Spatial Point Pattern Analyses and its Use in Geographical Epidemiology

Murat Yazici*

Data Scientist& Researcher, Turkey

Submission: March 31, 2017; **Published:**May 15, 2017

*Corresponding author: Murat Yazici, Data Scientist & Researcher, Turkey, Email: muratyaz@bu.edu

Abstract

Spatial epidemiology is a subfield of health geography focused on the study of the spatial distribution of health outcomes. Point pattern analysis is the evaluation of the pattern, or distribution, of a set of points on a surface. It can refer to the actual spatial or temporal location of these points or also include data from point sources. It is one of the most fundamental concepts in geography and spatial analysis. In this study, geographical epidemiology, data for spatial analysis, disease clustering, disease mapping are introduced. This paper also includes applications in environmental epidemiology and geographical epidemiology with spatial point pattern analysis.

Keywords: Spatial analyses; Environmental epidemiology; Point process; Clustering; Modeling

Introduction

The analysis of spatial point patterns came to prominence in geography during the late 1950s and early 1960s, when a spatial analysis paradigm began to take firm hold within the discipline. Researchers borrowed freely from the plant ecol-ogy literature, adopting techniques that had been used there in the description of spatial patterns and applying them in other contexts: for example, in studies of settlement distributions [1,2], the spatial arrangement of stores within urban areas [3] and the distri-bution of drumlins in glaciated areas [4]. The methods that were used could be clas-sified into two broad types [5]. The first were distance-based techniques, using information on the spacing of the points to characterize pattern (typically, mean distance to the near-est neighbouring point). Other techniques were area-based, relying on various characteristics of the frequency distribution of the observed numbers of points in regularly defined sub-regions of the study area ('quadrats'). For many geographers, point pattern analysis will conjure up images of nearest-neighbour analysis applied inappropriately to data sets of doubtful relevance [6].

In this paper, geographical epidemiology is introduced in section 2. In section 3, spatial point pattern analysis and geographical epidemiology is explained. Conclusion is discussed in section 4. The last part of the paper includes references.

Geographical Epidemiology

Geographical epidemiology can be defined as the description of spatial patterns of disease morbidity and mortality, part of descriptive epidemiological studies, with the aim of formulating hypotheses about the aetiology of diseases [7]. One can identify different branches in geographical epidemiology, which is a reflection of the different needs of public health specialists and epidemiologists in the assessment of ill-health aetiology. Predominant among the methods of geographical epidemiology are the following: disease mapping, disease clustering and ecological analysis [8]. There is usually a close relationship between these branches. However, as almost all geographical epidemiological studies are descriptive in nature and depend on scale, one should bear in mind that a more comprehensive picture of a spatial problem can be achieved when the results of geographical aggregatelevel data are combined with those at the individual level [9]. Multilevel modelling, hierarchical regression and contextual analysis are phrases describing one of the various statistical methods in which this combination is allowed. Multilevel modelling is a powerful, relatively new technique that can be used to determine how much of the ecological effect can be explained by variations in the distribution of individual-level risk factors, and recently attempts have been made to integrate this kind of analysis into geographical epidemiology [10,11]. There are also new developments incorporating time changes along with spatial variation. Such models are able to provide new insights into the aetiology of diseases that are otherwise unavailable [12,13].

Data for spatial analysis

There are usually two important types of spatial data: point and area data. Each item of health data (including population, environmental exposure, mortality and morbidity) may be

connected with a point, or precise spatial position such as a home, a street address or an area, which could be defined as a spatial region by postcode, ward, local authority, province and country [14]. A public health specialist may also come across spatial data in the form of continuous surface, such as the statistical surfaces of pollution interpolated from fixed-point characteristics [12]. As data for spatial analysis come from different sources, and have often been collected without taking into account the interests of the geographical epidemiologists, it is absolutely necessary to ensure that precise and complete point and/or area health data are used in spatial epidemiology [15,16]. In the developed world, most of the mortality and cancer incidence data have good quality. Nevertheless, other health data such as rates of suicide, congenital anomalies and hospital admissions may be subject to partial ascertainment (rates are underestimated). In addition, the diagnosis, collection, coding and reporting of a given health outcome may differ between geographical regions and over time [17].The danger of ignoring data-quality issues is that, because of missing cases or inaccurate baseline population data, one might arrive at a misleading (invalid) high or low estimated risk [18,19]. Confidentiality may also be an important issue. Breaching the confidentiality of spatial data may cause concern, especially when it discloses areas with high rates of morbidity/ mortality or high levels of pollutants [14].

