
THE GEOGRAPHICAL EPIDEMIOLOGY OF OCULAR DISEASES: SOME PRINCIPLES AND METHODS

A. C. GATRELL¹, J. R. O. COLLIN², R. DOWNES³, B. JONES² and T. C. BAILEY⁴
Lancaster, London, Nottingham and Exeter

SUMMARY

With the increasing availability of geographically referenced data in health research the time is ripe to review the use of particular geographical and spatial analysis techniques in ophthalmic research. Analysis of the geographical distribution of ocular diseases, particularly in Britain, has not had a high profile, but there are certain diseases, such as congenital eye malformations in children, where such analysis methods are particularly appropriate. We review the data requirements and then a variety of analytical techniques, some of which partition geographical space into areal units (such as counties or electoral wards), others of which treat space as continuous. We conclude with some comments on software that is available for such analyses.

Geographical epidemiology refers to the description and analysis of disease incidence in space. While the simple mapping of disease has a long and rich history (see, for example, Haggett and Cliff¹), the subject has been given fresh impetus in recent years, for two reasons. One reason is the increasing availability of health data that contain some form of 'spatial referencing'; a postcode, for example. The second motivation comes from the wish to see whether disease incidence shows any evidence towards 'clustering' or geographical aggregation, since this might give some aetiological clues. The most well-known example of this comes from the epidemiology of childhood cancer, especially the leukaemias and lymphomas, where both clustering² and the possible location of 'clusters' in the vicinity of nuclear installations (see, for example, Bithell and Stone³) have been subject to intensive research effort. Below, we review both data requirements and analytical methods for the detection of clusters and clustering.

From: ¹Department of Geography, Lancaster University, Lancaster; ²Moorfields Eye Hospital, London; ³Queen's Medical Centre, Nottingham; ⁴Department of Mathematical Statistics and Operational Research, University of Exeter, Exeter, UK.

Correspondence to: A. C. Gatrell, Department of Geography, Lancaster University, Lancaster LA1 4YB, UK.

In doing so we are adopting a 'spatial analysis' perspective; this means that we want to give prominence to the geographical configuration of the objects of study. In particular, we need to recognise that analysis of data in a spatial context does not simply mean applying standard, non-spatial statistical techniques. Typically, such techniques assume that observations constitute independent pieces of evidence. However, in studies where the observations are geographically located, we need to allow for the fact that disease incidence is likely to be spatially correlated; the incidence of ocular disease in one area, for example, will be broadly similar to that in nearby areas, but probably less similar to that in more distant areas. This kind of 'spatial correlation' (also known as 'autocorrelation') arises in many studies of the geographical incidence of disease; there is no reason to believe that eye disease is likely to behave differently.

SPATIAL DATABASES IN GEOGRAPHICAL EPIDEMIOLOGY

Before considering methods that can be brought to bear on the spatial analysis of ocular disease, we first consider the 'material' required; we pay less attention to the databases on disease (which will typically include diagnostic categories, age, gender, as well as treatment and so on), concentrating more on the data requirements for a *spatial* analysis. However, we first discuss the specific problem that has motivated the present review, and that brings a team of geographers, statisticians and ophthalmic surgeons together.

A Geographical Database on Anophthalmos and Microphthalmos

Our subsequent discussion of methods is motivated by a long-standing interest in geographical variations in the incidence of the rare conditions anophthalmos and microphthalmos. (The incidence of the former is 0.30 per 10 000 births in France; 0.60 in north-east Italy: see Stoll *et al.*⁴ and Clementi *et al.*⁵ respec-

tively). This interest was itself generated, in part, by anecdotal reports in newspapers (from 1986) of possibly elevated levels of the conditions around incinerators in Scotland and South Wales. Although Government reports^{6,7} suggested that there were no problems, the South Wales report was based on suspect data (see Gatrell and Lovett⁸ for a review). We considered it worth while to assemble our own database, comprising: names and addresses of cases; date of birth; gender; ophthalmic history (anophthalmos or microphthalmos, unilateral or bilateral, presence of other ocular abnormalities); family history of ocular abnormalities; maternal history during pregnancy, and any illness during pregnancy; and current management. A proforma was distributed to consultant ophthalmic surgeons in Britain and details of cases made available by them have been entered into a database.

