CONTINUING PROFESSIONAL EDUATION

Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary

Mohsen Rezaeian, Graham Dunn, Selwyn St Leger, Louis Appleby

....

We provide a relatively non-technical glossary of terms and a description of the tools used in spatial or geographical epidemiology and associated geographical information systems. Statistical topics included cover adjustment and standardisation to allow for demographic and other background differences, data structures, data smoothing, spatial autocorrelation and spatial regression. We also discuss the rationale for geographical epidemiology and specific techniques such as disease clustering, disease mapping, ecological analyses, geographical information systems and global positioning systems.

....

''Place'' can usually be applied as a surrogate for the interaction between genetic factors, lifestyle and environment.[1] Although the role of place in human health has been recognised historically,[2] the focus in public health research has mostly been on person and time, with little consideration of the implications of place.[3] Most public health specialists seem to have forgotten the space dimension of disease processes.[4] This is a pity, as comparison between places, together with comparisons between times and between individuals, is a useful means of formulating and testing aetiological hypotheses. In addition, from the perspective of public health practice, knowledge that a health problem is concentrated in identifiable places is essential for the efficient distribution of resources for prevention, treatment or amelioration. There have been several reasons for this apparent lack of interest in place, including a dearth of appropriate databases and insufficient appropriate software.[5] However, substantial recent advances in geographical information systems (GISs) now provide researchers and public health practitioners with an excellent environment in which to explore their data.[6] In addition, there is an increasing number of public health databases, in which the locations of the cases are recorded. It seems likely, therefore, that once they have understood its utility, scientists and public health practitioners will seek to use this spatial information.

At first glance, spatial analysis and its tools appear dauntingly complicated. This is not so, but there is a need for a glossary to explain common terms in geographical epidemiology, spatial analysis and GISs.

See end of article for authors' affiliations
....

Correspondence to:
Dr M Rezaeian, Social Medicine Department, Rafsanjan Medical School, Rafsanjan 7719617996, Iran; moeygmr2@yahoo.co.uk

Adjustment

Age, sex, socioeconomic and other variables vary from one place to another and may also influence the risk of the diseases. Observed differences in risk of illness or death are likely to be confounded by these variables and, therefore, comparisons of risk must take this important issue into account. The process of adjustment for potential confounding variables has an important role in the evaluation of the spatial variation in mortality and disease rates. The aim of an adjustment process is to produce a single summary value, such as the standardised incidence rate ratio (see below), which is unaffected by differences in the distributions of potential confounders.[7] The two most common approaches of adjustment are by direct and indirect weighting of stratum-specific rates. We illustrate the idea using adjustments for age differences, as age is almost always considered to be a confounding variable in epidemiological studies. However, we emphasise that the same procedure could be applied in adjustments to take account of other confounder variables.

In the direct approach, a weighted average of the age-specific rates from a study population is created, based on the age distribution of a reference population[8]—that is, the national population.[9] This is an estimate of the expected number of deaths in the reference population if the age-specific rates were the same as those that have been observed in the study population. An easily interpreted ratio is then obtained by dividing the expected number of deaths in the reference population by the observed number of deaths in the reference population over the same period of time.[10] This ratio is termed either the comparative mortality figure or standardised incidence rate ratio, and was first proposed in 1884.[11] In the indirect method, the crude rate in the study population is multiplied by a ratio known as the standardised mortality ratio (SMR).[8] The SMR is calculated by dividing the observed number (O) of cases within the study population by the expected number (E) of cases in the study population, assuming that the age-specific rates in the reference population also applied to the population under study.

