

John F. Bithell

Contents

37.1	Introduction	1415
37.1.1	The Nature of Geographical Epidemiology	1415
37.1.2	Scope of the Chapter	1415
37.1.3	Chapter Contents	1416
37.2	Statistical Models	1417
37.2.1	A Statistical Framework for Epidemiological Observations	1417
37.2.2	Statistical Models for Geographical Data	1419
37.3	Modeling Disease Risk in Relation to Geographically Referenced Factors.....	1420
37.3.1	Areal Data	1420
37.3.2	An Example of the Log-Linear Model for Areal Data.....	1422
37.3.3	Calculating the Expectations.....	1423
37.3.4	Continuous Data	1423
37.3.5	Spatial Structure in the Residual Variation.....	1424
37.4	Mapping Disease Risk	1425
37.5	The Detection of Generalized Heterogeneity	1428
37.5.1	The Assessment of Heterogeneity in Areal Data	1428
37.5.2	Detecting Heterogeneity in Poisson Data	1429
37.5.3	Spatial and Non-Spatial Analyses	1431
37.5.4	Heterogeneity Tests Based on the Risk Surface	1432
37.6	Clustering.....	1433
37.6.1	Methods Based on the RRF	1434
37.6.2	Knox's Test	1435
37.6.3	Other Space-Time Clustering Methods.....	1436
37.6.4	Space-Only Clustering	1436
37.6.5	Population Distance	1437
37.6.6	Choosing Scale Parameters	1437

J.F. Bithell
St Peter's College, University of Oxford, Oxford, UK

37.7	Predefined Sources of Risk	1438
37.7.1	Tests for Concentration of Risk	1439
37.7.2	Example: Childhood Leukemia Around UK Nuclear Installations	1441
37.7.3	Summary of Recommendations	1442
37.8	Conclusions	1443
	References	1444

37.1 Introduction

37.1.1 The Nature of Geographical Epidemiology

Although, at first sight, geographical epidemiology may appear to differ substantially from other areas of epidemiology, it has many features in common. In particular, a major objective of epidemiology – to infer etiological relationships from observed associations – applies also in geographical studies. The distinctive characteristic is of course that geographical location is an important explanatory variable, either because it reflects an environmentally determined element of risk or because people with similar risk attributes live together, so that risk varies from place to place. The two-dimensional nature of geographical location means that the standard statistical techniques for handling sets of essentially univariate variables need to be augmented by more sophisticated methods.

There are practical limitations to the scientific value of geographical studies. The data quality tends to be low – not least because population censuses are relatively infrequent – and any real effects may be attenuated by factors such as mobility, often to the point where they are not detectable. Consideration of these difficulties may lead to the conclusion that a lot of geographical epidemiology is, in scientific terms, of very limited value. Historically, however, there have been some spectacular successes: to the famous observation of Snow (1855) on the source of cholera infection may be added a number of more recent and equally dramatic observations, for example, the identification of the cause of an outbreak of asthma in Spain (Antó and Sunyer 1992) and the implication of erionite fibres in the etiology of mesothelioma from the very high localized rates in the Cappadocian region of Turkey (Baris et al. 1992).

37.1.2 Scope of the Chapter

This chapter attempts to sketch the statistical principles of the subject, with an indication of the kinds of analyses to which these principles lead quite naturally. There is a large literature on the methodology of geographical epidemiology, much of it employing a Bayesian standpoint and exploring hierarchical models analyzed by Markov chain Monte Carlo methods. It would be impossible to give a comprehensive review of the latter field, and we adopt the less ambitious objective of outlining the fundamentals of the subject, in the hope that this will in any case provide insight into more sophisticated analyses. Nevertheless we have attempted to provide some examples of the techniques discussed and, where possible, to make recommendations for practitioners, though this latter goal is difficult in view of the large number of different analyses that have been proposed but whose properties are relatively unknown.

Our presentation will in fact be almost exclusively frequentist. To some extent, the choice between Bayesian and frequentist methods in statistics is a matter of

philosophical standpoint. Frequentist arguments are undeniably limited in their scope and power and are frequently subject to misinterpretation. The limitations may, however, be argued to be intrinsic to the problem of inductive inference under uncertainty and such inference does not seem to this author to be more consistently clear-cut when derived from a Bayesian analysis. The modeling approach is admittedly more attractive than the mere detection of statistical significance, but it is not without its difficulties. For one thing, the amount of data in geographical studies may often not permit the estimation of numerous parameters, and to the extent that a model makes specific assumptions about underlying phenomena, there is a risk that it may inject spurious information into the analysis, leading to the overinterpretation of the data. The limitations of the hypothesis testing approach have not prevented its widespread use in practice, and an important part of the epidemiologist's role is to ensure that the tests that are carried out are chosen with due regard to maximizing their power against sensible alternatives. This at least is the standpoint from which we approach this topic here; in any case, the statistical framework underpins the more sophisticated analyses and forms a natural prerequisite for their understanding. See chapter ► [Statistical Inference](#) of this handbook for a discussion of the fundamental distinctions between Bayesian and frequentist inference, and chapter ► [Bayesian Methods in Epidemiology](#) of this handbook for an account of Bayesian modeling.

37.1.3 Chapter Contents

We start by considering ([Sect. 37.2](#)) the models that underlie statistical methods in geographical epidemiology in order to give insight into the justification for the methods that are discussed. A key feature is the duality that exists between the two approaches to epidemiological investigations generally. To be specific, we can elect to study either the occurrence of disease conditionally on case locations or vice versa, i.e., to regard case location as a random variable to be compared in fixed groups of affected and unaffected individuals. This duality precisely mirrors the distinction between the cohort and case-control approaches to epidemiological surveys. The case-control approach in geographical work has only recently been recognized and is particularly relevant for the analysis of data at the individual, as opposed to the areal, level. This important approach, though not yet fully exploited, has led recently to a number of new and interesting methodological developments.

In [Sect. 37.3](#), we develop the way in which risk may be modeled in relation to geographically referenced data, distinguishing between the analysis of areal data and data at the individual level, for which it is assumed that individual locations are known. As with any statistical modeling exercise, the objective is to explain as much of the variation as possible, up to the point where heterogeneity can be attributed to chance. There are numerous ways of approaching this subject, even within the compass of frequentist analyses, and some of the issues as to the best analysis are unresolved.

[Section 37.4](#) is concerned with mapping. From one point of view, mapping is an end in itself, and there are numerous methods available for producing maps. However, there is much scope for misinterpretation of data represented in this way, and we would argue that a map should be seen as the end product of some kind of modeling process, though possibly a very primitive one: no disease map can be constructed without assumptions about the underlying distribution of the disease it purports to represent.

[Section 37.5](#) addresses the question of heterogeneity in the distribution of risk. To some extent, this involves issues bound up with the problems of modeling. But the simple question of whether there is any non-uniformity of risk is a valid one that can be at least partially answered without reference to underlying models or alternatives.

In [Sect. 37.6](#), we address the problem of clustering. This may be seen as a violation of the twin assumptions of uniformity and independence discussed in [Sect. 37.2](#). However, we may well be more interested in detecting small clusters of cases that are related to one another, and to this extent it may be appropriate to use different methods from those in [Sect. 37.5](#).

Finally, [Sect. 37.7](#) considers the rather more specific problem of detecting an increase in risk near a putative point source of risk, and it is argued that analyses of this kind are essentially one dimensional, and perhaps for this reason, it is somewhat easier to determine good methods for doing so. This is in fact a problem of considerable interest, and many investigations of “clustering” are really of this kind. The issue is illustrated by the incidence of childhood leukaemia around nuclear installations in the UK using data introduced in [Sect. 37.3.2](#).

The concluding section summarizes the chapter and makes suggestions for further reading.

37.2 Statistical Models

In this section, we describe a statistical framework for the methods to be discussed. We start by explaining the elements that underlie the analysis of classical surveys and then show how the same starting point may be applied to geographical data.

37.2.1 A Statistical Framework for Epidemiological Observations

To describe a modeling framework for epidemiology, we start by supposing that the disease \mathcal{D} in which we are interested is an essentially dichotomous entity, i.e., it is the binary outcome – affected/not affected – of some biological process applied to a finite set of individuals. Such a starting point will serve irrespective of the temporal nature of the events we are studying, be they deaths or incident cases of a disease \mathcal{D} in a given time period or the prevalence of \mathcal{D} at a given epoch. We will be primarily interested in the association between \mathcal{D} and various covariates \mathcal{C} . Some of these may represent risk factors suspected of playing a causal role: we will describe these as

exposure variables and denote them by \mathcal{E} . Others may be of interest in their own right or because they are potential confounding variables for \mathcal{E} . We will treat \mathcal{E} as a subset of \mathcal{C} when this is convenient.

To take a specific geographical example, we cite the famous study of cardiovascular disease \mathcal{D} by Cook and Pocock (1983). The covariates \mathcal{C} included water hardness \mathcal{E} , whose etiological relationship to cardiovascular disease was of primary interest, and also various indicators of socioeconomic status, which played the role of a confounding factor: the gradients of mortality, water hardness, and socioeconomic status are highly correlated with latitude in the UK. The data were analyzed for males and females together, but they could equally well have been stratified by sex, which would be a covariate of interest in its own right, since one might be interested in the mortality of males and females separately.