Data have been received for 176 children born since 1970, but are in some cases incomplete. We have therefore not yet commenced any detailed statistical or geographical analyses. Indeed, in the wake of further newspaper reports (dating from 1993) suggesting 'clustering' of cases in rural areas and speculating on associations with pesticide exposure, we are now pooling our data with those from other sources in a collaborative study with the Environmental Epidemiology Unit at the London School of Hygiene and Tropical Medicine.

We note that the methods outlined below are to be used when we are as confident as possible in the completeness of case ascertainment. The methods are robust if the missing cases are randomly distributed geographically; but, of course, if there are geographical biases, such as under-reporting in one or more regions, then any evidence of geographical aggregations, or departures from randomness in spatial distribution, may well be entirely spurious. We simply make the point that we must ensure that we have adequate data in advance of analysis.

Geographical Data-handling

In Britain and North America, as well as in parts of Europe, it is possible to make use of the postal address of patients for geographical analysis. In Britain and Canada, for example, the 'unit postcode' is a collection of alphanumeric characters (e.g. BB2 6HG in Britain, L8S 4K1 in Canada), while in the USA the so-called Zip-code (a set of five digits, though currently being extended to nine) is used for address purposes. Gazetteers are available to link these postal codes to geographically defined, and mappable, locations. These are street blocks in North America. But in Britain there is available a machine-readable file, known as the Central Postcode Directory, which links each of the approximately 1.8 million unit postcodes to both an Ordnance

Survey grid reference and an electoral ward.⁹ The grid reference has a resolution of 100 m in England and Wales, 10 m in Scotland. Thus, given a database on eye disease in Britain, one that contains full unit postcodes, it is a simple matter to obtain precise locational information. A unit postcode does not identify the unique street address, since it is shared with, on average, about 15 households. But clearly it is more than adequate for most epidemiological purposes! The way in which such information has been used in health research, and the accuracy of the Central Postcode Directory, has been discussed elsewhere.¹⁰

Given that we have obtained a set of grid references for disease cases, how do we proceed? There are essentially two strategies. Either we conduct analyses on the grid references themselves (which may be displayed in map form as a 'point pattern', or collection of dots, each of which represents a grid-referenced case); or, we aggregate the individual events to form a count of the number of cases within fixed areal units. Such areal units might be electoral wards, Health Districts, or whatever the analyst decides is appropriate. There are advantages and disadvantages with both approaches. If we treat the point pattern itself we avoid well-known problems of dealing with areal units of arbitrary size and shape; the boundaries of these are drawn with no reference to local health or environment, and were a different set of areal units to be used any resulting maps and/or analyses would inevitably differ. However, simply mapping a point pattern on its own, for instance to detect 'clusters' of health events, is of little or no value, since we need to know something about the underlying distribution of population at risk. As we shall see below, this is not an impossible task. Adopting an areal unit perspective, especially for zones such as electoral wards, means that it is relatively straightforward to obtain denominators (populations at risk) from the decennial Census; moreover, it then becomes possible to relate measures of disease incidence (standardised, where appropriate, for age and gender) to socio-economic data, such as measures of social deprivation, themselves derived from the Census. (See Dale and Marsh¹¹ for details of the UK 1991 Census and its geographical relevance.)

Ideally, we would prefer the 'continuous space' approach of point pattern analysis, especially if we have other attribute data available for the individuals. Given such data it then becomes possible to conduct analyses at the individual level, rather than an aggregate, ecological level.¹² Such analyses can be both spatial, as reviewed below, as well as the conventional and powerful, but non-spatial, regression-based analyses.¹³

Before leaving this brief discussion of geographical

data-handling we draw attention to the burgeoning field of Geographical Information Systems (GIS), since this provides a natural environment within which to embed such geographical databases, and within which to perform spatial analysis. A GIS is a set of computerised tools for the collection, storage, integration, analysis and display of spatially referenced data.¹⁴ It becomes especially powerful when there is more than one database to be explored. For instance, we might wish to examine the relationship between disease and environment. This would require construction of databases on disease incidence (geographically referenced), on Census variables (measures of age–sex structure and socio-economic status, for example), and on air or water quality, for example. Postcodes and grid references provide the geographical link that binds these databases together in an integrated framework, while the GIS provides the structural framework within which such links may be constructed. We refer below to some proprietary GIS products, but note here that they are versatile and can, in general, handle data input derived from non-spatial database management systems, such as ORACLE, dBase, INGRES, and so on. The size of such databases is not a serious issue; tens of thousands of individual records of disease can readily be handled.¹⁰

ANALYTICAL TECHNIQUES

We now consider some of the analytical techniques that spatial analysts find useful in applications of geographical epidemiology. The treatment is inevitably brief, but references are given to more detailed accounts; in particular, a useful starting point is Bailey and Gatrell,¹⁵ which also provides software with which to gain familiarity with the approaches we outline.

