The Spatio-temporal Epidemiology of Bovine Spongiform Encephalopathy and Foot-and-Mouth Disease in Great Britain

A thesis presented
in partial fullfilment of the requirements
for the degree of Doctor of Philosophy
at Massey University

Mark Anthony Stevenson

Institute of Veterinary, Animal and Biomedical Sciences
Massey University
Palmerston North, New Zealand

2003
(Submitted April 2, 2003)

Institute of Veterinary, Animal and Biomedical Sciences
Massey University
Palmerston North, New Zealand
2003

Massey University
COLLEGE OF SCIENCES

EpiCentre
Institute of Veterinary, Animal and Biomedical Sciences
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 350 5270
Facsimile: 64 6 350 5716

CERTIFICATE OF REGULATORY COMPLIANCE

This is to certify that the research carried out in the Doctoral Thesis entitled

'The Spatio-temporal Epidemiology of Bovine Spongiform Encephalopathy and Foot-and-Mouth Disease in Great Britain' in the Institute of Veterinary, Animal and Biomedical Sciences at Massey University, New Zealand:

(a) is the original work of the candidate, except as indicated by appropriate attribution in the text and/or in the acknowledgements;

(b) that the text, excluding appendices/annexes, does not exceed 100 000 words;

(c) all the ethical requirements applicable to this study have been complied with as required by Massey University.

Candidate's Name: Mark Stevenson

Signature: M. Stevenson

Date: 2 April 2004

Supervisor's Name: Roger Morris

Signature: R S Morris

Date: 2 April 2004

Abstract

Great Britain suffered two of the most globally notable animal disease epidemics of recent decades — the bovine spongiform encephalopathy (BSE) epidemic which began in November 1986, and the foot-and-mouth disease (FMD) epidemic which lasted from February to September 2001. This thesis applies various analytical techniques to these two quite different epidemics: a rapidly spreading highly contagious disease for which urgent decisions are essential (FMD), and a feed-borne non-contagious disease with an exceptionally long incubation period (BSE). The BSE epidemic, in particular, presented major investigational challenges because its recent emergence meant that its epidemiological features were not yet fully clear.

The studies of BSE reported here showed that the control measures recommended as a consequence of the first epidemiological study of the new disease were remarkably effective. The July 1988 meat and bone meal ban resulted in a 60% reduction in BSE risk for cattle born in the first 12 months after its introduction. Descriptive spatial analyses, using kernel density and regression techniques showed a marked concentration of BSE risk in the south of Great Britain. Following the July 1988 meat and bone meal ban BSE risk shifted to the east of the country, an effect partly explained by cross contamination of cattle feed with high-protein concentrates destined for the pig and poultry industry.

Detailed investigation of the earliest BSE-exposed farm holdings identified the south of England as an area of excess exposure density. While interpretation of these findings is complicated by the fact that the disease must have been present for some years before it was first diagnosed, the evidence suggests initial amplification in the south provided risk material which progressively distributed the disease to the rest of the country.

In contrast to BSE, FMD presents different challenges, in that affected farms can be diagnosed rapidly, but it is difficult to accurately evaluate the relative importance of the different mechanisms of transmission, and hence determine how best to apply control efforts. Foci of FMD infection of matched size in the English counties of Cumbria and

Devon were compared to dissect out factors contributing to the two quite different epidemic patterns in these areas. This analysis showed evidence of strong spatio-temporal interaction of infection risk in Cumbria, due initially to cattle herds as the dominant influence, with a later growth in the role of sheep as a source of infection.

During the FMD epidemic a stochastic spatial simulation model was used extensively as an aid for decision making. After the epidemic was over the predictive accuracy of earlier real-time modelling was assessed for the whole of Britain and the most concentrated focus of disease in Cumbria. The model predicted the temporal epidemic curve closely at both levels, and predicted the national spatial pattern of infection with high specificities (over 99%) and useful sensitivities (37% to 71%). It was concluded that the model predicted FMD-infected locations within 0 to 14 days after simulation start date with sufficient accuracy to guide surveillance activities and to provide estimates of resources required for contingency planning. The spatial accuracy of predictions might be further improved through the use of a series of sub-regional models, better-capturing the characteristics of individual outbreak foci that typically emerge during extended large scale, multicentred epidemics.

The studies presented in this thesis demonstrate that the application of temporal, spatial and spatio-temporal analytical methods can enhance the understanding of the epidemiological features of diseases in animal populations. The value in applying these methods of analysis comes from the ability to identify high and low disease-risk time frames and locations, allowing more focussed allocation of investigative resources.

Acknowledgements

This work would not have been possible without the assistance provided by the following people:

Past and present colleagues from the EpiCentre — Colleen Blair, Greg Bolton, Todd Cochrane, Fiona Dickinson, Julie Dunlop, Kathy Goodwin, Jörg Henning, Cord Heuer, Ron Jackson, Dave Lawton, Daryl Lin, John Lockhart, Deb McCrae, Jo McKenzie, Solis Norton, Bryan O'Leary, Pranee Panichabhongse, Dirk Pfeiffer, Nigel Perkins, Daniel Russell, Robert Sanson, Carola Sauter-Louis, Duncan Schilling, Kevin Simmons, Gaolatlhe Thobokwe, and Simon Verschaffelt.

Geoff Jones, Steve Haslett, and Duncan Hederly from the Institute of Information Sciences and Technology, Massey University. Tony Gatrell from the Institute for Health Research, Lancaster University, and Paulo Ribeiro and Peter Diggle from the Department of Mathematics and Statistics, Lancaster University.

Judi Ryan, Linda Hoinville, Sarah Evans, Sophie Pascoe, Jane Archer, and staff at the Veterinary Laboratories Agency, Weybridge involved in maintaining the BSE database.

Special thanks to Andrew Lawson, now at the Department of Epidemiology and Biostatistics, University of South Carolina.

I am indebted to Roger Morris for his enthusiasm and vision, Peter Davies for his attention to detail, and John Wilesmith for his epidemiological insight. To identify meat and bone meal as the agent responsible for the transmission of BSE after 200 cases had been diagnosed ranks along side the work of John Snow. It has been a humbling experience to work with those directly responsible for this.

I thank the Department for Environment, Food and Rural Affairs (Defra) of the United Kingdom for funding this project.

Finally, my family and dear wife, Cathy — thanks for putting up with me during all of this.

Nomenclature

AR	Autoregressive (model)
BSE	Bovine spongiform encephalopathy
CAR	Conditional autoregressive (model)
CI	Confidence interval
CIGAR	Conditional intrinsic Gaussian autoregressive (model)
CPH	County-parish-holding (identifier)
FMD	Foot-and-mouth disease
ESDA	Exploratory spatial data analysis
GAM	Generalised additive model
GIS	Geographic Information System
HEPP	Heterogeneous Poisson process (model)
LISA	Local indicators of spatial association
MAFF	Ministry of Agriculture, Fisheries and Food
MPL	Maximum pseudolikelihood
DEFRA	Department for Environment, Food and Rural Affairs
SAR	Simultaneous autoregressive (model)
SERAD	Scottish Executive Rural Affairs Department
SMD	Standardised mortality difference
SMR	Standardised mortality ratio
SOAEFD	Scottish Office, Agriculture, Environment and Fisheries
TIN	Triangular irregular network

List of Publications

Wilesmith, J., Ryan, J., Stevenson, M., Morris, R., Pfeiffer, D., Lin, D., Jackson, R., & Sanson, R. (2000). Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: holding-associated risk factors for disease. *Veterinary Record*, 147, 319 – 325.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lockhart, J., Lin, D., & Jackson, R. (2000). Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: individual animal-associated risk factors for the disease. *Veterinary Record*, 147, 349 – 354.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lawson, A., Pfeiffer, D., & Lin, D. (2000). Descriptive spatial analysis of the bovine spongiform encephalopathy epidemic in Great Britain to June 1997. *Veterinary Record*, 147, 379 – 384.

Stevenson, M., Morris, R., Lawson A., Wilesmith, J., Ryan, J., & Jackson R. (2003). Area-level risks for BSE in British cattle before and after the July 1988 meat and bone meal feed ban. Submitted.

Morris R., Wilesmith J., Stern M., Sanson R., & Stevenson M. (2001). Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain. *Veterinary Record*, 149, 137 – 144.

Morris R., Sanson R., Stern M., Stevenson, M., & Wilesmith, J. (2002). Decision-support tools for foot and mouth disease control. *Revue Scientifique Et Technique De L'Office International Des Epizooties*, 21, 557 – 567.

Stevenson, M., Wilesmith, J., King, C., & Morris, R. (2002). Spatial, temporal and spatio-temporal epidemiology of foot-and-mouth disease in Cumbria, February to September 2001. In J. Rigby, C. Skelly, & P. Wigham (Eds.), *GeoHealth 2002 Proceedings of the Spatial Information Research Centre's 14th Colloquium* (p. 71 – 74). Wellington, New Zealand: Victoria University.

Wilesmith J., Stevenson M., King C., & Morris R. (2003). Spatio-temporal epidemi-

ology of foot-and-mouth disease in two areas of Great Britain in 2001. *Preventive Veterinary Medicine*, 61, 157 – 170.

Contents

List of Figures

List of Tables

'Winwood Reade is good upon the subject' said Holmes. 'He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty.'

Sir Arthur Conan Doyle
THE SIGN OF THE FOUR

CHAPTER 1

Introduction

Over the last decade there have been major enhancements to the methods available to examine patterns of disease in space and time and to relate those patterns to putative risk factors. In this thesis aspects of the epidemiology of bovine spongiform encephalopathy (BSE) and foot-and-mouth disease (FMD) in Great Britain are investigated. These are probably the two most complete epidemic data sets available in any species, including man. For the analyses of the BSE epidemic the case data set is comprised of all confirmed cases identified to 30 June 1997; for the analyses of the 2001 FMD epidemic the case data set is comprised of all farm holdings that were infected between February and September 2001. These analyses differ from the majority of earlier reports of these two epidemics in two ways: firstly, cases are considered in the context of the population of animals and/or farm holdings at risk during the relevant time periods, and secondly the spatial location of all study subjects (both cases and non-cases) have been defined. This thesis is presented as a series of papers, either published or prepared for publication. Each chapter shows the stage of preparation each paper has reached at the date of thesis submission.

With respect to BSE the motivation for this work has been to provide more detailed spatial and temporal descriptions of the evolution of the epidemic in Great Britain from 1986 to 1997. Chapters 3 and 4 provide (respectively) temporal analyses of the epidemic in terms of calendar date and in terms of the time of onset of disease in individual cases. Chapter 5 describes the spatial evolution of the epidemic in terms of farm holdings confirmed with the disease, and in terms of individual cases. Recognising that the risk of BSE was strongly related to birth cohort, Chapter 6 quantifies the influence of area-level risk factors for disease for cattle born before and after the July 1988 meat and bone meal feed ban. Chapter 7 seeks to more precisely describe the spatial and temporal features of the first documented holding-level exposures to the BSE agent.

In contrast to BSE, FMD is characterised by a short incubation period and rapid spread among susceptible farm holdings. During the course of FMD outbreaks it is essential that on-the-spot epidemiological analyses provide decision makers with evidence that supports field observations (and opinions) concerning factors that influence the spread of disease. Adopting an analytical approach in many ways similar to that used in Chapters 3 to 7, Chapter 8 describes the spatial and temporal epidemiology of FMD in the counties of Cumbria and Devon, firstly describing the epidemic in terms of calendar date then considering the spatial distribution of infected premises. Acknowledging that local 'contagiousness' is an important component of FMD spread Chapter 8 also describes the relative numbers of cases arising from spatio-temporal interaction,[1] showing how the relative importance of this aspect of infection risk changed throughout the course of the epidemic. Chapter 9 considers aspects of FMD control further, investigating the predictive accuracy of a stochastic spatial model of FMD that was used extensively throughout the 2001 British epidemic. In this chapter, the model's ability to correctly predict the time and place of infections during defined prediction windows is evaluated.

Drawing on the techniques reviewed in Chapter 2, the analyses presented in this thesis identify a collection of techniques suitable for describing and explaining the temporal, spatial and spatio-temporal aspects of animal health data. The value in applying these methods of analysis comes from the ability to identify high and low disease-risk time frames and locations, allowing more focussed allocation of investigative resources. In this way, although animal populations will continue to remain exposed to disease threats, the ability to derive more value from accumulated epidemic data means that authorities are better equipped to deal with them.

[1] Throughout this thesis the term *spatial and temporal* refers to two independent processes: one in space and one in time. The term *spatio-temporal* refers to the interaction between space and time.

CHAPTER 2

Literature Review

2.1 Introduction

An axiom of epidemiology is to describe and explain patterns of disease in terms of individual, place and time. Although well-proven methods exist for describing and explaining the occurrence of disease in relation to individual and time, techniques to describe and explain disease patterns in terms of space have been, until relatively recently, limited. In 1994, Bailey reviewed the state of spatial analytical methods and concluded that whereas Geographic Information Systems (GIS) were powerful tools for displaying and querying spatial information, their capabilities in terms of describing and explaining spatial patterns were lacking. Two reasons were cited for this imbalance: (1) an absence of data available in a format suitable for analysis, and (2) an absence of robust analytical techniques within commercially available GIS and statistical packages (Bailey 1994). Eight years on from Bailey's review, advances in hardware and software technologies have made GIS packages and spatial data more accessible and less expensive. The consequent widespread use of GIS has underpinned a burgeoning industry associated with the recording and management of spatial data. Concomitantly, advances in spatial statistical methods, and their incorporation within off-the-shelf statistics packages, have extended and facilitated the potential for analysing this information. With these critical elements in place it is anticipated that spatial analyses will proceed to play an increasingly prominent role in exploring the epidemiological triad of individual, place and time.

This chapter will review techniques used to visualise, describe and explain spatial data. Acknowledging that many of the statistical techniques for this purpose have been developed for use in other disciplines (e.g. mining, hydrology, and ecology) emphasis will be given to those that are of greatest use in a (veterinary) epidemiological setting. In

addition to the references cited, the reader is referred to the comprehensive introductions to this subject provided by Bailey & Gatrell (1995), Wakefield & Elliott (1999), Elliott et al. (2000), Lawson & Williams (2001), and Leyland (2001).

Three data sets are used to illustrate the techniques described in this review. The first relates to cases of laryngeal cancer diagnosed in Lancaster from 1974 to 1983 (Diggle et al. 1990) and is used to illustrate techniques applicable to point data. The second data set of lip cancer cases diagnosed in Scotland from 1975 to 1980 (International Agency for Research on Cancer 1985, Clayton & Kaldor 1987, Clayton & Benardinelli 1992, Breslow & Clayton 1993, Bell & Broemeling 2000) is used to illustrate techniques applicable to lattice data. The third data set is of piezometric-head measurements recorded at selected sites within the Wolfcamp aquifer in the Palo Duro Basin, Texas USA (Harper & Furr 1986, Ribeiro Jr & Diggle 2001). Piezometric-head measurements are used to determine the direction of flow of groundwater and are obtained by drilling narrow pipes into an aquifer and measuring how far groundwater rises within each pipe. This data set will be used to illustrate techniques applicable to spatially continuous data. Throughout this review the techniques described and discussed will be applied to these data sets, where appropriate.

2.2 Spatial data types

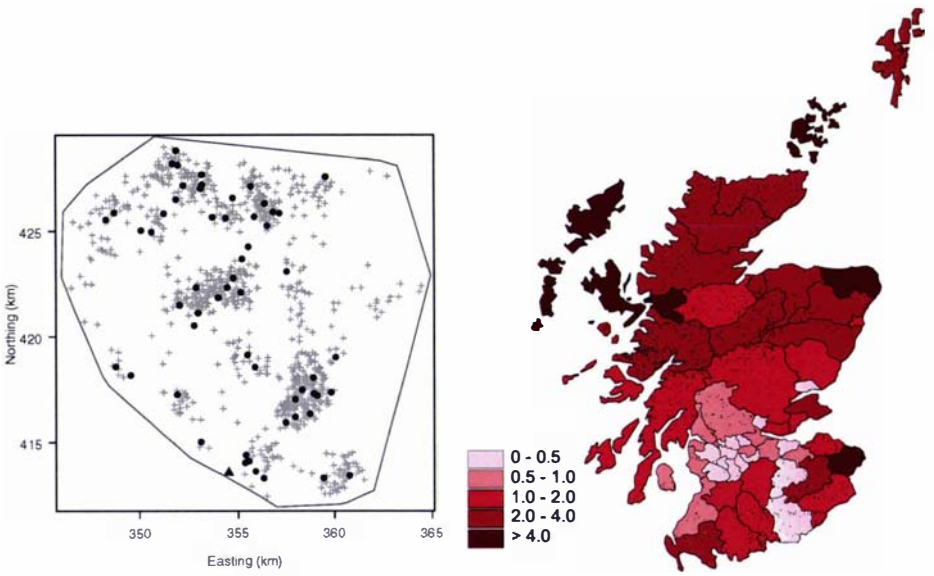
In addition to details recorded for each subject in a data set (attribute information), a distinguishing feature of spatial data is the inclusion of details defining location of the subject in space. Locational information can: (1) be point or areally referenced, (2) occupy a regular or irregular distribution across space, and (3) be continuous or discrete. Attribute data for each study subject can be measured on a discrete scale (e.g. presence or absence of disease, counts of disease events at each location) or continuous, meaning that a value for the attribute exists at all possible locations throughout the study area and the data set represents a 'sample' for predicting attribute values at other, non-sampled locations (e.g. soil mineral concentration, concentration of air pollution) (Table 2.1, Figure 2.1).

With these data types defined, three classes of spatial process have formed the framework around which many analytical techniques have been developed: (1) point, (2) lattice, and (3) continuous. Although this classification serves as a useful framework for

Table 2.1: Classification of data types commonly used in spatial epidemiological analyses.

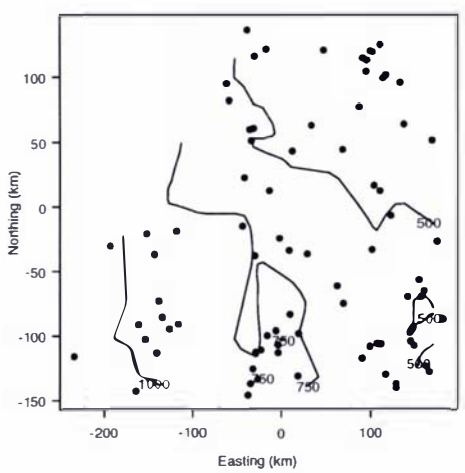
Data type	Location	Attributes
Point	Locations themselves are the variables of interest and a data set consists of a finite number of locations within a defined study region.	Recorded at each location. Point locations with associated attribute values are known as marked point processes.
Lattice (area or polygon)	Discrete areas regular or irregular in shape. Areas may be spatially referenced by an adjacency matrix defining each area's neighbours.	Counts of events within each area, or the value of some discrete or continuous variable defined for the area as a unit.
Continuous	All possible locations throughout a study area.	Vary continuously across space. A finite number of locations are sampled to estimate the value of the attribute at any position within the study area (for example, rainfall measurements recorded at monitoring stations).

categorising methodologies it is too simplistic to accommodate all applications. Rather, the different types of spatial processes should be viewed as building blocks for a range of possible hybrid models (Diggle & Gatrell 2001). For example, in some circumstances the distribution of point events is the sole focus of interest whereas in others interest might lie in the distribution of point events (for example cases of disease) in relation to a variable that is continuous throughout space (air pollution or soil contamination).



(a) Point data

(b) Lattice data



(c) Continuous spatial data

Figure 2.1: Data types used in spatial epidemiological analyses: (a) point data: easting and northing coordinates of the residence of laryngeal cancer cases (•) and lung cancer cases (+) diagnosed within an area of 400 km^2 (20 km $\times$ 20 km) in Lancashire, England (Diggle et al. 1990), (b) lattice data: standardised mortality ratios for lip cancer in each of 56 Scottish districts plotted as a choropleth map (International Agency for Research on Cancer 1985), (c) continuous spatial data: piezometric-head measurements (measured in metres above sea level) recorded at selected locations (•) within the Wolfcamp aquifer in the Palo Duro Basin, Texas USA plotted over an area of 62,500 km^2 (250 km $\times$ 250 km) (Harper & Furr 1986, Ribeiro Jr & Diggle 2001).