Next, we assume that occurrences of \mathcal{D} are independent. This does not preclude the possibility that individuals have probabilities p of \mathcal{D} that are related through their proximity, for example. Rather the condition stipulates that, *conditional on the values of \mathcal{C} and \mathcal{E}* , the occurrence of \mathcal{D} in one individual is independent of that in another, i.e., that the probability that individual A suffers from \mathcal{D} is unaffected by the *fact* (as opposed to the probability) that some other individual B also suffers from it. In practice, this is a reasonable mechanistic assumption for nearly all chronic disease epidemiology. It clearly breaks down for infectious diseases, for which more sophisticated models would be appropriate. In fact, little theoretical foundation for modeling the epidemiology of infectious diseases at the individual level exists. This is partly because the theory is intractable, partly because it is not necessary in setting up the null hypothesis of no contagion for the purposes of testing. It is only for formulating alternative hypotheses in this situation that statistical models for a contagious mechanism are necessary. Important though this is, we will not consider the problem in this chapter.

Under this independence assumption, the individual outcomes of \mathcal{D} are described by the very simple Bernoulli distribution. If all the probabilities p_i for the individuals in a group of n are the same, the number of occurrences out of the n will clearly follow the binomial distribution, while if all the p_i are different and supposed to depend on \mathcal{C} , we can model them through a (binary) logistic regression (Cox and Snell 1984).

Such analyses are becoming more common, but they require detailed information on individuals and are not without their technical difficulties. Much of epidemiology is in practice still conducted by the more traditional approach of grouping data according to disease status and to grouped values of \mathcal{C} . In this approach, the assumption is that the probabilities p_i within a particular group are indeed all the same, though in practice we know that this is unlikely to be true. However, this assumption is far less troublesome than appears at first sight. For one thing, as long as the p_i are small, the difference between a binomial distribution and that of a sum of slightly different Bernoulli variables will be negligible.

A typical analysis of epidemiological data proceeds by forming a cross-tabulation into a contingency table, whose rows, columns, and layers are labeled by components of \mathcal{D} , \mathcal{E} , and \mathcal{C} . The standard way of analyzing such a table is through a log-linear model, which implicitly assumes that the counts in the table are

values of Poisson distributed variables, conditioned by the requirements that certain subtotals in the table are deemed to be fixed. For details on log-linear regressions please refer to chapter ► [Regression Methods for Epidemiological Analysis](#) of this handbook.

The logistic regression and log-linear modeling approaches thus described have been constructed on the assumption that \mathcal{D} is a random response and the covariates \mathcal{C} are fixed, but we can also obtain useful analyses by conditioning on the numbers of “cases” affected by \mathcal{D} and unaffected or disease-free “controls” in a suitable control group and regarding one or more of the covariates as a random response. This leads to the so-called “case-control” study (formerly termed a *retrospective* study), in distinction to a “cohort” (or *prospective*) study. Thus, for example, it might be appropriate to use a normal linear regression to model the exposure \mathcal{E} of individuals to some risk factor – considered to be a continuous variable – as a function of the other variables, one of which would be an indicator for \mathcal{D} , the membership of the case or control group. We would then regard \mathcal{E} as the factor of primary interest and the other covariates would be fitted in order to control for their possible confounding effects.

37.2.2 Statistical Models for Geographical Data

Most of the ideas outlined above carry over quite naturally to data in which geographical location plays a role. We will preserve the assumptions that \mathcal{D} is a binary variable and that disease occurrences are independent conditionally on \mathcal{C} . We need to extend our conceptual notation to include geographical location, which we will denote by \mathcal{G} . There is a distinction between situations where we think of it as representing a pair of coordinates and those where it is an essentially two-dimensional location in the space representing a geographical region studied.

If \mathcal{G} is thought of as representing coordinates, such as Easting and Northing, it may be meaningful to treat them like other quantitative variables, perhaps to detect a trend with latitude, for example. Alternatively, it might be meaningful to consider polar coordinates from a specified point \mathcal{S} considered as a fixed origin, analyzing distance and direction from it. Typically, \mathcal{S} would be a point of some etiological significance, such as a putative source of pollution. We return to this topic in [Sect. 37.7](#) below.

However, this approach implicitly reduces our analyses to consideration of essentially one-dimensional variables, and it is useful to distinguish this from the *intrinsically spatial* case in which we regard two-dimensional space as a single entity. In this situation, a principal objective will be to depict the way in which risk varies over a region \mathcal{R} , usually by means of a map. It is unlikely that any kind of analytically determined trend surface, such as a polynomial, will be useful, though non-parametrically estimated surfaces might be. We return to the problems of mapping in [Sect. 37.4](#) below.

The distinctions we made in [Sect. 37.2.1](#) above apply for geographical data. For example, the majority of geographical analyses are effectively analyses of grouped data, in which observations have been grouped into k subregions A_1, A_2, \dots, A_k of

\mathcal{R} (which we shall refer to as “areas”). Within each area, we would hope to know the population to serve as a denominator and the number of occurrences Y_i of the disease \mathcal{D} would then follow a binomial or approximately a Poisson distribution, by the arguments outlined above. The areas may be regarded as analogous to the bins of a histogram, though they will nearly always be based on administrative areas with highly irregular boundaries, so that they do not share the attractive regularity properties of the more familiar histograms formed from quantitative variables. The identities of the areas themselves typically enter the analysis through the coordinates of their population centroids, and these may then be analyzed by incorporating them into the model as described above, though the analysis might well take account of spatial autocorrelation.

If instead of binning or grouping the observed cases into areas we record the exact locations of the occurrences of \mathcal{D} , we need a rather different modeling approach. The case-independence assumption implies that the cases are located according to a non-homogenous Poisson process (Diggle 2000), which is the standard probability model for events happening at random in a continuum, though not necessarily with a uniform pattern of risk. This model supposes that the probability of an event in a small area δA at the point (x, y) is $\lambda(x, y)\delta A$, where $\lambda(x, y)$ is the “intensity function” of the process giving the rate per unit area at (x, y) ; it also incorporates the crucial assumption that the occurrence of such a point is independent of occurrences outside δA .

It is well known, however, that when points occur according to a Poisson process in such a way that the total number is fixed at a value n , say, the pattern of points obtained is typically exactly the same as if we had sampled from a probability distribution with density function proportional to $\lambda(x, y)$. This enables us to describe the behavior in geographical space of a fixed sample of cases, with a view to estimating the risk at each point (x, y) or to compare the resulting risk function with that for a sample of controls. Thus we have moved to the “dual” or case-control approach, for we are effectively regarding the locations as realizations of a continuous bivariate random variable defined for our samples of cases and controls. Methods of analyzing data within this framework are discussed in Sect. 37.3.4 below.

37.3 Modeling Disease Risk in Relation to Geographically Referenced Factors

37.3.1 Areal Data

One of the commonest and most straightforward analyses of geographical data consists of modeling the counts Y_i of cases in areas A_i using a Poisson regression or, equivalently, a generalized linear model (GLM) with Poisson error and log link function; see McCullagh and Nelder (1989) and chapter ► [Regression Methods for Epidemiological Analysis](#) of this handbook. We start by assuming that we can

calculate “null expectations” e_i for the Y_i . In the simplest form, these could be obtained by multiplying some global reference estimate of risk p by the population sizes in the A_i . In practice, we will almost certainly wish to *standardize* for the age distribution and other known demographic factors such as socioeconomic status. Part of our objective is of course to modify the assumption that the risk is the same in every area, so we will incorporate a *relative risk* (RR) θ_i , to give the model for the counts as

$$Y_i \sim \text{Poisson}[\theta_i e_i].$$

We then model the θ_i in the usual manner for a GLM through

$$\log \theta_i = \sum_{j=1}^p x_{ij} \beta_j,$$

where the β_j are coefficients in the log-linear model and x_{ij} is the value of the j th covariate for the i th areal unit A_i .

Typical covariates in such an analysis might include intrinsically geographical features, such as altitude, geological composition, or levels of background radiation, or essentially demographic features, such as the age or socioeconomic composition of the population of each area. It should be emphasized that the units in such analyses are not the individuals with a disease \mathcal{D} but the areas within which they reside, and the covariates are also necessarily attributes of these areas. The object of such an analysis, however, will generally be to make inferences concerning individuals, and to ignore the distinction is sometimes described as perpetrating the “ecological fallacy.” Covariate values for the area as a whole are implicitly imputed to each individual member of the population, and this has the potential for introducing a number of different kinds of bias known variously as “ecological” or “aggregation bias.”

A genuinely “ecological” or geographical imputation would arise if a geographical feature (such as latitude) were averaged spatially without regard to population density (Diggle and Elliott 1995), and any such averaging should as far as possible be density-weighted, perhaps by using the relevant measurement at the centroid of the population. Demographic variables, such as age or socioeconomic deprivation will usually be averaged over the population anyway, and in this case grouping into areas has much the same effect as grouping by other factors; the problem may then be seen in the wider context of aggregation bias. Such a bias can result from concealed within-group confounding, and it is difficult to take account of this without individual-level data. An intrinsic bias may also arise from the non-linearity of the model used, though this is likely to be small when the disease risk is itself small and relatively uniform, since the logistic (or any similar) transformation used in the model will be nearly linear.

Many papers have addressed the issue of ecological bias. An early contribution by Greenland and Morgenstern (1989) was influential but may have painted too pessimistic a view of ecological studies, which can be very valuable for providing

pointers and which are often based on more objective data than case-control studies. Wakefield (2008) provides a useful overview and a review of the literature. Further discussion of the issues will also be found in chapter ► [Descriptive Studies](#) of this handbook.

37.3.2 An Example of the Log-Linear Model for Areal Data

An example of this use of the log-linear model is provided by the application to childhood leukaemia data described by Bithell et al. (1995). The dataset analyzed was from the UK National Registry of Childhood Tumours (NRCT) maintained by the Childhood Cancer Research Group in Oxford and related to 5,359 children diagnosed with leukaemia or non-Hodgkin lymphoma under the age of 15 years between 1966 and 1987. Each of the cases was located in one of 9,831 electoral wards, which are administrative areas with an average population of around 5,000.