Disease clustering

Searching for disease clustering is one of the branches of geographical epidemiology that involves an assessment of local or global accumulation of disease [20]. There are different types of clustering, including general and specific. General clustering involves the analysis of the overall clustering tendency of the disease incidence in a study region, and is paralleled by the assessment of global spatial autocorrelation, in which the exact location of clusters is not investigated. The second type of investigation of clustering uses specific disease-clustering methods, which are designed to examine the exact location of the clusters [21]. As we will discuss the importance of, and the ways of detecting, global and local clustering in areal data in the section below on spatial autocorrelation, here we will focus only on the detection of clusters in point data. Methods for the detection clusters in point format data are more numerous than those for areal format data, and are usually divided into the following three groups: global, localised and focused (ie, assesses clustering around a putative source [22]. There are a number of tests available that help to assess different kinds of clusters in point format data. However, we will discuss only three of them very briefly, and refer the readers to Bailey and Gatrell, and Gatrell et al for a complete discussion. Cuzick and Edwards method determines global clustering by examining the k nearest neighbours of each case. The geographical analysis machineand the spatial scan statistic assess the localised clustering by drawing circles of different sizes over the area of study and compare the risk of disease inside and outside of

each circle [23,24]. The spatial scan statistic has an advantage over geographical analysis machine in taking into account the problems of multiple testing [25].

Disease mapping

Data visualisation is the first step in disclosing the complex structure in data [26]. Data visualisation may not only create interest and attract the attention of the viewer but also provide a way of discovering the unexpected [27]. Although plots of data and other graphical displays are among the fundamental tools for analysts in general, for a spatial analyst, visualising spatial data usually means using a map [6].Disease mapping is one of the branches of geographical epidemiology fulfilling the need to create accurate maps of disease morbidity and mortality [28]. For instance, dot or dot-density maps are used to display point data, whereas choropleth maps are used for areal data, and contour or isopleth maps are used for continuous surface data [12]. The use of mapping in the medical context has developed so rapidly during recent decades that the presentation of maps is now established as a basic tool in the analysis of public health data [8,29].

There are two main classes of disease maps for areal data: maps of standardised rates and maps of statistical significance of the difference between disease risk in each area and the overall risk averaged over the whole map [7]. There are pros and cons for each of these classes. For instance, mapping rates in small areas tend to create a misleading picture (see the section Smoothing) while using statistical significance, particularly in areas with large populations, produce small p values indicating statistical significance, but do not disclose scientifically interesting differences [30]. The mapping of standardised rates is generally preferred to the mapping of p values, controlling for the influence of sampling variation by using a smoothing technique (see the section on Smoothing) [31].

Applications in environmental epidemiology

Epidemiology is concerned with the study of the distribution and determinants of health and diseases, morbidity, injuries, disability, and mortality in populations [32]. Epidemiologic studies are applied to the control of health problems in populations. Epidemiology is one of the core disciplines used to examine the associations between environmental epidemiology refers to the study of diseases and health conditions (occuring in the population) that are linked to environmental factors [33,34]. The exposures, which most of the time are outside the control of the individual, usually may be considered involuntary and stem from ambient and occupational environments [35]. According to this conception of environmental epidemiology, standard epidemioloc methods are used to study the association between environmental factors (exposures) and health outcomes. Examples of topics studied include air and water pollution, the occupational environment with its possible use of physical and

chemical agents, and the psychosocial environment (Rothman K.J.).Some of using techniques in environmental epidemiology are as follows;

    a. Spatial clustering

    b. Space-time interaction

    c. Modelling the raised incidence of disease

Spatial Point Pattern Analysis and Geographical Epidemiology

The behavior of a general spatial distribution process can be characterized in terms of its first-order and second-order properties. First-order properties describe the spatially varying intensity of a point pattern, in which intensity is defined as the expected (mean) value of the distribution at locations throughout the region of interest [36]. Second-order properties describe the covariance (or autocorrelation) structure of the point pattern and can be identified by analyzing the distribution of distances between those sample points [2,37,38].

