Title:

Using scan statistics for congenital anomalies surveillance – the EUROCAT methodology

Authors:

Teljeur C, Dept Public Health and Primary Care, Trinity College Dublin, Ireland.

Kelly A, Dept Public Health and Primary Care, Trinity College Dublin, Ireland.

Loane M, EUROCAT Central Registry, WHO Collaborating Centre for Surveillance of Congenital Anomalies, University of Ulster, UK.

Densem J, Biomedical Computing Ltd, UK.

Dolk H, EUROCAT Central Registry, WHO Collaborating Centre for Surveillance of Congenital Anomalies, University of Ulster, UK.

Corresponding author:

Professor Helen Dolk (h.dolk@ulster.ac.uk), EUROCAT Central Registry, WHO Collaborating Centre for Surveillance of Congenital Anomalies, Institute for Nursing and Health Research, University of Ulster, Shore Rd, Newtownabbey, BT370QB, UK. Tel: +44 (0)2890368540; Fax: +44 (0)28 90368341.

Suggested running head:

Scan statistics in congenital anomaly cluster detection

Abstract

Background

Scan statistics have been used extensively to identify temporal clusters of health events. We describe the temporal cluster detection methodology adopted by the EUROCAT (European Surveillance of Congenital Anomalies) monitoring system.

Methods

Since 2001, EUROCAT has implemented variable window width scan statistic for detecting unusual temporal aggregations of congenital anomaly cases. The scan windows are based on numbers of cases rather than being defined by time.

The methodology is imbedded in the EUROCAT Central Database for annual application to centrally held registry data. The methodology was incrementally adapted to improve the utility and to address statistical issues. Simulation exercises were used to determine the power of the methodology to identify periods of raised risk (of 1 to 18 months).

Results

In order to operationalize the scan methodology, a number of adaptations were needed, including: estimating date of conception as unit of time; deciding the maximum length (in time) and recency of clusters of interest; reporting of multiple and overlapping significant clusters; replacing the Monte Carlo simulation with a lookup table to reduce computation time; and placing a threshold on underlying population change and estimating the false positive rate by simulation. Exploration of power found that raised risk periods lasting 1 month are unlikely to be detected except when the relative risk and case counts are high.

Conclusions

The variable window width scan statistic is a useful tool for the surveillance of congenital anomalies. Numerous adaptations have improved the utility of the original methodology in the context of temporal cluster detection in congenital anomalies.

(254 words)

Introduction

In the context of public health, there is often an interest in identifying clusters of events in space and time as part of monitoring disease incidence. The existence of an unusual grouping of cases may indicate population exposure to some risk factor. Scan statistics have been used extensively to test for changes in the underlying distribution of the data, facilitating the identification of clusters of observed rare events. Scan statistics have been applied to detect clustering in a variety of diseases including cancers, Down Syndrome, suicide, SIDS, and HIV [1].

Congenital anomalies are a major cause of infant mortality, childhood morbidity and long-term disability [2]. They also represent a major cause of foetal death [2, 3]. Congenital anomalies carry a high burden for affected individuals, their families and the community in terms of quality of life, participation in the community and need for services. European Surveillance of Congenital Anomalies (EUROCAT) was established in 1979 as a prototype for European surveillance aiming to assess the feasibility of pooling data across national boundaries, in terms of standardization of definitions, diagnosis and terminology and maintaining confidentiality. A Central Registry manages the central database and develops methodology. Currently there are 37 participating registries in 20 countries covering more than 30% of the European birth population [4]. These are high quality multiple source registries, ascertaining terminations of pregnancy as well as births.

One of the objectives of EUROCAT is to provide the statistical tools required by registries to enable cluster and trend detection, whether that be in relation to medication taken in early pregnancy, exposure to environmental pollution or other potential teratogenic hazard. Monitoring for trends has been described in detail elsewhere [5]. The objective of this paper is to describe the methodology adopted by EUROCAT for cluster detection, and the specific adaptations that have been made over time to improve the utility of the tools and to address statistical issues that have arisen. The investigation of the clusters detected is described in an accompanying paper [6].