It should be noted that if, for instance, the age distributions of two areas differ, the comparison of
....

their SMRs (determined by reference to an external reference population) may have a potential bias comparable with statistical confounding.[10] In other words, when compared with an external reference population, the indirect method of standardisation (the SMR) yields different rate ratios for cohorts with a different demographic structure even though the incidence rates within the demographic strata are identical. Despite the cautions raised in the literature, however, SMRs have been recommended for mapping by well-known statisticians.[12] However, directly adjusted rates also have problems. For instance, they may provide less stable estimates because the standard error of the rates depends on variations in the age-specific number of cases rather than the total number of cases.[10] In practice, the choice of selecting one of these methods depends on the type of data available. The indirect method is the only choice, for example, when the age-specific incidence counts are unavailable for a reference population but its age-specific rates are available.[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.[15]

As data for spatial analysis come from different sources, and have often been collected without taking into account the interests of the geographical epidemiologists,[16] it is absolutely necessary to ensure that precise and complete point and/or area health data are used in spatial epidemiology.[17–20] 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.[20]

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.[21 22] 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.[23] 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.[24] 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).[25] 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,[6] and Gatrell *et al*[26] for a complete discussion. Cuzick and Edwards'[27] method determines global clustering by examining the k nearest neighbours of each case. The geographical analysis machine[28] and the spatial scan statistic[29] 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. The spatial scan statistic has an advantage over geographical analysis machine in taking into account the problems of multiple testing.

Disease mapping

Data visualisation is the first step in disclosing the complex structure in data.[30] Data visualisation may not only create interest and attract the attention of the viewer but also provide a way of discovering the unexpected.[31] 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.[23] 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.[15] The use of mapping in the medical context has developed so rapidly during recent decades[32] that the presentation of maps is now established as a basic tool in the analysis of public health data.[23]

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.[33] 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.[34] 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).[35]

There are also other important issues that need to be considered while creating a map. These include the selection of an appropriate administrative unit for mapping, the selection of an appropriate method of data classification in the map, and the selection of an appropriate colour scheme or collection of hatching patterns. We will not discuss these issues in detail here. We will cover the optimum choice of mapping regions very briefly in the following section, and for the other issues refer the readers to other sources, detailed in the reference list.[18 36 37]

Ecological analysis

Ecological analysis is defined as the assessment of the associations between disease incidence (eg, suicide) and variables of interest (eg, social or environmental covariates).[23 38] These variables in an ecological analysis are defined on aggregated groups of individuals rather than the individuals themselves.[39] The reason for focusing on the comparison of groups rather than individuals is that individual-level data on the joint distribution of two or more variables within each group are usually missing. Therefore, an ecological study may be considered to be based on an incomplete design.[40]

An ecological analysis can be crucially dependent on scale (ie, the region based on the hypotheses under study).[1] The optimum choice of scale is a trade-off between making the

groups (regions) large enough to have stable rate estimates and also small enough to make them homogeneous in terms of their socioeconomic and other important characteristics.[19] If the regions are large, there is a greater possibility that associations measured at the aggregate level will differ from the same association measured at individual level. This can lead to a problem known as ecological fallacy,[41] or cross-level or ecological bias[42]—a situation in which one mistakenly infers an individual-level association from one that is actually only observed at the regional level. At the same time, if the regions chosen are too small, the results may show spurious spatial patterns due to random variation in small numbers of events.[43] [44]

The scale dependency of data may, therefore, cause what is known as the modifiable areal unit problem, which arises from the uncertainty induced by the aggregation procedure.[15] For this reason, it is important to take the scale of analyses into account and, if possible, to analyse the data at two or more levels of aggregation.[45] It is also possible to overcome both scale dependency and ecological bias by adopting a multilevel approach using the individual-level and group-level data together (see the next section).[46] For instance, to understand the effect of income on suicide,[47] we have to have data on the individual income and/or household income and area-level income, the latter of which may be ''compositional'' (ie, fully explained by the individual-level data) or ''contextual'' (irreducible to the individual—ie, an effect that persists even after allowance for the individual-level data).[48]

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.[49] 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.[23] Predominant among the methods of geographical epidemiology are the following: disease mapping, disease clustering and ecological analysis.[23] There is usually a close relationship between these branches.[50]

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 aggregate-level data are combined with those at the individual level.[51] Multilevel modelling, hierarchical regression and contextual analysis are phrases describing one of the various statistical methods in which this combination is allowed.[52] Multilevel modelling is a powerful, relatively new technique[53] 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,[52] and recently attempts have been made to integrate this kind of analysis into geographical epidemiology.[54] [55] 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.[15]