Area-Based Approaches

In this section we give a brief overview of some of the methods that may be brought to bear on a spatial analysis of eye disease if data are analysed by areal unit. Useful additional, and more detailed, descriptions are provided elsewhere.^{1,15,16}

Given data on counts of disease incidence by area it is appropriate to derive rates by dividing these counts by the population at risk, age–sex standardised where necessary. Such rates may be mapped in order to detect, visually, evidence of spatial pattern. However, there are a number of problems and issues that arise. Among these, we consider first the assessment of pattern; and second the ‘stability’ of the estimated rates. But we should also recognise that fundamental cartographic issues arise in the visual representation of areal data; for example, the number of class intervals used to display the incidence rates, as well as the particular class

interval scheme. These decisions can have profound implications for the ‘message’ conveyed by the map.

One question that arises in exploratory spatial data analysis is whether the spatial arrangement of rates shows departure from randomness. In other words, is the collection of rates that are mapped simply one of a number of permutations of values that could arise on a purely random basis, or is there a tendency for high rates to be ‘clustered’ near other high rates while areas with low rates are adjacent to areas also with low rates? Our visual perception is ‘trained’ to look for pattern and we need methods that can determine whether this is genuine or spurious. The methods that are used to detect departures from randomness are called autocorrelation (or spatial correlation) tests. There is a variety of indices that can summarise this, ranging from Moran’s *I* coefficient for spatially continuous data (such as the incidence rates we might be using), to indices based on the ranked values or a simple classification of rates into above and below average. Such methods are described in full in Cliff and Ord¹⁷ and have been used extensively in cancer epidemiology (see, for example, Walter¹⁸).

One problem with using incidence rates, standardised or otherwise, is particularly pronounced if we are interested in spatial variation in disease incidence among small areas. This is because the populations forming the denominators, and the counts of cases themselves, will often be small, especially in rural areas. The rates may thus be inherently variable or ‘unstable’. For example, if we are mapping measures of relative risk (i.e. observed divided by expected numbers of case, where the expected numbers are based on region-wide incidence rates and the age–sex composition of the small areas) and we have an expected number of cases in a particular area of, say, 0.5, then observed counts of 0, 1, 2 or 3 cases will generate relative risks (or standardised ratios if we multiply by 100 as is conventional) of 0, 200, 400 and 600 respectively! We need some way of taking into account the ‘reliability’ of the estimated expected number of cases. One possibility is to use Poisson probabilities,¹⁹ where we estimate and map the probability of getting a count that is more extreme than that observed, under the assumption that the count in each area is Poisson distributed. Low probabilities indicate that an area’s rate is unusually high or low. However, statisticians are wary of this approach, since it assumes an independence of values among different areal units; as we have noted, some spatial dependence is to be expected in areal data.

An alternative that is gaining widespread acceptance in geographical epidemiology is to use what are called ‘empirical Bayes estimates’. We do not have space to describe these in great detail (see chapter 8

of Bailey and Gatrell¹⁵ for a full account, and chapters in Elliott *et al.* 1992¹⁶). Briefly, however, such methods allow us to estimate the true incidence in an area as a weighted sum of both the observed incidence rate in an area and the mean rate in all areas (or a local neighbourhood of the zone in question). The issue then becomes one of trying to estimate the weights. The idea is to make these weights a function of the population at risk in an area. If the population is quite large (and therefore the estimate of disease risk is quite reliable) then the 'raw' incidence rate is given a high weight. Conversely, if the population is small (and hence we cannot be too confident in our estimate of incidence) we give it less weight and give more weight instead to the 'global' mean; in other words, we 'shrink' the estimate of incidence so that it is closer to the mean for all areas. Details of how to obtain these weights, and use of such estimates in geographical epidemiology, are given in Elliott *et al.*¹⁶ and chapter 8 of Bailey and Gatrell.¹⁵

These methods have yet to be widely applied in the epidemiology of eye disease. However, Gatrell and Lovett⁸ took published data from the Office of Population Censuses and Surveys on the incidence of congenital eye malformations (specifically, anophthalmos and microphthalmos) by Health Authority in England and Wales and estimated Poisson probabilities for 1983–1985. This highlighted a few areas of significantly elevated incidence. However, we should caution that the data are far from perfect in a diagnostic sense, and that the number of areas deemed significant is no more than would be expected on a chance basis; that is, if we conduct say 200 Poisson tests we would expect to observe 10 'significant' probabilities at the 0.05 level purely on a random basis. Moreover, we would prefer, of course, to be working with individual-level data rather than data for quite large areal units.