Methods

EUROCAT defines a temporal cluster as "an unusual aggregation of cases in time", adapted from the CDC definition [7]. Thus, the term cluster is used in a statistical sense rather than for describing a set of cases with a known common causal agent.

A number of methodologies are available for cluster detection, a number of which rely on the definition of a base-line or expected rate [8]. Defining a base-line can pose difficulties, particularly when the methodology is intended to be used across many countries for a rare outcome. Appropriately resetting the base-line after each cluster test run may be challenging and contentious. To avoid such issues, it was decided to adopt a methodology that would not require a regular base-line adjustment. Discussions amongst a group of interested statisticians from the registries during 1997 considered several alternatives [9], and then, after rigorous testing, opted to adapt the variable window scan technique in 2001.

In general the primary goal for a technique to detect temporal clusters is to determine if there is a statistically significant excess (or, in some instances, deficit) of cases in a given time period assuming relatively constant geographic and population coverage. This can be tested against a null hypothesis that the distribution of cases is random using a scan statistic [10]. The scan statistic can use a fixed window – testing the alternative of clusters of a fixed size – or a variable window – testing the alternative of clusters of varying size [1,10-12]. If a researcher wished to test the statistical significance of a hypothetical or suspected cluster during a stated time frame, a fixed window scan statistic could be employed. However, in surveillance applications where there may be no prior expectation of a cluster, a variable window width scan offers greater flexibility for routine application and reduces the problem of multiple testing.

Scan statistic

Following agreement in 2001, the variable window width scan statistic described by Nagarwalla [11] was initially adopted and subsequently adapted for routine cluster detection in the EUROCAT central registry and by the individual registries. This method involves the computation of a scan statistic, λ, for a set of dated events

in a given time period (see Appendix for further details). Monte Carlo simulation is used to establish λ_{sig} – the 95% percentile of maximum λ recorded in each simulation. Due to the use of a variable window a large number of events will involve a large number of tests and the method used is self-adjusting for multiple testing through Monte Carlo simulation. The benefit of using variable window width scan statistics rather than single fixed window scan statistics, is the greater sensitivity for detecting a change in the underlying distribution of the observed data [13].

The scan algorithm (adapted from Nagarwalla [11]) was coded in Visual Basic for Applications (see supplementary online material) and embedded into the EUROCAT Central Database (ECD) for annual application to registry data held centrally. The software was made available to individual registries in the EUROCAT Data management Program (EDMP) for local ad-hoc investigations [4,5,14]. The system was designed to run on a routine basis via the software, without a statistician at the interface, and to communicate results clearly and simply. Although a temporal scan statistic is available in the SaTScan software [15], a module implementing the scan statistic could readily be incorporated into the existing data management system and adapted according to user needs.

For an analysis to be possible in the ECD/EDMP software, all cases in the dataset must have complete date of birth information. The scanning period is the most recent five full years of data. The final year is censored as analyses are based on estimated date of conception, rather than birth, thus the scan test includes dates of conception from 1st January in the first year to 31st March in the fifth year. The registry must also have a stable population, with no more than a 10% change in the population size during the scan period.

The ECD/ EDMP software evolved over time to enable it to deal with a variety of technical issues as these were identified following routine application by the central registry and arising from feedback from registry leaders.

Results

Multiple clusters

Early testing revealed that in some datasets in which a statistically significant cluster had been identified, additional clusters with a λ greater than λ_{sig} (where λ_{sig} is the lamda value required to reach the prespecified significance level) might also be found. By reversing the procedure to first establish λ_{sig} before analysing the case data, it was possible to record all clusters with a λ greater than λ_{sig}. These additional clusters might be variants of the most significant cluster (i.e. $\pm$ 1 or more cases and hence a slightly different timeframe) or they could be entirely different clusters. As an example, the method was applied to the Knox data (oesophageal atresia and tracheo-oesophageal fistula) used by Nagarwalla [11] to illustrate the variable window scan (Figure 1). A total of 19 instances in which $\lambda > \lambda_{sig}$ were detected. In this case, all detected clusters include some of the cases included in the most significant cluster.