In addition to the above comments it should be noted that most spatial phenomena will vary in appearance depending on the scale at which they are viewed. At a low level of resolution the location of a farm may be represented by a single point. At a higher level of resolution individual farms may be better-represented as lattice data where the exact boundaries of the farm area are explicitly defined. Depending on the objectives of specific analyses and the level of resolution required by end users, different families of analytical techniques may need to be applied to the same data set.

A structured approach to the analysis of the spatial patterns of disease has been proposed by Bailey & Gatrell (1995). Firstly, one can visualise the data of interest in order to identify patterns and generate plausible hypotheses about the process under investigation. Secondly, the data can be explored to numerically describe it, again helping to generate plausible hypotheses. Thirdly, the data can be statistically modelled to identify explanatory variables which explain or predict the spatial distribution of the process. The sequential approach (visualise, explore, and model) is useful as a starting point, but in many circumstances there will be a close interplay of the three. Data tends to be visualised initially, aspects of the data are explored, which in turn leads to modelling being undertaken. The results of the spatial model may then be visualised, explored and then further refined. The three stages should therefore be seen not as distinct, separate entities but rather interlinked in an iterative fashion.

2.3 Point data

2.3.1 Visualising point data

Perhaps the oldest and most frequently used method to visualise point data is to plot the locations of study subjects using their Cartesian coordinates (Figure 2.1a). John Snow's account of the Golden Square cholera epidemic in 1854 bears testimony to the usefulness of this technique: high numbers of cases of cholera around a public well provided powerful support to the hypothesis that the disease was transmitted via contaminated drinking water (Snow 1855, Vinten-Johansen et al. 2003). Whereas point maps provide a general impression of the spatial distribution of disease event data they present problems where there are large numbers of events or where there are multiple events at the same location since no indication of event density can be appreciated. Because of this, point maps are

best suited for displaying location information for small numbers of events, for example to show the spatial distribution of rare diseases.

2.3.2 Exploration of point data

In any epidemiological analysis exploration of data is a critical step that helps to identify patterns, identify unusual or interesting features (including errors) and for developing hypotheses. Exploratory spatial data analysis (ESDA) is the extension of exploratory data analysis to spatial applications (Haining et al. 1998), allowing one to visualise patterns in the data and distinguish between broad spatial trends (first order properties) and the tendency for similar values to be found in close spatial proximity (second order properties). Computationally, ESDA techniques can involve a process known as 'brushing' in which locations on a map are interactively selected and graphical or statistical results for this subset of locations are displayed (Haslett et al. 1991). A variety of software packages with this functionality have been developed including SAGE (Haining et al. 1998) and the DynESDA extension to ArcView for Windows (Anselin & Smirnov 1998). The XGobi package (Swayne et al. 1998) allows brushing to be carried out on non-spatial data and has been adapted for spatial applications (Symanzik et al. 1998). Developing the ESDA technique further, Carr et al. (2000) describe the use of conditioned choropleth maps for demonstrating relationships between a dependent variable (represented in a classed choropleth map) and explanatory variables. These initiatives have provided tools that facilitate the identification of relationships and hypothesis generation.

Kernel smoothing point data

Kernel smoothing techniques provide a means for representing the density of plotted point locations (Rosenblatt 1956, Whittle 1958, Parzen 1962, Silverman 1986, Bowman & Azzalini 1997) overcoming some of the limitations of plotting disease events as point maps. The technique involves the placement of a regular grid over a region of interest and the construction of an area (known as a kernel) around each event $(y_1, y_2, \ldots, y_n)$ to define a distance-decaying density estimate. The kernel density estimate for each grid cell equals the sum of the density estimates that fall within the cell. Formally, the intensity of a spatial

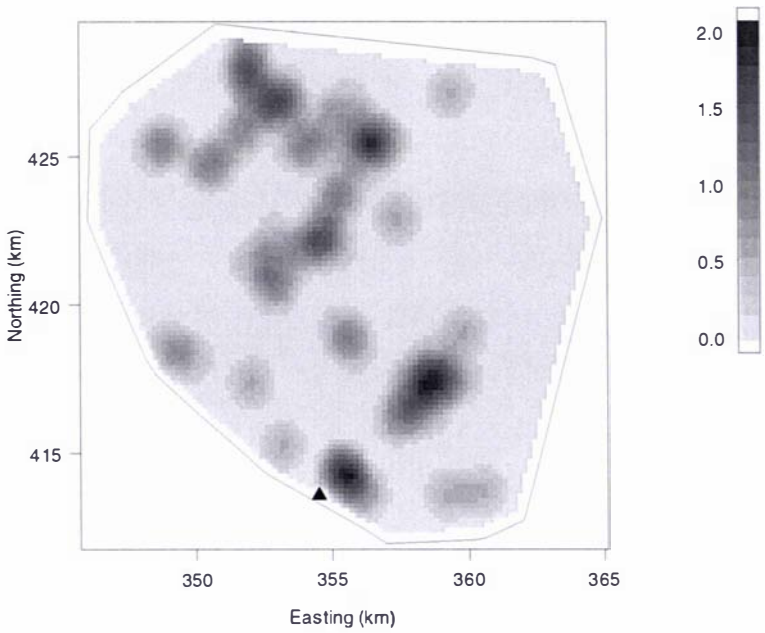


Figure 2.2: Image plot of kernel density estimates for the laryngeal cancer data set, showing the density of laryngeal cancer cases per unit area. A quartic kernel function has been used with a bandwidth of 0.9 kilometres, chosen by cross validation. An industrial incinerator in the south of the study area is marked (▲).

point pattern at point y, $f(y_h)$, is given by:

$$f(y_h) = \frac{1}{nh^2} \sum_{i=1}^{n} k\left(\frac{(y - y_i)}{h}\right) \tag{2.1}$$

In Equation 2.1 the term h defines the radius of the kernel (known as the bandwidth or smoothing parameter) and the term k defines the shape of the kernel structure to be used. Various types of kernel structure may be used including Gaussian, triangular, parzen and quartic with the Gaussian being the most commonly used (Kelsall & Diggle 1995a, Bowman & Azzalini 1997). A kernel density surface for the case events shown in Figure 2.1a is shown in Figure 2.2. In this example a quartic kernel function has been used with a bandwidth of 0.9 kilometres.

In a veterinary setting, consideration should be given to the nature of the density surface to be estimated. With marked point processes where attribute variables are recorded at each location (for example, a count of the number of stock present on a farm holding) the density of the attribute value is estimated by constructing a non-parametric kernel *regression* surface where the Cartesian coordinates provide the covariates of the regression and the attribute values (in this case, stock counts) provide the vector of responses. The resulting surface provides an estimate of stock density expressed as numbers of stock

per unit area. Where no attribute values are provided (for example, where interest lies in determining the density of farm holdings in an area) a kernel *density* surface may be constructed using the Cartesian coordinates of each farm holding.

Kernel density surfaces provide simple and aesthetically pleasing raster images from which other data sets can be derived. Contour lines delineating areas of high or low density can be used on their own for display or used to define the boundaries of 'hot spot' areas that can be analysed in their own right. An additional strength is that the technique lends itself well to temporal analyses where raster images of density can be used as inputs into correlation or time series analyses (Koussoulakou & Kraak 1992).

Effective use of kernel density estimation requires the selection of an appropriately-sized bandwidth. When the selected bandwidth is too small the resulting density or regression estimate will resemble the original point pattern and will contain spurious features that are artefacts of the sampling process, as shown in Figure 2.3a. When the bandwidth is too large the data will be over-smoothed and important features of the underlying spatial structure obscured, as shown in Figure 2.3b. Several methods are available to objectively determine the most appropriate bandwidth for a given data set including: (1) optimal smoothing, (2) normal optimal smoothing (Hogg 1979), and (3) cross-validation (Rudemo 1982, Bowman 1984). In some instances the optimal bandwidth determined by numerical algorithms may prove unsuitable, in which case the bandwidth should be selected subjectively. Stoyan & Stoyan (1994) provide guidelines in this situation, concluding that a reasonable approach is to assess the sensitivity of conclusions to bandwidths of different sizes. Where there is marked variability in the intensity of events across a study region, a fixed bandwidth for the entire study region may be inappropriate because an interval appropriate for sparse areas will result in obscured details for dense areas. In this case the bandwidth may be adjusted locally using adaptive kernel estimation in which h is replaced by $w(h)$ where w represents a weighting factor that depends on the number of events in the neighbourhood of each point y_i (Silverman 1986, Gatrell et al. 1996).

Any technique for statistical estimation will have a measure of variability (or reliability) associated with it. Variability plots of kernel density surfaces identify areas where a sparsity of data makes density estimates less reliable. A variability plot of the kernel density surface computed for the Lancashire laryngeal cancer data set shown in Figure 2.2a is shown in Figure 2.4. Viel et al. (1995) provide an example of kernel density estimation

and its interpretation in conjunction with a variability plot from a study of leukaemia incidence in under 25 year-olds living around a nuclear waste reprocessing plant in France. In this study these authors noted an area of relatively high spatial variability of estimated density adjacent to the area of peak leukaemia density, concluding that this reflected either true variability or a paucity of data.

In an epidemiological setting smoothed plots of the intensity of counts of cases may not be appropriate to support decision making because of variability in the spatial distribution of the underlying population at risk. Large numbers of cases present at a given location may either represent a true elevation in incidence or reflect the underlying population structure with large numbers of cases observed in areas where population counts are highest (Kelsall & Diggle 1995a, Kelsall & Diggle 1995b, Lawson 2001b). To account for variation in the spatial distribution of the population at risk two kernel density surfaces can be constructed: the first representing the density of cases $f(y)$, and the second representing the density of non-cases $g(y)$. A relative risk function $\rho(y)$ can then be calculated as the ratio of $f(y)$ to $g(y)$ to produce, what has been termed by Lawson & Williams (1993) an extraction map. Bithell (1990) proposed the transformation:

$$\frac{\rho(y)}{1+\rho(y)} = \frac{f(y)}{f(y)+g(y)} \tag{2.2}$$

which represents the conditional probability of a case appearing in a small region near y given that a randomly selected case or non-case is to appear there. Alternatives to the above specification may be more appropriate in some applications: for example, a density difference surface may be plotted as $f(y) - g(y)$, which has an expected value of zero rather than one (as in the case of the density ratio measure defined in Equation 2.2).

With extraction mapping the issue of selecting an appropriately-sized bandwidth has additional significance since it does not necessarily follow that optimal bandwidth estimates for the cases and non-cases leads to an optimal degree of smoothing of the ratio. Small changes in the denominator (the density of the population at risk) in areas where its value is small can produce unacceptable variations in the ratio. Bailey & Gatrell (1995) recommend that it is preferable to over smooth the kernel estimate of the background intensity by adopting a larger bandwidth than would normally be appropriate if interest lay in estimating the density of the population at risk alone. Bowman & Azzalini (1997) provide conflicting advice, recommending the use of a common smoothing parameter for

both the case and non-case surfaces. In addition to these issues, care must be taken to consider the effects of study region edges on the interpretation of the ratio. Some edge-effect compensation may be necessary if there is considerable influence of study region edges in the final interpretation of the map (Lawson, Biggeri & Dreassi 1999).

Where location and covariate data for the entire population at risk are unavailable, a suitable proxy may be obtained by using a case event map of a disease which represents the background population but is not affected by the aetiological process of interest (Lloyd 1982, Diggle et al. 1990, Diggle & Rowlingson 1994, Viel et al. 1995). An extraction surface computed using the location of laryngeal cancer diagnoses as cases and lung cancer diagnoses as controls $\hat{\rho}(y)$ (Figure 2.5) shows a relatively high density of laryngeal cancer in the area surrounding an industrial incinerator, providing informal support to a hypothesis that proximity to the industrial incinerator increased the risk of upper airway neoplasia.

Whereas the use of control populations provides a practical solution to a difficult (and frequently encountered) problem in epidemiology, care should be taken to avoid confounding. In the example cited here, cases of cardiovascular disease would be unsuitable controls because both respiratory cancer and cardiovascular disease are associated with smoking. In addition, the estimated extraction surface will have poor precision in areas where there are few cases or controls (that is, sparsely populated areas of the study region). To avoid overinterpretation of features of the $\hat{\rho}(y)$ it is important to estimate its precision — Kelsall & Diggle (1995a) and Kelsall & Diggle (1995b) provide theoretical expressions for the mean square error of $\hat{\rho}(y)$ and propose a Monte Carlo method for assessing non-constancy of $\hat{\rho}(y)$ across an area of study (Diggle 2000).

Compared with the considerable number of studies discussing the technical issues associated with kernel smoothing, applications of the technique in a practical setting are few. Zeman (1997) used extraction mapping to describe the spatial distribution of tick-borne encephalitis and Lyme borreliosis in the Central Bohemian region of the Czech Republic, identifying differences between the spatial distributions of the diseases. At the time of writing no published examples of kernel smoothing applied to disease mapping have been found in the veterinary literature.

Spatial filters

A technique similar in concept to kernel density and regression estimation is that of spatial filtering and ratio smoothing (Kafadar 1996, Rushton & Lolonis 1996). In its simplest form spatial filtering involves applying a grid over a study region, and an estimated disease rate at a given grid point is defined as the observed rate of disease for the population that lies within fixed distances. Distances are large enough so that neighbouring grid points share observations. Once rates are defined, contour lines are drawn to create isorithmic maps in which areas with a constant range of values can be identified. Kafadar (1996) proposed a general class of smoothing functions where observations closer to the centre of each grid point were weighted more heavily than those further away. Talbot et al. (2000) developed the technique further by allowing the size of the spatial filter to be defined in terms of constant population size (rather than constant geographical size) analogous to the use of adaptive bandwidths in the density estimation literature. Rushton & Lolonis (1996) used spatial filtering to identify areas of elevated birth defects in Des Moines, Iowa. Talbot et al. (2000) used the population-size modification of the spatial filtering technique to describe the spatial distribution of low birth weights in New York state from 1994 to 1995.

The K-function

The K-function or reduced second moment function (Bartlett 1964, Ripley 1976, Ripley 1977) has been used extensively in the ecological literature to describe the second order characteristics of spatial point patterns (Cliff & Ord 1981, Ripley 1981, Diggle 1983, Cressie 1993). For an isotropic process with an intensity of λ points per unit area, the K-function at distance s may be defined as $K(s)$ such that $\lambda K(s)$ equals the expected number of events with a distance s of an arbitrarily-chosen event. For clustered patterns, each event is likely to be surrounded by further members of the same cluster and, for small values of s, $K(s)$ will be relatively large. Conversely, if events are regularly spaced, each event is likely to be surrounded by empty space and, for small values of s, $K(s)$ will be small. The validity of the K-function as a descriptor of second order properties of a point process depends on the assumption of an underlying isotropic process and is problematic to use in the presence of first order effects. If there is variation in intensity throughout the study region, then the first order effect may unduly influence the computed K-function. In

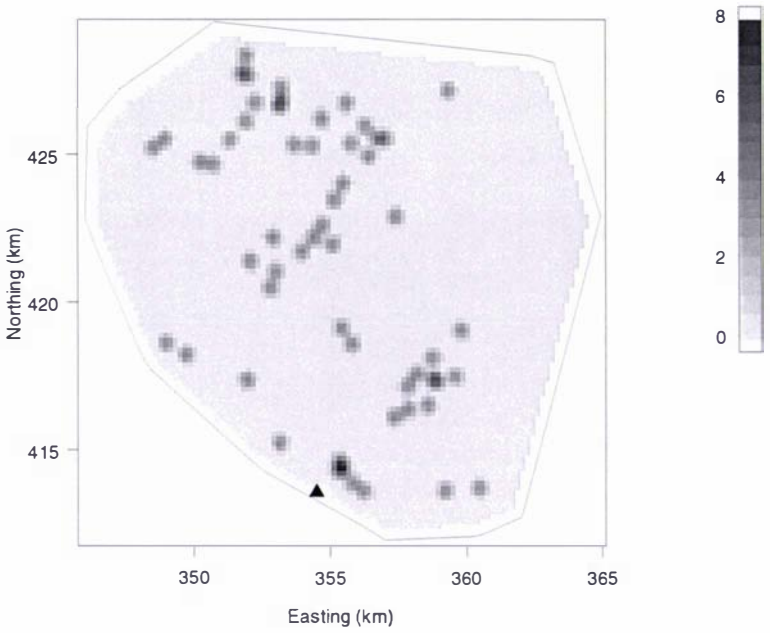
this case analysis of selected sub-regions where isotropy is expected has been suggested as a reasonable approach (Bailey & Gatrell 1995, Diggle & Gatrell 2001). The K-function computed for the cases events shown in Figure 2.1a is shown in Figure 2.6.

As with smoothing techniques, K-functions computed for case data alone may not be informative because of a heterogenous spatial distribution of the underlying population at risk. To account for this, separate K-functions may be computed for case events and compared with the K-function computed for either non-case events or a set of controls. The observed difference function, $D(s) = K_{case}(s) - K_{control}(s)$ may be interpreted as a measure of the extra aggregation of case events over and above that observed for the controls. For hypothesis testing, the locations of cases can be randomly permuted and values of the observed difference function computed for each permutation (Chetwynd & Diggle 1998, Chetwynd et al. 2000). The upper and lower bounds of these simulations can be plotted and any deviation of the observed difference function $D(s)$ above the limits of this envelope would indicate significant aggregation of cases at given scales of distance, relative to the controls (Figure 2.7).

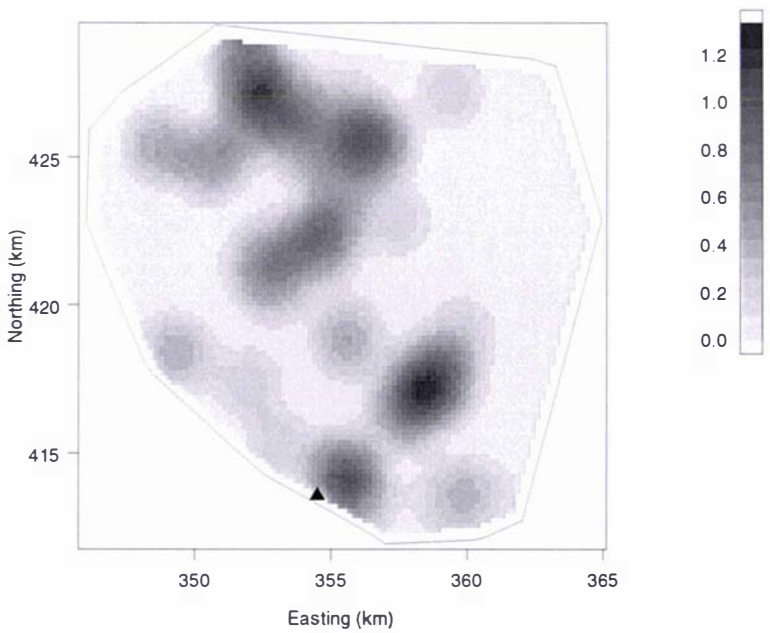
It should be noted that the variance of $K(s)$ increases with increasing values of distance, limiting the ranges of s over which $K(s)$ can be evaluated. Diggle & Gatrell (2001) recommend to restrict the range of s to no greater than 50% of the length of the shorter side of a rectangular study area. In addition, the sampling distribution of $K(s)$ is unknown if the number of events is large and an appreciation of the variability of $K(s)$ can be gained by dividing the study region into two or more sub-regions and calculating $K(s)$ for each sub-region. Silman et al. (1997) provide an example of this approach in a study of spatial and temporal clustering of rheumatoid arthritis in East Anglia, England. These authors found evidence of non-random distribution of arthritis in two of the six sub-regions that were examined, concluding that focused investigations in these areas would be of value.

Examples of the use of the K-function in the veterinary literature are few. In a retrospective study of canine cancer diagnoses made at a university teaching hospital from 1964 to 1994, O'Brien et al. (1999) compared the spatial distribution of the household of residence of dogs diagnosed with four types of neoplasia with the distribution that would be expected under complete spatial randomness. The tendency for canine neoplasia to cluster varied among the three counties examined, though the authors acknowledge that referral bias, an uneven dog population density, and variations in the age structure of the

dog populations investigated made interpretation of the study findings difficult. In a further study O'Brien et al. (2000) compared the distribution of canine cancer cases with comparable human cancers over the same time frame with a view to evaluating the use of dogs as a sentinel species for humans. These authors concluded that although the processes that determined the aggregation of cases in dogs and humans were not independent of each other they did not effect the two species equally. Abernethy et al. (2000) used the K-function and random simulation envelopes to confirm the presence of spatial clustering of Newcastle Disease affected poultry flocks in Northern Ireland during an outbreak that occurred in the province during 1996 and 1997. Foley et al. (2001), in a survey of 1082 clinically normal dogs in California, used the K-function to characterise the second order spatial characteristics of seropositivity to the causative agent of granulocytic ehrlichiosis.