Geographical information systems

GISs can be defined as software systems for the automatic capture, storage, retrieval, analysis and display of spatial data.[56] The development of GIS technology dates back to the 1960s.[57] GISs have dramatically changed the ability of epidemiologists and public health specialists to work with spatial data.[4] The advantages of GISs are many, and include an ability to operate repetitive tasks, quickly compare spatial data and handle large volumes of data.[4] Other advantages include the ability to ask ''what if'' questions (asking conditional questions—ie, what if we locate a given hospital in place ''A'' rather than ''B''), Boolean searches (finding places that fulfil two or more criteria—ie, areas of high mortality and poverty), creation of ''buffer zones'' (circle a point data or centroid of an area data— eg, a 5-km circle around a putative source) and using data from remote sensing and global positioning systems (GPSs).[4]

However, historically, GISs have relied on their mapping capabilities rather than performance of statistical analyses. This is clear from the limited number and types of statistical analyses that most GISs are able to perform.[3] Until the full integration of spatial statistical analysis into a GIS environment is achieved, other solutions should be applied.[3] These include developing a ''loose coupling'' between a statistical package and a GIS, or a ''close coupling'' designing either some statistical functions within the GIS or adding some GIS tools into an analytical package.[6] For instance, one of the best known GIS software is ArcView,[58] for which a number of links with some statistical packages have been developed.[59] [60] When GISs are combined with spatial analytical methods, the result could provide a helpful tool in the study of public health issues.[61] Nevertheless, the users of GISs and readers of the output should not study the attractive maps produced by the software uncritically, and they should always remember the rules of good data management, analysis, presentation and interpretation.[37] [62]

Global positioning systems

A GPS consists of a system of at least 24 and up to 32 solar-powered satellites orbiting Earth every 12 h and transmitting radio pulses at very precisely timed intervals.[63] To determine a position in three dimensions (latitude, longitude and elevation), a receiver needs signals from at least four satellites.[64] GPS has become a standard method for data capturing in geographical epidemiology and public health studies.[25] Moreover, as the different components of a GPS receiver work efficiently under severe weather conditions such as sandstorm, torrential rain and high temperature, they could have a key role in combination with GISs, especially in emergency humanitarian activities.[65]

Smoothing

Mapping disease mortality or morbidity, especially in the smaller geographical areas, or when the given disease is somewhat rare, may give rise to the problem of small numbers, which in turn produces unstable rates. Although greater stability of rates may be achieved by choosing larger areas, simple mapping of the raw data is unattractive in that it still yields sudden changes at geographical boundaries.[66] In such circumstances, it is advantageous to ''smooth'' the local risk estimate on the basis of the overall pattern of rates.[6]

The basis of this technique is that when the underlying population of a given area is large and as a result the statistical error of the rate estimate is small, the adjusted rate will be close to the observed rate. However, when the underlying population is small and, therefore, the statistical error correspondingly large, the observed rate is shrunk by smoothing towards a value representing the overall mean of the map.[67] If spatial autocorrelation tests confirm that there is a spatial dependency (see the next section), the rates can be adjusted towards averages of neighbouring rates rather than the overall mean.[67] When we wish to improve the quality of a rate estimate for an area with an unstable rate by ''borrowing strength'' from its neighbours, a Bayesian analysis may be applied.[68]

When using smoothing, we are in effect making a prior assumption that a rate estimate for a given area is better if it in some way makes a combination of data from the area itself and those from the surrounding areas. A Bayesian analysis is one

way of achieving this combination. In a Bayesian analysis, an assumed (prior) probability distribution for the values of a parameter (the area rate) is converted (under the influence of the observations—ie, the observed rates) to a posterior (ie, after using the observed data) distribution for the values of that parameter. This posterior distribution is then used to provide an estimate for the parameter(the estimated rate for a given area) together with a standard error for this estimate. With such an approach, the prior distribution can be based on the results of previous studies or on background knowledge. It is also possible to base this distribution on particular global aspects of the data currently at hand. The latter approach is usually referred to as empirical Bayes estimation.[6] There is a non-iterative empirical Bayes method of moments for smoothing the rates towards either local or global mean.[69] Although this method is useful to estimate the relative risks of a given disease and in this sense it functions in a similar way as the fully Bayesian approach, the latter produces more informative interval estimates.[70 71] The technical details of Bayesian methods are well beyond the scope of this article but, in essence, the Bayesian approach uses the observed data to update prior knowledge. If there is a large number of observations, then the prior knowledge has little influence (ie, the observed rates provide good estimates); if not, the prior knowledge is used to reduce (smooth) the sampling fluctuations between the unreliable observed rates.