To distinguish between variants of the same cluster and distinct clusters, all clusters were allocated to groups based on their temporal overlap. The cluster with the largest λ is identified and is allocated to the first group. Each ungrouped cluster is then tested for overlap with the first cluster. Overlap is calculated as the number of cases appearing in both clusters divided by the number of cases that appear in either cluster. If the overlap is greater than or equal to 0.75 then both are given the same group label. If any clusters remain ungrouped then the cluster with the largest lambda is identified and given a new group label and the procedure is repeated until all clusters are grouped. By applying this method to the Knox data, the 19 clusters were allocated to four groups (indicated in Figure 1). It should be noted that when a very small significant cluster is wholly included in a very large cluster, these are considered as separate events as the overlap value of 0.75 would not be reached.

The overlap value of 0.75 was chosen on pragmatic grounds, and it was anticipated that a visual inspection of the overlap would also be used to inform any further analysis.

Computational burden

An early issue that arose was that for small datasets the p-value for a cluster could vary quite substantially across repeated analyses of the same dataset. To overcome this it was stipulated that for datasets with fewer than 50 cases, the scan statistic would be calculated using 9999 rather than 999 simulations. However, with small numbers of cases (e.g. 10) there were still problems with variation in the p-values reported. An initial solution was to bootstrap the p-values. This was achieved by sampling with replacement from the simulated λ values to generate 1000 p-values for a given observed λ. The mean p-value was reported along with 2.5[th] and 97.5[th] percentiles to give the user a sense of the uncertainty associated with the p-value.

Computation time followed a power curve as the total number of cases in the dataset increased. When a dataset contained hundreds of cases, as would occur for large registries when considering relatively common congenital anomalies such as Down Syndrome, the computation time ran to several hours and was prohibitive. The problem of time-consuming simulations was resolved by developing a look-up table to compute the p-value for a given λ for a given total number of cases. This was possible as the computation of λ is independent of the duration of the scanning period. A table of λ values was generated for case total of 7 cases, and from 10 to 700 cases in increments of 5, and p-values from 0.005 to 0.05 (in increments of 0.005) and for p=0.1 (see Table 1 and supplementary online material). Linear interpolation was used to estimate for case counts between tabulated values. The values were based on 50,000 λ values computed for each increment of total number of cases. The introduction of the look-up table solved both the problem of p-value variation and slow computation. Confidence bounds for the p-value are no longer reported.

False-positives

As a Monte Carlo approach is used to determine significance with an alpha level of 0.05, we could expect that approximately 5% of tests would return significant clusters by chance alone. In an effort to produce results that were both relevant and interpretable, it was decided to limit the output to clusters that were no longer than 18 months in duration and overlapped with the most recent two years of data. By simulation it was determined that approximately 2.5% of datasets will produce clusters by chance that meet the restriction criteria.

Power to detect clusters

To determine the power of the methodology to identify periods of raised risk, a simulation exercise was developed. Sample datasets were generated with a random uniform distribution of cases. For a set period of 1, 3, 6 and 18 months, the relative risk of anomalies was increased as a step function by two, three, and four-fold. This was defined as the raised risk period. For a given case count, duration of raised risk and relative risk, 1000 sample datasets were generated and the scan statistic applied to see if the raised risk period was identified. In line with the methodology used in ECD/EDMP, the raised risk period was considered identified if the overlap between a detected cluster and raised risk period was at least 75%. That is, at least 75% of the cases occurring in either are shared by both. The total number of cases over the scanning period was varied in steps of 20 from 10 to 270 cases, to represent the typical range of case counts observed in anomaly subgroups in most registries over a 5 year period. It was found that raised risk periods lasting 1 month are unlikely to be detected except when the relative risk and case counts are high (Figure 2).