The explanatory variables fitted were “Standard Region,” a classification of Britain into ten regions, and the Townsend Index, an areal index of social deprivation which is a function of unemployment, housing ownership, and other socioeconomic indicators. As shown in [Table 37.1](#), there was a significant reduction in the deviance associated with each of these factors: the p -values shown in the first two lines are based on the chi-square approximation to the deviance reductions. It is interesting, incidentally, to note that the direction of the association is negative for the Townsend Index, i.e., the disease is slightly commoner in less deprived families. This is a feature of childhood leukaemia that differentiates it from most other diseases.

The goodness of fit of the model can in principle be tested by the residual deviance, but because the expected numbers of cases per ward in this analysis were small (less than half on average), the chi-square approximation is unreliable. However, the theoretical mean and variance of the deviance for Poisson observations with a specified set of expectations can be calculated straightforwardly. We can therefore obtain an approximate test of the residual deviance as follows:

1. Compute the values for the expectations predicted by the model for each ward.
2. Compute the mean μ and variance σ^2 of the deviance statistic D as defined by

$$D = 2 \sum_i [Y_i \log(Y_i/e_i) - (Y_i - e_i)] \quad (37.1)$$

as if the contributions to D were independent.

3. Refer the statistic $(D - \mu)/\sigma$ to the standard normal distribution.

Table 37.1 Analysis of deviance of childhood leukaemia data

Variation due to	d.f.	Deviance	p -value
Standard region	9	23.1	0.0060
Townsend index	1	23.6	10^{-6}
Residual	9,820	8610.6	0.025

The assumption of independence should be approximately true in view of the large number of degrees of freedom. Bithell et al. check the two-sided p -value by simulation of data from the fitted model and found a very good degree of approximation to the calculated value of 0.025. These results may be interpreted as meaning that the model fits much better than if the explanatory factors had not been taken into account (for which the equivalent p -value was 0.00042); though there is some evidence of residual heterogeneity, it must be remembered that this is a large data set and the level of significance observed is not indicative of a large degree of variation. We return to the issue of testing residual variation in [Sect. 37.5](#).

37.3.3 Calculating the Expectations

The model described above involves the expectations e_i , which appear as an “offset” term in the model, i.e., $\log(e_i)$ is added to the linear function of the covariates defining $\log \theta_i$. These may be calculated from externally calculated rates, for example, from national statistics. If such rates are not easily available, the data can be internally standardized by supplying the sizes of the populations at risk; any factor representing the overall risk will appear in the intercept term of the model. The expectations predicted by the model can then be used as expectations for subsequent analyses, and this is a useful by-product of the modeling process. The method can be seen as an elegant and more consistent alternative to classical standardization, permitting the flexible inclusion of covariates according to their importance, as indicated by the modeling process.

Indeed, the analysis described by Bithell et al. is part of a larger one designed to produce expected numbers of childhood leukaemias for the areal analysis of incidence near nuclear installations; this is briefly described in [Sect. 37.7.2](#).

37.3.4 Continuous Data

Following the discussion in [Sect. 37.2.2](#) above, we suppose that we have a sample of exact locations of cases of disease \mathcal{D} and that we denote their density function over \mathcal{R} by $\psi(x, y)$. We need an analogue of the denominators in an areal analysis to serve as a measure of how many individuals there are at risk at each point (x, y) of \mathcal{R} . This is provided in principle by knowledge of the population density, which we will consider to be continuous and which we will denote by $\pi(x, y)$. Then our problem becomes one of comparing the density function for the incident cases with that of the population. For a rare disease, the population density (which strictly speaking includes diseased as well as healthy individuals) will be very similar to that for all non-diseased individuals, which can in turn be estimated by a suitable *sample* of controls. The natural way to make this comparison is through the ratio, and it is easily seen that this ratio

$$\theta(x, y) = \psi(x, y)/\pi(x, y)$$

defines a *relative risk function* (RRF) that gives the risk of being affected by \mathcal{D} at each point (x, y) of \mathcal{R} relative to the mean for the whole of \mathcal{R} (see Bithell 1990).

A natural estimate $\hat{\theta}(x, y)$ of $\theta(x, y)$ is provided by the ratio of estimates of $\psi(x, y)$ and $\pi(x, y)$. These may be obtained using one of the modern methods available for estimating a density function (see the books by Silverman (1986) and Scott (1992), e.g.). The process is not without difficulties but it can be used to provide meaningful estimates of the RRF over \mathcal{R} , in effect providing a map of it. We return to the problem of mapping in Sect. 37.4 below.

A more ambitious objective than merely mapping the RRF is to model it as a function of covariates \mathbf{x} , say. These may be geographically defined at every point of \mathcal{R} , or they may be attributes of the cases and controls in the samples. An elegant modeling approach is due to Diggle and Rowlingson (1994) and proceeds by analogy with classical case-control studies. We condition on the coordinates of the n cases and m controls and consider the probability that, under random allocation of the cases and controls to the $m + n$ locations, an individual sampled at a given location (x, y) is a case rather than a control. This probability can then be modeled logistically as a function of \mathbf{x} . If there appears to be unexplained variation in the RRF, it can in principle be modeled by adding a non-parametric function of (x, y) to the linear predictor. The numerical problems of the latter approach appear not to be trivial.

The inclusion of attributes of the individuals in the analysis is particularly attractive, since it provides the possibility of controlling for them within the geographical analysis. In practice, it is not always straightforward to obtain suitable controls for analyses of this kind, partly because the current emphasis on data protection makes it difficult to access individual records and partly because of the number of combinations of categories with respect to which we may wish to match. Nevertheless, this methodology, though still in its infancy, would seem to have considerable potential.

37.3.5 Spatial Structure in the Residual Variation

The object of fitting a model of the kind discussed is to obtain a satisfactory explanation of the data, i.e., a residual deviance that is not statistically significant. This is not always very easy, since the risk of disease may depend on factors that we have been unable to measure. Large data sets – for example, of national mortality rates – may also demonstrate a statistically significant deviance resulting from unobserved factors that are scientifically unimportant simply because of the large numbers of cases involved.

Unfortunately, conclusions about the importance of individual explanatory variables in a model are strictly valid only if the model fitted is correct. In practice, we will believe a model to be correct if it appears to fit reasonably well, i.e., if the residual deviance is not statistically significant. This raises the question of how to proceed if there is a degree of residual variation that we cannot explain.

In geographical studies, it is quite likely that such variation will be due to unobserved variables that are spatially autocorrelated, and in this case we can include terms in the model designed to reflect this autocorrelation. Typically, this is done for data in areal form using a conditional autoregression (CAR) model (Wakefield et al. 2000) while, for continuous data, Kelsall and Diggle (1998) use a generalized additive model (GAM) which effectively gives an extra term in the model estimating residual variation non-parametrically. These ideas are important but are somewhat beyond the scope of this chapter; see Pfeiffer et al. (2008) for an introductory account of spatial models and Diggle (2000) for a good overview of the field. We only remark that the issue may not always be as significant as some authors maintain. The deviances of the terms that are fitted in a model will still be a reliable indication of their importance unless they are confounded with the unobserved variables that are inflating the deviance; in this case, fitting a spatial model merely tells us that this confounding has a spatial structure – it does not help us to identify the variable or determine its scientific importance.

37.4 Mapping Disease Risk

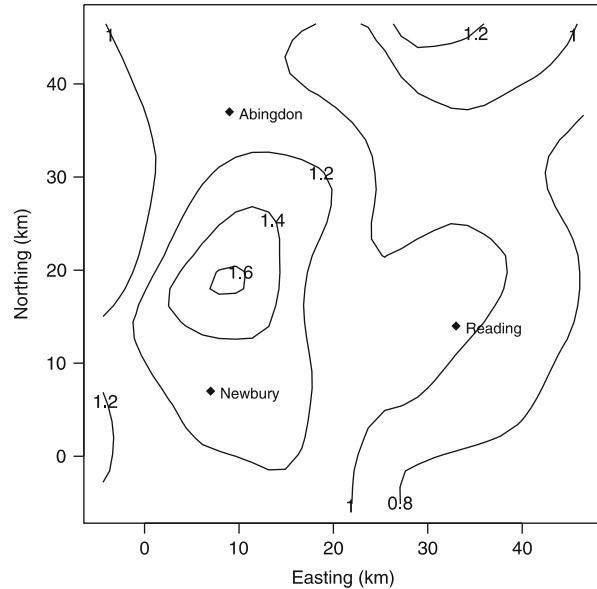
The mapping of disease risk is a central endeavor of geographical epidemiology: a map is as convenient for portraying such location-specific information as it is for indicating the geography of the land to which it relates. It is therefore no surprise to discover that mapping has a long history predating any systematic development of the statistical principles that underlie it.

As with other areas of geographical epidemiology, many methods have been proposed. Broadly speaking, these can be divided into two classes, model-based and non-parametric. Methods in each of these classes can be applied to data in either areal or continuous form. It is important to appreciate, however, that, whatever method is applied, there is inevitably a degree of smoothing involved that is to some extent arbitrary and under the control of the investigator.

For example, the simplest form of map is the so-called choropleth map, which uses a gray or color scale to depict the risk of \mathcal{D} in each of a number of areas, usually administratively defined so that denominators are easily available. Here, the degree of smoothing is determined by the size of the areas A_i , since the process represents the risk as being the same throughout each given area. An example of a choropleth map is given in chapter ▶ [Descriptive Studies](#) of this handbook.

Similarly, data in continuous point form can be mapped using the methods described in [Sect. 37.3.4](#) by plotting the RRF $\hat{\theta}(x, y)$. Here, the smoothing is determined by the degree of smoothing used in the estimation of the densities: it is a commonplace of this methodology that some smoothing parameter always has to be used, though there are data-driven methods for estimating the most appropriate value. See Bithell (1990) for an early example of this method applied to small numbers of cases and controls, and Davies and Hazelton (2010) for a more recent development of the methodology.