Spatial autocorrelation

Lack of independence of data from neighbouring areas gives rise to spatial autocorrelation. The correlation or dependency implies that rates for geographically close areas are more highly related than those from areas that are geographically distant.[72] For instance, suicide rates in neighbouring areas are likely to be more similar than those in distant ones.[15] This is because neighbouring areas may have similar underlying social, economic and cultural characteristics that trigger suicidal behaviour. Detecting spatial dependency, which is accomplished by the use of spatial autocorrelation statistics, would help researchers to justify their selected regression models in an ecological analysis, or their smoothing techniques when mapping a rare disease or when mapping in small boundaries (see the sections on ecological analysis, spatial regression and smoothing).[73]

Spatial autocorrelation statistics provide very useful summary information about the spatial arrangement of data in a map.[73] In fact, some of these statistics compare neighbouring area values to assess the level of large-scale or global clustering. Whenever a large number of neighbouring areas have either reasonably large or small values, large-scale clustering may be detected.[4] The two most commonly used spatial autocorrelation statistics for detecting global clustering in continuous areal data (ie, morbidity and mortality rates) are the I statistic, developed by Moran,[74] and Geary's c statistic.[75] There is also a number of spatial autocorrelation statistics available (Getis and Ord's G* statistic), which measure the amount of local clustering (ie, hot spots of high or low values) by finding any association between a value at a particular area and values of adjacent or nearby areas.[25 76]

Spatial regression

Typically, we use some form of regression model in an ecological analysis to predict, for example, suicide rates in given areas with the area's other attribute data such as poverty or social cohesion[15] (see the section on ecological analysis). In such situations,[77] we usually divide the study region into a set of non-overlapping, administratively defined areas, and model the counts of the number of cases within each area.[78] These will be accompanied by information on the population at risk in different relevant age and sex groups in addition to other factors such as socioeconomic status. If the risk within a given area is constant, the distribution of the count for that area is clearly a binomial distribution.[34] However, if the risk is small (eg, suicide mortality), one may approximate the binomial distribution by the Poisson distribution.[17] If there is evidence that the Poisson model does not fit well, this indicates that there is a component risk that has not been incorporated into the model. This is an example of "extra-Poisson" variation or overdispersion.[79] The overdispersion may arise for several reasons, including large numbers of cells having zero counts, which may happen when a rare disease is being studied.[36] To cope with this problem, the Poisson model can be replaced by the negative binomial regression model.[80]

In case the exploratory techniques uncover the nature of spatial dependence (see the section on spatial autocorrelation), a richer model involving spatial autocorrelation may be fitted.[34] There are two empirical Bayesian models in which the spatial dependency of the data can be taken into account.[81] The first model is a simultaneous autoregressive model,[82] and the second is a conditional autoregressive model.[83] The conditional autoregressive model provides a more general framework with less complexity, and is therefore preferred to the simultaneous autoregressive model.[77] More recently, the WinBUGS software has provided a fully Bayesian analysis of the conditional autoregressive model.[84]

. . . .

Authors' affiliations

Mohsen Rezaeian, Graham Dunn, Biostatistics Group, Division of Epidemiology & Health Sciences, The University of Manchester, Manchester, UK
Selwyn St Leger, Evidence Based Population Health Unit, Division of Epidemiology & Health Sciences, The University of Manchester, Manchester, UK
Louis Appleby, Division of Psychiatry, The University of Manchester, Manchester, UK

Competing interests: None declared.