Population changes

As the null hypothesis is that cases are distributed at random, there is an explicit assumption that a case is equally likely at any point during the scan period. Two factors directly affect the validity of this assumption: the congenital anomaly detection or ascertainment rate; and the population size of the registry and the corresponding birth rate. The scan period can be restricted to begin after a registry has achieved full ascertainment, which is dependent on completion of a prescribed set of quality-related procedures. In terms of birth rate, some registries have shown substantial year-on-year population growth with a corresponding increase in births while others, due to changes in administrative boundaries, have experienced quite sudden increases in population. The latter case is more difficult to adjust for when basing monitoring on date of conception, since the analysis then needs to take into account whether births were from the original geographic area or the extension.

A generalised scan statistic approach can adjust for the change in underlying population [16]. To apply this methodology the relevant population needs to be known – in this case the number of conceptions, which can be crudely estimated using births' data and subtracting 39 weeks. Although feasible when the population change is gradual, in instances of changing administrative boundaries it becomes very difficult to accurately estimate the appropriate population. In practice, a generalised scan statistic requires the underlying population for periods, such as months. Sudden monthly changes in the population cause edge effects in the adjustment that can make the results difficult to interpret, particularly where an apparent cluster straddles a large change in population size within the registry catchment.

By not adjusting the scan statistic for population change, the main impact is to reduce or increase the p-value of a cluster in periods of lower and higher population, respectively. If a population change has occurred, it is advisable to restrict the scanning period to one of minimal population change. It may also be pragmatic to be less prescriptive about the significance level and perhaps investigate clusters with p-values <0.06 in periods of lower population. In an analysis of simulated case data with a steady year-on-year increase in population and cases, it could be seen that a much larger annual rate of increase is required when there are small numbers of cases annually (Figure 3). When the expected numbers of cases annually is 100, there would have to be a year-on-year population increase of 15.8% to find a cluster when there has been no change in the rate and cases are distributed uniformly. A decision was taken in EUROCAT to exclude registries where there had been a population change of at least ±10% in the most recent 5 years period on the grounds that detected clusters may be due to the population increase alone. This decision led to the exclusion of very few registers, and the main reason for population change larger than this was change in administrative boundaries. Other systems experiencing other patterns of population change might need to make other adjustments.

Length of scan period

The methodology allows the start and end points of the scan period to be specified arbitrarily. However, the choice of start and end points can affect the probability of a cluster being detected. There are reasons to restrict the scan period: substantial changes in population or the case ascertainment rate; a change in the incidence of a given anomaly; and a significant cluster at the beginning of a scan period could effectively mask a cluster later in the scan period. An excess of cases at any point during the scan period increases the baseline rate so that a secondary cluster with a small excess of cases may not be identified as statistically significant. Conversely, using only the most recent data may provide insufficient data to reliably detect a cluster. The plot in Figure 4 shows the largest detectable cluster as a proportion of the scan period. With more cases a longer cluster is detectable but the interpretation of such a long cluster may not be meaningful.

Scan windows are defined by numbers of cases rather than periods of time. As multiple cases can occur on the same day, it is possible to generate a one day cluster. Importantly, the scan period should be chosen *a priori*, as manipulating the scan period *a posteriori* can make a cluster appear more or less significant. To maintain consistency and avoid the arbitrary specification of the scan period, within the EUROCAT system the scan period is always defined by date of conception as January 1st in year 1 to March 31st in year 5 using the most recent 5 birth years of complete data. The EUROCAT software (EDMP) only reports clusters of up to 18 months in length within the most recent 2 years [6].

"Clusters" with a deficit of cases

The scan statistic defined by Nagarwalla specifically defines a cluster as an excess of cases compared to the expected. However, by removing the requirement to find an excess of cases it becomes possible to observe the inverse – periods with a statistically significant lack of cases, or a 'deficit'. This can point to potential data quality issues but may also be a chance deficit. Very few deficits were observed.

Discussion

The methodology implemented for cluster detection in EUROCAT identifies periods with unusually high numbers of events. The primary purpose is to flag possible clusters that require further investigation to

understand whether or not there is a genuine reason for concern [6]. Given that a certain proportion of chance clusters will be reported, the cluster detection method is a filtering tool that draws attention to potentially linked events.