(a) $h = 0.45$ kilometres



(b) $h = 1.35$ kilometres

Figure 2.3: Image plots of kernel density estimates for the laryngeal cancer data set, showing the density of laryngeal cancer cases per unit area: (a) kernel density estimates computed using a bandwidth of 0.45 kilometres, 0.50 times that used in Figure 2.2, (b) kernel density estimates computed using a bandwidth of 1.8 kilometres, 2.0 times that used in Figure 2.2. An industrial incinerator in the south of the study area is marked (▲).

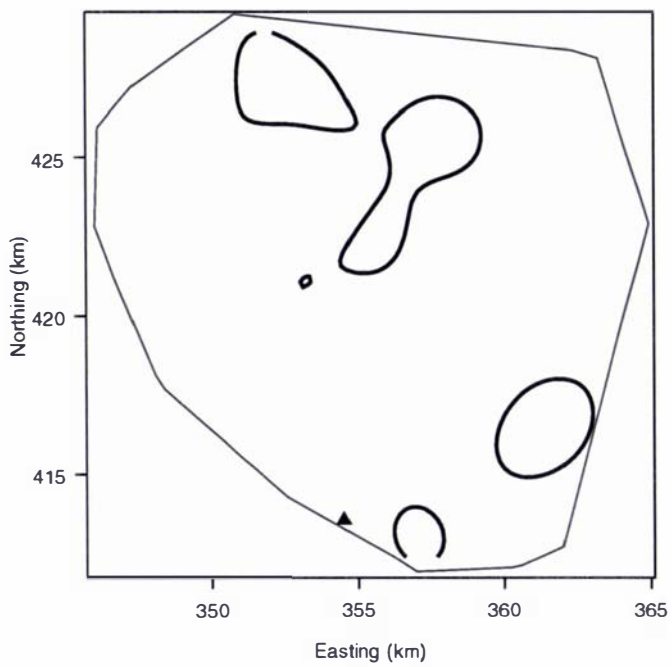


Figure 2.4: Contour plot showing the upper 90th percentile of the variability of the kernel density estimates for laryngeal cancer cases. An industrial incinerator in the south of the study area is marked (▲).

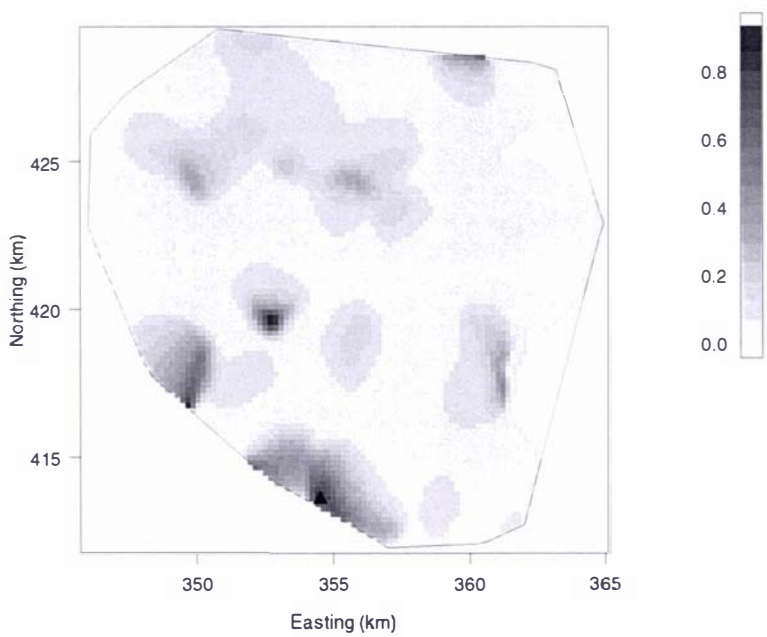


Figure 2.5: Image plot showing the estimated spatial variation in the density of laryngeal cancer risk constructed using the location of lung cancer diagnoses as controls. An industrial incinerator in the south of the study area is marked (▲).

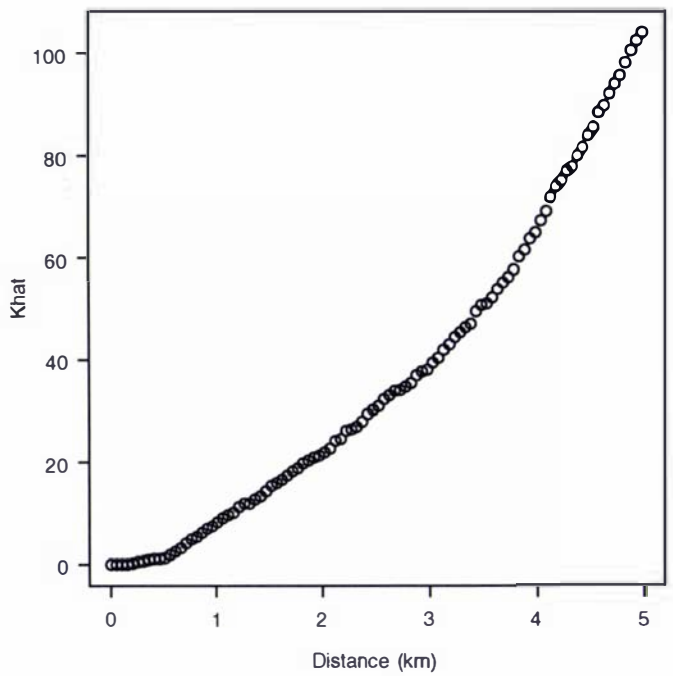


Figure 2.6: K-function plot based on the location of laryngeal cancer diagnoses shown in Figure 2.1a.

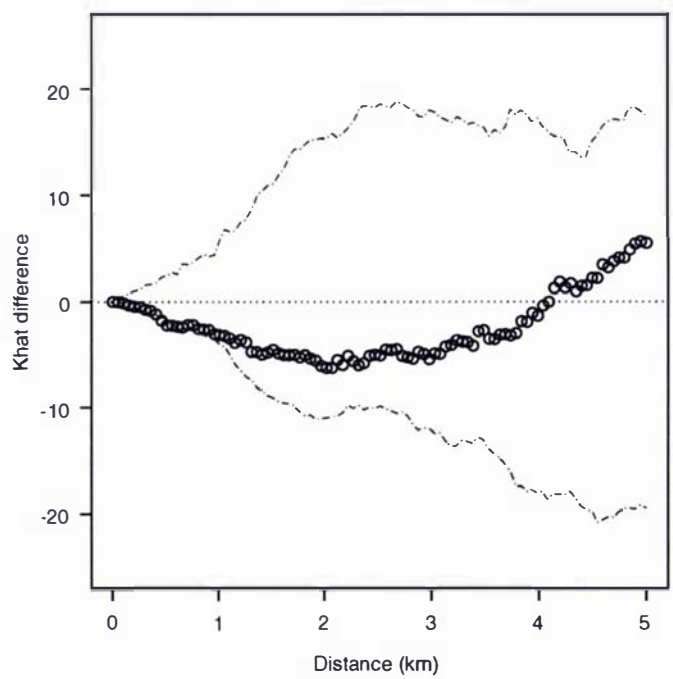


Figure 2.7: K-function difference plot $D(s)$ (∘) and upper and lower bounds of 99 simulation envelopes (-) based on the location of laryngeal cancer diagnoses (cases) and location of lung cancer diagnoses (controls). The observed K-function difference plot lies wholly within the bounds of the simulation envelopes and it can be concluded that there is no tendency for case events to be spatially aggregated, over and above that exhibited by the controls.

2.3.3 Modelling point data

Models of spatial point data seek to quantify the influence of a set of explanatory variables on the occurrence of events throughout a study region. Typically these techniques have been applied to putative sources of hazard where the number of disease events in a region of study are enumerated for a defined period and the spatial distribution of cases in relation to a hypothesised hazard is assessed (see for example Diggle & Rowlingson 1994, Lawson & Williams 1994, Diggle & Elliott 1995, Viel et al. 1995, Lawson & Clark 1999).

Heterogenous Poisson process (HEPP) techniques provide one means by which the spatial distribution of point events may modelled (Diggle 1983, Lawson 1989, Baddeley & Turner 2000). HEPP models are derived on the basis of three assumptions: (1) individuals within a specified study population behave independently with respect to disease propensity, (2) the population at risk from which cases arise has a continuous spatial distribution, and (3) case events are unique in that they occur as single, spatially separate events. Within this framework the density of case events at any point $\lambda(y)$ can be parameterised as:

$$\lambda(y) = \varrho \cdot g(y) \cdot f(y;\psi) \tag{2.3}$$

In Equation 2.3 ϱ represents an overall region-wide rate of disease, $g(y)$ is a function representing the spatial distribution of the population at risk and $f(y;\psi)$ is relative risk function (which may include covariates quantifying proximity to a hypothesised pollution source). The specification of $f(y;\psi)$ determines the relationship between the hypothesised source and a disease event occurring at a given location. A variety of relationships may be defined. Firstly, a distance decline function may be used which accounts for the reduction in risk with distance r from the source:

$$f(y;\psi) = 1 + exp^{(-\beta r)} \tag{2.4}$$

This model includes an additive link between the $g(y)$ function and the excess risk due to the source, thereby preserving the background risk at large distances (that is, $f(y;\psi) = 1$ when r is large). In addition, it is possible to parameterise the effect of direction from a hypothesised source as well as distance (Lawson & Williams 1994, Lawson 1995, Viel

et al. 1995, Le et al. 1996):

$$f(y;\psi) = 1 + exp^{(\beta_1 log_e r - \beta_2 r + cos\theta + sin\theta)} \tag{2.5}$$

Lawson, Biggeri & Williams (1999) warn that summarising disease risk in terms of distance only from a putative source can lead to erroneous conclusions and recommend that the influence of direction and distance be routinely evaluated in putative source analyses.

An attractive feature of the HEPP approach is that it can be utilised for focused clustering assessments, where a cluster centre is identified *a priori* and a model developed to determine the significance of this location as a cluster centre. In this situation, Equation 2.3 can be re-parameterised as:

$$\lambda(y) = \varrho \cdot g(y) \cdot m\left\{\sum_{j=1}^{k} h_1(y - y_j)\right\} \tag{2.6}$$

which describes the intensity of events around k centres located at y_j. Here, the function $f(\cdot)$ is replaced by a link function $m\cdot$. The distribution of events around each hypothesised cluster centre is defined by the cluster distribution function $h_1(\cdot)$ which may be specified for each cluster separately. Lawson & Clark (1999) provide an example of this approach using infant lymphoma and leukaemia diagnoses in Humberside, England. These authors found little support for a positive number of cluster centres, in agreement with the findings of earlier analyses of the same data (Cuzick & Edwards 1990, Diggle & Chetwynd 1991).

HEPP models can be fitted within standard generalised linear model packages using special integration schemes (Berman & Turner 1992, Lawson 1992, Baddeley & Turner 2000). The Spatstat package (Baddeley & Turner 2002) implemented within R (Ihaka & Gentleman 1996) offers a comprehensive range of functions for fitting HEPP models. Although primarily intended for ecological applications, this package offers a set of tools eminently suitable for applications in spatial epidemiology.

An alternative to HEPP techniques is provided by generalised additive models (GAM) (Hastie & Tibshirani 1990). For binary responses, typical in point pattern analyses, a

generalised additive model may be expressed as an additive logistic model:

$$log[p(x)/\{1-p(x)\}] = \alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + S(y) \quad (2.7)$$

In Equation 2.7 α represents the intercept, β_m represent the regression coefficients for m explanatory variables and $S(y)$ provides a measure of the risk of disease at location y. In a GAM setting the only assumption about S is that it is a smooth function of y. Kelsall & Diggle (1998) used kernel regression and a GAM approach to investigate the spatial distribution of cancer diagnoses in Walsall, an area in the north of England. These authors concluded that although the GAM approach was computationally more demanding it offered the advantage of allowing any number of explanatory variables to be explicitly controlled-for. Examples of applications of the GAM methodology in the medical and veterinary epidemiological literature are non-existent. In contrast, the technique appears to be relatively frequently used in marine biology to quantify aspects of fish abundance (see for example Borchers et al. 1997, Bellido et al. 2001, Zheng et al. 2002).

As with any statistical model, it is possible (and frequently the case) that there are factors not included in the analysis that influence the risk of disease. This could be because either the background population hazard is not directly available or the disease displays a tendency to cluster (perhaps due to unmeasured covariates). This heterogeneity could be structured (spatially correlated) or could lack correlation in which case it could be regarded as a type of overdispersion, or unstructured heterogeneity. At the time of writing, there are no widely-available methodologies available for including both structured and unstructured heterogeneity terms within a Poisson process or generalised additive model framework, representing a shortfall of these techniques in general. In contrast, Bayesian approaches (Section 2.4.3) readily accommodate this extension and as a result offer a flexible and analytically tractable means of dealing with these issues that are not altogether unique to spatial epidemiology.

2.4 Lattice data

With lattice (area) data the geographical space under investigation is comprised of a set of sampling units, typically made by partitioning the region into fixed areas that are regularly or irregularly defined in space. An example of regular lattice data is information obtained

from remote sensing satellites. The earth's surface is divided into a series of small rectangles (or pixels) and information for each pixel is arranged in a regular grid within an area of interest. An example of irregular lattice data is cancer counts corresponding to census tracts throughout a country.

2.4.1 Visualising lattice data

Choropleth maps are a common method for visualising lattice data (Figure 2.1b, Figure 2.8). Areas within a region of interest are shaded according to a discrete scale based on the recorded value of the attribute of interest for that area. Although choropleth maps are probably the most widely-used method for illustrating the spatial distribution of disease data, they have three inherent problems. Firstly, component polygons of the study region that are large tend to dominate the display and may induce bias in interpretation (Monmonier & De Blij 1996). Secondly, patterns that are observed across zones may be as much a function of the zone boundaries chosen as of the underlying spatial distribution of the attribute of interest — an effect known as the modifiable area unit problem (Openshaw 1984). The third problem relates to the distribution of the data value being plotted: highly skewed distributions being difficult to display using a finite number of colour shading scales.

One approach to the problem of physical dominance of large areas is to geometrically transform each of the areas of interest so as to make its area proportional to the corresponding attribute value whilst at the same time maintaining the spatial contiguity of each area. The result is known as a density equalised map or cartogram and here the map projection technique sacrifices geographic detail in order to better represent the distribution and magnitude of each area's attribute value (Dorling 1995). Selvin et al. (1988) and Schulman et al. (1988) proposed that areal units may be transformed on the basis of population counts to yield density equalised areas, allowing the density of disease cases to be compared among areas. These authors applied this methodology to investigate the hypothesis that childhood cancers were clustered around a microwave tower in San Francisco, California concluding that the spatial patterns of the childhood cancers investigated were random with respect to the point source (Selvin et al. 1992). Further applications of this technique include investigations into the spatial distribution of breast cancer (Selvin et al. 1998) and adult-onset leukaemia (Selvin & Merrill 2002).

With respect to the modifiable area unit problem a general rule of practice is to analyse the data at hand on the basis of the smallest areal units for which information is available. Aggregation to larger areas should be avoided unless there are good reasons for doing so. Re-analysis of the same data set using different polygonal boundary definitions is advised, if practical (Arlinghaus 1995, Lawson & Williams 2001). Alternatively, irregular area (or point location) data may be re-aggregated to fine regular lattices, a technique used by Pfeiffer et al. (1997) in a logistic regression model developed to predict the occurrence of theileriosis outbreaks in Zimbabwe.

For highly skewed distributions, transformations of the data can be applied. However in doing so some degree of interpretability is lost since large values become truncated into (typically) a single category. Dynamic exploratory data analysis techniques are useful in this situation where a histogram of the variable of interest may be displayed in conjunction with choropleth maps.

2.4.2 Exploring lattice data

If data are collected over areal units that are relatively small in relation to the overall size of a study region, values for each areal unit may be applied to a specific point within the area (for example, the centroid) and the exploratory methods described for point data such as kernel smoothing can be used to explore and describe the process under investigation. More typically however, disease count data is recorded for larger areas and in these circumstances spatial descriptions of disease count data are typically expressed in terms of choropleth maps of standardised mortality (or morbidity) ratios (SMRs).

Standardised mortality ratios

Standardised mortality ratios are calculated by comparing the observed disease count in each area with the expected counts, calculated on the basis of a standardised population. If there are a total of N areas in a study region and O_i represents the observed count of disease in area i and n_i is the population size in area i, the expected disease counts for each area, E_i are given by:

$$E_i = n_i \left(\frac{\sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} n_i} \right) \tag{2.8}$$

If the population within each area can be broken down (for example) by age and gender the expected total number of cases in an area i can be calculated by combining known rates for the disease in each of the separate age-gender groups within the individual area populations. Using the above notation, the SMR for each area and its standard error is given by:

$$SMR_i = \frac{O_i}{E_i} \tag{2.9}$$

$$SE(SMR_i) = \frac{\sqrt{n_i}}{E_i} \tag{2.10}$$

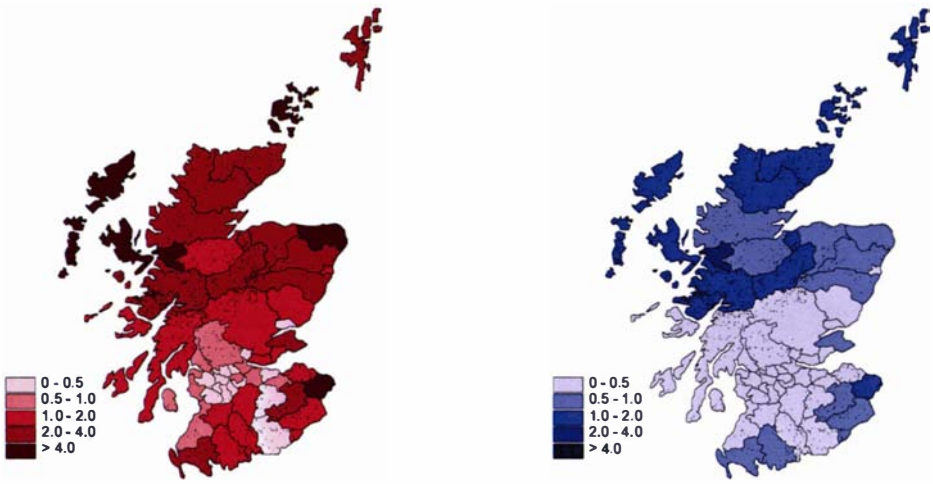
An alternative to computing the standard error of a SMR estimate using Equation 2.10 is to regard O_i as a Poisson variable and determine a 95% confidence interval for the SMR estimate based on the confidence interval for O_i (Altman et al. 2001).

Whereas SMR plots are satisfactory for most disease mapping situations, they are not suitable when the expected disease count in each area equals zero (as in the case of rare diseases) or for differentiating the relative risk in areas where there has been no observable disease. In addition, areas with low population counts may display high risk due to the high variability of the estimated SMRs. Mapping significance levels provides one means of overcoming this problem, although areas with large populations are likely to attain significance even if the excess risk is small (Clayton & Kaldor 1987). As recommended earlier (for kernel density plots), choropleth maps of SMRs should be accompanied by plots that provide some indication of the variability of the estimate quoted for each areal unit, as shown in Figure 2.8.

To overcome problems associated with mapping rare diseases (where disease counts within areas may legitimately equal zero) a standardised mortality difference (SMD) may be calculated as the difference between the observed number of disease counts and the expected number for each area. In this situation the expected value is zero, rather than one, as in the case of standardised mortality ratios.