Ripley's K function is regarded as a suitable tool to characterize second-order properties of a point pattern. It is the expected number of points in a circle of radius d with a random point at center, and is formally defined as [39]:

$$K(d) = \lambda^{-1} E(f)$$

where $E(f)$ is the expected number of other sample points within distance d of a sample random point λ and is the intensity of sample points per unit area. Since the expected number of sample points within a distance d of a chosen random point in a process with no spatial dependence is $\lambda \pi d^2$, $K(d)$ for a spatially random process can be defined as [40,41]:

$$K(d) = \pi d^2$$

If the points display a clustered pattern, an excess of sample points at short distances can be shown. The empirical function $\hat{K}(d)$ is defined as:

$$\hat{K}(d) = n^{-2} A \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \varpi_{ij}^{-1} I(u_{ij})$$

where is the number of events in the analyzed plot(m^2), A I is the area of the plot , is a counter variableu_{ij} , is the distance between events i and jϖ_{ij} , and is a weighting factor to correct for edge effects. In our study, toroidal edge correction was used to avoid the edge effects by treating the rectangular study plot encompassing the study region as a torus, that is, the part of a sample outside the rectangle is made to appear at the corresponding opposite border [42]. Points at opposite sides of the plot are now close to each other and the boundary does not exist [43] (Figure 1).
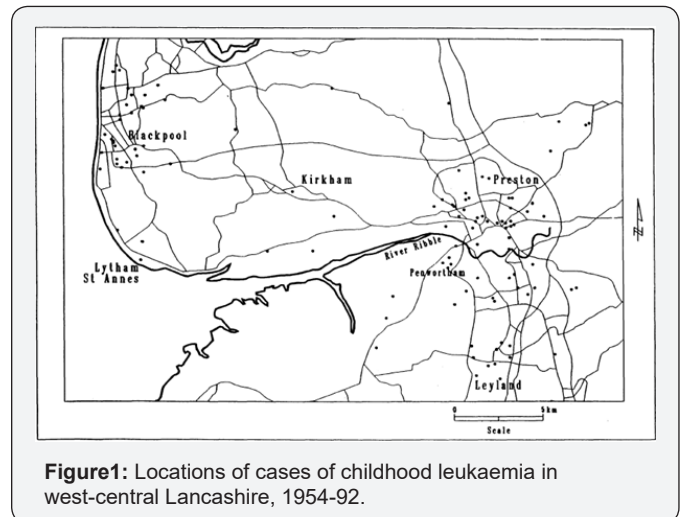

Figure1: Locations of cases of childhood leukaemia in west-central Lancashire, 1954-92.

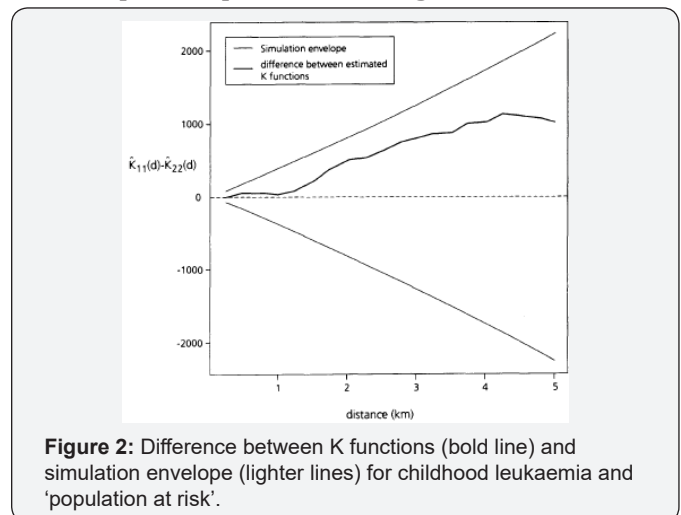## An Example for Spatial Clustering

Figure 2: Difference between K functions (bold line) and simulation envelope (lighter lines) for childhood leukaemia and 'population at risk'.

The question of whether the geographical inci-dence of disease shows any tendency towards clustering in geographical space has a long and rich history Figure 2. Do cases of disease tend to occur in proximity to other cases? The problem has become more urgent in recent years in the light of concerns raised about possible links between disease inci-dence and potential sources of environmental con-tamination, such as nuclear installations. Evidence of clustering might also lend support to other theories of disease incidence, such as a viral aetiol-ogy. For example, exposure to a common, persist-ent viral infection, either during gestation or as a young child with an immune system that had been protected at a very early age, might provide clues to explaining possible leukaemia clustering [44-49].