REFERENCES

1 **Elliot P**, Best N. Geographical patterns of disease. In: Armitage P, Colton T, eds. *International encyclopaedia of biostatistics*. Chichester: Wiley, 1998:1694–701.
2 Hippocrates. On Airs Waters, and Places. Translated and published in *Med Classics* 1938;**3**:19–42.
3 **Glass GE**. Update: spatial aspects of epidemiology: the interface with medical geography. *Epidemiol Rev* 2000;**22**:136–9.
4 **Moore DA**, Carpenter TE. Spatial analytical methods and geographical information system: use in health research and epidemiology. *Epidemiol Rev* 1999;**21**:143–61.
5 **Rezaeian M**, Dunn G, St Leger S, *et al*. Spatial epidemiology of suicide in England and Wales. *First European Conference, Geographical Information Sciences in Public Health*, Sheffield, England, 2001. http://gis.sheffield.ac.uk/ Library/Downloads/Resources/2001_ConferenceAbstracts.pdf.
6 **Bailey TC**, Gatrell AC. *Interactive spatial data analysis*. Harlow: Longman, 1995.
7 **Selvin S**. *Statistical analysis of epidemiological data*. Oxford: Oxford University Press, 1996.
8 **Gerstman BB**. *Epidemiology kept simple. An introduction to classic and modern epidemiology*. USA: Willey-Liss, 1998.
9 **Inskip H**. Standardisation methods. In: Armitage P, Colton T, eds. *International encyclopaedia of biostatistics*. Chichester: Wiley, 1998:4237–50.
10 **Breslow N**, Day N. *The design and analysis of cohort studies*, Vol 2.IARC Scientific Publication No 82. Lyon: International Agency for Research on Cancer, 1987:65–73.
11 **General Register Office**. *Annual report of the Registrar General for England and Wales*. London: HMSO, 1884.
12 **Pickle L**, White AA. Effects of the choice of age-adjustment method on maps of death rates. *Stat Med* 1995;**14**:615–27.
13 **Waller LA**, Gotway CA. *Applied spatial statistics for public health data*. Hoboken, NJ: Wiley, 2004.
14 **Elliott P**, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect* 2004;**112**:998–1006.

15 **Jerrett M**, Burnett RT, Goldberg MS, *et al.* Spatial analysis for environmental health research: concepts, methods, and examples. *J Toxicol Environ Health A* 2003;**66**:1783–810.

16 **Staines A**, Jarup L. Health event data. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial epidemiology, methods and applications*. Oxford: Oxford University Press, 2000:15–29.

17 **Wakefield JC**, Best NG, Waller L. Bayesian approach to disease mapping. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial epidemiology, methods and applications*. Oxford: Oxford University Press, 2000:104–27.

18 **Lawson AB**, Williams FLR. *An introductory guide to disease mapping.* Chichester: Wiley, 2001.

19 **Walter SD**, Birnie SE. Mapping mortality and morbidity patterns: an international comparison. *Int J Epidemiol* 1991;**20**:678–89.

20 **Jarup L**. Health and environment information systems for exposure and disease mapping, and risk assessment. *Environ Health Perspect* 2004;**112**:995–7.

21 **Best N**. Bayesian ecological modelling. In: Lawson AB, Bohning D, Biggeri A, Lesaffre E, Viel JF, eds. *Disease mapping and its uses.Disease mapping and risk assessment for public health*. Chichester: Wiley, 1999:193–201.

22 **Walter SD**. Assessing spatial patterns in disease rates. *Stat Med* 1993;**12**:1885–94.

23 Lawson AB, Bohning D, Biggeri A, *et al*, eds. *Disease mapping and its uses. disease mapping and risk assessment for public health*. Chichester: Wiley, 1999.

24 **Besag J**, Newell J. The detection of clusters in rare disease. *J R Stat Soc A* 1991;**154**:143–55.

25 **Cromley EK**, McLafferty SL. *GIS and public health*. New York: The Guilford Press, 2002.

26 **Gatrell AC**, Bailly TC, Diggle PJ. *et al*, Spatial point pattern analysis and its application in geographical epidemiology.*Trans Inst Br Geogr*, 1996;**21**:256–74.

27 **Cuzick J**, Edwards R. Spatial clustering for inhomogeneous populations. *J R Stat Soc B* 1990;**52**:73–104.

28 **Openshaw S**, Charlton M, Wymer C, *et al.* A mark 1 geographical analysis machine for the automated analysis of point data sets. *Int J Geogr Inf Syst* 1987;**1**:335–58.