Bayesian smoothing

Choropleth maps of raw rates or standardised mortality ratios per area may be misleading because the addition of a small number of cases in areas where the population at risk is

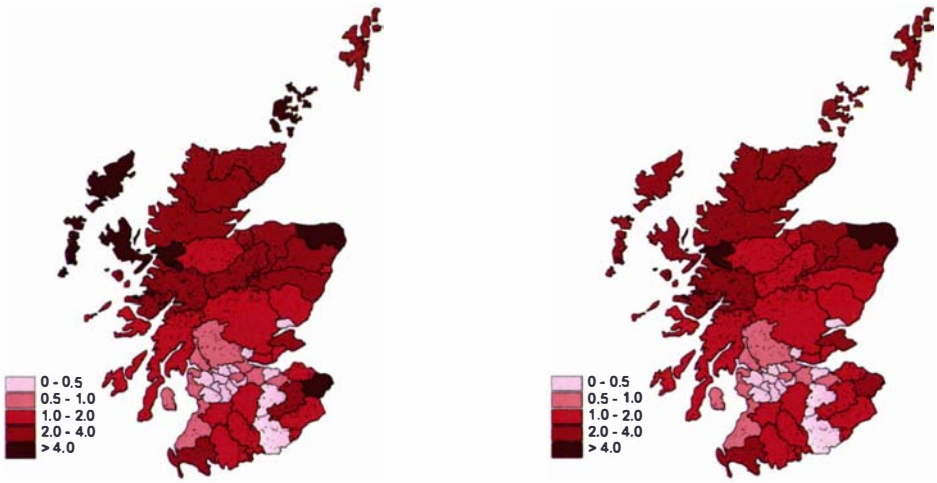


(a) SMR (b) Standard error of SMR

Figure 2.8: Choropleth maps of: (a) standardised mortality ratios (SMRs) and (b) standard errors of the SMRs for lip cancer in Scotland from 1975 to 1980. Although district-level SMRs for lip cancer are highest in the northern-most districts there is greater uncertainty of the SMR estimates recorded for these areas.

small can dramatically increase the reported rate of disease for that area. Conversely, the addition of the same number of cases of disease in areas where the population at risk is high has little effect on the reported rate of disease for that area. Bayesian approaches have been applied to deal with this issue (Clayton & Kaldor 1987, Marshall 1991*b*, Langford 1994). Here, disease rates are adjusted by combining the observed rate with a knowledge of the rates observed in surrounding areas. When the population of an area of interest is large (and the statistical error of the rate estimate is small) higher credibility is given to the observed estimate and the Bayes-adjusted rates are the same (or similar) to the observed rates. When the population is small (and the statistical error of the rate is large) little credibility is given to the observed rate, with it being 'shrunk' towards the mean rate observed over the entire study area. If there is evidence of first order effects in the spatial process under investigation, the rate can be adjusted towards averages of the neighbouring rates rather than the overall mean (Cromley & McLafferty 2002).

There is now a considerable body of literature devoted to the application of Bayesian methodologies to disease mapping (see for example chapters in Lawson, Biggeri, Böhning, Lesaffre, Viel & Bertollini 1999 and Elliott et al. 2000). At its simplest, empirical Bayesian smoothing methods may be applied to observed disease rates or SMRs, as shown in Figure 2.9 and 2.10. At the other end of the spectrum full Bayesian methods using



(a) SMR

(b) Empirically Bayes-smoothed SMR

Figure 2.9: Choropleth maps of: (a) standardised mortality ratios (SMRs) and (b) empirically Bayes-smoothed SMRs for lip cancer in Scotland from 1975 to 1980. Empirical Bayes smoothing has resulted in relatively little change in the raw SMRs since in this example the population counts in each district are large.

sampled distributions may be used to develop models that describe and explain the spatial distribution of observed disease processes.

Most descriptions of empirical Bayes smoothing in the literature illustrate refinements or variations of the methodology, using selected data sets as examples. Marshall (1991*b*) provides a description of empirical Bayes methods smoothing using child mortality data from Auckland, New Zealand as an example. Martuzzi & Elliott (1996) applied empirical Bayes smoothing to non-rare disease conditions using the distribution of respiratory symptoms in school children in Huddersfield, England as an example. Several authors have compared empirical and full Bayes-smoothed estimates (Benardinelli & Montomoli 1992, Heisterkamp et al. 1993, Cislaghi et al. 1995). In the veterinary literature empirical Bayes smoothing was used to describe the prevalence of *Echinococcus multilocularis* in red foxes in Lower Saxony, Germany (Berke 2001). In that study inherent instability in the data was smoothed allowing first order trends to be more readily appreciated.

Autocorrelation statistics

With lattice data, the spatial relationship between areas within a region may be described in terms of a proximity matrix, W. Spatial proximity matrices can be constructed on the

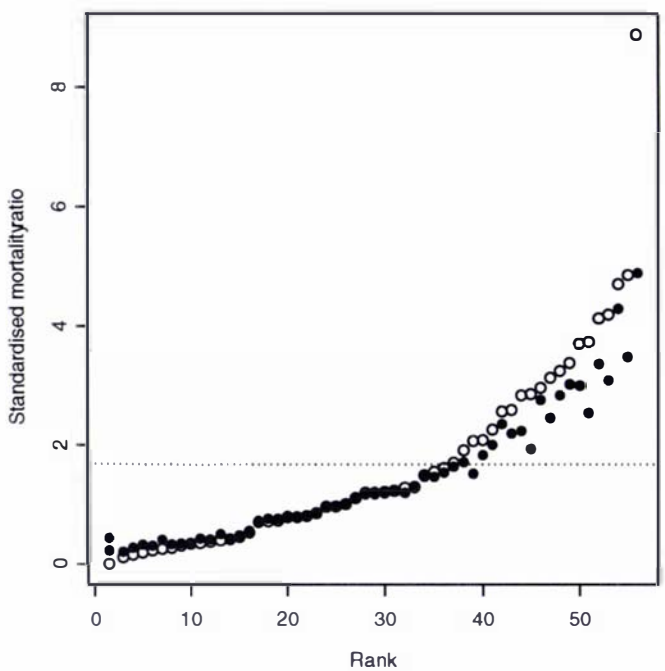


Figure 2.10: Explanation of empirical Bayes smoothing. Raw standardised mortality ratios (SMRs) (○) and empirically Bayes-smoothed SMRs (●) for lip cancer in Scotland from 1975 to 1980 have been plotted in order of rank of the raw SMR estimates. The horizontal reference line shows the raw mean SMR value (across all districts). Bayes-smoothing has resulted in the large raw SMR estimates being 'adjusted' towards the mean.

basis of areal units sharing a common boundary, the distance between area centroids or boundaries, or on the length of a common boundary. Alternative measures of proximity might include travel time or the flow of traffic (animals, people, or goods) between areas. An advantage of distance-based specifications over contiguity-based specifications is that missing areas (e.g. areas with zero denominators) and discontinuities (e.g. bodies of water) are easily accounted-for (Pascutto et al. 2000). A key feature of spatial proximity matrices is that the relationship between areas is defined by relative (rather than absolute) location.

In an exploratory sense it is useful to assess the level of spatial autocorrelation present in the data (that is, characterise how similar attribute values are among neighbouring areas). Implicit in the concept of spatial proximity matrices is that the relationship between two areas is discrete, an assumption that may not be fully appropriate in some situations (for example, in the case of infectious diseases where airborne spread plays an important role in transmission). This issue, where the strength of the relationship between areas varies continuously through space, rather than discretely, will be discussed more fully in

Section 2.5.2.

Various measures are available to quantify spatial autocorrelation among areas. Moran's *I* statistic (Moran 1950) is similar to a conventional correlation coefficient quantifying the similarity of an outcome variable among areas that are defined as spatially related (on the basis of a specified proximity matrix). Where N equals the number of areas within a study region Moran's *I* has an expected value of $-1/(N-1)$. The expected value of Moran's *I* is negative and is a function of sample size, approaching zero as the sample size increases. A Moran's I coefficient larger than its expected value indicates positive spatial autocorrelation; spatial clustering of similar values (either high or low). Values of Moran's I less than 0 indicate negative spatial autocorrelation; spatial clustering of dissimilar values (either high or low). Geary's c statistic (Geary 1954) is closely related to Moran's *I*. A Geary's c statistic close to zero indicating spatial dependence; a Geary's c statistic close to one indicating an absence of spatial dependence.

Moran's *I* or Geary's c can be computed at each of k spatial lags (W^k) and the value of *I* (or c) and its 95% confidence interval can be plotted as a function of W^k to produce a correlogram (Figure 2.11). For exploration of spatial data the correlogram quantifies the degree of similarity between neighbouring variables at given scales of proximity. Glick (1979) investigated spatial autocorrelation in cancer diagnoses in Pennsylvanian counties from 1950 to 1969. Using neighbourhood definitions based on contiguity Glick (1979) identified positive autocorrelation among stomach cancer diagnoses that persisted to the second order spatial lag. The conclusion from this study was that the findings were consistent with stomach cancer rates being grouped into relatively homogeneous clusters of 15 or more counties, in agreement with aetiologic hypotheses linking stomach cancer to diet, ethnicity, soil characteristics and air pollution. Lung cancer rates, on the other hand, showed a tendency to be autocorrelated only as far as the first order spatial lag, reflecting the role of physical distance (rather than contiguity) in determining the spatial distribution of this disease. Perez et al. (1994), in a study of the epidemiology of bovine anaplasmosis and babesiosis in Costa Rica, used Moran's I to identify two foci of seropositivity in the country for *Babesia bovis*: one located in dry tropical forest and the other located in moist tropical forest.

Moran's *I* provides a global summary of the degree of linear association between a vector of observed values and a weighted average of the neighbouring values. To provide

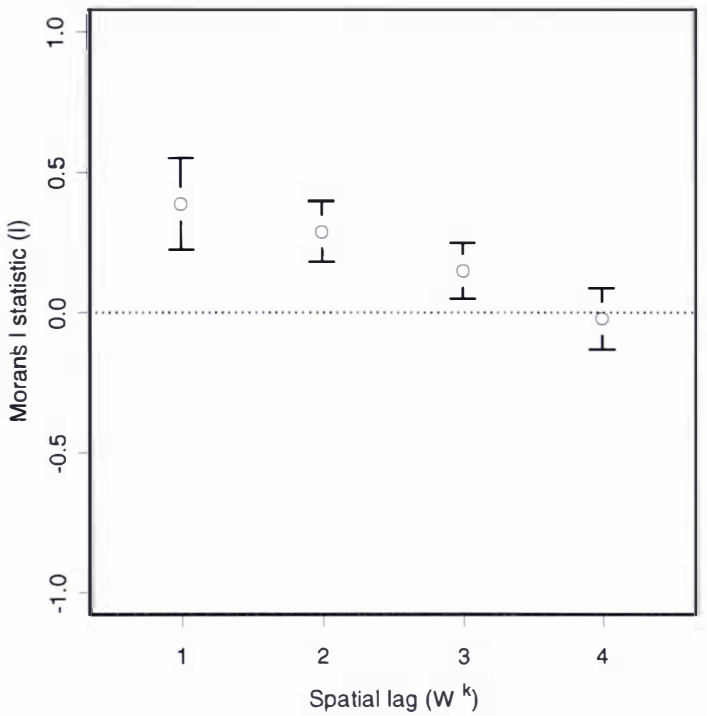


Figure 2.11: Moran's I statistic (and 95% confidence interval) computed on the basis of the standardised mortality ratios for lip cancer diagnoses in Scotland from 1975 to 1980. Moran's I statistics have been computed at first, second, third and fourth order spatial lags. This plot shows that lip cancer SMRs are significantly autocorrelated up to the third order spatial lag.

further insight into the composition of this statistic, local indicators of spatial association (LISA) may be determined for each area such that the sum of the LISA statistics for all zones is proportional to the global Moran's I. LISA statistics may be presented as a Moran scatter plot (Anselin 1995, Anselin 1996), as shown in Figure 2.12. Here, the horizontal axis of the plot represents the vector of observed values and the vertical axis specifies the weighted average of neighbouring values. The extent of the 'mix' of pairs between the four types of association (high-high, low-low, high-low, and low-high) provides an indication of the stability of the spatial association throughout the observed data. It may also suggest the existence of different regimes of association in different subsets of the data: positive association in one area and negative association in another. In addition, the Moran scatter plot is useful for identifying outliers within the global pattern of association.

Two important conditions hold for the autocorrelation statistics described: (1) the value of the computed autocorrelation statistic is sensitive to the specification of the spatial proximity matrix for the region under investigation, and (2) tests of significance are reliant on the variable of interest having constant variance across all areas. Standardised mortality ratios are unsuitable for determining summary autocorrelation statistics unless

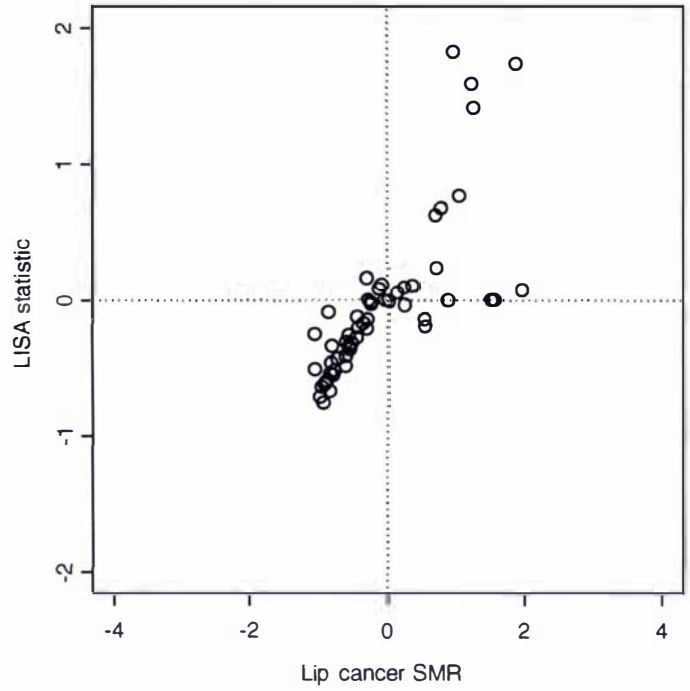


Figure 2.12: Moran scatter plot of the standardised mortality ratios for lip cancer diagnoses in Scotland from 1975 to 1980, based on a first order adjacency matrix. This plot shows that the neighbours of districts with low lip cancer SMRs have similar, low SMR values (bottom left quadrant). This relationship is not as marked for those districts with high lip cancer SMRs (top right quadrant), reflecting greater spatial instability in districts with high SMRs (see Figure 2.8).

they are based on approximately equal expected numbers. To deal with areas of different population size a Moran's I statistic adjusted for population size has been described (Oden 1995). To avoid these inherent problems variogram analyses may be applied to lattice data (Cressie & Chan 1989, Wakefield & Morris 1999), as described in Section 2.5.2.

2.4.3 Modelling lattice data

Typically there are two goals in ecological (lattice) modelling: firstly, to quantify the influence of fixed-effects on the level of disease within each area of a region under study and secondly to identify areas where there are higher than expected counts of disease after influence of specified fixed-effects have been accounted for.

To achieve these goals a number of techniques have been described. These extend from frequentist approaches where a global summary of the strength of area-to-area effect is provided in addition to fixed-effect regression estimates (see for example Walter et al. 1999) to full Bayesian mixed-effects models where area-level spatially correlated random

effect terms are determined (see for example Toledano et al. 2001, Jarup et al. 2002). This discussion will briefly focus on frequentist approaches then provide a more detailed description of Bayesian methods which have become widely applied to spatial epidemiological problems in recent years (Lawson, Böhning, Biggeri, Lesaffre & Viel 1999). Before doing so, some comments are made concerning the aims and pitfalls of ecological analyses.

For descriptive studies of spatial patterns of disease using lattice data, inference is made at the area level. On the other hand analytical studies based on lattice data tend to focus inference at the individual rather than the area level. There exists the possibility of a difference between risk estimates at the individual and area level, a difference referred to as cross-level or ecological bias. Richardson & Monfort (2000) discuss this issue at length and recommend that meaningful interpretation of ecological regression studies requires consideration of: (1) how to relate individual-level and aggregated level dose-response relationships, (2) the importance of considering confounders, including the use of appropriate summaries of confounder information at the aggregated level, and (3) the influence of unmeasured confounders, some which may vary spatially, and which are implicitly modelled as spatially-varying random effects.

As a starting point for modelling area data it is acceptable to use ordinary least squares regression to quantify first order trends in the data of interest. To do this, the coordinates of the centroids of each component zone of the study region may be introduced as covariates in addition to other variables that are thought to be influential. It should be remembered that the assumptions implicit with the use of ordinary regression (that is the error terms produced have a mean of zero and have constant variance) are frequently violated when applied to spatial data and these assumptions should be routinely checked. To identify the presence of a spatial effect not accounted-for by an ordinary regression analysis it is recommended to test for the presence of spatial autocorrelation in the model error terms using global Moran's I statistics and/or LISA statistics. Where spatial autocorrelation in the error terms is absent (and the error terms meet all of the assumptions of an ordinary regression) it may be concluded that the ordinary least squares model provides an adequate description of the data. Conversely, the presence of spatial autocorrelation in the error terms indicates that the model should be re-parameterised to account for these second order (spatial) effects.

Simultaneous Autoregressive models

Given the general expression for a generalised least squares regression model as:

$$Y_i = \alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + \epsilon_i \tag{2.11}$$

we can define the relationship between the outcome variable in area i and its neighbours using an interaction scheme:

$$\epsilon_i = \rho W \epsilon_j + \xi_i \tag{2.12}$$

In Equation 2.12 ρ is an interaction parameter estimated from the data, W is a spatial proximity matrix defined for the study region, ϵ_j is the vector of error terms recorded for area i's neighbours, and ξ_i is a vector of independent random errors with zero mean and constant variance. Using this notation, the expression for a Simultaneous Autoregressive (SAR) model is:

$$Y_i = \alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + \rho W Y_j - \rho W \beta x_j + \xi_i \tag{2.13}$$

In this context the outcome in each area Y_i depends on: (1) the magnitude of covariates measured in the area of interest $\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}$, (2) the magnitude of Y in neighbouring areas $\rho W Y_j$, (3) covariates measured in neighbouring areas $\rho W \beta x_j$, and (4) a 'well behaved' error term ξ_i. The spatial autocorrelation coefficient ρ should be thought of as a global measure of the residual spatial structure in the data after accounting for the effects of the explanatory variables included in the model. Where the proximity matrix is standardised to have row sums of unity, ρ takes a value from -1 to $+1$. This parameterisation may be simplified to:

$$Y_i = \alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + \rho W Y_j + \xi_i \tag{2.14}$$

where the second order variation is represented as being dependent only on the magnitude of Y in neighbouring areas. A further simplification results if the first order term is removed completely, yielding a purely autoregressive (AR) model:

$$Y_i = \rho W Y_j + \xi_i \tag{2.15}$$

Examples of the use of SAR models in the epidemiological literature include studies of regional variations in cancer incidence in Ontario, Canada (Walter et al. 1999), the relationship between exposure to air pollution exposure and socioeconomic status in Hamilton, Canada (Jerrett et al. 2001) and early childhood mortality in Sao Paulo, Brazil (Antunes & Waldman 2002).

Conditional Autoregressive models

Whereas SAR models have historically been the most commonly-used model parameterisation in spatial statistics, alternative classes are available. One such class is based on a Conditional Autoregressive (CAR) spatial covariance structure (Sun et al. 1999) where the conditional density of the outcome variable Y_i is dependent on values recorded in neighbouring areas Y_j. Besag et al. (1991) proposed a special 'intrinsic' form of the CAR model for the case where elements of the spatial proximity C equal zero unless areas i and j are adjacent. If areas i and j are defined as contiguous then $C_{ij} = C_{ji}$ are positive weights. MacNab & Dean (2000) have proposed extensions to this model, allowing rates of disease within defined areas to depend not only on their immediate neighbours but also on their neighbour's neighbours.

Bayesian approaches to modelling lattice data

A considerable literature has developed around Bayesian approaches to modelling disease counts within arbitrarily-defined regions (Manton et al. 1981, Tsutakawa 1988, Besag et al. 1991, Marshall 1991*a*, Clayton & Benardinelli 1992, Breslow & Clayton 1993, Lawson 1994, Ghosh et al. 1998). Central to this method is the inclusion of area-specific random effect terms to account for unobserved spatial features within the data. The appeal of this approach is that areas of disease excess can be identified after controlling for the influence of known risk factors. These locations could be targeted for further epidemiological investigations or used by policy makers for the purposes of funding allocation. For discussions of the methodological issues related to this subject the reader is referred to Clayton & Kaldor (1987), Besag et al. (1991), Clayton & Benardinelli (1992), Cressie (1992), Devine et al. (1994a), Devine et al. (1994b), Pickle et al. (1996), Waller et al. (1997), Xia et al. (1997), and Conlon & Louis (1999).