The cluster detection methods used in EUROCAT have been implemented using a transparent and user-friendly interface. The software is straightforward to use by non-specialists and minimises the number of decisions to be made by the user. Rather than being inflexible, the software guides the user to a consistent analytical approach. Some sensible restrictions have been imposed on the output from the program. For example, proposed clusters exceeding 18 months in duration or where the last case in the cluster occurred more than 2 years before the end of the scan period are suppressed and the grounds that they do not provide useful new information. Also, as the scan method is regularly applied, there is little benefit in displaying statistical clusters that have been previously identified and assessed.

A number of adaptations have been made to the original published methodology [11] to address the evolving needs of EUROCAT. For example, multiple clusters can be detected rather than simply the cluster generating the largest lambda value. A strict focus on a single detected cluster might well disguise the fact that other relevant clusters occur at other points in the dataset. In addition, to avoid the substantial computational burden of large numbers of simulations, a lookup table was prepared that facilitates much faster computation. For registries with large populations this adaptation had made an immense difference to the time taken to obtain results.

In the last decade, interest in cluster detection has increased, particularly with regard to bio-surveillance [17]. The focus is frequently on spatio-temporal clusters and early event detection (EED), with a view to pinpointing emerging clusters. Newer methods are also putting greater emphasis on the integration of multiple data sources, particularly for syndromic surveillance [17,18]. Prospective detection is necessary for the timely detection of outbreaks of disease or cases of bioterrorism and reliance on a single data source may not be sufficient [18]. In terms of congenital anomalies, the use of complete diagnostic data leads to a time lag before full case ascertainment is achieved, thereby necessitating a retrospective approach. A spatial dimension is currently only available in EDMP as sub-regions within registries. Use of this facility is limited to local registry use only. A finer spatial resolution would incur data protection issues if applied centrally.

The implemented cluster detection method can also be used as a *de facto* data quality assessment tool. The process of investigating a cluster requires careful data checking that can reveal data entry errors. In many instances, once data errors were corrected the statistical anomaly disappeared [6].

Other solutions are available for computing temporal scan statistics. For example, SaTScan uses a Poisson model to implement the variable window width scan statistic and may output multiple clusters from a single dataset [15]. Although producing similar results, SaTScan could not easily be integrated into the ECD.

Limitations

The scan statistic method used in EUROCAT requires the assumption of negligible changes in the size of the at-risk population over the time horizon used. The use of a relatively short time horizon (no more than 5 years) should limit the amount of likely change to the population size, and the availability of total population data make it possible to monitor where changes are occurring that might impact on the methodology.

The main limitation of the implemented surveillance approach is that it lends itself towards detecting clusters where the underlying anomaly rate suddenly increases and then, after a period, returns to the previous baseline rate. An alternative situation is where there is a gradual increase in the rate of anomalies possibly followed by a gradual decrease and a return to baseline, also referred to as a clinal cluster [10]. While methods exist for the detection of the latter type of cluster [19], to date it has not been attempted to incorporate these into the ECD/EDMP system. The choice of detection method must also be linked to expectations about what causal factors might be giving rise to the clusters and whether a short or prolonged exposure is plausible. At present, the EUROCAT system also carries out an associated trend analysis. To date, formal linking of the results of the

cluster and a separate trend analysis has not been implemented, although results from the trend analysis are involved in the interpretation of the clinical significance of the cluster results. Ideally an overarching methodology would be developed that would incorporate both and possibly adjust one for the other. This may become an area for future development of the system.

Some of the registries involved in EUROCAT cover substantial geographic areas with large populations. Cases in a detected cluster may have occurred at a similar point in time but may be far apart spatially. The software does not offer any spatio-temporal cluster detection and this could only be considered if detailed spatial data were also collected. An individual registry must determine whether cases in an identified cluster might be linked by exposure to a common agent. Cases do not have to be in close spatial proximity to potentially be linked depending on the nature of the exposure and therefore a spatio-temporal approach may not provide improved detection of clusters.

