualness of events. Our choice of a 0·05 decision level was arbitrary, but we expect that choosing a different arbitrary level would still show large differences between adjusted and unadjusted approaches. We do not discuss the question of what to do when more than one cluster is detected, i.e. when more than one cluster exceeds the arbitrary decision level. In the extreme, all clusters with excess risk might be investigated, starting with the most extreme and progressing down the list as time and resources

allow. We expect that in this case, the date-and-tract adjustment would seem even more distinct from the simpler date adjustment. It would probably result in different orderings of clusters and indeed altogether different clusters than the date adjustment.

This article introduces a model-based space and time adjustment for surveillance using the space–time scan statistic. The modelling technique adjusts the expected cases so that more cases are expected under the null hypothesis when the model suggests they may naturally occur. More generally, we demonstrate the need for time and space adjustment in the important context of surveillance, regardless of the statistical method used.

In summary, for a variety of applications including surveillance for bioterrorism, failure to adjust for seasonality, weekly trends, and local variability can produce so many alarms as to render the surveillance of little use, and can mask some events, leading to undetected true events. However, it should be noted that the relationship between statistical signals generated from this data and aetiology of illness was not investigated in the example. It is possible to adjust scan-statistic-based syndromic surveillance, substantially affecting cluster detection. Ultimately, such systems generate signals that can be evaluated and investigated using traditional epidemiological and public health techniques.

ACKNOWLEDGEMENTS

The authors recognize the improvements in the manuscript prompted by the insightful comments of anonymous reviewers.

APPENDIX. Details of date adjustment

We fit a logistic regression model to the probability of a case on a given day. With 1399 days of observation between 1 January 1996 and 31 October 1999, there were 1399 observations used to fit the model. Covariates were characteristics of days: month, day of week, holiday status, and secular time. The model can be represented as: $E(\sum y_{st}) = n_t P_t$ and $\text{logit}(P_t) = x_t\beta$, where y_{st} is an indicator of whether subject s is a case on day t and the summation is across the n_t subjects eligible on day t. Since n_t is approximately 240 000, it is computationally efficient to do this via the sum than through the individual y_{st} which all share a probability P_t of being a case on day t under the model. The covariate vector x_t contains indicator variables describing day t as above, plus an intercept and a continuous term reflecting the secular time trend.

After estimating the parameters β of the model, the logit is inverted to get the estimated probability of being a case on each given day $= e^{x_t\hat{\beta}}/(1+e^{x_t\hat{\beta}})$. Note that this simple model excludes the possibility of extra-binomial variation due to autocorrelation between consecutive days.

REFERENCES

1. **Lazarus R, Kleinman K, Dashevsky I, et al.** Use of automated ambulatory-case encounter records for detection of acute illness clusters, including potential bioterrorism events. Emerg Infect Dis 2002; **8**: 753–760.
2. **CDC Emergency Preparedness and Response**, Bioterrorism Agents/Diseases (http://www.bt.cdc.gov/agent/agentlist-category.asp). Accessed 24 September 2004.
3. **Kulldorff M.** Prospective time periodic geographic disease surveillance using a scan statistic. J Roy Statist Soc (Series A) 2001; **164**: 61–72.
4. **Kulldorff M and Information Management Services, Inc.** SaTScan: Software for the spatial and space-time scan statistics (computer program). Version 2.1. Bethesda, MD: National Cancer Institute, 1998 (available from http://www.satscan.org).
5. **Kulldorff M, Feuer EJ, Miller BA, et al.** Breast cancer clusters in the northeast United States: a geographic analysis. Am J Epidemiol 1997; **146**: 161–170.
6. **Sheehan TJ, DeChello LM, Kulldorff M, Gregorio DI, Gershman S, Mroszczyk M.** The geographic distribution of breast cancer incidence in Massachusetts 1988–1997, adjusted for covariates. Int J Health Geographics 2004; **3**: 17.
7. **Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D.** Syndromic surveillance in public health practice: the New York City emergency department system. Emerg Infect Dis 2004; **10**: 858–864.
8. **Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V.** Dead bird clustering: a potential early warning system for West Nile virus activity. Emerg Infect Dis 2003; **9**: 641–646.
9. **Lazarus RL, Kleinman KP, Dashevsky I, et al.** Using automated records for rapid detection of illness syndromes: the example of lower respiratory disease. BMC Public Health (serial online) 2001; **1**: 1–9 (http://www.biomedcentral.com/1471-2458/1/9). Accessed 10 December 2003.
10. **Kreiger N, Waterman P, Lemieux K, Zierler S, Hogan JW.** On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. Am J Public Health 2001; **91**: 1114–1116.
11. **Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR.** Evaluating cluster alarms: a space-time scan statistic

and brain cancer in Los Alamos, New Mexico. Am J Public Health 1998; **88**: 1377–1380.

12. **Dwass M.** Modified randomization tests for non-parametric hypotheses. Ann Math Statist 1957; **28**: 181–187.

13. **Kleinman K, Lazarus R, Platt R.** A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. Am J Epidemiol 2004; **159**: 217–224.

https://doi.org/10.1017/S0950268804003528 Published online by Cambridge University Press