Drawing on these technique-based studies, hierarchical Bayesian methods have re-

cently been used to investigate the spatial distribution of testicular and prostate cancer in Britain (Toledano et al. 2001, Jarup et al. 2002), breast cancer in Greece (Vlachonikolis et al. 2002), insulin dependent diabetes mellitus in Austria (Schober et al. 2001), stroke and cardiovascular disease in Britain (Maheswaran et al. 2002), multiple sclerosis in Italy (Pugliatti et al. 2002), low birthweights in Papua New Guinea (Müller et al. 2002) and malaria in South Africa (Kleinschmidt et al. 2002). The predominance of European-based studies is noteworthy and most likely reflects the influence of European research groups that have been at the forefront of developing these techniques. At the time of writing, there are no published examples of full Bayesian approaches to disease mapping in the veterinary literature.

An outline of Bayesian hierarchical spatial modelling of disease count data is as follows. When a disease process is non-contagious and rare the observed count in each area O_i is assumed to follow an independent Poisson distribution with mean μ_i:

$$O_i \sim Poisson(\mu_i) \tag{2.16}$$

If E_i denotes the expected disease count and ψ_i denotes the relative risk of disease in area i then:

$$\mu_i = E_i \psi_i \tag{2.17}$$

This model can be extended to include area-specific covariates x_i such that:

$$log(\mu_i) = log(E_i) + (\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}) + \epsilon_i \tag{2.18}$$

In a Bayesian context non-informative prior distributions can be assumed for the intercept α and each of the regression coefficients β_m. This formulation represents a standard log-linear fixed-effects model where the covariate effects act multiplicatively on overall relative risk. The exponent of ϵ_i represents the residual relative risk in area i after adjusting for covariates included in the model. It is often convenient to interpret ϵ_i as reflecting the residual between-area variability due to unknown or unmeasured risk factors.

Unknown risk factors will typically vary in space, which in turn induces spatial correlation between the observed disease counts in nearby areas. To account for the presence of spatial correlation it can be assumed that the unexplained variation in the log-linear

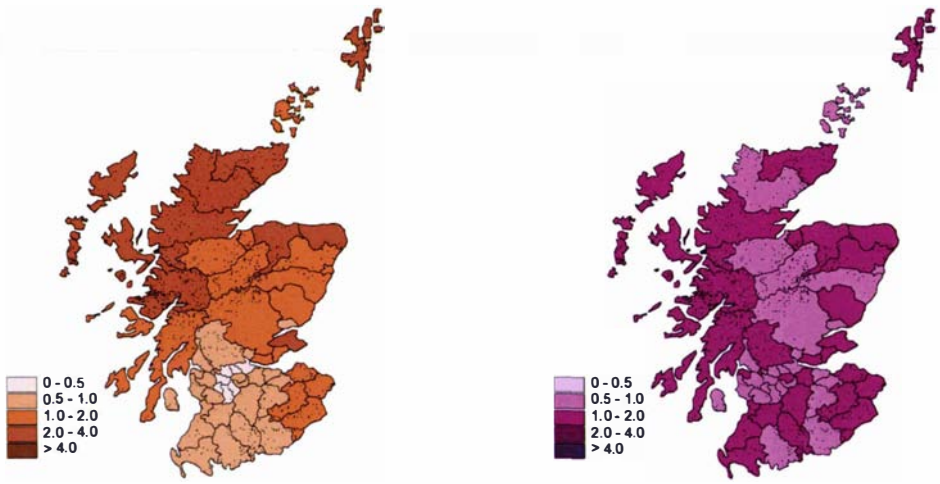
model are comprised of two parts: (1) a structured (spatially correlated) component, and (2) an unstructured component. An intermediate distribution of the log relative risks that ranges from prior independence (unstructured heterogeneity) to prior local dependence (structured heterogeneity) is known as a convolution Gaussian prior (Besag 1989, Besag & Mollié 1989, Besag et al. 1991, Mollié 1996). In this context the parameterised model becomes:

$$log(\mu_i) = log(E_i) + (\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}) + U_i + S_i + \xi_i \qquad (2.19)$$

In Equation 2.19 S_i represents the structured (spatially correlated) random effects and U_i represents the unstructured random effects for each area i. Here it is usual to assume a spatially structured prior distribution for the structured component of the random effects. Various choices exist, but the most popular is a special case of the conditional intrinsic Gaussian autoregressive (CAR) model (Besag et al. 1991, Besag & Kooperberg 1995). This models the log relative risk in area i conditional on the risks in all other areas as being normally distributed about the weighted mean of the log relative risks in the remaining areas, with the sum of the weights being inversely proportional to the variance σ^2. The unstructured heterogeneity component U_i is typically parameterised as being normally distributed with mean zero and variance τ^2.

The strength of the 'mix' of structured and unstructured heterogeneity components depends on depends on the priors specified for σ^2 and τ^2. Large values of σ^2 allow S_i to show wide variation (resulting in little spatial smoothing) whereas small values of σ^2 forces all of the spatial heterogeneity terms to be similar (resulting in greater spatial smoothing). In practice, σ^2 and τ^2 are assigned prior distributions (hyperpriors) which are typically assumed to follow a gamma distribution with a shape parameter a and inverse scale parameter b. Benardinelli, Clayton & Montomoli (1995), Best et al. (1999), and Richardson & Monfort (2000) have discussed the issues associated with the parameterisation of hyperpriors, concluding that the sensitivity of proposed models be tested against a range of hyperprior specifications.

Plots of the structured and unstructured heterogeneity terms for the Scottish lip cancer data set (expressed as relative risks) are shown in Figure 2.13. Relative risk estimates computed for the structured heterogeneity terms show that after accounting for the effect of a single covariate (the proportion of the workforce in each district that are involved in



(a) Structured heterogeneity (b) Unstructured heterogeneity

Figure 2.13: Choropleth maps showing: (a) the distribution of spatially-correlated relative risk of lip cancer that remains after covariate adjustment, and (b) the distribution of unstructured relative risk of lip cancer that remains after covariate adjustment. The spatial distribution of the structured heterogeneity terms show that after accounting for the effect of the proportion of the workforce involved in agricultural industry the risk of lip cancer is greater at higher latitudes.

agricultural industry) there is an increased risk of lip cancer associated with high latitudes. The unstructured heterogeneity terms on the other hand show a more irregular spatial distribution.

In addition to plotting the relative risk estimates attributable to structured and unstructured heterogeneity it is also informative to consider how much more variable the spatially correlated terms are relative to the uncorrelated terms. Although the variance parameters σ^2 and τ^2 are not directly comparable, Best et al. (2000) cite the ratio of the standard deviation of the posterior estimates of S_i to the standard deviation of the posterior estimates of U_i as a reasonable summary measure. In the Scottish lip cancer example the standard deviations of the structured and unstructured heterogeneity terms were 0.61 and 0.01, respectively. This means that, based on the first order adjacency matrix specified, the structured (i.e. spatial) random effects were roughly fifty times more variable than the unstructured random effects. This implies that the 'spatial effects' are large in some areas and not in others. The unstructured random effects, in contrast, are relatively uniform across all areas.

Mixture models described by Böhning & Schlattmann (1992), Schlattmann (1996b) and Schlattmann (1996a) are a variation of the Bayesian hierarchical models described and assume that areas within a study region can be grouped into discrete homogeneous risk classes. These models force those areas within a group to have the same structured heterogeneity terms imposing a discontinuity in spatially-structured risk between one group and the next. While these models address the issue of discontinuities that may legitimately exist in some situations (for example at urban-rural fringes) they do not account for the possibility that areas of a study region may also have smooth rate transitions. Developing this concept Lawson & Clark (2002) describe a special type of spatial mixture model which admits different forms of spatial variation, with these linked together by a spatially varying weight matrix. Models of this type therefore provide a more flexible compromise between the fixed spatial weighting scheme used by Besag et al. (1991) and the homogeneous risk class scheme used by Böhning & Schlattmann (1992), Schlattmann (1996b) and Schlattmann (1996a).

2.5 Continuous spatial data

Spatially continuous data includes phenomena such as rainfall, humidity, air pollution and soil mineral concentrations: variables that may be measured at all possible locations within a study area. When dealing with continuous spatial variables the aim in most situations is to collect data at a series of fixed sampling points and to use that data to predict the value of the variable of interest at other, unsampled locations. In epidemiology continuous variables of the type cited above are typically used as covariates for predicting the risk of disease (see for example Perry et al. 1991, Diggle et al. 2002, Hammond et al. 2001). Continuous spatial data are also commonly used in fate and transport models that investigate the distribution of agents released into the environment (Cromley & McLafferty 2002). These models require geographic and physical descriptions of the source and information on the rate of release into the atmosphere, surface water and land. Given the location of the source, the meteorology of the receiving air and the hydrology and hydrogeology of the receiving water features, distribution of the agent throughout the environment can be evaluated. A recent and widely publicised application of fate and transport modelling applied in an epidemiological context was the *Escherichia coli* O157:H7 and

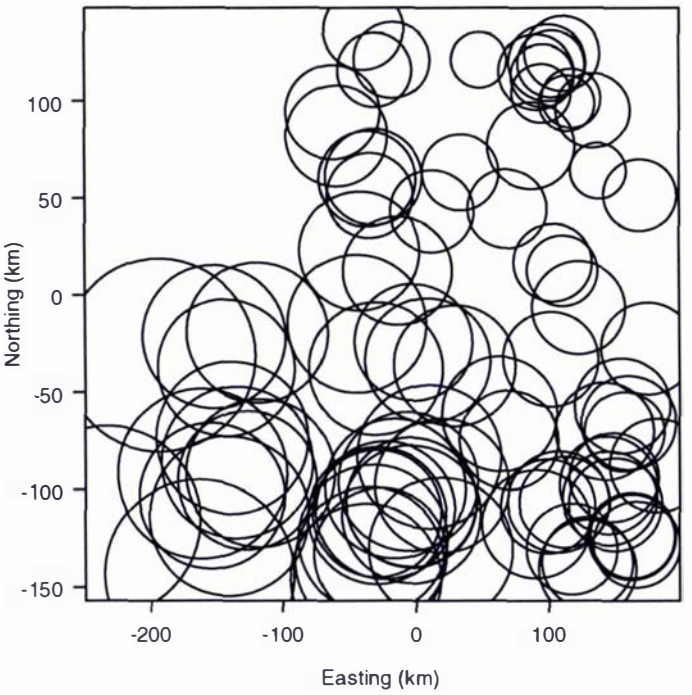


Figure 2.14: Bubble plot constructed using the aquifer data set shown in Figure 2.1c. Open circles have been drawn at each of the sampling locations with the size of each circle proportional to the value of the sample recorded at that location.

Campylobacter outbreak that occurred in the town of Walkerton, Ontario, Canada in May 2000 (Meyers et al. 2002). In this case heavy rainfall resulted in bacteriological contamination of the town's water supply resulting in 2,300 clinical cases of gastrointestinal disease and seven deaths. Hydraulic modelling tools were used to trace contamination from the source to the household and to provide confirmation of the outbreak source.

2.5.1 Visualising continuous spatial data

In the simplest situation continuous spatial data can be visualised by plotting values recorded at each sampled location onto a map. To improve interpretation proportional symbol maps may be used where geometric symbols (for example circles) are plotted with their area proportional to the recorded data value (Figure 2.14). Alternatively, map symbols may be of fixed size and shaded according to their recorded value.

2.5.2 Exploration of continuous spatial data

While proportional symbol maps are useful for preliminary visualisation of data they give no appreciation of spatial continuity. To achieve continuity it is necessary to predict the

mean of the variable of interest across the entire area under investigation. This entails the construction of a 'surface' representing the attribute of interest and deriving contour plots from this surface for purposes of interpretation. The two principal methods for estimating the mean of a variable of interest across space are: (1) tessellation methods, and (2) kernel estimation techniques.

Tessellation methods

Tessellation is the geometric process of partitioning an area into smaller units that do not overlap but completely fill the entire area of investigation (Arlinghaus 1994). When the partitions are the same size and shape (for example, a square, triangle or hexagon) the tessellation is said to be regular. Where the partitions do not have the same size, shape or orientation the tessellation is said to be irregular. Given a defined number of sampling locations, one can assign to each sampled location s an 'area' defined as that part of the study area which is closer to s than any other sampled location. Lines identifying the border of these areas constitutes a Dirichlet tessellation of the sampled locations within the study area (Figure 2.15). Each area constructed by this process is known as a Voronoi or Thiessen polygon. Thiessen polygons that share a common boundary are declared contiguous. The lines that join all pairs of contiguous point locations (contiguous, in the sense that the Thiessen polygon that covers each point shares a boundary) form a Delaunay triangulation network (Figure 2.16). With this set of non-overlapping triangles defined, a vertical line can be projected from each sampled point, with the height of the point proportional to its recorded value. The Delaunay triangulations can then be projected from the sampling points to yield a set of tilted planes over the study region. The slope of each of these tilted planes enables a value to be estimated at any point, and a series of contour lines constructed, as shown in Figure 2.17.

Kernel smoothing continuous spatial data

Kernel estimation techniques that were described in Section 2.3.2 can also be applied to spatially continuous data. In this case, interest lies in determining $\mu(s)$ the mean value of an attribute whose values y_i have been sampled at locations s_i. For any chosen bandwidth, values of $\mu(s)$ can be examined at locations on a fine grid over the study region. The bandwidth can be altered to vary the smoothness of this estimate, as for point data. A

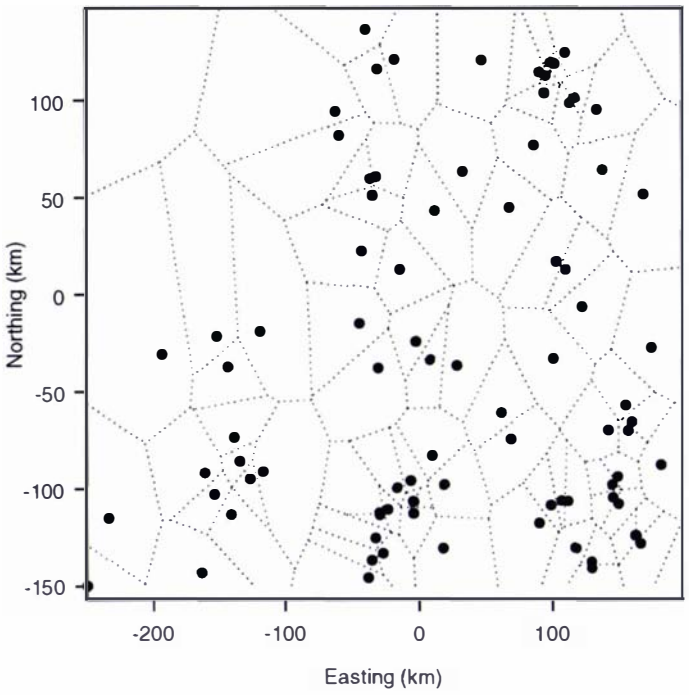


Figure 2.15: Dirichlet tessellation plot constructed using the aquifer data. Each area defined by the tessellation process is known as Voronoi or Thiessen polygon.

kernel-smoothed regression surface for the aquifer data set is shown in Figure 2.18.

Variograms

Variogram functions characterise the second order properties of a continuous spatial process and are analogous to the K-function techniques described for point pattern analyses. A variogram $\gamma(s)$ is estimated empirically as the average squared difference of the values recorded for each pair of sampled points where the distance between the points falls within a given lag band. A variogram starts from 0 at distance $s = 0$ and increases to a maximum of σ^2 (the sill). Variogram functions where the sill is reached at large values of s indicate a high degree of spatial autocorrelation. Variograms functions where the sill is reached at small values of s indicate little spatial autocorrelation. If a variable shows little spatial dependence its variogram will be without pattern. A variogram computed for the aquifer data set is shown in Figure 2.19.

Although in theory $\gamma(0) = 0$, sampling errors and small scale variability may often cause sample values with small separations to be dissimilar. This produces a discontinuity at the origin of the sample variogram. The size of the discontinuity at $s = 0$ is termed the nugget (τ^2) and represents micro-scale variation or measurement error. As part of a

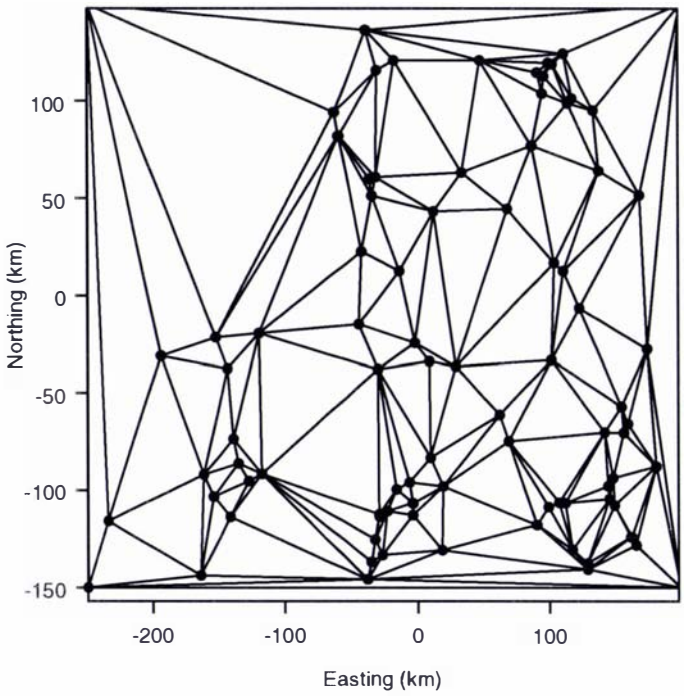


Figure 2.16: Delaunay triangulation network constructed using the aquifer data.

descriptive analysis it is informative to determine the ratio of τ^2 to σ^2, the noise to signal ratio. A large noise to signal ratio indicates weak spatial autocorrelation whereas a small noise to signal ratio is indicative of strong spatial autocorrelation.

Similar to the approach adopted for point and lattice data, after assessing continuous spatial data for first order trends one would begin an exploratory analysis of the second order properties or covariance structure by estimating an isotropic variogram (that is, a variogram that ignores direction). Subsequent analyses would then proceed to calculate directional variograms in order to identify directional effects or anisotropy. This allows an informal assessment of the degree of spatial dependence in the data and identifies any strong directional effects that may be present. In addition, a useful exploratory technique is to take all possible pairs of points and plot the squared difference of the values recorded for each pair $(v_i - v_j)^2$ versus their distance separation, producing a variogram cloud. This may reveal extreme outlying points that dominate the estimate of the sample variogram. It may also reveal skewness in the distribution of the differences at any lag which shows that the variogram will be a poor estimate of the true covariance structure at that lag. A variogram cloud computed for the aquifer data set is shown in Figure 2.20.

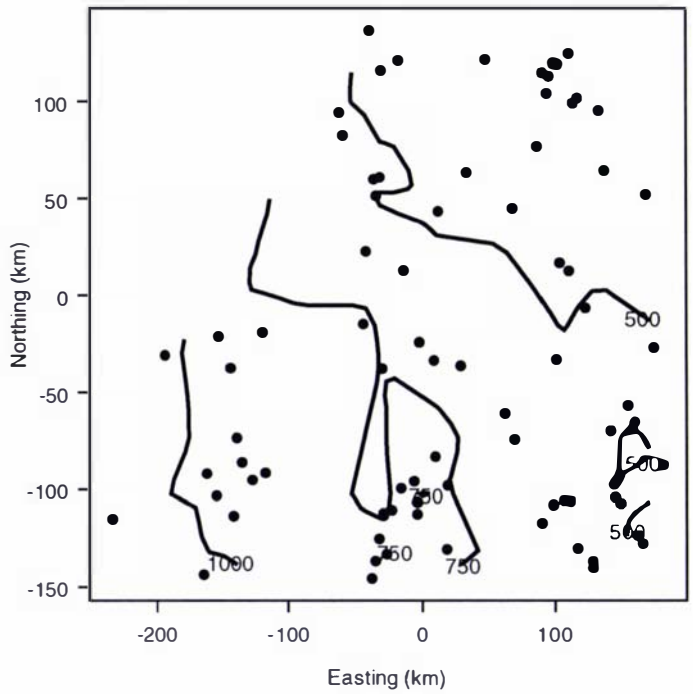


Figure 2.17: Triangular irregular network contours constructed using the aquifer data.