The cluster detection method described here and implemented in EUROCAT follows the Nagarwalla suggestion of a minimum window size of five cases and a maximum window size of n-2 [11]. This minimum has implications when applied to datasets of rare conditions, as is the case for numerous subsets of congenital anomalies [20]. The maximum window implies that seven is the minimum total number of cases accepted for analysis. A larger minimum case count may be warranted, although additional research would be required to determine what might constitute an acceptable minimum. In developing the software, recognition was given to the fact that it would be used primarily by clinicians and others with potentially limited statistical expertise. Furthermore, it was designed to support surveillance rather than individualised analyses. To standardise the approach and minimise the reliance on statistical support, pragmatic decisions were made to create a generic tool. If the scan statistic was used by someone with extensive statistical experience, they may adapt the methodology to suit the circumstances (e.g., by adopting a smaller minimum window size for small datasets). We believe these adaptations will have had a negligible impact on the robustness of the outputs.

An underlying assumption of the scan statistic is that events are equally likely to occur throughout the period of interest. Events are re-scaled to a uniform distribution where cases can take place at any point on the scale between zero and one. In reality, cases occur on discrete days (their estimated date of conception) and, particularly for more common congenital anomalies, multiple cases may occur on the same day by chance. Where the number of cases in a single day is equal to or greater than the minimum window size, adjustments must be made to ensure the scanning window does not span a date fraction of zero, as this would result in a lambda value of infinity. The solution we have employed is to add a nominal value (to represent the hour of the day) to each case. Simulation exercises suggested that this is an acceptable statistical solution. The alternative of distributing same-day cases uniformly across 24 hours would, in theory, be preferred. Clusters of less than a day have no relevance in this surveillance context. Whereas further research could determine the impact of different approaches, the more important issue is that the day of conception (here equivalent to date of Last Menstrual Period) is imprecisely known and calculated from gestational age measured in weeks, requiring introduction of this uncertainty into the model [6]. Clusters of 1-2 days have in practice been rarely detected [6].

Conclusions

The variable window width scan statistic is a useful tool for the surveillance of congenital anomalies. Numerous adaptations to the original methodology have improved the utility of the scan statistic in the context of EUROCAT and the participating registries.

Conflict of Interests:

None.

References

(1) Naus J, Wallenstein S. Temporal surveillance using scan statistics. Stat Med 2006; 25(2):311-324.

(2) Dolk H, Loane M, Garne E. The prevalence of congenital anomalies in Europe. Adv Exp Med Biol 2010; 686:349-364.

(3) Khoshnood B, Greenlees R, Loane M, Dolk H. Paper 2: EUROCAT public health indicators for congenital anomalies in Europe. Birth Defects Res A Clin Mol Teratol 2011; 91 Suppl 1:S16-S22.

(4) Boyd PA, Haeusler M, Barisic I, Loane M, Garne E, Dolk H. Paper 1: The EUROCAT network--organization and processes. Birth Defects Res A Clin Mol Teratol 2011; 91 Suppl 1:S2-15.

(5) Loane M, Dolk H, Kelly A, Teljeur C, Greenlees R, Densem J. Paper 4: EUROCAT statistical monitoring: identification and investigation of ten year trends of congenital anomalies in Europe. Birth Defects Res A Clin Mol Teratol 2011; 91 Suppl 1:S31-S43.

(6) Dolk H, Loane M, Teljeur C, Densem J, Greenlees R, McCullough N et al. Detection and investigation of temporal clusters of congenital anomaly in Europe: seven years of experience of the EUROCAT surveillance system. Epidemiology 2015 [Epub ahead of print, PMID: 25840712).

(7) Centres for Disease Control and Prevention. Guidelines for Investigating Clusters of Health Events. Morbidity and Mortality Weekly Report 1990; 39(RR-11):1-16.

(8) Sego LH, Woodall WH, Reynolds MR. A comparison of surveillance methods for small incidence rates. Statist Med 2008; 27(8):1225-1247.

(9) Quataert PK, Armstrong B, Berghold A, Bianchi F, Kelly A, Marchi M, Martuzzi M, Rosano A. Methodological problems and the role of statistics in cluster response studies: a framework. Eur J Epidemiol. 1999 Oct;15(9):821-31.