Fig. 37.1 Relative risk function for childhood cancer in a region of Oxfordshire, estimated from areal data. The three town centers are shown only approximately. The ASH smoothing parameter used was 8 (See Bithell 1999 for details)



It may be noted that the RRF method can easily be adapted to areal data by suitably modifying the customary density estimation methods (Bithell 1999). Figure 37.1 depicts the incidence of childhood cancer in a 50-km square region of Oxfordshire using data from the UK National Registry of Childhood Tumors maintained by the Childhood Cancer Research Group in Oxford. They consist of 279 cases of childhood cancer (other than leukaemia and non-Hodgkin lymphoma) registered under the age of 15 years between 1966 and 1987. Each case was located in one of 150 electoral wards for which expected numbers of cases were calculated using similar methods to those for the leukaemia data described in Sect. 37.3.2. The point observations for the cases were used to construct a density estimate $\hat{\psi}(x, y)$ using the *average shifted histogram* (ASH) method due to Scott (1992). For the controls, the density estimate $\hat{\pi}(x, y)$ was constructed by treating the centroids of the wards as point locations weighted by the expectations and using a version of ASH modified accordingly.

The basis of the ASH method is to count the numbers of cases in the cells of a square grid; these are then smoothed by slightly shifting the grid a number of times and averaging the resulting counts; this process effectively smoothes the surface by spreading out the contributions of the points through neighboring grid squares.

The RRF was then obtained by dividing the density estimates for the cases and controls to give $\hat{\theta}(x, y) = \hat{\psi}(x, y) / \hat{\pi}(x, y)$. This is depicted in Fig. 37.1 as a contour plot with a scale in km and an origin located in South West Oxfordshire.

The methods sketched above may be regarded as empirical or non-parametric, in that there is nothing underlying them that is more sophisticated than the division of one number by another (specifically a count by a denominator or one density

estimate by another). It is generally difficult to see how to determine the appropriate degree of smoothing by any objective process, as distinct from using intuitively plausible and aesthetically pleasing values.

The necessity for a degree of smoothing can easily be seen by considering a choropleth map, for which we have areas A_i with small numbers of cases, either because we have chosen small areas or because \mathcal{D} has low incidence. In this case, the estimates of the risk in each A_i will be subject to large sampling errors; our belief about the true risk in A_i will be determined in part by the observed rate, but it will also rely on information from the region as a whole, to the extent that we believe there will be some comparability between the areas.

This idea has led to the development of model-based approaches using Bayesian arguments to integrate area-specific information with information from the whole region, using a statistical model for the underlying variation of the true risk. In a classical treatment of this problem, Clayton and Kaldor (1987) suppose that the true risk θ_i in A_i is distributed over the areas as a whole according to a gamma distribution with mean μ and variance σ^2 . It can then be shown that the posterior distribution of θ_i has mean

$$\tilde{\theta}_i = \frac{y_i + \mu^2/\sigma^2}{e_i + \mu/\sigma^2},$$

where y_i is the observed value of the count Y_i in A_i . This formula can be seen to be a form of average of the maximum likelihood estimate $\hat{\theta}_i = y_i/e_i$ of each θ_i and the overall mean μ , which can be estimated by $\sum y_i / \sum e_i$. The value of σ^2 can also be estimated from the data as a whole, though this requires an iterative method.

This method and variants of it provide *empirical Bayes* estimates, in that the prior distribution of the θ_i can be estimated from the data. The method is essentially non-spatial, in the sense that the true θ_i is supposed to vary independently. In practice, it is likely that rates in neighboring areas will be consistently more similar to one another than those in more separated areas. If this were not so, it would be essentially fruitless to attempt to produce a smoothly varying map. The Bayesian methodology has been extended to permit the prior distribution of the θ s to depend on the values in neighboring areas. These more complicated models involve a greater number of arbitrary assumptions, however. They are gaining ground in popularity and appear to be used quite successfully. The reader is referred for more details and references to Clayton and Bernardinelli (1992) and to chapter [► Bayesian Methods in Epidemiology](#) of this handbook.

Attractive though these ideas are, the maps they produce need careful interpretation, since they have imposed a degree of spatial autocorrelation, and this process is capable of making adjacent areas look more similar than they really are. In a sense, this is true of all mapping methods and is a feature as intrinsic as the implicit smoothing itself.

In a challenging paper, Gelman and Price (1999) discuss the issue and illustrate the phenomenon of induced spatial pattern by means of simple modeling paradigms.

They point out that the probability that a particular area rate $\hat{\theta}_i$ exceeds a given value increases with decreasing population size, n_i , say. The effect of this is that high observed rates of disease tend to be observed predominantly in low population areas; since these tend to be spatially aggregated – i.e., low population areas are more likely to occur next to other such areas – observed rates also appear to be spatially related even when in fact no such relationship exists for the underlying risk.

They further demonstrate that plotting the posterior means from a Bayesian analysis produces observed rates that are likely to exceed a particular value with probabilities that are *decreasing* functions of n_i , so that such plots overcorrect in some sense. Although scores exist – at least for continuous observations – that are not subject to these artifacts, they have no direct interpretation as estimates of the θ_i .

One is driven to the conclusion that disease maps are potentially misleading when used as anything except what Gelman and Price call “look-up tables,” i.e., as a convenient way of depicting the rate in a given area *without reference to neighboring areas*. It is the temptation to use the map to generalize about the spatial pattern of rates that can be misleading, and it is probably better to formulate such questions within the context of a statistical model rather than to attempt to portray spatial relationship graphically. However, one suspects that this timely caution is unlikely to diminish the enthusiasm for constructing and overinterpreting disease maps.

37.5 The Detection of Generalized Heterogeneity

37.5.1 The Assessment of Heterogeneity in Areal Data

Heterogeneity is the key to epidemiology, in the sense that a uniform risk in observed data gives no possibility for associating differences with factors that may have etiological significance. We have already touched on the issue of modeling in [Sect. 37.3.1](#), and our objective there is to find a model that appears to fit well in the sense that the residual deviance is not statistically significant – i.e., it is consistent with chance deviations from the predictions of the model.

As long as we have Poisson data with reasonably large means, we can assess the residual deviance as if it had a chi-square distribution with a number of degrees of freedom (d.f.) determined by the model – specifically the number of units minus the number of parameters fitted. It is important to remember, however, that this is based on asymptotic theory which, roughly speaking, supposes that the total number of cases is large compared with the number of units – areal or otherwise – in the analysis. A rule of thumb suggests that the expectations of the counts in a Poisson regression should mostly be in excess of 5. When the average expectation falls below this, we should expect the distribution of the deviance in a correct model to depart progressively from a chi-square distribution, which of course means that

a corresponding statistical test of goodness of fit of the model based on the chi-square distribution would not be valid.

In this situation, we can obtain an approximate assessment of the value of the deviance – and hence the goodness of fit of the model – by simulation. Typically, we would generate, say, s new samples of data from Poisson distributions with means obtained from the model $\mathcal{M}_{\text{fitted}}$ fitted to the actual data. For each simulated sample, we would refit the same model and compute the residual deviance. The s values of the deviance thus obtained provide an estimate of the distribution of the deviance. This in turn provides a means of calibrating the deviance observed for our actual data. A formal test of goodness of fit would only be approximate since we are simulating from $\mathcal{M}_{\text{fitted}}$ rather than the true model with the true (unknown) parameter values. This situation is typical of “bootstrapping,” and the theory of this subject could in principle lead to better approximations. For an account of bootstrapping see, for example, Efron and Tibshirani (1993).

37.5.2 Detecting Heterogeneity in Poisson Data

A special case arises when we have expectations, provided, for example, by some prior analysis or by simple calculation from population data and we merely wish to detect whether the Poisson distribution fits well with the assumed e_i , without reference to any model fitting. This is sometimes seen as a problem of detecting “clustering,” though there are qualifications to this interpretation that we discuss below: for the moment, we prefer to regard this as the problem of assessing heterogeneity, i.e., variations in risk between areas without reference to a possible geographical origin for the phenomenon.

Relating this to the deviance of a Poisson model suggests that the deviance of the observations, defined in Eq. 37.1, Sect. 37.3.2, would be a sensible test statistic. The fact that this test is a likelihood ratio test means that it is *asymptotically* fully efficient – i.e., its power approaches that of the best possible test against a Poisson alternative in which the relative risks are different from unity.

Popular alternative contenders include Pearson’s chi-square statistic

$$X^2 = \sum (Y_i - e_i)^2 / e_i ,$$

and the Potthoff–Whittinghill statistic (Potthoff and Whittinghill 1966)

$$PW = \sum Y_i(Y_i - 1) / e_i ,$$

which is regarded by some authors as a test of clustering. The former is, at least in simple cases, asymptotically equivalent to the deviance but is easier to compute and to study analytically. The asymptotic requirement, however, implies that the

Table 37.2 Expected significance levels (ESL) % and their standard errors for Pearson's X^2 , the deviance, and the Potthoff–Whittinghill tests: k wards each with expectation e under H_0 and an alternative expectation dispersion with variance σ^2

e	σ^2	k		Pearson	Deviance	Potthoff
5.0	0.05	200	ESL	6.6	3.5	44.5
			s.e.	0.22	0.18	0.50
1.0	0.2	500	ESL	3.5	2.7	6.3
			s.e.	0.18	0.16	0.24
0.2	1.0	1,000	ESL	14.2	23.8	3.3
			s.e.	0.35	0.43	0.20

expectations should be large and the theoretical properties give rather little guidance on which test is best for small expectations.