Conclusion

3In this paper, geographical epidemiology, data for spatial analysis, disease clustering, disease mapping are explained. Also,

applications in environmental epidemiology and geographical epidemiology with spatial point pattern analysis including spatial clustering are introduced. Example is given Gatrell et al.'s studies. I would like to thank you them for their enlightening information. For future work, spatial point process analysis can be examined for geographical epidemiology.

References

1. Dacey MF (1962) Analysis of central place and point patterns by a nearest neighbour method. Lund Studies in Geography Series B Human Geography 24: 55-75.

2. King LJ (1962) A quantitative expression of the pattern of urban settlements in selected areas of the United States. Journal of Economic and Social Geography 53: 1-7.

3. Rogers A (1965) A stochastic analysis of the spatial cluster-ing of retail establishments. Journal of the American Statistical Association 60(312): 1094-1102.

4. Trenhaile AS (1971) Drumlins: their distribution, orien-tation and morphology. Canadian Geographer 15(2): 113-26.

5. Haggett P, Cliff AD, Frey AE (1977) Locationalm ethods in human geography. Edward Arnold, London.

6. Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Harlow: Longman, London.

7. Clayton D, Bernardinelli L (1996) Bayesian methods for mapping disease risk. In: Elliott P, Cuzik J, English D and Stern R (Eds.), Geographical and environmental epidemiology-methods for small area studies. Oxford: Oxford University Press, USA.

8. Best N (1999) Bayesian ecological modelling. In: Lawson AB, et al (Eds.), Disease mapping and its uses. Disease mapping and risk assessment for public health. Chichester: Wiley, 1999: 193-201.

9. Dooley D, Catalano R, Rook K (1989) Economic stress and suicide: multilevel analyses, Part 1: aggregate time-series analyses of economic stress and suicide. Part 2: cross-level analyses of economic stress and suicidal ideation. Suicide Life Threat Behav 19: 321-36.

10. Morgenstern H (1998) Ecological studies. In: Rothman KJ, Greenland S (eds.), Modern epidemiology. Philadelphia: Lippincott-Raven Publisher pp. 459-80.

11. Merlo J, Yang M, Chaix B, Lynch J, Råstam L (2005) A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. J Epidemiol Community Health 59(9): 729-736.

12. Jerrett M, Burnett RT, Goldberg MS, Sears M, Krewski D, et al. (2003) Spatial analysis for environmental health research: concepts, methods, and examples. J Toxicol Environ Health A 66(16-19): 1783-1810.

13. Rezaeian M, Dunn G, St Leger S, Appleby L (2004) The production and interpretation of disease maps: a methodological case-study. Soc Psychiatry Psychiatr Epidemiol 39(12): 947-954.

14. Elliott P, Wartenberg D (2004) Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 112(9): 998-1006.

15. Staines A, Jarup L (2000) Health event data. In: Elliott P, et al. (Eds.), Spatial epidemiology, methods and applications. Oxford: Oxford University Press, USA, pp. 15-29.

16. Elliott P, Wakefield JC, Best NG, et al. (2000) Spatial epidemiology: methods and applications. In: Elliott P, Wakefield JC, Best NG, Briggs DJ (Eds.), Spatial epidemiology, methods and applications. Oxford: Oxford University Press, UK, pp. 3-14.

17. Jarup L (2004) Health and environment information systems for exposure and disease mapping, and risk assessment. Environ Health Perspect 112(9): 995-997.

18. Lawson AB (2001) Tutorial in biostatistics: disease map reconstruction. Stat Med 20(4): 2183-2204.

19. Walter SD (1993) Assessing spatial patterns in disease rates. Stat Med 12(19-20): 1885-1894.

20. Lawson AB, Bohning D, Biggeri A (1999) Disease mapping and its uses. In: Lawson AB, Bohning D, Biggeri A(Eds.), Disease mapping and risk assessment for public health. Chichester: Wiley, US.

21. Besag J, Newell J (1991) The detection of clusters in rare disease. J R Stat Soc A 154(1): 143-155.

22. Cromley EK, McLafferty SL (2002) GIS and public health. The Guilford Press New York, USA.

23. Openshaw S, Charlton M, Wymer C, et al. (1987) A mark 1 geographical analysis machine for the automated analysis of point data sets. Int J Geogr Inf Syst 1(4): 335-358.