(10) Tango T. Tests for temporal clustering. In: Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, editors. Statistical Methods for Disease Clustering. New York: Springer; 2010. 49-70.

(11) Nagarwalla N. A scan statistic with a variable window. Stat Med 1996; 15(7-9):845-850.

(12) Naus JI. The Distribution of the Size of the Maximum Cluster of Points on a Line. Journal of the American Statistical Association 1965; 60(310):532-538. DOI:10.1080/01621459.1965.10480810

(13) Glaz J, Zhang Z. Multiple Window Discrete Scan Statistics. Journal of Applied Statistics 2004; 31(8):967-980. DOI:10.1080/0266476042000270536

(14) McCullough N, Loane M, Greenlees R, Dolk H. EUROCAT Statistical Monitoring Report – 2009. 2012. Newtown abbey, Northern Ireland, EUROCAT Central Registry.

(15) Kulldorff M. SaTScan User Guide for version 9.3. 2014. Boston, USA.

(16) Weinstock MA. A generalised scan statistic test for the detection of clusters. Int J Epidemiol 1981; 10(3):289-293.

(17) Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2012; 175(1):49-82. DOI: 10.1111/j.1467-985X.2011.00714.x

(18) Fricker RD. Some methodological issues in biosurveillance. Statist Med 2011; 30(5):403-415.

(19) Tango T. A test for spatial disease clustering adjusted for multiple testing. Stat Med 2000; 19(2):191-204.

(20) Cucala L. A hypothesis-free multiple scan statistic with variable window. Biom J 2008; 50(2):299-310.

Appendix: The scan statistic defined by Nagarwalla [11]

An analysis requires the dates of a set of cases, and the specification of the start and end dates for the analysis period. For a given subset of n cases occurring between $date_{start}$ and $date_{end}$, it is possible to calculate the test statistic lambda as follows:

$$\lambda = \left(\frac{n}{r}\right)^n \left(\frac{r-n}{n}\right)^{r-n} \left(\frac{1}{d}\right)^n \left(\frac{1}{1-d}\right)^{r-n}$$

Where r is the total number of cases, and

$$d = \frac{date_{last} - date_{first}}{date_{end} - date_{start}}$$

The dates $date_{first}$ and $date_{last}$ are the dates of the first and last cases in the scanning window, respectively. Typically a minimum scanning window of 5 is used. Lambda is calculated for every possible subset of n consecutive cases for $n = 5,\ldots, r$

The lambda for each subset of cases is calculated and the largest found is recorded along with the details of the start and end dates. Next, simulated datasets are generated each with r uniform random numbers in the interval [0,1]. For each simulation, the lambda for each subset is calculated and the largest lambda is recorded. After, for example, 999 iterations the lambdas are ordered by size from smallest to largest and the 95th percentile is noted as λ_{sig}. If the lambda from the real dataset is greater than to λ_{sig} then the cases are designated as a significant cluster.

The method assumes a uniform distribution of cases and hence no seasonal variation.

Multiple cases in the observed data can occur on the same day. Where the number of cases in a single day is equal to or greater than the minimum window size, the d value for that window will equal zero and hence lambda will equal infinity. A nominal value equivalent to an hour is added to each case that occurs on the same day, thereby distributing multiple cases. An upper limit of 24 cases within a day is a reasonable assumption in our context. The alternative of distributing same-day cases uniformly across 24 hours would, in theory, be preferred.

It should be noted that different statistical terminology was adopted in EUROCAT to avoid confusion with variable names in computer coding. Hence our notation uses 'r' in place of 'N'. To maintain consistency between the code and this article, we have used same terminology in the code and the paper.

Figures

Figure 1. Detection of multiple clusters in oesophageal atresia and tracheo-oesophageal fistula

Notes:

Vertical dashed lines indicate start and end cases included in most significant cluster.

Numbers indicate cluster group that each detected cluster is assigned to.

Figure 2. Probability of detecting a fixed period of increased risk as a function of baseline risk and relative risk

Figure 3. Population change required to find a cluster in uniformly distributed cases

Figure 4. Largest detectable cluster as a proportion of the scan period