Table 37.2 shows the results of a simulation study, designed to provide such guidance, in which the expected significance level (ESL) of each test has been estimated in each of three conditions. (The ESL is a convenient alternative criterion to power (Dempster and Schatzoff 1965): a smaller ESL corresponds to a more powerful test.) In each case, the ESLs were estimated from 10,000 simulations performed under varying conditions. These were chosen to produce values in a critical range corresponding to situations where the test would be quite likely to lead to different conclusions at conventional significance levels. In each case, a specific number k of wards were supposed to have the same expectations e under the null hypothesis, while under the alternative hypothesis, these expectations were multiplied by a set of RR factors θ_i sampled from a gamma distribution with mean one and variance σ^2 .

In interpreting this table, we suppose that the key parameter is the size of the expectation e . Because the test statistic will be roughly proportional to the number of wards k , this latter parameter represents the amount of information and was chosen to bring the ESLs into an interesting range; it would not be expected to change the relative ordering of the three tests. The variance σ^2 represents the distance between the null and alternative hypotheses, and the values were chosen to be typical of the sort of discrepancy that one could reasonably expect to detect in practical situations. It could conceivably affect the relative properties of the different tests but is less likely to do so than e .

It will be seen that, with an expectation of $e = 5$, the deviance is indeed the best test, while the Potthoff–Whittinghill test trails behind Pearson's chi-square test. The difference between X^2 and D becomes marginal around $e = 1$ while, for smaller expectations, the ordering is reversed and the Potthoff–Whittinghill test appears to be superior. These results suggest that it would be wise to carry out simulations in particular marginal cases to determine the best test to use. It should also be emphasized that one should evaluate the significance of the chosen statistic using simulation when the e_i are small, since the Pearson and deviance statistics are then likely to have distributions markedly different from the chi-square.

37.5.3 Spatial and Non-Spatial Analyses

A test of heterogeneity in areal data of the kind described above provides only a non-spatial test of the heterogeneity of our observations. Whether this is appropriate depends on whether or not the areal units are defined by essentially geographical criteria. If, for example, they are defined by simply dividing our region \mathcal{R} into urban and rural areas, then a factor associated with the degree of urbanization could be expected to induce heterogeneity into the areas irrespective of their spatial positions.

More frequently, however, areas are merely convenient administrative subdivisions of \mathcal{R} . In this case, we might expect a factor that raises the incidence in one area to do so in adjoining areas also. Then, a test that takes no account of the spatial relationship of the areas will be less powerful than one that does.

To take a simple hypothetical example, suppose that \mathcal{R} consists of two subregions: \mathcal{R}_1 with n areas each having expectation $e_i = 9$ and \mathcal{R}_2 with n areas each having expectation $e_i = 11$. A dispersion test based on Pearson's chi-square statistic would use the variance of the observations to test the null hypothesis H_0 that all the expectations are the same:

$$X_{2n-1}^2 = \sum_{i=1}^{2n} (Y_i - e)^2 / e,$$

where $e = \sum_1^{2n} Y_i / n$ is the (estimated) expected count based on all $2n$ observations. To a good approximation, this statistic would have a chi-square distribution with $2n - 1$ degrees of freedom under H_0 . If, however, we knew which areas belonged to \mathcal{R}_1 and which to \mathcal{R}_2 , we would base the test on the equivalent statistic for testing the difference between the totals for the two subregions:

$$X_1^2 = \frac{\left(\sum_1^n Y_i - ne\right)^2 + \left(\sum_{n+1}^{2n} Y_i - ne\right)^2}{ne},$$

and it is fairly obvious that this would be a much more powerful test of H_0 . This idealized situation is analogous to isolating sources of variation in an analysis of variance.

In practice, of course, we will almost certainly not be in a position to divide \mathcal{R} into high- and low-risk areas a priori, but this example does suggest that the detection of non-uniformity of risk should take account of the spatial structure of the data. A classical account of tests of spatial autocorrelation is given by Cliff and Ord (1981), who establish some theoretical properties of their sampling distributions, particularly in the case of normally distributed observations. In one of the few comparative studies published, Walter (1993) examines the power empirically for three of the most popular tests against a variety of geographically plausible alternatives. The three considered were as follows:

- The I statistic of Moran (1948), which is analogous to a correlation coefficient and is defined by

$$I = \frac{n \sum_{ij} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{ij} w_{ij} \sum_i (x_i - \bar{x})^2}$$

- The c statistic of Geary (1954), which is similar to I
- A non-parametric test statistic which uses only the ranks of the observations

The first two statistics used as observations $x_i = y_i/e_i$, the standardized incidence ratios for the different areas, and spatial weights w_{ij} chosen to be one if A_i and A_j are adjacent and zero otherwise. Walter's Table II shows that, in each of the situations he considered, Moran's I had the highest power of the three, and it would seem that this should be the method of choice, at least for detecting generalized spatial relationship as opposed to isolated peaks in the risk. The question of whether higher power could be achieved by using more sophisticated weighting than a simple adjacency matrix, or by weighting the pairs of observations according to the amount of information they contain (in terms of sample size, for example), has not been much considered. Walter concludes that "the precise type of spatial pattern involved may have a major impact on the spatial power of the analysis" and that "more experience is needed to better understand the potential of these methods, and their limitations." Nevertheless this study was a useful contribution, and the use of Moran's I to detect spatial autocorrelation is probably a good choice.

37.5.4 Heterogeneity Tests Based on the Risk Surface

If we have continuous data – i.e., observations at the individual level – we can base a test of uniformity on the RRF $\hat{\theta}(x, y)$ as estimated by the methods described in Sect. 37.3.4. We may regard a test statistic as being defined by a functional of $\hat{\theta}(x, y)$, and there are various possibilities.

A natural choice is the weighted variance of $\hat{\theta}(x, y)$:

$$T_{\text{var}} = \iint_{\mathcal{R}} \pi(x, y) \{\hat{\theta}(x, y) - 1\}^2 dx dy .$$

In the absence of any reliable theory, it is necessary to resort to Monte Carlo methods to test the statistic. For case-control data, we use a permutation method that is straightforward though laborious:

1. Construct a map of the risk function $\theta(x, y)$ by a suitable method, using a degree of smoothing which is determined as a function of the data.
2. Evaluate the chosen test statistic for the observed data t_{obs} .
3. Choose a new sample of "cases" by choosing at random n points from the set of $m + n$ cases and controls combined.
4. Compute the value of the statistic t_1 , say, for the simulated data, using the same procedure as in step 1.

5. Repeat steps 3 and 4 $s - 1$ more time so that there are s simulated values altogether.
6. Reject at level $\alpha = m/(s + 1)$ the null hypothesis of uniformity of cases if t_{obs} is greater than all but $m - 1$ of the simulated observations.
7. Alternatively estimate the p -value of the test as the number of $\{t_i \geq t_{\text{obs}}\}/s$.

This general Monte Carlo procedure is applicable in very general circumstances, and it is especially useful in the analysis of spatial data, where construction of suitable models is difficult. We must remember, however, that a hypothesis test is, by itself, of very little inferential value without some idea of how probable the observed results would be under a plausible alternative.

The method can easily be adapted to a test based on a risk surface constructed from areal data as described in Sect. 37.4. The simulation would take the form of sampling areal counts from Poisson distributions with expectations e_i and computing the variance over a square grid as before. In either the continuous or areal data case, the degree of smoothing used in the density estimation process determines the scale of aggregation for which the test is most sensitive and is analogous to the choice of weights w_{ij} in Moran's statistics.

The use of tests of this sort is still in its infancy, but the underlying philosophy is attractive and increasing computing power is making them more practicable even for large data sets.

37.6 Clustering

Closely related to the idea of heterogeneity is the concept of clustering, with which much of geographical epidemiology is preoccupied. There is a large literature on the subject, not all of which is very clear on the issue of what we actually mean by the words "cluster" and "clustering." We may conveniently define a cluster as a localized aggregation of disease cases greater than can easily be explained by chance. Clustering may be regarded as the tendency to form clusters or, more generally, as any departure from the assumptions of uniform risk and independence of case occurrences as discussed in Sect. 37.2.1. We will continue to use the word heterogeneity to refer to a departure from uniformity and reserve the word clustering as far as possible to refer to mechanisms in which case occurrences are not independent. This kind of clustering may be supposed to act locally, whereas heterogeneity is more likely to be observed throughout \mathcal{R} and is sometimes referred to as "generalized clustering." For further discussion of the issues the reader is referred to a useful paper by Diggle (2000).

We can give here only the briefest of accounts. We will distinguish between methods based on increased levels of risk and methods based on the proximity of neighbors. First, however, we make two general points about clustering.

In the first place, it is a well-accepted fact of spatial statistics that it is not possible to distinguish on the basis of a single realization of observed data from a spatial process whether any non-uniformity of the distribution of points (relative to an expected population distribution) is due to a variation of underlying risk, with cases

occurring independently (i.e., points generated by a non-homogeneous Poisson process or its equivalent), or to a mechanism in which existing cases induce others nearby, such as would happen in a contagious process. Secondly, we remark that, from an abstract point of view, clustering may take place in any continuum and, in the geographical context, we may observe clustering in space, time, or the “product-space” of time and geographical space. This mathematical commonality means that tests can be adapted from one problem to another, with very fruitful consequences.