29 **Kulldorff M**. Statistical methods for spatial epidemiology: tests for randomness. In: Gatrell A, Loytonen M, eds. *GIS and health*. London: Taylor & Francis, 1998:49–62.

30 **Cleveland WS**. *Visualising data*. Summit, NJ: Hobart Press, 1993.

31 **Everitt BSE**, Dunn G. *Applied multivariate data analysis*. London: Arnold, 2001.

32 **Cliff AD**. Analysing geographically related disease data. *Stat Methods Med Res* 1995;**4**:93–101.

33 **Clayton D**, Bernardinelli L. Bayesian methods for mapping disease risk. In: Elliott P, Cuzik J, English D, Stern R, eds. *Geographical and environmental epidemiology–methods for small area studies*. Oxford: Oxford University Press, 1996.

34 **Bithell JF**. Geographical analysis. In: Armitage P, Colton T, eds. *International encyclopaedia of biostatistics*. Chichester: Wiley, 1998:1701–16.

35 **Cartwright RA**, Alexander FE, McKinney PA, *et al. Leukaemia and lymphoma: an atlas of distribution within areas of England and Wales 1984–1988*. Leeds: Leukaemia Research Fund, 1990.

36 **Lawson AB**. Tutorial in biostatistics: disease map reconstruction. *Stat Med* 2001;**20**:2183–204.

37 **Rezaeian M**, Dunn G, St Leger S, *et al.* The production and interpretation of disease maps: a methodological case-study. *Soc Psychiatry Psychiatr Epidemiol* 2004;**39**:947–54.

38 **Gatrell AC**, Bailly TC. Interactive spatial data analysis in medical geography. *Soc Sci Med* 1996;**42**:843–55.

39 **Richardson S**. Statistical methods for geographical correlation studies. In: Elliott P, Cuzik J, English D, Stern R, eds. *Geographical and environmental epidemiology – methods for small area studies*. Oxford: Oxford University Press, 1996:181–204.

40 **Kleinbaum DG**, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. New York: Van Nostrand Reinhold, 1982.

41 **Morgenstern H**. Uses of ecological analysis in epidemiologic research. *Am J Public Health* 1982;**72**:1336–43.

42 **Richardson S**, Monfort C. Ecological correlation studies. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial epidemiology, methods and applications*. Oxford: Oxford University Press, 2000:205–20.

43 **Miyawaki N**, Chen SC. A statistical consideration on the mapping of mortality. *Soc Sci Med* 1981;**15D**:93–101.

44 **Aylin P**, Maheswaran R, Wakefield J, *et al.* A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. *J Public Health Med* 1999;**21**:289–98.

45 **Rezaeian M**, Dunn G, St Leger S, *et al.* The ecological association between suicide rates and income and employment indices of deprivation in the North West region of England – the importance of the size of the administrative unit. *J Epidemiol Community Health* 2006;**60**:956–61.

46 **Greenland S**. Ecological versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol* 2001;**30**:1343–50.

47 **Rezaeian M**, Dunn G, St Leger S, *et al.* The ecological association between suicide rates and indices of deprivation in English local authorities. *Soc Psychiatry Psychiatr Epidemiol* 2005;**40**:785–91.

48 **Krieger N**. A glossary for social epidemiology. *J Epidemiol Community Health* 2001;**55**:693–700.

49 **English D**. 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, 1996:3–13.

50 **Elliott P**, Wakefield JC, Best NG, *et al.* Spatial epidemiology: methods and applications. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial epidemiology, methods and applications*. Oxford: Oxford University Press, 2000:3–14.

51 **Dooley D**, Catalano R, Rook K. *et al.* 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*, 1989;**19**:321–36.

52 **Morgenstern H**. Ecological studies. In: Rothman KJ, Greenland S, eds. *Modern epidemiology*. Philadelphia: Lippincott-Raven Publisher, 1998:459–80.

53 **Verheij RA**. Explaining urban-rural variations in health: a review of interactions between individual and environment. *Soc Sci Med* 1996;**42**:923–35.

54 **Merlo J**, Yang M, Chaix B, *et al.* A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health* 2005;**59**:729–36.