24. Kulldorff M (1998) Statistical methods for spatial epidemiology: tests for randomness. In: Gatrell A, Loytonen M(Eds.), GIS and health. Taylor & Francis, London, pp. 49-62.

25. Rezaeian M, Dunn G, Leger S, Appleby L (2007) Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. J Epidemiol Community Health 61(12): 98-102.

26. Cleveland WS (1993) Visualising data. Summit, Hobart Press, NJ, USA.

27. Everitt BSE, Dunn G (2001) Applied multivariate data analysis. In: Arnold (Ed.), London.

28. Lawson AB, Williams FLR (2001) An introductory guide to disease mapping. In: Wiley (Ed.), Chichester, England.

29. Cliff AD (1995) Analysing geographically related disease data. Stat Methods Med Res 4(2): 93-101.

30. Bithell JF (1998) Geographical analysis. In: Armitage P & Colton T (Eds.), International encyclopaedia of biostatistics. Wiley, Chichester, England, pp. 1701-1716.

31. Cartwright RA, Alexander FE, McKinney PA, (1990) Leukaemia and lymphoma: an atlas of distribution within areas of England and Wales. Leukaemia Research Fund, Leeds, pp. 1984-1988.

32. Friis RH, Sellers TA (2009) Epidemiology for Public Health Practice (4th edn), Sudbury, Jones and bartlett Publishers, MA.

33. Pekkanen J, Pearce N (2001) Environmental epidemiology: challenges and opportunities. Environ Health Perspect 109(1): 1-5.

34. Terracini B (1992) Environmental epidemiology: a historical perspective. In: Elliott P, Cuzick J, English D, Stern R (Eds.), Geographical and Environmental Epidemiology: Methods for Small-area Studies. Oxford University Press, New york, USA.

35. Acquavella JF, Friedlander BR, Ireland BK (1994) Interpretation of low to moderate relative risks in environmental epidemiologic studies. Ann Rev Public Health 15: 179-201.

36. Diggle PJ (1983) Statistical Analysis of Spatial Point Patterns. Oxford University Press Inc, New York, USA.

37. Wiegand T, Moloney KA (2004) Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104(2): 209-229.

38. Juan P, Mateu J, Saez M (2012) Pinpointing spatio-temporal interactions in wildfire patterns. Stoch. Environ Res Risk Assess 26(8): 1131-1150.

39. Ripley BD (1977) Modelling spatial patterns. J R Stat Soc Series B Stat Methodol 39(2): 172-212.

40. Diggle PJ, Chetwynd AG, Haggkvist R, Morris S (1995) Second-order analysis of space-time clustering. Statistical Methods in Medical Research 4(2): 124-136.

41. Getis A (1983) Second-order analysis of point patterns: The case of Chicago as a multi-center urban region. 35(1): 73-80.

42. Wiegand T, Kissling WD, Cipriotti PA, Aguiar MR (2006) Extending point pattern analysis for objects of finite size and irregular shape. J Ecol 94(4): 825-837.

43. Haase P (1995) Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. J Veg Sci 6(4): 575-582.

44. Gatrell AC, Bailly TC, Diggle PJ, Barry S (1996) Spatial point pattern analysis and its application in geographical epidemiology. Trans Inst Br Geogr 21(1): 256-274.

45. Wakefield JC, Best NG, Waller L (2000) Bayesian approach to disease mapping. In: Elliott P, Wakefield JC, Best NG, Briggs DJ (Eds.), Spatial epidemiology, methods and applications. Oxford: Oxford University Press, USA, pp. 104-127.

46. Rothman KJ (1993) Methodologic frontiers in environmental epidemiology. Environ Helath Perspect 101(suppl 4): 19-21.

47. Merlo J, Chaix B, Yang M, Lynch J, Råstam L (2005) A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. J Epidemiol Community Health 59(12): 1022-1028.

48. Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations. J R Stat Soc B 52(1): 73-104.

49. English D (1996) Geographical epidemiology and ecological studies. In: Elliott P, Cuzik J, English D, Stern R (Eds.), Geographical and environmental epidemiology- methods for small area studies. Oxford: Oxford University Press, 3-13.

Your next submission with Juniper Publishers
will reach you the below assets
- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php