37.6.1 Methods Based on the RRF

Clustering is likely to be observed as an increase in risk in some locality and it follows that we can use the estimated risk surface $\hat{\theta}(x, y)$ to provide an appropriate test. What functional of $\hat{\theta}(x, y)$ we use will depend on the alternative we have in mind or, equivalently, the pattern we would most like to detect. If, for example, we are content to demonstrate a single cluster or aggregation of cases we could choose as our test statistic, the maximum of the $\hat{\theta}(x, y)$ over the whole region \mathcal{R} :

$$T_{\max} = \max_{x, y \in \mathcal{R}} \{\hat{\theta}(x, y)\}.$$

This does not, of course, preclude the possibility that we would detect multiple clusters, but it is likely that our test would be most powerful in the situation where there are in fact very few. We could of course extend the statistic to consider, for example, the mean of the r largest peaks in $\hat{\theta}(x, y)$, but it is unlikely that we would have good a priori grounds for fixing r . Tests based on peaks of incidence must also be expected to be quite sensitive to the scale of the clustering phenomenon and to the degree of smoothing we employ in constructing $\hat{\theta}(x, y)$.

A statistic likely to have similar properties to T_{\max} is based on a scanning window, typically a square that moves over \mathcal{R} . At each point of a fine grid the observed number of cases is compared with its expectation; the test statistic is defined as the maximum discrepancy using a suitable criterion such as the incidence ratio. Here, the size of the window plays the role of a smoothing parameter; the main difference from T_{\max} is that a peak incidence is weighted according to its radial extent; it seems likely that it behaves in a similar manner to T_{\max} for suitably chosen smoothing parameters. Anderson and Titterton (1995) describe a version of this method that varies the window size to keep constant the expected number of cases under the null hypothesis. Much subsequent work describing similar tests has been published; see Tango (2010) for a recent summary. Some of these have considered windows of different shapes, notably elliptical, but the usefulness of these is limited, not least because we are unlikely to know a priori what shape a cluster might have.

In fact, the scanning window is a two-dimensional version of an approach originally used for detecting clustering in time; even this one-dimensional version is notoriously intractable analytically and simulations or other numerical methods would seem to be unavoidable.

37.6.2 Knox's Test

The use of what we may call pairing methods is historically older than the methods based on the risk surface discussed above; they have the attraction of being very simple to describe and understand.

The earliest such test is due to Knox (1964), who counted the number, Z , of pairs of children with leukaemia diagnosed within 60 days and 1 km of each other in Northumberland and Durham, two counties in the North East of England (see Table 37.3, taken from Knox (1964)). The study used local registration and hospital records, as well as death certificates to ascertain 185 children with an onset of leukaemia under 15 years of age between the years 1951 and 1960 inclusive. However, certain cases were excluded, and Table 37.3 refers just to children under the age of six, a restriction that needs to be borne in mind when interpreting the results; in fact, older children showed no effect.

The rationale for this test is explicitly related to the non-independence of the cases, namely, that a contagious mechanism passing a disease from one individual to another would be likely to lead to cases that are closer to one another in space and time than would be expected by chance. This in turn leads to the idea of considering pairs of cases.

Knox refers this statistic to its expectation calculated on the assumption that the spatial locations and times of occurrence of the disease are independent. This is given by

$$\mathbf{E}[Z] = \frac{N_T N_S}{\binom{n}{2}},$$

where N_T , N_S are the numbers of pairs of cases close in time and close in space, respectively, and the denominator is the total number of pairs out of the n cases.

In effect, this becomes a test of the independence of these two variables, and it uses their *marginal* distributions to determine the null distribution of Z . Knox conjectures that Z should follow a Poisson distribution approximately; this is shown to be true in certain circumstances in work reported by David and Barton (1966), who give a formula for the variance of Z . It is wise to calculate this or to use a Monte Carlo test in which the times of occurrence of the cases are randomly permuted relative to the space coordinates, and the statistic Z is recomputed a large number of times. For Knox's data, the value of $\mathbf{E}[Z]$ is 0.83, for which $Z = 5$ has a p -value of 0.0017 when tested as a Poisson observation. David

Table 37.3 Pairs of cases of childhood leukaemia classified according to their closeness in space and time (see text)

		Distance apart (km)		Total
		0–1	Over 1	
Time apart	0–59	5	147	152
(days)	60–3,651	20	4,388	4,408

and Barton report an early simulation experiment for Knox's data carried out by M.C. Pike; the latter finds $Z \geq 5$ in 4 out of 2,000 simulations. This leads to an estimated significance level of 0.002 which is very close to that based on the Poisson approximation.

The choice of the critical distance and time separation is of course crucial. It determines the scale of clustering likely to be detected, and it should ideally be fixed in advance for the formal validity of the testing procedure. In particular, it is certainly not formally valid to test at a large number of different critical distances and times and then select the most significant result without allowance for this selection. If we really have no idea of the time and distance scales that would be appropriate, we need to use a data-driven method of identifying the most promising values (see [Sect. 37.6.6](#)).

37.6.3 Other Space-Time Clustering Methods

An alternative test based on the proximity in space and time of pairs of cases is proposed by Jacquez (1996). This is based on the number out of the l nearest neighbors in space of a given case that are also among the l nearest neighbors in time. Like the Knox test, it can be adapted to provide a test of space-only clustering. Jacquez claimed superior power to that of the Knox test, though in practice, this is likely to depend on the alternative being considered. Here, the parameter l serves as a kind of scale parameter since it determines how far we look for association between cases.

Knox's very elegant idea permits us to dispense with the need to estimate the marginal distributions, though only under the assumption that space and time are in fact independently distributed in the population. This assumption applies of course to Jacquez' test also. It will clearly be violated by population drift, i.e., a change of population distribution with time. Kulldorf and Hjalmar (1999) examine the size of this effect and conclude that it can be "a considerable problem." They recommend that space-time clustering should be tested using the joint space-time distribution of the population size, but this is of course rather hard to obtain with good accuracy and resolution. It seems likely that the use of the interaction tests will remain popular.

37.6.4 Space-Only Clustering

Knox's idea of counting pairs has been very fruitful and has been adapted to a number of related situations, including the use of a sample of controls to provide a reference distribution when testing for space-only clustering (Pike and Smith 1974). The essential idea here is to regard the controls as being similar to the cases, except that they are considered to have occurred at different "pseudo-times," while the cases are considered to have occurred simultaneously. The statistic computed is

then the number of pairs of cases that are close in space, and it is not hard to see that this is formally equivalent to Z , with identical distributional statistical properties. Knox's test is not the only test that can be adapted to detecting space-only clustering using controls. Other possibilities are explored by Rogerson (2006) in a paper giving some analytical results but no power study.

37.6.5 Population Distance

One alternative to this adaptation of Knox's test for case-control data is a kind of dual approach due to Cuzick and Edwards (1990). This is based on the count of the number of individuals among the l nearest neighbors of each case that are also cases (as opposed to controls). The quantity l in the Cuzick–Edwards test serves as a determinant of the scale of clustering to be detected in this method. It is given in terms of the number of individuals likely to be within a region of contagion, rather than by a distance.

This may be seen as more relevant for some, though not for all, mechanisms of disease spread. Indeed, for any given pair, we can think of closeness in terms of distance or in terms of the number of other members of the population residing between the two members of the pair. The choice between these two metrics is crucial, though which is the more appropriate will presumably depend on the supposed etiology of the disease.

The idea of a population distance lies behind another method of testing, due to Besag and Newell (1991), who consider each case in turn and aggregate the areas around it that are necessary to include the r th nearest case. The expectation for the aggregate of these areas is then compared with r in the usual way. This can be regarded as a kind of inverse sampling, and again the number of cases considered, r , is a parameter that determines the scale of clustering to which the procedure is sensitive.

37.6.6 Choosing Scale Parameters

Every clustering phenomenon has an implied scale of the clustering effect and it is clearly desirable to have some idea of this before attempting to detect it. When we have no idea, the temptation to perform multiple testing arises and it is important to make allowance for this. A method for testing a range of distances and times in the Knox test is proposed by Abe (1973); effectively, this examines a multi-way table for association between space and time, making due allowance for the non-independence of the pairs. This statistic is sensitive to association over the whole range of distances and times rather than attempting to identify the most interesting scale. To identify the scale of maximal clustering effect, we can use a general data-driven procedure that can be constructed along the following lines:

1. Test the data at each of a number of critical space and time distance pairs.
2. Form a single test statistic, either using some aggregate over different values of the scale parameters or using some measure of the maximum degree of clustering; call this statistic t_{obs} .
3. Simulate further data sets under the null hypothesis: for Poisson data, this will probably involve sampling Poisson-distributed counts, while for case-control data, it may involve pooling all the cases and controls and randomly selecting a subset to serve as simulated “cases.”
4. Rank the simulated values of the statistic t_1, t_2, \dots, t_s and compare the ranked values with t_{obs} .

This Monte Carlo procedure is of general applicability and provides a way of getting round the problem of unknown scale. It does of course sacrifice power by comparison with a test that correctly focuses on the true degree of clustering, so that the more carefully alternative hypotheses can be framed a priori the better.

Faced with this wide variety of tests, it is difficult for the researcher to know which to use. Each new test published typically is claimed to be more powerful than previously existing tests, but there is a wide variety of alternatives to uniformity of risk that could be considered, and it is certain that no one test is uniformly most powerful against all alternatives. In principle, it is open to the researcher to examine competing tests to see which would be best for the data and the alternative hypothesis in question, but this can be an arduous exercise. This is an area where we badly need more insight into which tests are preferable.