2.5.3 Modelling continuous spatial data

Trend surface analysis

Trend surface analyses involve the application of a polynomial function of the spatial co-ordinates of sample sites to the observed data values using ordinary least squares regression. Covariates other than location may be included in the model to further understand or explain spatial variation. A simple linear trend surface for the aquifer data set is shown in Figure 2.21. The surface is described by three parameters: β_1, β_2, and β_3 relating to (respectively) the intercept of the plane at the origin of the coordinate system, the slope in the coordinate direction s_1 and the slope in the coordinate direction s_2. The assumptions of ordinary least squares regression state that the errors produced by a model at different points have a mean of zero and a constant variance. With spatial data these assumptions are often violated since the error terms tend to be spatially autocorrelated and the variance of the error terms is not constant throughout the region of investigation. Furthermore, if higher order terms are included in an ordinary least squares model in order to explain local (second order) variation in the data, then the regression coefficients will be highly correlated and their estimates correspondingly unstable. Thus, ordinary least squares regression techniques should only be used to indicate broad trends and relationships in a

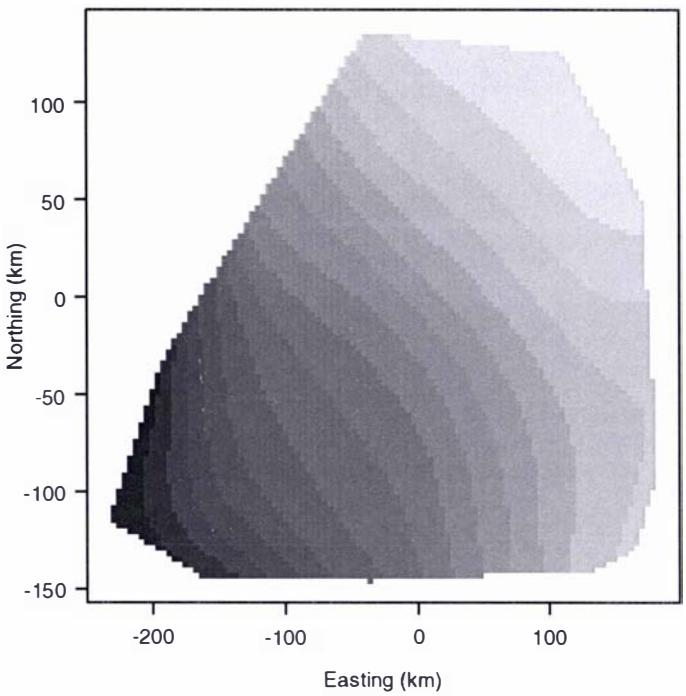


Figure 2.18: Image plot of kernel regression estimates computed using the aquifer data.

data set without emphasis on formal hypothesis tests for parameter values and model fit.

Moore (1999) used trend surface analyses to identify the major direction and speed of diffusion of an epidemic of raccoon-rabies in Pennsylvania USA from 1982 to 1986. Acknowledging the caveats involved in applying this technique to spatial data, Moore (1999) concluded that the technique was useful for removing the inherent noise in the reported data and helped to identify geographic 'corridors' within Pennsylvania that were associated with higher rates of diffusion of the disease. Hanchette & Schwartz (1992) used trend surface analyses to assess geographic patterns in prostate cancer mortality in the United States identifying a north-south trend in prostate cancer mortality rates, thought to be consistent with the hypothesis that exposure to ultraviolet radiation is protective against the disease.

Generalised least squares models

To avoid the implausible assumptions required by a trend surface analysis it is necessary to relax the assumption of independence of residuals by using a generalised least squares approach. Given the general expression for a generalised least squares regression model shown in Equation 2.11, the $\beta_m x_{mi}$ terms represent the first order component of the spatial process and ϵ_i represents a zero mean vector of errors with variance-covariance matrix C.

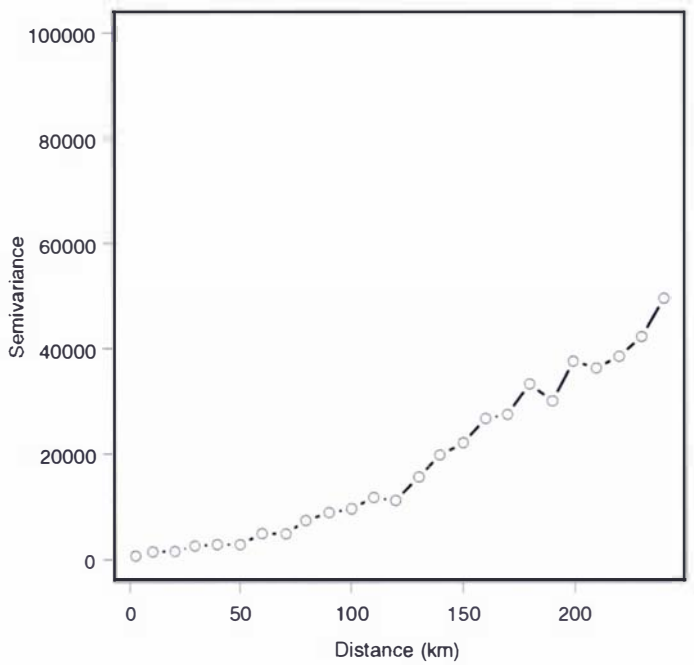


Figure 2.19: Empirical variogram plot computed using the aquifer data.

Under this parameterisation, while ϵ_i has a mean of zero, the values of ϵ_i at different locations are not necessarily independent, having a covariance function defined by the term C. In order to produce an appropriate spatial generalised least squares regression model for continuous data one needs to estimate the elements of C directly. This is done by considering the variogram functions that were developed to explore the second order properties of the data.

Three variogram models are commonly-used for stationary spatial processes: (1) spherical, (2) exponential, and (3) Gaussian. Nugget effects can be introduced into any of these forms by the inclusion of a constant term on the understanding that this will represent a discontinuity at the origin and that $\gamma(0)$ still remains zero. Having obtained a satisfactory variogram model for the error terms produced by a generalised least squares regression, an estimated covariance matrix can be constructed between the sample sites C with elements $C(s_i, s_j)$. In effect this 'corrects' the parameter estimates and standard errors of the generalised least squares regression for second order effects.

If the objectives of a study are to describe and understand the nature of the variation in the observed attribute value throughout a study region, then knowledge of the trend component and the form of the covariance structure of the process is necessary. If, in addition, there is interest in prediction or interpolation of attribute values at locations that have not been sampled, then kriging techniques (Matheron 1965, Matheron 1971) are

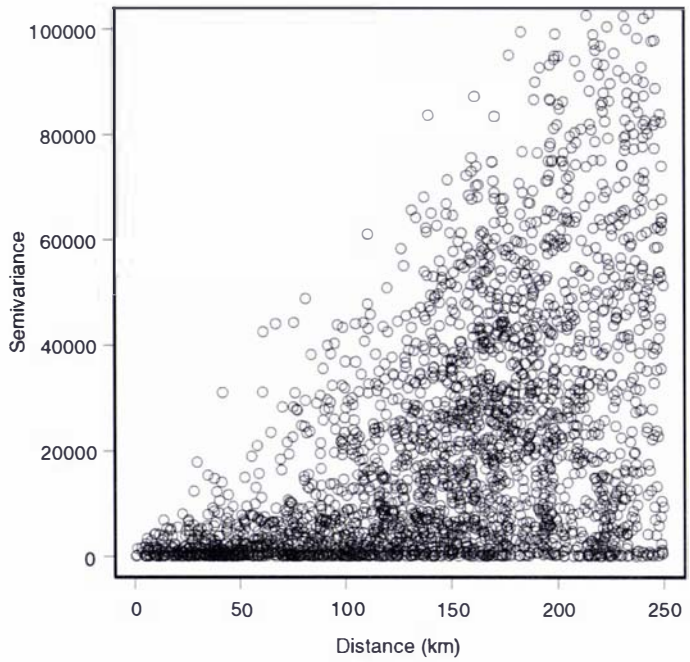


Figure 2.20: Variogram cloud plot computed using the aquifer data. At short distance separations (0 to 50 kilometres), piezometric-head measurements tend to be similar. Dissimilarity in measured values increases with increasing distance.

appropriate.

Kriging

Kriging involves predicting a zero mean spatially correlated process at each location in a study area using an optimal weighted linear combination of neighbouring observed values of the spatial process under investigation (Matheron 1965, Matheron 1971). Two variations of this technique are ordinary kriging and universal kriging. Ordinary kriging is applied in the situation where the first order component of the spatial process has a constant mean and does not vary geographically. Universal kriging is applied in the more general situation where a first order trend exists.

Because continuous spatial data are not commonly used as outcome variables in epidemiological studies, kriging techniques have not featured highly in this area of the literature. Carrat & Valleron (1992) used ordinary kriging to describe the spatio-temporal evolution of an influenza-like epidemic in France during the winter of 1989 and 1990. Counts of influenza cases seen by medical practitioners across the country were collated and a series of contour maps were produced demonstrating a north-to-south progression of the epidemic over time. More recently, Campos et al. (2002) used ordinary kriging (in

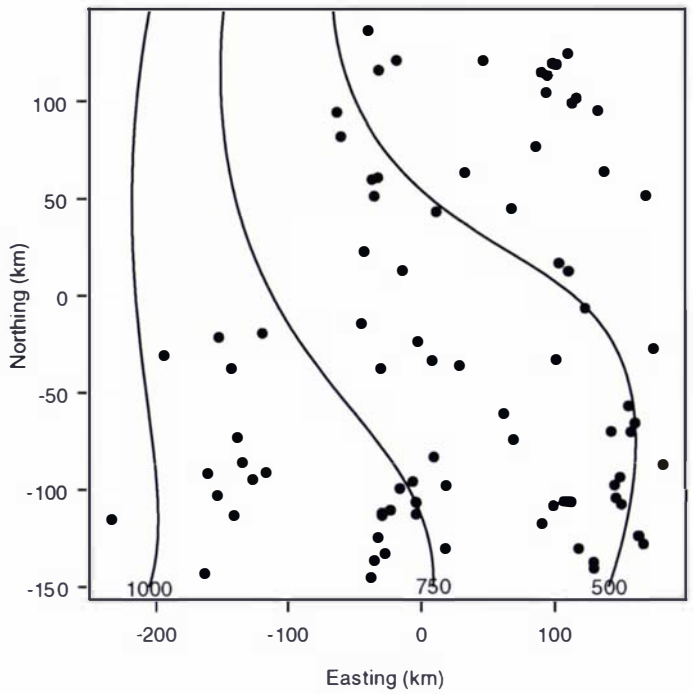


Figure 2.21: Contour plot showing variation in piezometric-head measurements throughout the Wolfcamp aquifer, computed using trend surface techniques.

addition to variogram analyses) to identify risk areas for *Ascaris lumbricoides* infection in selected census districts of Rio de Janeiro, Brazil. A sample of 1,664 children between 1 and 9 years of age from 19 census districts were tested for the parasite and a risk map produced, identifying high risk areas for infection.

The utility of kriging methods for epidemiologic applications shows some promise of advance as a result of ongoing work by Diggle and others (Diggle et al. 1998, Ribeiro Jr & Diggle 2001, Christensen et al. 2002). These authors have extended generalised linear mixed models to include spatially continuous covariance structures. In this context, Poisson log-linear geostatistical models of the form:

$$log(\mu_i) = log(E_i) + (\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}) + U_i + S(y_i) + \xi_i \qquad (2.20)$$

can be fitted. In Equation 2.20 $S(y_i)$ represents a structured (spatially correlated) heterogeneity term that is allowed to vary continuously through space (rather than discretely, as in the case of models generally used for lattice data) and is based on a zero mean Gaussian process with variance σ^2. The term U_i represents unstructured heterogeneity with a mean of zero and variance ν^2. The GeoR (Ribeiro Jr & Diggle 2001)

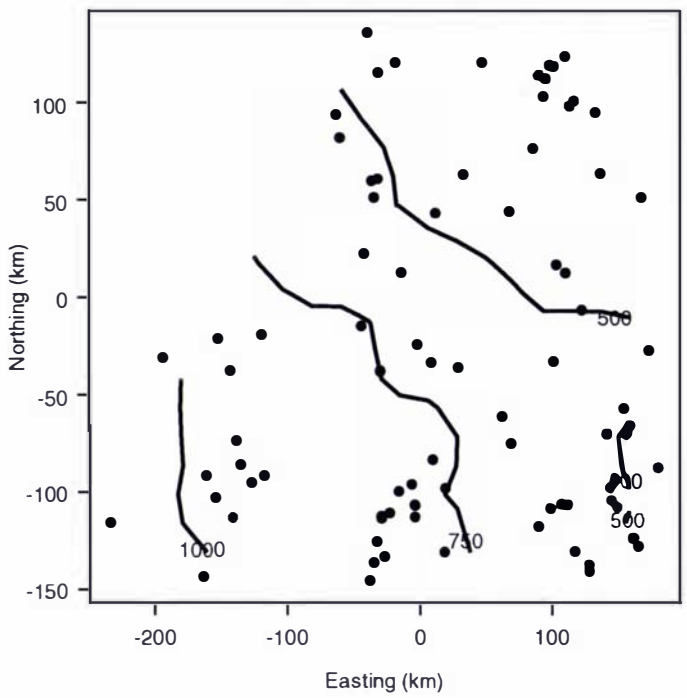


Figure 2.22: Contour plot showing variation in piezometric-head measurements throughout the Wolfcamp aquifer, computed using kriging.

and GeoRglm (Christensen et al. 2002) packages implemented within R (Ihaka & Gentleman 1996) provide functionality for fitting Poisson log-linear geostatistical models and binary logistic-linear geostatistical models, respectively. Diggle et al. (2002) applied a binary logistic-linear geostatistical model to a study of risk factors for childhood malaria in Gambia. Here the presence of malarial parasites in a blood sample was parameterised in terms of child-level covariates, village-level covariates and separate components for residual spatial and non-spatial extra-binomial variation. These authors concluded that the dominant component of extra-binomial variation in this data was spatially structured suggesting that unexplained risk of malaria was due to environmental factors rather than non-spatial factors (such as familial susceptibility).

2.6 Dynamic spatial models

The spatial models for point, lattice and continuous spatial data that have been described so far are empirical in the sense that the relationship between the outcome of interest and a set of explanatory variables is parameterised by a mathematical relationship and the influence of each explanatory variable included in the model is quantified by a regression coefficient, usually determined by maximum likelihood methods. Within this framework,

variables used to explain or predict the outcome of interest may have no biological relationship with the outcome of interest and may simply represent proxies for other (often more difficult to obtain) biologically-meaningful variables. As a result, models of this type tend to be applicable only for the defined area and time frame of investigation and are best applied to the environment in which they were first derived.

In contrast, dynamic models mathematically define the key features of the epidemiology of a disease process in order to simulate the spread of a disease condition throughout an animal population. Dynamic models may be: (1) deterministic, where the relationship between input and output variables is fixed, or (2) stochastic, where some or all of the various input parameters are sampled from distributions of possible values. In an outbreak situation, once parameter settings of a dynamic model are set to replicate the behaviour of an observed epidemic, various 'what if' scenarios can be applied to simulate the effect of alternative control strategies. Addition of a stochastic element provides a further advantage by providing and indication of the variability associated with model predictions. In veterinary science a stochastic modelling approach has been used to develop a decision support tool for the management of foot-and-mouth disease (Sanson et al. 1991, Sanson 1993, Morris et al. 2001, Morris et al. 2002) and to evaluate control strategies for African trypanosomiasis (Yu et al. 1995), classical swine fever (Jalvingh et al. 1999, Nielen et al. 1999, Mangen et al. 2001), bovine herpesvirus (Noordegraaf et al. 2000) and rabies (Thulke et al. 2000). The relatively large number of veterinary studies that have adopted a dynamic approach to disease modelling (rather than empirical) is noteworthy, and is most likely a reflection of: (1) the absence of widely-available methods for empirical disease modelling, and (2) the need for 'general purpose' models to support decision making.

2.7 Spatial clusters

In general usage, clustering refers to spatial aggregations of case events. Acknowledging that cases of disease diagnosed at a given location may simply be a function of the distribution of the population at risk and risk factors (both measured and unmeasured), a more robust definition might be that a disease is said to be clustered if there is residual spatial variation in risk after known influences have been accounted for (Wakefield et al. 2000).

Diseases may have many causes and because we never have information on all relevant risk factors, some degree of residual spatial variation in risk, and hence clustering, will be present. The important consideration is whether this clustering is epidemiologically significant and whether the data that is available allows this variation to be detected.

This section reviews commonly used techniques for identifying the presence and location of clusters of disease. Before doing so however, some terms should be introduced to help classify the different approaches. Firstly, cluster detection techniques may simply identify whether or not there is spatial aggregation in the data. These techniques are called global (or non-specific) clustering tests. Secondly, techniques may be focused meaning that the location of a cluster is identified *a priori* and the likelihood of that location truly being a cluster centre is determined. Non-focused cluster techniques are those where the most likely cluster centres are identified and ranked by the method. Further discussion of cluster detection techniques are provided by Lawson & Kulldorff (1999) and Wakefield et al. (2000). A review framed within a veterinary context is provided by Ward & Carpenter (2000a).

Clusters may be detected prospectively (as part of routine disease surveillance) or retrospectively. In both situations it is vital that the sensitivity and specificity of surveillance methods are assessed. Where clusters are detected retrospectively, aetiological clues of disease causation may be obtained by comparing exposures in areas of elevated risk. In terms of the residual risk surfaces developed it may also be beneficial to compare areas of low risk since these may correspond to protective factors or areas within which exposures are at low levels.

Unknown risk factors may be of various types including area-level environmental factors and factors influential at the herd, flock or individual animal level. This distinction is important since the nature of the clustering identified will be dependent on these risks. For example, environmental risk factors are commonly hypothesised to be responsible for clustering and may act through a variety of media (e.g. air, water, and soil). Point sources of air pollution produce exposures that may be approximated by a circular region. In this case distance and direction-based methods may be appropriate to detect the local scale clustering that may result (Lawson & Williams 1994, Lawson 1995). It is desirable to have a clustering hypothesis available *a priori* in which case a model-based approach may be adopted. As far as group or individual-level factors are concerned it is expected

that these will lead to broader-scale variability, since factors influential at this level may or may not be spatially dependent.

As with all epidemiological investigations it is essential to consider the quality of the data when a cluster is identified. Underascertainment will lead to a loss of power, but perhaps more importantly, may be spatially biased and may induce artefactual clustering. This may be on a broad geographical scale or be localised (perhaps reflecting the coding practices and diagnostic capabilities in individual practices and laboratories). Diagnostic inaccuracies may also lead to apparent spatial variation in risk, possibly at relatively local scales (associated with a local diagnostic laboratory, for example). This problem may be reduced to some extent by only choosing those diagnoses that are less prone to error or using multiple criteria to establish a case definition. The locations of cases and controls (or the population at risk) may not be known exactly, so location of events may be taken as being at the centroid of a post code or parish area. The resulting loss of spatial accuracy may lead to attenuation of clustering, particularly if the location of cases is known reliably, the location of controls has been estimated and the size of the areal units are relatively large in comparison to the size of the entire study area.

2.7.1 Traditional methods

In the following descriptions, the notation defined in previous sections of this review will be used. A hypothetical study region is divided into N non-overlapping areas with the observed number of cases in each area over a defined period given as O_i and the size of the population at risk as n_i. The expected number of cases for each area (as defined in Equation 2.8) is E_i. For rare and non-infectious diseases it is assumed that the number of cases of disease in area i is a function of a relative risk estimate ψ_i so that $O_i \sim Poisson(E_i\psi_i)$. The null hypothesis is $H_0 : \psi_1 = \psi_2 \cdots = \psi_N = 1$.

Pearson's chi-squared test

A simple test for global clustering is provided by Pearson's chi-squared test:

$$T = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{2.21}$$

Under the null hypothesis, T follows a chi-squared distribution with $N - 1$ degrees

of freedom. Large values of T result if there is heterogeneity in observed differences. The significance of T may be assessed by randomly simulating observations O_i under the null hypothesis and calculating the test statistic under each simulation. Comparison of the distribution of simulated test statistics with the observed value of T leads to a Monte Carlo test of significance. As with all classical hypothesis testing procedures the null hypothesis will be rejected for large sample sizes even for slight departures from H_0. Although Pearson's chi-squared test is a global test for clustering, choropleth maps showing large positive values of $O_i - E_i$ may help to identify cluster locations.

2.7.2 Adjacency methods

Adjacency methods may be used to indicate spatial dependence between counts (or points). As for Pearson's chi-squared test adjacency methods are global tests, identifying a general trend for the outcome variable to be spatially autocorrelated.

