CA_Ch1: Introduction to investigation of spatial clusters of disease

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

1.1 Background

"Epidemiologists frequently investigate spatial patterns of disease to identify clusters of cases for control purposes and to develop hypotheses about factors that contribute to their occurrence. Many factors may determine the spatial distribution of disease occurrences, such as the density and species of animal populations, movement and trade in these populations, geographic and climatic factors that affect the animals, their environment or disease vector species, or existing disease control programmes. These factors should be considered when applying methods to evaluate the clustering of disease.

The history of methods to detect clusters of disease has grown since the 1980's out of increased concerns about actual or potential adverse environmental effects of contaminants or toxins on public health, for example, aerial deposition of particles downwind from nuclear power stations. However, many diseases will show geographic (and possibly temporal) clustering for other reasons that are associated with the disease, and not the one postulated. Hence, in some situations, diseases may appear to cluster, even when the known aetiology doesn't suggest it should be observed."

CONTINUE

1.2 Reasons for studying disease clustering

Study of disease aetiology

Differential disease rates in large scale or localised geographic areas have long been used to study disease aetiology. The same methods can be used to detect areas with high rates of disease in which to conduct further epidemiologic studies (especially cohort studies for which it is important to enrol subjects with a relatively increased risk of disease to achieve sufficient statistical power in a cost-efficient way).
A key feature of these methods for purely spatial or space-time data is that they are able to pinpoint the location of clusters. However, the interest may alternatively be in general nature of distribution of disease, for example, to understand whether the disease is likely to be infectious (may exhibit both space-time interaction and spatial clustering), or has risk factors that vary geographically (only spatial clustering). When assessing the association between geographic risk variables and disease risk in ecological analysis, it is important to adjust for any spatial autocorrelation not explained for by the known variables.

Population surveillance and outbreak response

The detection of clusters of disease with known risk factors may prompt a government-led health response, where those factors can be controlled, for example from water-borne diseases or environmental pollution, or an effective vaccination programme in the face of an infectious disease. Additionally, geographic areas with high mortality rates may be searched for, to identify areas of substandard treatment or screening. Alternatively, areas of reduced risk may indicate successful treatment or screening programmes from which others can learn, or they may alternately reflect areas of under-reporting of cases.

"In each setting, it is important to account for the population at risk of the disease in the study. This is because the distribution of cases alone may just reflect the underlying distribution of the population at risk of being a case. Therefore, the population at risk needs to be taken into account when calculating the actual incidence or risk of disease."

CONTINUE

1.3 Definition of clustering

When the exact extent or form of the clusters to be studied are unknown, a cluster is:

"A geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance" Lawson and Kulldorff (1999)

Or less formally- an area within the study region of significantly elevated risk of disease over a given period of time

Or clustering is also known as a ‘hot-spot’

CONTINUE

1.4 Spatial statistical concepts

- [ ] A spatial process is stationary when the dependence between measurements of the same (outcome) variable is the same for all locations in the study area
- [ ] A spatial process is isotropic when it is affected by distance from a location, but the effect is the same in all directions
- [ ] A first-order effect describes large-scale variations in the mean pattern of occurrence of a variable across the study region
- [ ] A second-order effects describes small-scale variation due to interactions between neighbours

CONTINUE

1.5 Methods to investigate spatial clustering of disease

"A range of methods exist to investigate spatial and temporal clustering of disease because of the various data types that might be obtained from epidemiologic studies, for example point locations or area-level aggregated data, and because of different objectives of the analyst.

The types of cluster detection methods have been categorised by Besag and Newell (1991) and Tango (1999) as either general (or global or non-specific) or localised, which in turn are further categorised as localised non-focused or localised focused (Table 1.1)."