37.7 Predefined Sources of Risk

One of the epidemiological questions most often asked in a geographical context is whether there appears to be an aggregation of cases around a putative source of risk \mathcal{S} such as an industrial plant. For example, there has been much interest in the UK, as in other countries, in the possibility of an elevated risk of childhood leukaemia around nuclear power stations. This results in part from the finding of an unusually large aggregation of cases near the nuclear reprocessing plant at Sellafield, which is situated on the coast of Cumbria in the North West of England. In fact, ordinary nuclear generating stations have little in common with the reprocessing plant and the experimental reactor at Sellafield, nor is there evidence of significant releases of radioactivity into the environment from generating stations. Nevertheless, public anxiety persists about the safety of the plants, partly perhaps because of the difficulty in comprehending the nature of nuclear power and partly because of sensational reporting in the news media. In fact, there is little evidence of a general increase in risk (Bithell et al. 1994), but it is highly desirable that the best statistical procedures are used to test the data that come under scrutiny. The public may not have a very sophisticated understanding of statistics, but it is obvious even to the uninitiated that some of the procedures used in the past have not been well-chosen from the point of view of maximizing the chance of detecting a real effect.

Aggregations around \mathcal{S} are sometimes referred to as “clusters,” but it is not generally supposed that the cases involved are related, only that the risk to individuals in the vicinity of \mathcal{S} is elevated. Analyses could therefore proceed using the methods described in Sect. 37.3.1, with the obvious qualification that geographical variables clearly represent spatial relationship to \mathcal{S} . In practice, this nearly always means using distance from \mathcal{S} or some function of it, so that the analysis is implicitly one-dimensional. Moreover, analyses are often required in situations where the number of cases is very small, and in this situation, the fitting of GLM’s tends to be unstable and to lead to parameter estimates with large standard errors and unknown distributional properties.

37.7.1 Tests for Concentration of Risk

In this situation, it is probably better to rely on a formal significance test and the issue then becomes that of selecting the most powerful test against a suitable hypothesis or range of hypotheses. The resulting analyses are likely not to be very powerful in any case, but choosing the most powerful test at least increases the chance that a significant result can be attributed to a genuine departure from the null hypothesis of uniform risk.

The method of early investigators of simply comparing the risk in the area around \mathcal{S} with a reference or “control” rate outside the area defines a test procedure that is in fact powerful only against an alternative hypothesis prescribing a uniform excess risk within the area, which drops to zero on the boundary. This is clearly implausible and critically dependent on the size of the area chosen; one inevitably concludes that a better test would be one designed for some systematic relationship between the risk and the distance from \mathcal{S} . We may reasonably suppose that this relationship is monotonic, but the rate of decay and the shape of the RRF (expressed now as a function of distance) will determine the power of the test.

An ingenious class of tests designed to be powerful against general monotonic alternatives was proposed by Stone (1988). His “MLR test” compares the ratio of the maximum of the likelihood under the null hypothesis of uniform risk against the likelihood of the observations maximized subject to the restriction that the risk is a non-increasing function of distance from \mathcal{S} , i.e.,

$$H_1 : \theta_1 \geq \theta_2 \geq \dots \geq \theta_k \quad (\geq 1), \quad (37.2)$$

where θ_i is the relative risk in the i th area in order of increasing distance from \mathcal{S} . Stone’s test has become very popular in the UK epidemiological literature, though it is known that it is never the most powerful test against a specific hypothesis, this being provided by a *linear risk score* (LRS) test of the form

$$T = \sum_j \ln(\theta(d_j)),$$

where d_j is the distance of the j th case from \mathcal{S} and $\theta(d)$ is the risk at a distance d from \mathcal{S} as specified by the alternative hypothesis (Bithell 1995).

Unfortunately, knowing the most powerful test against a specific alternative hypothesis does not greatly help if we do not know what that alternative is. However, it provides a benchmark against which we can judge other tests, and in particular, it enables us to determine the sensitivity of the power to variations in the alternative. It turns out that statistics of the form

$$T = \sum_j 1/\phi(d_j),$$

for monotonic functions $\phi(\cdot)$, define a class of *canonical* tests which can come close to optimal power in many circumstances. In particular,

$$\phi(d_j) = d_j \text{ and } \phi(d_j) = \sqrt{\text{rank}(d_j)}$$

behave well in areas with a reasonably uniform population distribution. However, the latter affects quite strongly which of the canonical tests actually is most powerful and it is wise to check the performances of the competing tests in each different study using simulation. In addition to their simplicity, the canonical tests have the great advantage that they are not dependent on any parameters in the RRF; the test based on the reciprocal of distance, for example, is most powerful against all alternatives for which the RRF is of the form $a \exp(b/d_j)$, for any parameters a (which governs the overall degree of risk) and b (which governs the rate of decay). The fact that risk is unbounded at zero is a small price to pay for this advantage, which means that there is no need to perform multiple tests with different values of a and b .

Because the LRS test statistics are sums, they should in principle have an approximately normal distribution, and it is easy to compute their moments. In small samples, this asymptotic normal approximation will not necessarily apply, and it is advisable to use simulation also to carry out the tests, i.e., to carry out Monte Carlo tests. In doing so, it is easy to see that the way the samples are drawn can be either to fix the total number of cases and use the multinomial distribution or to use unconstrained Poisson distributions to determine the counts in the areas A_i . Which of these two sampling schemes is used is very important and will typically affect the results quite substantially. The first method defines a conditional test which might be appropriate if the expectations e_i for the rates in the different areas are unreliable in absolute terms (though possibly still all right relatively); it is important to note though that, if the expectations are correct, a conditional test could reject the null hypothesis because of a deficit of cases near the boundary of the region rather than an excess near \mathcal{S} . The second, unconditional, test is appropriate if the rates are reliable and in this case the test statistic combines the evidence from the overall relative risk in the area with that from the spatial distribution. In this case,

the appropriate form of Stone's test should also include the last (bracketed) inequality in Eq. 37.2 above.

Many other tests have been proposed for testing the concentration of risk around a point source; these are sometimes referred to as "focused tests." Some of these are in the class of LRS tests, though this is not always recognized. Some have been designed to use polar coordinates (Lawson 1993) and so to test for a directional effect; unfortunately no equivalent of the canonical test appears to be available for this problem, and so the maximal direction of the effect is a nuisance parameter that has to be estimated from the data unless there is a clear a priori reason for choosing one specific direction. Such tests have been applied to rather few datasets in practice. Tango (2010) gives a recent review of the field and reports extensive power calculations; these confirm that the canonical LRS tests do reasonably well for non-directional alternatives corresponding to a smooth monotonic RRF.

37.7.2 Example: Childhood Leukemia Around UK Nuclear Installations

The tests described above were developed partly in conjunction with analyses of the distribution of childhood leukaemia around nuclear installations. An analysis of all major sites in England and Wales is described by Bithell et al. (1994) using the data on leukaemia and non-Hodgkin lymphoma described in Sect. 37.3.2. The sites were examined separately using the LRS test with the reciprocal of distance rank as the primary test, though Stone's MLR test was also used for comparison. As remarked above, the results were largely negative.

However, public interest in the possibility of a raised risk persists, and two subsequent updated unconditional analyses have been published (COMARE 2005, 2011). In the first of these, the analyses were carried out in the light of a large simulation study that identified which of a number of tests would be most powerful at each of the sites. Experience of these analyses suggests that the power does indeed depend on the population distribution, but it has been found that, for the majority of test sites studied, the most powerful test against the alternatives considered was the LRS test based on $1/\sqrt{\text{distance rank}}$.

Table 37.4 shows the average power averaged over 75 alternative hypotheses and the significance levels achieved by each of five tests for one of the datasets from the 1994 analysis. It will be noticed that the smallest p -value was the Poisson maximum (often known as "Pmax"); this is in effect the maximum value of the cumulative relative risk as we move out from \mathcal{S} . The most powerful test, on the other hand, gives a non-significant result. This analysis is a timely warning against judging a test by the significance level achieved in a real dataset. More details and discussion of this analysis are given in Bithell (2003).

In the later study of nuclear power plants in Britain (COMARE 2011), the analysis was restricted to children under the age of 5 years and distances closer to the installations. This reduced the numbers to the point where it was necessary

Table 37.4 Average power of five tests and significance levels achieved for the 80 wards within 25 km of Hinkley Point, in which there were 57 cases observed against 57.2 expected

Test	MLR	Pmax	1/rank	1/distance	$\sqrt{1/\text{rank}}$
Power	0.359	0.204	0.421	0.649	0.630
<i>p</i> -value	0.150	0.020	0.108	0.357	0.341

to combine all 13 plants, and there were then sufficient cases to perform a Poisson regression on 1/distance. The resulting estimate of the risk coefficient was positive, but not statistically significant.

37.7.3 Summary of Recommendations

In summary, this is an area of geographical epidemiology where some progress has been made in identifying efficient procedures, perhaps because the problem is essentially one dimensional. Because data sets are usually small, it is an especially important aim to use tests of maximum power, and this criterion seems to be sensitive to the population distribution as well as the precise alternative considered. As far as areal data are concerned, it is recommended that a study should be guided by the following considerations:

1. First and foremost, thought should be given to the patterns of risk that it is desired to detect; these can be expressed in terms of the RRF and may reasonably be supposed to be monotonic decreasing unless special circumstances prevail. The more specifically this can be linked to a biological hypothesis, the more convincing a positive result will be.
2. Next, a circular region of radius R around S should be chosen and the observed and expected numbers of cases for the areas A_i in R obtained. There is no great advantage for testing purposes in calculating the numbers within fixed distance bands from S . The magnitude of R is important since, if it is much greater than the distance of any conceivable risk, the analysis will inevitably lose power. As a guideline, it would seem sensible to choose the radius R so that the excess relative risk might reasonably be supposed to have declined to half its value at distance $R/2$.
3. A Poisson regression should be used only if the total number of observations is large enough to ensure convergence of the estimation procedure and to provide reliable estimates of the parameters. It is difficult to provide guidelines, but an analysis with fewer than 20 cases in R should be treated with caution. The alternative of a non-parametric test may then be preferable.
4. For a non-parametric test, the first choice to make is between the conditional and the unconditional versions. This will depend largely on the perceived reliability of the expectations and whether it is desired to detect an overall excess in the area as well as spatial pattern.
5. Among tests of either kind, the LRS canonical tests will be reasonably powerful against most monotonic hypotheses, and it is recommended that 1/distance or

- $1/\sqrt{\text{distance rank}}$ be used unless the population distribution is very unusual or unless a very non-standard RRF is suspected. In either case, it is recommended that a simulation study be undertaken to determine the most powerful test for the suspected alternatives.
6. The analysis should then proceed with the test identified as best, using simulation to perform a Monte Carlo test unless the expectations are quite large, in which case normal approximations can be used for assessing the significance.