Point-Based Approaches

We commented earlier that there are some advantages to adopting an approach that treats the disease cases as point 'events' rather than aggregating them to areal units of arbitrary size and shape. Of course, if data are only provided as counts or incidence rates by areal unit we have no choice in the matter; however, it is obviously preferable to work with individual-level data if at all possible. This avoids the problem of so-called modifiable areal units, as well as the risk of committing the 'ecological fallacy', where inferences about individuals are made from aggregate data.

As noted above, simple 'dot mapping' the cases of disease is of little value, despite the fact that some epidemiologists have recently 'discovered' the idea!²⁰ If we were to map the distribution of population (or

households) as points we would observe pattern or 'clustering', simply because people arrange themselves in clustered settlements. If we wish to detect disease clustering, we shall need some way of filtering out this natural tendency of the population at risk to cluster. For example, if we wish to see whether children with leukaemia are 'clustered', or are clustered around nuclear installations, we shall require to know something about the distribution of children as a whole.

Note that we have in fact asked two separate questions, both of which might be of interest in an ocular epidemiological context. First, we might want to ascertain whether cases of eye disease show any tendency towards clustering; do they tend to occur together more than we might expect on the basis of population distribution? Second, we might want to identify the locations of 'clusters', or perhaps to determine whether there is an elevated risk of eye disease in the vicinity of possible pollution sources (which might themselves be point locations, such as the incinerators mentioned above, or perhaps linear features such as major roads or rivers). Fortunately, methods are available to solve both problems²¹ and from the substantial literature we comment briefly on two recent techniques.

It should be clear from the above remarks that if we wish to detect spatial clustering in a point pattern that displays disease incidence we need another point pattern as a 'benchmark'; in other words, along with a set of cases we need a set of controls. Given a single point pattern, there is a variety of methods available to determine the presence or absence of randomness. Most of these are known as 'second-order' methods, assessing whether the density of points in one small region is correlated with the density in a neighbouring region. From a practical point of view, assessment of spatial pattern is based on measuring distances between point events; intuition suggests that a clustered pattern should show an excess of short distances, while a dispersed pattern would be characterised by many longer distances. Formally, statisticians use the so-called *K* function²² to describe this spatial structure. This is defined as the expected number of point events within a fixed distance of an arbitrarily chosen event, scaled by the overall density of point events. Details of its estimation are given elsewhere.^{15,22} However, for a bivariate point pattern, comprising cases and controls, we need to estimate separate *K* functions for each class of point. The estimated *K* functions may then be subtracted and the result plotted as a function of distance; if the cases are clustered, over and above that of background population (reflected in the distribution of controls) then this will be displayed graphically by peaks in the plot. This method, due to Diggle and Chetwynd,²³ has been

used in other epidemiological applications;²⁴ for a related approach see Cuzick and Edwards.²⁵ Once case ascertainment is complete, or without obvious bias, it can be put to good effect in studying the research problem mentioned earlier: that of whether there is clustering of eye malformations over and above that of births in general.

We might, in addition, have further information on the temporal ordering of cases; for example, the date of notification. This information can also be used analytically, in studies of space-time clustering.^{15,26} Here, interest centres on whether cases that are 'close' together in space are also 'close' together in time. This is particularly important if we suspect that some infective mechanism is at work. Several attempts have been made, for example, to see whether Burkitt's lymphoma in parts of tropical Africa exhibits space-time clustering (e.g. Siemiatycki *et al.*²⁷), since there are grounds for believing that infection with the Epstein-Barr virus plays a role in aetiology.²⁸