Autocorrelation statistics

If Z_i denotes the standardised mortality ratio for an area i and W_{ij} is a measure of the spatial association between areas i and j, Moran's I (Moran 1950) and Geary's c (Geary 1954) may be used to quantify spatial dependence. To formally test the magnitude of observed autocorrelation, the means and variances of I or c may be computed and a Monte Carlo test of significance performed. A major problem with these statistics (Besag & Newell 1991) is that the unequal variances of Z_i are not allowed for. As discussed in Section 2.4.2, a Moran's I statistic adjusted for population size may be used in this instance (Oden 1995). Ward & Carpenter (1995) investigated bluetongue virus infection in Queensland cattle herds using autocorrelation statistics.

Compared with autocorrelation methods, join count statistics (Ebdon 1985) provide a somewhat cruder method for identifying the presence of global clustering. The join count test statistic is based on comparison of the observed number of adjacent regions (joins) with dissimilar classifications (discordant joins) to the number of discordant joins expected for a spatially random pattern. The expected number is based on the total number of joins and the total number of regions with each classification. Hungerford (1991) and Hungerford & Smith (1996) investigated clustering of bovine anaplasmosis in Illinois, USA using join count statistics. These authors identified significant spatial clustering of

the disease, thought to be associated with woodland areas. Mannelli et al. (1998) used join count statistics to identify clustering of African swine fever-affected municipalities in Sardinia.

The K-function

As discussed in Section 2.3.2 the K-function difference, $D(s)$ can be regarded as a measure of the extra clustering among cases compared with the clustering of controls. For pre-defined distances, $s_1, \ldots, s_m$, an overall clustering test statistic is given by:

$$D = \sum_{k=1}^{m} w_k \hat{D}(s_k) \tag{2.22}$$

with weights $w_k = [var\hat{D}(s_k)]^{-1/2}$. The variance in this expression is evaluated under the null hypothesis of random re-labelling of cases and controls. Under the null hypothesis, Monte Carlo methods may be used to evaluate w_k and the sampling distribution of D. A plot of $\hat{D}(s)$ versus s with tolerance limits under random re-labelling is useful for assessing at which distances there are departures from random labelling (Figure 2.7).

Whereas the K-function has been widely used in the ecological literature, examples of its use in the medical and veterinary literature are relatively few. In the medical literature Morrison et al. (1998) used the K-function as part of a series of analyses of the spatial and temporal incidence of dengue fever cases in Florida during 1991 and 1992. Abernethy et al. (2000) used the K-function and random simulation envelopes to confirm the presence of spatial clustering of Newcastle disease affected poultry flocks in Northern Ireland during an outbreak that occurred in 1996 and 1997.

Given that the K-function performs poorly in the presence of first order spatial trends (Section 2.3.2) it is apparent that in an epidemiological setting it is a technique suitable for a relatively small range of conditions, being most informative where there is a generalised tendency for disease-positive locations to be located close to each other (as in the case of Newcastle disease and foot-and-mouth disease where there may be a high level of local, farm-to-farm spread). In contrast, the technique performs poorly where there is a single cluster centre within a study region. In this case strong spatial autocorrelation in one area is typically diluted by the absence of autocorrelation in other, disease-free areas.

2.7.3 Moving window and related methods

So called 'moving window' tests for clustering superimpose a number of circular evaluation areas over a study region and determine the significance of the number of cases that fall within each area. Evaluation areas may be defined in terms of the number of cases (Besag and Newell's method, Besag & Newell 1991) or population size (Kulldorff & Nagarwalla 1995, Kulldorff 1997).

Besag and Newell's method

Besag & Newell (1991) describe a method for identifying clusters of events, based on (in principle) Openshaw's 'geographical analysis machine' (Openshaw et al. 1987). Each case in a defined study region is considered in turn and a circle is drawn around that case such that the kth nearest case neighbour is included within the circle. The expected number of cases is calculated for each circle (given the size of the circle and the overall density of cases in the region as a whole) and those circles with an excess of cases are identified as clusters. Monte Carlo testing with random re-labelling of cases may be used to test significance, avoiding problems associated with performing multiple non-independent tests. A critical issue in this method is the choice of k, the cluster size parameter. The value of k must be large enough to identify real clusters of cases, not just isolated groupings. Conversely it must be small enough to permit the identification of distinct clusters within a study region. Values of k between two and five are common, representing cluster sizes of three to six cases including the centroid case. Besag & Newell (1991) recommend trying several k values and assessing the sensitivity of the findings to the choice of k.

Timmander & McLafferty (1998) used a modification of Besag and Newell's method to search for spatial clustering of breast cancer cases among long-term residents of West Islip, New York. This study identified no strong evidence of spatial clustering of breast cancer. Although four significant overlapping clusters were identified in the south central portion of the town, these were not evident after controlling for known risk factors for breast cancer (family history of the disease and age at first pregnancy).

Scan statistics

A spatial scan statistic has been proposed by Kulldorff & Nagarwalla (1995) and Kulldorff (1997). This method may be applied as either a focused or non-focused test and operates by imposing a series of circular windows around each of a series of possible cluster centroids positioned throughout a study region. For each centroid the radius of the window varies continuously in size from zero to an upper limit (expressed as a percentage of the total population at risk) creating an infinite number of distinct geographical circles with each being a possible candidate for a cluster. For each location and size of the scanning window, the alternative hypothesis is that there is an elevated rate within the window, compared with outside. Once the window with the greatest likelihood ratio statistic is identified the sampling distribution of the likelihood ratio is evaluated using a Monte Carlo test. The method allows for either a Poisson- or Benoulli-based model suitable for count or case-control data, respectively (Kulldorff et al. 1998). A limitation of the method is that it uses circular scanning windows resulting in low power to detect long, narrow clusters (such as those that may occur around a main road or river). Figure 2.23 shows, for the Lancashire laryngeal cancer data, the most likely cluster detected by the spatial scan statistic. Consistent with the analyses already presented, Figure 2.23 shows a non-significant, most likely cluster 1.5 kilometres in diameter adjacent to the hypothesised pollution source.

In the medical literature the spatial scan statistic has been applied to investigations of childhood leukaemia (Hjalmars et al. 1996), breast cancer (Kulldorff et al. 1997), soft-tissue sarcoma (Viel et al. 2000), breast cancer therapy (Gregorio et al. 2001, Gregorio et al. 2002), childhood mortality (Sankoh et al. 2001), and prostate cancer (Jemal et al. 2002). Viel et al. (2000) examined the spatial distribution of soft-tissue sarcomas in France to identify the most likely clusters around a municipal solid waste incinerator. While these authors did not control for known risks for soft-tissue sarcomas they concluded that the clustering of cases was unlikely to have been confounded by socioeconomic status, urbanisation, or patterns of medical referral. In the veterinary literature the technique has been used to investigate spatial clustering of sheep blowfly strike in Queensland, Australia (Ward & Carpenter 2000a), bovine spongiform encephalopathy in Switzerland (Doherr et al. 2000, Doherr et al. 2002), enzootic bovine leukosis in New Zealand (Teekayuwat et al. 2000) and bovine tuberculosis in Argentina (Perez et al. 2002).

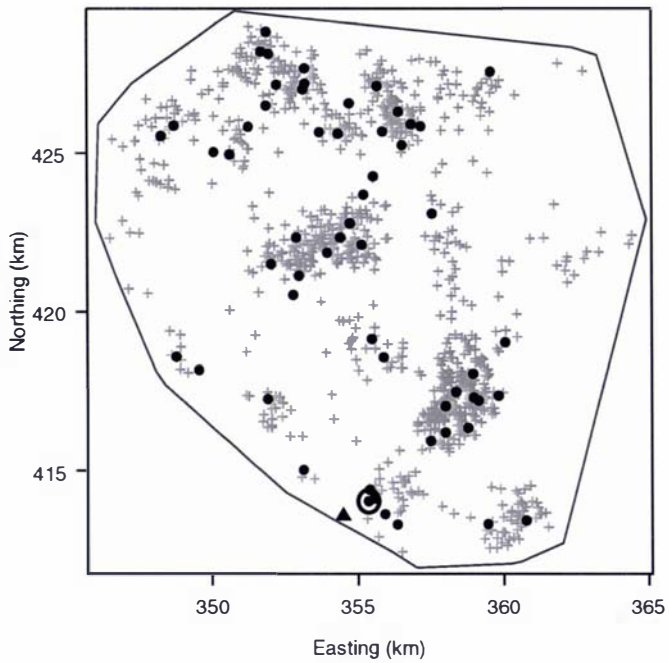


Figure 2.23: Easting and northing coordinates of the residence of laryngeal cancer cases (•) and lung cancer cases (+) diagnosed within a 400 km^2 area in Lancashire, England. The superimposed contour line delineates the boundaries of the most likely cluster identified by the spatial scan statistic (likelihood ratio statistic = 8.59; P = 0.158). The location of an industrial incinerator in the south of the study area is marked (▲).

Cuzick and Edward's method

Cuzick & Edwards (1990) developed a global clustering test based on case data and a sample of controls (non-cases). If there are $n = \sum_{i=1}^{N} O_i$ cases in a study area and $m_i(k)$ denotes the number of cases among the k nearest neighbours of case i, then a test statistic for global clustering is:

$$T_k = \sum_{i=1}^{n} m_i(k) \tag{2.23}$$

When population data are available a modification of T_k is the statistic:

$$U_k = \sum_{j=1}^{n} (O_j - E_j) \tag{2.24}$$

where circular regions are centred on each case and the radius of each circle are chosen so that E_j is taken to be as close to k as possible and O_j is the number of cases within each circle. Under the null hypothesis $E[U_k] = 0$ and the variance may be calculated. Cuzick & Edwards (1996) suggest that a range of nearest neighbour counts should be evaluated and that the overall significance level subject to a Bonferroni-type adjustment

(Simes 1986). Cuzick and Edward's test was used by Ward & Carpenter (1995) and Ward et al. (1996) to assess the clustering of bluetongue virus in Queensland cattle herds; by Rodríguez-Lainz et al. (1996) to examine clustering of papillomatous digital dermatitis in southern California dairy farms; and by Singer et al. (1998) to assess spatial clustering of *Pasteurella multocida* and *Pasteurella haemolytica* isolated from cattle in California.

2.7.4 Risk surface estimation

The methods described in this section are model-based and are potentially more flexible than the cluster detection methods that have been described so far. Many of the techniques have been described in earlier sections of this chapter, and here will only be discussed in the context of cluster detection.

Kernel smoothing

As discussed in Section 2.3.2, kernel density surfaces can be constructed for the spatial distribution of case and control populations and a risk surface computed. To identify areas of increased or decreased risk Bowman & Azzalini (1997) suggest plotting the standardised density difference as:

$$\frac{\sqrt{\hat{f}(y)} - \sqrt{\hat{g}(y)}}{\sqrt{\alpha(\omega)^2/(4nh^2)}} \tag{2.25}$$

In Equation 2.25 $\hat{f}(y)$ represents the density of cases, $\hat{g}(y)$ represents the density of non-cases, and the expression under the square root sign in the denominator is the bivariate version of the variance using a multivariate approximation. Locations where the standardised density difference exceeds two are defined as locations of excess or decreased risk. This method offers an advantage over moving window techniques in that it readily accommodates irregularly-shaped clusters (Figure 2.24).

Generalised additive models

Kelsall & Diggle (1998) show how generalised additive models (Hastie & Tibshirani 1990) may be used for spatial analyses, providing the ability to adjust for covariates. As discussed in Section 2.3.2, where the outcome of interest is a binary response variable (0 representing controls and 1 representing cases), a generalised additive model can

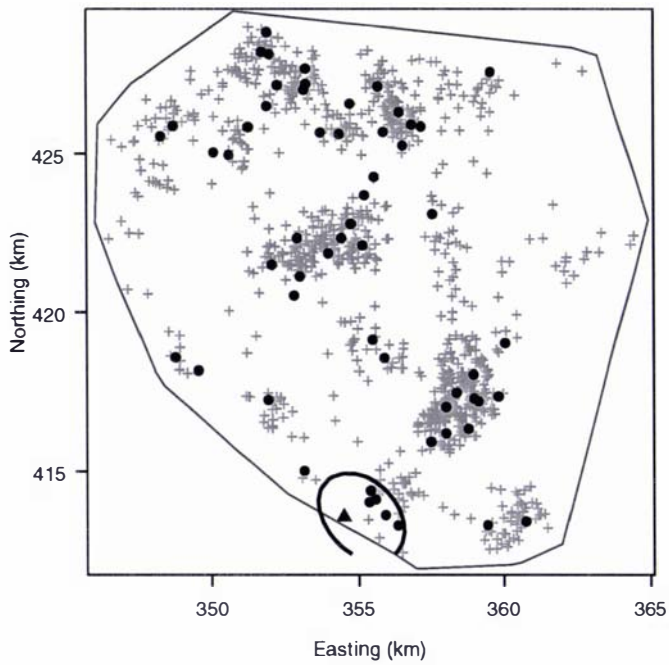


Figure 2.24: Easting and northing coordinates of the residence of laryngeal cancer cases (•) and lung cancer cases (+) diagnosed within a 400 km^2 area in Lancashire, England. The superimposed contour line delineates the boundaries of an area of excess disease risk, defined as locations where the standardised kernel density difference of cases and non-cases exceeds two in absolute value. This area is similar to that identified by the spatial scan statistic shown in Figure 2.23. The location of an industrial incinerator in the south of the study area is marked (▲).

be expressed as an additive logistic model as defined in Equation 2.7. With regard to cluster detection the residual risk surface $\hat{S}(y)$ may be examined for spatial variation and a Monte Carlo test of constant residual risk overall may also be performed (Kelsall & Diggle 1995b, Kelsall & Diggle 1998).

Disease mapping

The models proposed by Besag et al. (1991) (Section 2.4.3) may be used to investigate components of variability due to heterogeneity and clustering. From Equation 2.19 the terms U_i and S_i account for unstructured and structured (spatially correlated) heterogeneity, respectively. Values of S_i expressed in terms of relative risk may be mapped and tests of autocorrelation (for example, Moran's I, Section 2.4.2) applied to ascertain the significance of areas of residual spatial risk. Of the techniques discussed this approach is perhaps closest to the spirit of the definition of clustering provided by Wakefield et al. (2000): firstly it allows one to control for (and quantify the effect of) variables known to influence the risk of disease, and secondly one can test the significance of areas of

unexplained, spatially autocorrelated risk.

Extending the work of Ranta et al. (1996), Ranta & Penttinen (2000) used Bayesian hierarchical models to investigate spatial clustering of the incidence of insulin dependent diabetes mellitus in under 15 year-olds in Finland from 1987 to 1996. These authors demonstrated both spatial and temporal trends in the incidence of diabetes, identifying areas of the country where there was a significantly increased risk of disease. These authors were not able to support the hypothesis of a permanent environmental agent being causative and, because elevated risk was present at different locations at different times, they concluded that an infectious aetiology seemed likely.

2.8 Space-time interaction

Understanding the influence of the interaction between space and time extends spatial analyses into a further dimension and, with respect to infectious disease processes, helps to provide insight into the characteristics of subject-to-subject spread or 'contagiousness'. Techniques for assessing spatio-temporal clustering may be categorised as either global or non-global tests. Global tests identify the presence of spatio-temporal clustering in a data set; non-global tests assess the significance of explicitly-defined spatio-temporal cluster centres.

Ward & Carpenter (2000b) reviewed techniques available for detection of space-time clustering in a veterinary context, concluding that several techniques are best used in parallel to maximise analytical power.

2.8.1 Descriptive techniques

As outlined in Section 2.3.2, kernel smoothed maps of disease density estimated for a series of consecutive time intervals offer an informative means for describing the evolution of a disease process over space and time (Koussoulakou & Kraak 1992). Using similar logic, kriging has been used to estimate the underlying spatial process of an influenza-like epidemic in France (Carrat & Valleron 1992). A series of contour maps were produced, characterising the spread of the disease over time. Torok et al. (1997) used the same approach to characterise the incidence of rotavirus infections across the United States.

Time-to-event analyses conducted for a disease process (for example survival analy-

ses) can augment descriptive spatial analyses by providing a means of visualising subtleties in the temporal pattern of disease evolution. If an extraction surface defines an area of high risk (using for example, the methodology outlined in Section 2.7.4) a time-to-event analysis for the population present within the high-risk area can be compared with the population located outside of the area. Spatio-temporal clustering might be considered to be present if time to onset for subjects within the high-risk area is significantly shorter than that recorded for those outside. Although both of these approaches provide informative descriptive analysis of disease processes in terms of space and time, they do not quantify space-time interaction at any given location and time frame.

2.8.2 Spatio-temporal clustering

Knox and Mantel's test

Knox (1964) proposed a test for space-time interaction using critical space and time distances. Ordered pairs are formed from an observed set of N events. For each pair the distance and the time interval separating them is determined and, according to pre-defined cut points (s and t respectively), each event pair is allocated to one of four groups: close in space and close in time; close in space and far in time; far in space and close in time; and far in space and far in time. Knox's test statistic for spatio-temporal clustering is:

$$X = \sum_{i=1}^{N} \sum_{j=1}^{i-1} w_{ij} t_{ij} \tag{2.26}$$

Where w_{ij} is a spatial adjacency indicator equal to one if the distance between events i and j is $< d$ and zero otherwise, t_{ij} is a temporal adjacency indicator equal to one if the time between events i and j is $< t$ and zero otherwise. The null distribution of X is constructed under an approximate randomisation which permutes the row-column elements of the time adjacency matrix while holding the space adjacencies constant. This is equivalent to scrambling the time observations across localities, and calculating X each time. A P value is obtained by comparing the test statistic to this null distribution. The null hypothesis of no spatio-temporal clustering is rejected if Knox's X is 'unusually large' when compared with this theoretical distribution. Typically the mean distance and time interval between event pairs is used to determine the proximity measures, although it is recommended to assess the sensitivity of the choice of proximity by repeating the

analysis using a range of distance and time intervals.

An extension of the Knox test is Mantel's (1967) test which uses a continuous (rather than categorical) definition for spatial and temporal separation. Each event pair is taken in turn and a time and space separation measure is determined as follows:

$$x_{ij} = \frac{1}{(k_s + s_{ij})} \text{ and } y_{ij} = \frac{1}{(k_t + t_{ij})} \tag{2.27}$$

Where s_{ij} and t_{ij} are the Euclidean distance and time interval between events i and j and k_s and k_t are arbitrary constants chosen to allow for events that have identical space or time coordinates. The Mantel test statistic is computed as:

$$Z = \sum_{i \neq j} \sum x_{ij} y_{ij} \tag{2.28}$$

The null distribution of Z is constructed using a Monte Carlo method, permuting the elements of one of the distance matrices while holding the other constant. Similar to the construction of the null distribution for the Knox test, this is equivalent to scrambling the time observations across localities and calculating Z each time. A P value is obtained by comparing the test statistic to this null distribution.

White et al. (1989) used both the Knox and Mantel tests to confirm the presence of spatio-temporal clustering of winter dysentery outbreaks among Canadian dairy farms over a 2 year period. Carpenter et al. (1996) applied the Knox test to spatio-temporal clustering of fowl cholera in California over a 12 month period. Ward & Carpenter (2000b) used the Knox and Mantel test (in addition to Barton's test and nearest neighbour methods) to detect space-time interaction of blowfly catches in Tasmania, Australia. Tinline et al. (2002) categorised records of the time and place of identified raccoon rabies cases into a time-space matrix to estimate the distribution of the incubation period for the disease in this species. In a study of spatio-temporal clustering of scabies in Styrian chamois, Fuchs & Deutz (2002) used variogram analyses to select an appropriate distance cut point for the Knox test.

Space-time K-function

Diggle et al. (1995) describe a method for detecting space-time interaction using an extension of the K-function. Here they define $K(s, t)$ as the expected cumulative number

of cases that occur within distance s and time interval t of an arbitrarily-selected case, scaled by the expected number of cases per unit area and per unit time (λ). If an observed process operates independently in space and time (that is, there is no space-time interaction) then $\hat{K}(s,t)$ will be the product of two K-functions — one in space $\hat{K}_S(s)$ and one in time $\hat{K}_T(t)$. The observed difference:

$$\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}_S(s)\hat{K}_T(t) \tag{2.29}$$

provides a measure of the cumulative excess number of expected cases to be found from an arbitrarily-selected case that are attributable to the interaction between space and time, scaled by λ. In an epidemiological setting a plot of $\hat{D}_0(s,t) = \hat{D}(s,t)/\hat{K}_S(s)\hat{K}_T(t)$ is more readily interpretable, providing an estimate of the proportional increase in cases attributable space-time interaction.