Table 1.1: Summary of methods to detect clustering of spatial data

Feature	General or global or non-specific	Localised non-focused	Localised focused

Feature	General or global or non-specific	Localised non-focused	Localised focused
Aim	To investigate the overall tendency of of a disease to cluster in a study region, regardless of where that clustering is located	To define the location and intensity of any clusters if they exist	To investigate whether one or more clusters occur in relation to a pre-determined focus point
	Is analogous to and uses the same methods used for the assessment of autocorrelation of spatial data		
Provide information on	Overall or global aspects of clustering	The spatial locations of clusters	The spatial locations of clusters
		Local variability of clustering	Local variability of clustering
Don't provide information on	The spatial locations of clusters		
	Local variability of clustering		
Assumptions	The disease process is relatively uniform (stationary) throughout the study region	Analyst needs to set parameters for the shape and size of the search window	The source of infection of disease is a pre-determined point
Practical example of use	Preliminary analysis of spatial disease data before using localised cluster detection methods	Evaluation of FMD outbreak reporting to identify locations of clusters of cases	To investigate whether livestock markets or transport routes are possible sources of infectious diseases

CONTINUE

Several problems are likely to be met when attempting to evaluate whether clustering occurs in spatial disease data.

1. Cases of disease may be rare or distributed over an extended period of time, which means that they may not be detected.

2. The information on the population at risk may be unavailable or of poor quality so that the non-cases or controls can't be adjusted for. This is more likely the case when the analysis is at the farm-level which may not be geo-located, rather than an administrative location like a village which is often mapped.

3. Searching for spatial clusters involves decisions about the size and shape of the search window within which local clusters are detected. Finally, when dealing with endemic diseases, the background occurrence of disease in the population may be unknown so as to determine whether the currently observed occurrence is elevated above what might be expected normally.

CONTINUE

"A range of statistical methods are used to investigate and describe spatial clustering of disease. However, in this course we will only consider visual exploratory spatial data analysis (ESDA) methods and one statistical method to identify and describe any localised non-focused spatial clusters of disease.

This is because this category of spatial clustering analysis is likely to be most commonly used in disease surveillance, and because the other methods require more advanced statistical training and experience with additional computer software applications which are beyond the scope of this course."

CONTINUE

"It is important to consider the structure of hypotheses that could include clustering components in any analysis of geographic health data. If the disease of interest naturally clusters (beyond that explained by the background population), then this form of clustering should be investigated. This form of clustering may arise from unobserved covariates and should be considered as

heterogeneity, and can often be modelled through the use of random effect regression models, although these models that aim to explain the observed patterns of disease won't be taught in this course."

CONTINUE

1.6 Exploratory spatial data analysis methods to detect clusters of disease

"The first step in an investigation of spatial patterns of disease occurrence is exploratory data analysis (ESDA), initially by simply creating a map of disease case locations. Such a map provides readily-understood information of the overall distribution of disease in the study area, and in particular, any variation in the number of cases between regions within the study area to locate where the main disease burden lies.

When there are a large number of cases of disease, the point locations may be plotted over one another which obscures any pattern. An additional GIS step to aid interpretation of a map in these cases is to smooth the distribution of cases by kernel-smoothing methods."

CONTINUE

"A second ESDA step is to plot the distribution of both case and non-case locations to gain an impression of whether the distribution of cases merely reflects the distribution of the underlying population at risk rather than an actual local increase in risk compared to other regions in the study area.

However, the visual appearance of maps drawn in this way may be difficult to interpret and is subject to user-defined parameters such as colour and size of points on the map, and particularly the diameter of areas over which the kernel-smoothing function is set to operate on."

CONTINUE

Congratulations - end of lesson reached

References

1. Besag, J., and J. Newell. 1991. The detection of clusters in rare diseases. Journal of the {R}oyal {S}tatistical {S}ociety {S}eries {A}-{S}tatistics in {S}ociety 154:143–155. doi:10.2307/2982708.
2. Lawson, A.B., and M. Kulldorff. 1999. A Review of Cluster Detection Methods. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J.-F. Viel, and R.

Bertollini, ed. John Wiley & Sons, Ltd, Chichester, United Analyst (Cambridge, United Kingdom).

3. Tango, T. 1999. Comparison of general tests for spatial clustering. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J.-F. Viel, and R. Bertollini, ed. John Wiley & Sons, Ltd, Chichester, United Analyst (Cambridge, United Kingdom).

NEW ZEALAND
FOREIGN AFFAIRS & TRADE

OIE

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Australian Government
Department of Agriculture, Water and the Environment