If we wish to detect the locations of clusters, as opposed to clustering as an overall phenomenon, we need other methods. By 'cluster' we mean a localised aggregation of cases. We might undertake a purely inductive approach to the problem, screening different parts of a study region for evidence of such aggregations. This is the approach advocated by Openshaw²⁹ and refined by Besag and Newell.³⁰ Alternatively, we might have some *a priori* hypothesis that there is raised incidence of eye disease, or an increased relative risk, in the vicinity of a suspected point or linear source of pollution. In this case we need methods that will evaluate the research hypothesis against the null hypothesis of no elevated risk. For example, Diggle³ (see also Diggle²¹ and chapter 4 in Bailey and Gatrell¹⁵) has devised a model to relate the intensity of disease incidence to both background population and proximity to a suspected point source of pollution; this allows the proximity effect to be quantified. A more recent approach, which allows other explanatory variables (covariates) to be incorporated into the model, is outlined in Diggle and Rowlingson.³² Such methods could be put to good effect in any test of the hypothesis that the incidence of anophthalmos and/or microphthalmos is elevated around existing or closed high-temperature incinerators.

It is worth commenting that the approaches discussed in this section make some rather crude assumptions about human behaviour and mobility. We have assumed that 'exposure' is adequately summarised by address at diagnosis. This will be acceptable for diseases with a short latency period, but in other cases, particularly for older populations, we shall need to recognise that people have individual migration histories and that address at

diagnosis may disguise exposures in earlier residential and occupational environments. The same issue can arise in younger populations; for example, when considering the possible link between retinoblastoma in children and nuclear radiation it was important to recognise that cases were occurring in children whose mothers had been resident in Seascale, west Cumbria (the site of the Sellafield nuclear reprocessing plant), even though the children had been born elsewhere.³³

SOFTWARE ENVIRONMENTS

We now review briefly what software is available to perform some of these geographical analyses. First, note that standard statistical packages, such as SPSSx, MINITAB or BMDP, are not appropriate since they have no facilities for recognising the inherently spatial nature of the data, and allowing for this in the analysis. In addition, most of these common packages have little or no facility for viewing the data in map form. There are numerous packages available for producing computer-drawn maps, and if the object is simply to draw maps, one of these will suffice. However, we have stressed in this paper the need for spatial *analysis* and not simply mapping; ideally, we would like a software tool that allows us: to visualise (e.g. draw maps and graphs of) data; to explore such data (e.g. by deriving new pictures of our data, or alternative spatial representations); and perhaps to model our data (using spatial statistical models). What facilities are currently available? We discuss three possible approaches and give some details of software availability in the Appendix.

As noted earlier, a variety of commercial Geographical Information Systems (GIS) have emerged in the last 10 years that are of enormous value in handling spatially referenced data, especially where the data are voluminous. Some of these systems, such as IDRISI,³⁴ are reasonably easy to use; others, such as ARC/INFO,³⁵ are more demanding, containing several hundreds of functions. Increasingly, however, the systems are using menu structures for interacting with the software; the days of entering lengthy commands from the keyboard are fast disappearing. Unfortunately, while some progress is being made, historically these systems have been weak on spatial analysis; their real strength has lain in linking together large, often disparate, data sets. Thus, for example, such systems would allow one to place a 'buffer zone' around a water course, and to overlay point data on, for example, 'river blindness' in order to see whether there is any association between disease incidence and proximity to water. But if one wanted to go beyond this to test any hypotheses more formally, such systems have been of little value. While the situation is changing,³⁶ there is still much that remains to be done.

In the absence of spatial analytical power within GIS, a second possibility is to exploit those statistical software environments that do allow some spatial analysis. Of these, S-Plus is of particular interest, some work having been done on embedding the point-based approaches outlined earlier into the system.³⁷ S-Plus contains the usual facilities for exploratory data analysis and modelling that are to be expected of any statistical package, but also high-level functions to perform other tasks, together with excellent graphics. It includes a programming language that can be used to link some fundamental operations into more complex tasks, which then become new functions. The suite of procedures, known as 'SPLANCS', for point pattern analysis³⁷ is just such an example. It offers plots of point maps, spatially smoothed estimates of intensity (density), methods (such as the K function estimation mentioned above) for the description and estimation of spatial structure in point patterns and tests for space-time clustering, among other options. It provides a flexible and very valuable environment within which to perform the point-based methods outlined above.