55 **Merlo J**, Chaix B, Yang M, *et al.* 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* 2005;**59**:1022–8.

56 **Clarke KC**, McLafferty SL, Tempalski BJ. On epidemiology and geographical information systems: a review and discussion of future directions. *Emerg Infect Dis* 1996;**2**:85–92.

57 **Schuurman N**. *GIS: a short introduction*. Oxford: Blackwell Publications, 2004.

58 **ArcView**. *ArcView GIS V.3.1*. Redlands, CA: Environmental System Research Institute, 1998.

59 **Bao S**, Anselin L, Martin D, *et al.* Seamless integration of spatial statistics and GIS: the S-PLUS for ArcView and the S+Grassland links. *J Geogr Syst* 2000;**2**:287–306.

60 **Anselin L**, Bao S. Exploratory spatial data analysis linking SpaceStat and ArcView. In: Fischer M, Getis A, eds. *Recent developments in spatial analysis*. Germany: Springer-Verlag, 1998:35–9.

61 **Gesler W**. The uses of spatial analysis in medical geography a review. *Soc Sci Med* 1986;**23**:963–73.

62 **Paterson AD**. Problems encountered in the practical implementation of geographical information systems (GIS) in veterinary epidemiology. *Presented at the meeting of the Society of Veterinary Epidemiology and Preventive Medicine*, Edinburgh, United Kingdom, 1995.

63 **Kennedy M**. *Global positioning system and GIS*. London: Taylor and Francis, 2002.

64 **Longley PA**, Goodchild MF, Maguire DJ, *et al. Geographical information system and sciences*. Chichester: Wiley, 2005.

65 **Kaiser R**, Spiegel PB, Henderson AK, *et al.* The application of geographic information systems and global positioning systems in humanitarian emergencies: lessons learned, programme implications and future research. *Disasters* 2003;**27**:127–40.

66 **Boyle P**, Muir CS, Grundmann E. *Cancer mapping*. Berlin: Springer-Verlag, 1989.

67 **Olsen FS**, Martuzzi M, Elliot P. Cluster analysis and disease mapping–why, when, and how? A step by step guide. *BMJ* 1996;**313**:863–6.

68 **Bell BS**, Broemeling LD. A Bayesian analysis for spatial processes with application to disease mapping. *Stat Med* 2000;**19**:957–74.

69 **Marshall R**. Mapping disease and mortality rates using empirical Bayes estimators. *Appl Stat* 1991;**40**:283–94.

70 **Mollie A**. Bayesian mapping of disease. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov chain Monte Carlo in practice*. London: Chapman and Hall, 1996:359–76.

71 **Bernardinelli L**, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Stat Med* 1992;**11**:983–1007.

72 **Tobler WR**. A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 1970;**46**(Suppl):234–40.

73 **Odland J**. *Spatial autocorrelation*. California: Sage Publications, 1988.

74 **Moran PAP**. The interpretation of statistical maps. *J R Stat Soc B* 1948;**10**:243–51.

75 **Geary RC**. The contiguity ratio and statistical mapping. *Incorporated Stat* 1954;**5**:115–45.

76 **Getis A**, Ord K. The analysis of spatial association by use of distance statistics. *Geogr Anal* 1992;**24**:189–206.

77 **Cressie N**. *Statistics for spatial data*. New York: Wiley, 1991.

78 **Haining R**. Spatial statistics and analysis of health data. In: Gatrell A, Loytonen M, eds. *GIS and health*. London: Taylor & Francis, 1998:29–47.

79 **Breslow NE**. Extra-Poisson variation in log-linear models. *Appl Stat* 1984;**33**:38–44.

80 **Johnson NL**, Kotz S, Kemp AW. *Univariate discrete distributions*. New York: Wiley, 1992.

81 **Mollie A**. Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. *Stat Med* 1991;**10**:95–112.

82 **Whittle P**. On stationary process in the plane. *Biometrika* 1954;**41**:434–49.

83 **Besag J**. Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc B* 1974;**36**:192–236.

84 **Spiegelhalter D**, Thomas A, Best N, *et al.* BUGS: Bayesian inference using Gibbs sampling, Version 0.50, Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health. Cambridge: Cambridge University, 1995.