37.8 Conclusions

In this chapter, we have attempted to give a simple but unifying overview of the statistical methods that underlie geographical epidemiology. We have been able to refer to only a small proportion of the very large number of methods that have been proposed for different aspects of the subject. For further reading, we refer to edited volumes by Elliott et al. (1992, 2000), Lawson et al. (1999), and to the *Encyclopedia of Biostatistics* edited by Armitage and Colton (1998), for example the review article by Bithell (1998).

It will be clear that the rational choice of method is not an easy matter. Although the classical theory of statistics provides a number of principles leading to optimal procedures, there are areas of geographical epidemiology where they do not apply. In the first place, they apply essentially to the frequentist paradigm: the increasingly popular Bayesian methods raise essentially new optimality issues that are not easy to resolve. Secondly, many optimality results are asymptotic: when observations are effectively widely distributed throughout two-dimensional space, asymptotic results are less likely to be applicable even in moderately large datasets. Thirdly, many methods are essentially non-parametric and the classical optimality theory applies less directly to these. Lastly, the theoretical results apply mostly to situations where there is a large degree of independence in the structure of the data; they are therefore less applicable to models for the contagious processes needed to model alternatives to the null hypotheses in studies on clustering.

It follows that evaluating the relative merits of different methods has in practice to proceed by largely empirical methods, making extensive use of simulation. This makes appraisal difficult because of the large number of parameters that can be varied in the simulation experiments. It is important that any general principles suggested by the underlying theory are used to direct the empirical investigations, as exemplified, for example, by the discussion of methods for predefined sources of risk in Sect. 37.7. We conclude that geographical epidemiology, despite its practical limitations, can in principle provide useful pointers to the etiology of disease but that the methodology would be much more convincing if we knew more about its behavior in various plausible situations.

References

- Abe O (1973) A note on the methodology of Knox's tests of "Time and Space Interaction." *Biometrics* 29:67–77
- Anderson NH, Titterton DM (1995) Some methods for investigating spatial clustering, with epidemiological applications. *J R Stat Soc Ser A* 160:87–105
- Antó JM, Sunyer J (1992) Soya bean as a risk factor for epidemic asthma. In: Elliott P, Cuzick J, English D, Stern R (eds) *Geographical environmental epidemiology: methods for small-area studies*. Oxford University Press for World Health Organization, Oxford, pp 323–341
- Armitage P, Colton T (eds) (1998) *Encyclopedia of biostatistics*. Wiley, Chichester
- Baris YI, Simonato L, Saracci R, Winkelmann R (1992) The epidemic of respiratory cancer associated with erionite fibres in the Cappadocian region of Turkey. In: Elliott P, Cuzick J, English D, Stern R (eds) *Geographical environmental epidemiology: methods for small-area studies*. Oxford University Press for World Health Organization, Oxford, pp 310–322
- Besag J, Newell J (1991) The detection of clusters in rare diseases. *J R Stat Soc Ser A* 154:143–155
- Bithell JF (1990) An application of density estimation to geographical epidemiology. *Stat Med* 9:691–701
- Bithell JF (1995) The choice of test for detecting raised disease risk near a point source. *Stat Med* 14:2309–2322
- Bithell JF (1998) Geographical analysis. In: Armitage P, Colton T (eds) *Encyclopedia of biostatistics*. Wiley, Chichester, pp 1701–1716
- Bithell JF (1999) Disease mapping using the relative risk function estimated from areal data. In: Lawson AB, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertollini R (eds) *Disease mapping and risk assessment for public health decision making*. Wiley, Chichester, pp 247–255
- Bithell JF (2003) Selecting a powerful test for detecting risk near point sources. *Bulletin of the international statistical institute 54th session*. The Institute, Berlin/Germany, pp 97–98
- Bithell JF, Dutton SJ, Draper GJ, Neary NM (1994) Distribution of childhood leukaemia and non-Hodgkin's lymphomas near nuclear installations in England and Wales. *Br Med J* 309(6953):501–505
- Bithell JF, Dutton SJ, Neary NM, Vincent TJ (1995) Use of regression methods for control of socio-economic confounding. *J Epidemiol Community Health* 49(Suppl 2):S15–S19
- Clayton D, Bernardinelli L (1992) Bayesian methods for mapping disease risk. In: Elliott P, Cuzick J, English D, Stern R (eds) *Geographical environmental epidemiology: methods for small-area studies*. Oxford University Press for World Health Organization, Oxford, pp 205–220
- Clayton D, Kaldor J (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671–682
- Cliff AD, Ord JK (1981) *Spatial processes: models and applications*. Pion, London
- COMARE (2005) Tenth report. The incidence of childhood cancer around nuclear installations in Great Britain. HMSO, London
- COMARE (2011) Fourteenth report. Further consideration of the incidence of childhood cancer around nuclear power plants in Great Britain. HMSO, London
- Cook DG, Pocock SJ (1983) Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. *Biometrics* 39:361–372
- Cox DR, Snell EJ (1984) *The analysis of binary data*, 2nd edn. Chapman and Hall, London
- Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations (with discussion). *J R Stat Soc Ser B* 52:73–104
- David FN, Barton DE (1966) Two space-time interaction tests for epidemics. *Br J Prev Soc Med* 20:44–48
- Davies TM, Hazelton M (2010) Adaptive kernel estimation of spatial relative risk. *Stat Med* 29:2423–2437
- Dempster AP, Schatzoff M (1965) Expected significance levels as a sensitivity index for test statistics. *J Am Stat Assoc* 60:420–436

- Diggle PJ (2000) Overview of methods for disease mapping and its relationship to cluster detection. In: Elliott P, Wakefield JC, Best NG, Briggs DJ (eds) *Spatial epidemiology: methods and applications*. Oxford University Press, Oxford, pp 87–103
- Diggle PJ, Elliott P (1995) Disease risk near point sources: statistical issues for analyses using individual or spatially aggregated data. *J Epidemiol Community Health* 49(Suppl 2):S20–S27
- Diggle PJ, Rowlingson BS (1994) A conditional approach to point process modelling of elevated risk. *J R Stat Soc Ser A* 157:433–440
- Efron B, Tibshirani R (1993) *An introduction to the bootstrap*. Chapman and Hall, New York
- Elliott P, Cuzick J, English D, Stern R (eds) (1992) *Geographical environmental epidemiology: methods for small-area studies*. Oxford University Press for World Health Organization, Oxford, pp 323–341
- Elliott P, Wakefield JC, Best NG, Briggs DJ (eds) (2000) *Spatial epidemiology: methods and applications*. Oxford University Press, Oxford, pp 87–103
- Geary RC (1954) The contiguity ratio and statistical mapping. *Inc Stat* 5:115–145
- Gelman A, Price PN (1999) All maps of parameter estimates are misleading. *Stat Med* 18:3221–3234
- Greenland S, Morgenstern H (1989) Ecological bias, confounding, and effect modification. *Int J Epidemiol* 18(1):269–274
- Jacquez GM (1996) A k nearest neighbour test for space-time interaction. *Stat Med* 15:1935–1949
- Kelsall JE, Diggle PJ (1998) Spatial variation in risk: a nonparametric binary regression approach. *Appl Stat* 47:559–573
- Knox EG (1964) The detection of space-time interactions. *Appl Stat* 13:25–29
- Kulldorf M, Hjalmar U (1999) The Knox method and other tests for space-time interaction. *Biometrics*: 55:544–552
- Lawson AB (1993) On the analysis of mortality events associated with a prespecified fixed point. *J R Stat Soc Ser A* 156:363–377
- Lawson AB, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertollini R (eds) (1999) *Disease mapping and risk assessment for public health decision making*. Wiley, Chichester
- McCullagh P, Nelder JA (1989) *Generalized linear models*, 2nd edn. Chapman and Hall, London
- Moran PAP (1948) The interpretation of statistical maps. *J R Stat Soc Ser* 10:243–251
- Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers D, Clements A (2008) *Spatial analysis in epidemiology*. Oxford University Press, Oxford
- Pike MC, Smith PG (1974) A note on a ‘close pairs’ test for space clustering. *Br J Prev Soc Med* 28:63–64
- Potthoff RF, Whittinghill M (1966) Testing for homogeneity II. The Poisson distribution. *Biometrika* 53:183–190
- Rogerson PA (2006) Statistical methods for the detection of spatial clustering in case-control data. *Stat Med* 25:811–823
- Scott DW (1992) *Multivariate density estimation*. Wiley, London
- Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman and Hall, London
- Snow J (1855) *On the mode of communication of cholera*, 2nd edn. Churchill, London
- Stone RA (1988) Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. *Stat Med* 7:649–660
- Tango T (2010) *Statistical methods for disease clustering*. Springer, New York
- Wakefield JC (2008) Ecological studies revisited. *Ann Rev Public Health* 29:75–90
- Wakefield JC, Best NG, Waller L (2000) Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best NG, Briggs DJ (eds) *Spatial epidemiology: methods and applications*. Oxford University Press, Oxford, pp 87–103
- Walter SD (1993) Assessing spatial patterns in disease rates. *Stat Med* 12:1885–1894