Finally, we mention a package, INFO-MAP, written to accompany the text *Interactive Spatial Data Analysis*.¹⁵ This PC package was prepared as an educational aid and allows the user to create, explore and analyse spatial data sets of modest size (for example, typically fewer than 1000 areal units). It is written with the analysis of *spatial* data explicitly in mind, permitting the user to create maps and other views (graphs, histograms, scatter plots, and so on) of data and to perform a variety of both non-spatial statistical analyses (such as least squares regression and principal components analysis) and analyses that recognise the spatial relationships present in the data. Many of its functions are menu-driven, so that one simply defines the 'base file' or study region, together with a data file, and then selects a particular function. In addition, its very simple command language allows an extensive set of techniques to be implemented, such as the second-order analysis of spatial point patterns (K function estimation), autocorrelation tests, and empirical Bayes estimation discussed above.

CONCLUSIONS

In this paper we have sought to give an up-to-date account of the 'material and methods' that those working in ocular epidemiology should find appropriate for spatial analysis. We discussed the database requirements, particularly the way in which geographical information could be derived from postcoded data, but also commenting on a specific long-term study into the geographical incidence of anophthalmos and microphthalmos. We argued that a Geographical Information Systems framework pro-

vides a natural vehicle within which to construct and analyse such databases. Next, we considered a variety of analytical techniques. First, those that are appropriate when data are collected by, or analysed within, an areal unit framework; second, those that may be employed when the data are to be treated as a spatial point pattern. In particular, the issue of detecting 'clustering' of disease was given prominence. Last, we reviewed a number of possible software environments that could be used as frameworks within which to use such techniques. These will change, of course, as new developments in computing come on stream; however, the principles and methods remain more permanent and provide a potentially valuable set of tools to be exploited by the epidemiologist interested in ocular disease.

APPENDIX: DETAILS OF SOFTWARE AVAILABILITY

Here, we give further information and addresses if readers wish to seek further information on some of the software mentioned in the paper. There are many other available products on the market.

ARC/INFO

ARC/INFO is one of the leading proprietary GIS systems and is available for both PC and workstation environments, though the latter is recommended. The command language is sophisticated and the software demanding, though 'add-ons' are available that aid the novice user.

Contact: ESRI (UK), 23 Woodford Road, Watford WD1 1PB, UK.

IDRISI

IDRISI is a low-cost, PC-based, GIS that is mainly 'raster-based', that is, suitable where data are in gridded form. However, 'vector', or coordinate-based data (such as those for points or irregularly shaped areal units) can be imported into the package, which offers good display and analysis capabilities.

Contact: IDRISI Project, Department of Geography, Clark University, Worcester, MA 10160, USA.

SPlus

SPlus is available for the PC, but highly recommended is the UNIX workstation version. Splus is a sophisticated statistical programming environment, containing all the statistical analysis functions one would expect, together with excellent graphics facilities.

Contact: Statistical Sciences (UK), Sandfield Road, Oxford, UK. For details of the SPlus library SPLANCS contact: Professor P. Diggle, Depart-

ment of Mathematics, Lancaster University, Lancaster LA1 4YB, UK.

INFO-MAP

INFO-MAP is PC-based software designed as an educational aid in teaching spatial data analysis, comprising statistical graphics, a range of statistical, logical and mathematical functions, and spatial analysis functions, for handling several types of spatial data, including point patterns and areal data. The software is available as part of the text *Interactive Spatial Data Analysis* by Bailey and Gatrell,¹⁵ together with appropriate instructional materials.

Contact: Longman Geoinformation, 307 Cambridge Science Park, Milton Road, Cambridge CB4 4ZD, UK.

Key words: Ocular diseases, Spatial point patterns, Spatial clustering.

REFERENCES

- Haggett P, Cliff AD. Atlas of disease distributions. Oxford: Blackwell, 1989.
- Draper G. (ed.) The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83. Studies in Medical and Population Subjects no. 53. London: OPCS, 1991.
- Bithell JF, Stone RA. On statistical methods for analysing the geographical distribution of cancer cases near nuclear installations. *J Epidemiol Community Health* 1989;43:79–85.
- Stoll C, Alembik Y, Dott B, Roth MP. Epidemiology of congenital eye malformations in 131 760 consecutive births. *Ophthalmic Paediatr* 1992;13:179–86.
- Clementi M, Turolla L, Mammi I, Tenconi R. Clinical anophthalmia: an epidemiological study in north-east Italy based on 368 256 consecutive births. *Teratology* 1992;46:551–3.
- Scottish Office. Report into the incidence of microphthalmos in Forth Valley area. London: HMSO, 1988.
- Welsh Office. The incidence of congenital malformations in Wales, with particular reference to the district of Torfaen, Gwent. London: HMSO, 1985.
- Gatrell AC, Lovett AA. Congenital eye malformations and incineration in Britain: a geographical analysis. In: Akhtar R, editor. *Environment and health*. Delhi: Ashish Publishing House, 1991:417–30.
- Raper J, Shepherd J, Rhind DW. Postcodes: the new geography. Harlow: Longman Geoinformation, 1992.
- Gatrell AC, Dunn CE, Boyle PJ. The relative utility of the Central Postcode Directory and Pinpoint Address Code in applications of Geographical Information Systems. *Env Planning A* 1991;23:1447–58.
- Dale A, Marsh C. (eds.) The 1991 census user's guide. London: HMSO, 1993.
- Walter SD. The ecologic method in the study of environmental health: I. Overview of the method. *Environ Health Perspectives* 1991;94:61–5.
- Kelsey JL, Thompson WD, Evans S. Methods in observational epidemiology. Oxford: Oxford University Press, 1986.
- Maguire DJ, Goodchild MF, Rhind DW. Geographical information systems: principles and applications. Harlow: Longman, 1991.
- Bailey TC, Gatrell AC. Interactive spatial data analysis. Harlow: Longman, 1995.
- Elliott P, Cuzick J, English D, Stern R. (eds.) Geographical and environmental epidemiology: methods for small area studies. Oxford: Oxford University Press, 1992.
- Cliff AD, Ord JK. Spatial processes: models and applications. London: Pion, 1981.
- Walter SD. Assessing spatial patterns in disease rates. *Stat Med* 1993;12:1885–94.
- Choynowski M. Maps from probabilities. *J Am Stat Assoc* 1959;54:385–8.
- Barreto ML. The dot map as an epidemiological tool: a case study of *Schistosoma mansoni* infection in an urban setting. *Int J Epidemiol* 1993;22:731–41.
- Diggle PJ. Point process modelling in environmental epidemiology. In: Barnett V, Turkman KF, editors. *Statistics for the environment*. Chichester: Wiley, 1993.
- Diggle PJ. Statistical analysis of spatial point patterns. London: Academic Press, 1983.
- Diggle PJ, Chetwynd AG. Second-order analysis of spatial clustering. *Biometrics* 1991;47:1155–63.
- Gatrell AC, Mitchell J, Gibson H, Diggle PJ. Tests for clustering in epidemiology: with special reference to motor neurone disease. In: Clifford Rose F, editor. *Advances in ALS research*. London: Smith-Gordon, 1991:67–77.
- Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *JR Stat Soc*, B1990;52:73–104.
- Bhopal R, Diggle PJ, Rowlingson BS. Pinpointing clusters of apparently sporadic Legionnaires' disease. *BMJ* 1993;304:1022–7.
- Siemiatycki J, Brubaker G, Geser A. Space-time clustering of Burkitt's lymphoma in east Africa: analysis of recent data and a new look at old data. *Int J Cancer* 1980;27:209–20.
- Lenoir GM. Burkitt's lymphoma: a human cancer model. Lyon, France: IARC, 1985.
- Openshaw S, Charlton M, Wymer C, Craft A. A mark 1 geographical analysis machine for the automated analysis of point data sets. *Int J Geogr Inf Syst* 1987;1:335–58.
- Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc A* 1991;154:143–55.
- Diggle PJ. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a pre-specified point. *J R Stat Soc*, A1990; 153:349–62.
- Diggle PJ, Rowlingson BS. A conditional approach to point process modelling of elevated risk. *J R Stat Soc*, A 1995;157:489–503.
- Morris JA, Butler R, Flowerdew R, Gatrell AC. Retinoblastoma in children of former residents of Seascale. *BMJ* 1993;306:650.
- Eastman JR. IDRISI: a geographical information system. Worcester, Mass.: Department of Geography, Clark University, 1992.
- ESRI. Understanding GIS: the ARC/INFO method. Harlow: Longman Geoinformation, 1992.
- Gatrell AC, Rowlingson BS. Spatial point process modelling within a GIS environment. In: Fotheringham AS, Rogerson PA, editors. *Spatial analysis and GIS*. London: Taylor and Francis, 1994:147–63.
- Rowlingson BS, Diggle PJ. SPLANCS: spatial point pattern analysis code in S-Plus. *Comp Geosci* 1993;19:627–55.