Formal assessment of the significance of the observed values of $\hat{D}(s,t)$ may be made by performing m simulations in which each of the n recorded events are re-labelled with the observed n time 'markers'. A total of m estimates of $\hat{D}(s,t)$ is obtained and the observed sum of $\hat{D}(s,t)$ over all s and t is then compared with the empirical frequency distribution of the m estimates of $\hat{D}(s,t)$. An extreme value of the observed $\hat{D}(s,t)$ compared with this distribution provides evidence of significant space-time interaction.

In the medical literature the space-time K-function has been used in epidemiological investigations into Legionnaires' disease (Bhopal et al. 1992), Burkitt's lymphoma (Williams et al. 1978, Bailey & Gatrell 1995) and rheumatoid arthritis (Silman et al. 1997). French et al. (1999) and French et al. (2000) applied the technique to an investigation of sheep scab outbreaks in Great Britain from 1973 to 1992, demonstrating that disease occurrence in a selected area of the country was clustered in both space and time.

Scan statistics

In addition to the spatial scan statistic Kulldorff & Nagarwalla (1995) defined a space-time scan statistic for spatio-temporal clustering. Extending the explanation provided in Section 2.7.3, the space-time scan statistic uses a cylindrical window with a circular geographic base and height corresponding to time. This cylindrical window is moved so that for each possible geographical location and cluster size it also considers each possible time period. An infinite number of overlapping cylinders of different size and

shape are generated where each cylinder reflects a possible cluster. A likelihood ratio statistic is computed for all cylinders and the cylinder with the greatest likelihood ratio statistic constitutes the most likely spatio-temporal cluster. As described in Section 2.7.3 the distribution of the likelihood ratio statistic is evaluated repeating the same analytic process on a large number random of replications of the data set generated under the null hypothesis, in a Monte Carlo simulation.

Hjalmars et al. (1999) applied the space-time scan statistic to investigations of paediatric brain tumours in Sweden. In the veterinary literature the technique has been applied to investigations of acute respiratory disease in Norwegian cattle herds (Nörstrom et al. 2000, Nörstrom & Jarp 2000), blowfly strike in Australian sheep flocks (Ward 2000), tuberculosis breakdowns in New Zealand cattle herds (McKenzie 1999), and leptospirosis among dogs in the United States and Canada (Ward 2002). The reason for this technique's relative popularity (compared with other spatial analytical methods) is perhaps related to its ability to identify the specific location and time frame of most likely clusters — an important attribute for the investigation of point-source outbreaks.

2.8.3 Spatio-temporal modelling

A number of methods for modelling spatio-temporal patterns of disease incidence in small areas have been described including linear mixed-effects models (Pickle 2000), mixture models (Böhning et al. 2000) and Bayesian hierarchical models (Benardinelli, Clayton, Pascutto, Montomoli & Ghislandi 1995, Knorr-Held & Besag 1998, Waller et al. 1997, Xia et al. 1997, Sun et al. 2000). The Bayesian approach defines a Poisson distribution for the observed disease counts with a log linear model for the spatial and spatio-temporal components. Using the notation introduced in Equation 2.19, Benardinelli, Clayton, Pascutto, Montomoli & Ghislandi (1995) proposed:

$$log\,\mu_i(t) = log(E_i(t)) + (\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}) + \gamma t + S_{i1} + S_{i2}t \qquad (2.30)$$

where γ represents a coefficient for temporal trend, S_{i1} is a structured (spatially correlated) heterogeneity term and $S_{i2}t$ is a heterogeneity term representing the interaction between space and time. This model allows a temporal trend to vary across areas. Waller

Table 2.2: Summary of spatial epidemiology studies in the veterinary literature, classified by the primary analytical technique used.

Data type	Technique	Study
Point	Dot maps	Numerous
	K-functions	O'Brien et al. (1999), O'Brien et al. (2000), Abernethy et al. (2000), Foley et al. (2001)
	Deterministic models	Yu et al. (1995), Jalvingh et al. (1999), Nielen et al. (1999), Noordegraaf et al. (2000), Thulke et al. (2000), Mangen et al. (2001), Morris et al. (2001), Gerbier et al. (2002)
Lattice	Choropleth maps	Numerous
	Empirical Bayes smoothing	Berke (2001)
	Autocorrelation statistics	Perez et al. (1994)
Continuous	Variograms	Fuchs & Deutz (2002)
	Trend surface analyses	Moore (1999)
Clustering	Autocorrelation statistics	Hungerford (1991), Ward & Carpenter (1995), Hungerford & Smith (1996), Mannelli et al. (1998), Perez et al. (2002)
	K-function	Abernethy et al. (2000)
	Spatial scan statistic	Teekayuwat et al. (2000), Doherr et al. (2002), Ward & Carpenter (2000b), Perez et al. (2002)
	Cuzick and Edward's method	Ward & Carpenter (1995), Rodríguez-Lainz et al. (1996), Ward et al. (1996), Singer et al. (1998), Perez et al. (2002)
Spatio-temporal	Knox, Mantel tests	White et al. (1989), Carpenter et al. (1996), Ward & Carpenter (2000a), Fuchs & Deutz (2002), Tinline et al. (2002)
	Space-time K-function	French et al. (1999)
	Space-time scan statistic	McKenzie (1999), Nørstrom et al. (2000), Ward (2000), Ward (2002)

et al. (1997) considered an extension of this model where spatial random effects are nested within time such that for each time period t there are i structured heterogeneity terms. In describing this parameterisation, the authors considered lung cancer diagnoses in Ohio and demonstrated increased residual spatial clustering of annual lung cancer mortality over a 21 year period. Pickle (2000) used linear mixed effects models to describe spatial and temporal changes in breast cancer deaths across the USA, assuming that small area effects were random within larger regions. Pickle (2000) concluded that this approach was computationally less intensive than Markov chain Monte Carlo methods, while at the same time allowing the flexibility to include various spatial and temporal covariance structures.

In contrast to the variety of space-time model specifications suggested for small area

studies the literature on individual case-event space-time modelling is sparse. Lawson (2001a) provides some insight. Extending the model specified in Equation 2.3 Lawson defined a general framework for $\lambda(y, t)$ the intensity of cases at location y at time t:

$$\lambda(y,t) = \varrho \cdot g(y,t) \cdot f_1(y;\psi_y) \cdot f_2(t;\psi_t) \cdot f_3(y;t;\psi_{yt}) \tag{2.31}$$

where ϱ is a constant background rate (in space-time units), $g(y, t)$ is a modulation function describing the spatio-temporal change in the population at risk, f_k are appropriately-defined functions of space, time and space-time with ψ_y, ψ_t and ψ_{yt} parameters relating to the temporal, spatial, and spatio-temporal components of the model.

2.9 Conclusion

This review has outlined the major techniques available for describing and explaining spatial data in an epidemiologic context and has shown that the subject of spatial statistics is rapidly evolving with considerable progress made in recent years in the development of methodologies related to disease mapping and ecological analyses — particularly in the fields of hierarchical modelling and Bayesian analysis. Table 2.2 lists recently-published studies in the veterinary field that have used spatial data, categorised according to the primary analytical technique used. Clearly evident is a paucity of descriptive and explanatory (model-based) studies and a predominance of studies focused on cluster identification. Given the increasing availability of GIS packages, spatial data and statistical methodologies for analysing spatial data it is anticipated that this gap in the literature will be overcome in future.

Whereas new techniques provide greater flexibility and allow complex spatial models to be constructed, further questions arise regarding the robustness of chosen methods to mis-specification, the choice of priors used for model parameters and so on. While many new methods have been developed the challenge now is to evaluate their role and appropriateness across a broad range of applications. In addition, further challenges arise from the interrelated problems of induction periods of disease conditions and migration bias and their effect on spatial inferences. Whereas some disease events develop quickly after exposure to the causative agent (for example, clinical signs of foot-and-mouth disease occur four to seven days after exposure) others (for example a variety of cancers and

bovine spongiform encephalopathy) may take years to develop. Thus, while most spatial analyses work with single locations at single points in time, study subjects (both humans and animals) move frequently meaning that the current location of residence may bear little relationship to the environmental exposures experienced over the life span. Although (to a limited extent) these issues can be accounted-for during analysis, there is a need to devise systems that record details appropriate for a given location at a given time for all stages of the life path and to devise appropriate analytical methods that fully account for these details. In this sense a population is no longer geographically defined since it encompasses people and animals at different locations, at different points in time.

CHAPTER 3

Holding-associated risk factors for bovine spongiform encephalopathy in Great Britain

3.1 Introduction

Bovine spongiform encephalopathy (BSE) was first recognised as a pathological entity in November 1986 (Wells et al. 1987) and soon after that, cases began to be recognised throughout Great Britain. Geographically, marked variation in incidence has been observed, with the highest number of cases being diagnosed in the southern counties of England and the lowest in Scotland. The geographical pattern of incidence has been suggested to be a result of variation in the risk of exposure to BSE-contaminated meat and bone meal included in ruminant feedstuffs (Wilesmith et al. 1991). Whereas the existence of holding-associated risk factors such as holding type (dairy, beef suckler or mixed) and holding size have been alluded to in epidemiological analyses conducted to date (Wilesmith 1994, Wilesmith 1998), they have not been quantified against a temporally and spatially precise population at risk, nor has the important issue of confounding between these factors been adequately examined. The BSE epidemic in Great Britain is now clearly in decline, and given the accumulated epidemiological data it is timely that detailed analyses of the entire epidemic be undertaken in order to more fully understand its spatial and temporal features.

In this paper the temporal pattern of the BSE epidemic is analysed against the population of cattle holdings recorded on the annual agricultural censuses conducted for England, Scotland and Wales between 1986 and 1996. Case holdings were those that experienced at least one confirmed case of BSE up to 30 June 1997. Targeted slaughter of at-risk cattle after this date has created biases, which makes inclusion of later cases in an epi-

demiological analysis inappropriate. Survival analysis techniques were used to describe the temporal pattern of occurrence of the first confirmed BSE case on affected holdings, in relation to the population of holdings at risk. Cox proportional hazards regression was used to identify holding-associated risk factors that influenced the date of onset of clinical signs of a holding's first confirmed BSE case.

3.2 Materials and methods

Study population

In this study the unit of interest was every cattle holding that had details recorded on the annual agricultural censuses conducted by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998) in England and Wales and the Scottish Office, Agriculture, Environment and Fisheries Department, Scotland (SOAEFD 1987 – 1997) between 30 June 1986 and 30 June 1996. Census data for each holding included a unique identifier and the number of adult dairy and beef suckler animals present on 30 June of the respective census year. For holdings in England and Wales, census data was collected for all holdings employing at least one labour unit and a random sample of holdings employing less than one labour unit. In Scotland, census data was collected for all holdings.

On the basis of census data recorded over the entire study period, each holding was assigned a type classification according to the following algorithm. The total number of cattle recorded on each holding at each census was summed and where the number of dairy cattle exceeded 80% of the total, the holding was classified as dairy; where the number of suckler cattle exceeded 80% of the total, the holding was classified as beef suckler and those holdings fitting neither category were classified as mixed. A summary data set was constructed recording the region in which the holding was located, the date of the holding's first recorded census, the date of last observation (assumed to be 12 months after the last recorded census date), the mean holding size (referred to in the discussions that follow as 'holding size') and the calculated holding type.

The BSE database (Wilesmith, Ryan, Hueston & Hoinville 1992, Sanson & Ryan 1997) provided details of the outcome of interest for this study, the date of onset of clinical signs of the first confirmed case of BSE recorded for a holding. In this study we use the terms 'BSE-positive' to describe a holding that had at least one BSE case confirmed by

histopathology (Wells et al. 1989) and 'BSE-negative' to describe those holdings that either never reported a suspect BSE case or reported a suspect case but had none confirmed. The term 'BSE index case' is used to describe the first confirmed BSE case recorded on a holding and 'BSE onset date' identifies the date on which clinical signs were first observed for a BSE index case.

Those BSE-positive holdings which could not be matched with data present on either the MAFF or SOAEFD census rolls were presumed to be holdings employing less than one labour unit (for those located in England and Wales) or holdings where there was a discrepancy in the unique identifier recorded in the census database and the BSE database. For these BSE-positive holdings the BSE database was used to provide details of holding size and holding type. For these holdings it was assumed that the first census date was 30 June 1986.

Since BSE is thought to have occurred principally through exposure to the causative agent early in life (as a result of the feeding of meat and bone meal supplement in the period between birth and first calving) (Wilesmith, Ryan & Hueston 1992, Wilesmith, Ryan, Hueston & Hoinville 1992) cases recorded as purchased were included with data for the natal holding, if it was known. The BSE database provided details of the natal holding via a separate variable specifically to record the natal holding for purchased cases or via the herdmark descriptor of the animal's ear tag. Purchased cases where the natal holding was unable to be identified were included with the data for the holding of diagnosis.

Statistical analyses

Univariate and multivariate survival analysis techniques were used to identify factors influencing BSE onset date. Data for holdings were left-truncated up until their first recorded census date. Holdings that never experienced a case of BSE were right-censored 12 months after the date of their last recorded census. We investigated the influence of three holding-associated risk factors on BSE onset date: (1) geographical region in which the holding was located (Eastern, Mid and West, Northern, Scotland, Southeast, Southwest and Wales, as defined in Figure 3.1), (2) holding size (mean adult cattle numbers) and (3) holding type (dairy, mixed or beef suckler). Region and holding type were analysed as categorical variables and holding size was re-classified into quartiles and analysed as a categorical variable. The effect of each stratum of each hypothesised risk factor on

BSE onset date was assessed using the Kaplan-Meier technique in the LIFETEST procedure in SAS (The SAS System for Windows, Release 6.12, SAS Institute Inc., Cary, NC, USA). Kaplan-Meier survival curves for each stratum of a hypothesised risk factor were plotted and the homogeneity of the curves between strata was tested using the log rank statistic. Risk factors that showed an association with BSE onset date (that is, a difference in the Kaplan-Meier survival curves that was significant at $P < 0.20$) were selected for inclusion in the multivariate analysis.

Regression coefficients for the Cox proportional hazards model were determined using a forward-stepping approach using the PHREG procedure in SAS Release 6.12. The Efron approximation was used to handle tied event times (Efron 1977). The significance of the addition of explanatory variables into the model was tested using a likelihood ratio test. Those explanatory variables that were significant at $P < 0.05$ remained in the model. To verify that the proportional hazards assumption of the Cox model was valid for each prognostic variable a plot of $log[-logS(t)]$ against time was constructed for each stratum (Collett 1994). Covariate adjusted survival curves, based on the results of the Cox proportional hazards model, were computed using the modified estimated risk score approach described by Hosmer & Lemeshow (1999).

3.3 Results

Table 3.1 presents counts of the total number of holdings included in these analyses and cattle holding density (expressed as number of cattle holdings per 100 hectares), stratified by region. Table 3.2 presents counts of BSE-positive and BSE-negative holdings to 30 June 1997, stratified by region. In total 22,299 (74%) of the 30,042 BSE index cases were able to be located to their natal holding. The proportion of BSE index cases that were unable to be located to their natal holding differed significantly between regions (χ^2 294; df 6; $P < 0.01$) (Table 3.2).

Descriptive statistics of holding size of BSE-positive and BSE-negative holdings stratified by region are shown in Table 3.3. In all regions the median size of BSE-positive holdings was greater (by inspection) than the median size of BSE-negative holdings. Table 3.4 shows counts of BSE-positive and negative holdings stratified by holding type and region.

Figure 3.2 shows the cumulative proportion of cattle holdings that had not experienced a BSE index case, as a function of calendar time. Figure 3.3 shows the cumulative proportion of cattle holdings that had not experienced a BSE index case stratified by region. The regional proportions of holdings that were BSE-positive by 30 June 1997 (as shown in Figure 3.3) differ slightly from the result obtained by dividing the number of BSE-positive holdings by the total number of holdings present throughout the study period as a result of censoring. In the discussion that follows we use the Kaplan-Meier estimate to describe the final proportion of holdings BSE-positive by 30 June 1997.

Table 3.5 presents results of the Cox proportional hazards regression for BSE onset date. Based on this regression analysis, Figure 3.4 shows the cumulative proportion of cattle holdings that had not experienced a BSE index case stratified by region and adjusted to account for the effect of mean holding size and holding type. Figure 3.5 shows the cumulative proportion of cattle holdings that had not experienced a BSE index case stratified by mean holding size and adjusted to account for the effect of location and type. Figure 3.6 shows the cumulative proportion of cattle holdings that had not experienced a BSE index case stratified by holding type and adjusted to account for the effect of location and mean holding size.

After adjusting for the effect of holding size and holding type, holdings in the East, Southeast and Southwest of England (respectively) had 2.2 (95% CI 2.1 – 2.4), 2.4 (95% CI 2.3 – 2.6) and 2.4 (95% CI 2.3 – 2.5) times the monthly hazard of having a BSE index case compared with holdings in the reference category (those located in Scotland). All other regions had between 1.7 and 1.8 (95% CI 1.6 – 1.9) times the monthly hazard of having a BSE index case compared with holdings located in Scotland.

After adjusting for the effect of region and holding type, those holdings with greater than 53 adult cattle had 5.9 (95% CI 5.6 – 6.2) times the monthly hazard of having a BSE index case compared with holdings in the reference category (those with 7 – 21 adult cattle). Dairy holdings had 3.1 (95% CI 3.0 – 3.2) times the monthly hazard of having a BSE index case compared with beef suckler holdings.

Table 3.1: Counts of the total number of holdings present in Great Britain between 30 June 1986 and 30 June 1997, estimates of the area (in hectares) of the seven regions of Great Britain, and holding density (expressed as the number of holdings per 100 hectares).

Region	Total holdings included	Area (hectares $\times 10^6$)	Holdings per 100 hectares
Eastern	5978	2.4	0.25
Mid and West	23116	2.6	0.89
Northern	16393	2.8	0.59
Scotland	19285	7.8	0.25
Southeast	7779	1.5	0.52
Southwest	24661	2.4	1.03
Wales	20845	2.0	1.04
Total	118057	22	0.54

Table 3.2: Counts of BSE-positive holdings confirmed up to 30 June 1997, counts of BSE-negative holdings to 30 June 1997 and counts of the total number of cattle holdings recorded on MAFF and SOAEFD censuses between 30 June 1986 and 30 June 1996, stratified by region. BSE-positive holdings have been classified according to whether or not the natal holding of the index case was known. Percentages shown in brackets refer to the number in each group as a proportion of total holdings in the same region.

Region	BSE-positive		BSE-negative	Total holdings
	Natal known	Natal unknown		
Eastern	862 (14%)	329 (6%)	4787 (80%)	5978 (100%)
Mid and West	5293 (23%)	1606 (7%)	16217 (70%)	23116 (100%)
Northern	3268 (20%)	1126 (7%)	11999 (73%)	16393 (100%)
Scotland	1648 (9%)	1002 (5%)	16635 (86%)	19285 (100%)
Southeast	1610 (21%)	479 (6%)	5690 (73%)	7779 (100%)
Southwest	6528 (26%)	1940 (8%)	16193 (66%)	24661 (100%)
Wales	3090 (15%)	1261 (6%)	16494 (79%)	20845 (100%)
Total	22299 (19%)	7743 (7%)	88015 (75%)	118057 (100%)

Table 3.3: Descriptive statistics of holding size for BSE-positive and BSE-negative holdings, stratified by region. This table is based on MAFF and SOAEFD census data recorded between 30 June 1986 and 30 June 1996 and data derived from the BSE database, to 30 June 1997.

Region	BSE-positive holdings					BSE-negative holdings				
	n	Mean	SD	Median	Q1, Q3	n	Mean	SD	Median	Q1, Q3
Eastern	1191	79	66	67	33,104	4787	18	28	8	2, 21
Mid and West	6899	81	59	69	44,102	16217	24	30	12	4, 32
Northern	4394	72	51	61	40,92	11999	26	30	16	5, 35
Scotland	2650	93	62	84	52,120	16635	36	43	20	5, 52
Southeast	2089	94	72	80	44,124	5690	19	29	9	3, 23
Southwest	8468	81	63	68	43,102	16193	21	31	10	3, 28
Wales	4351	61	46	51	32,78	16494	20	23	12	4, 28
Total	30042	79	59	67	41,101	88015	25	33	13	4, 33

SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.

Table 3.4: Counts of BSE-positive and BSE-negative holdings stratified by holding type and region. This table is based on MAFF and SOAEFD census data recorded between 30 June 1986 and 30 June 1996 and data derived from the BSE database, to 30 June 1997.

Region	BSE-positive holdings				BSE-negative holdings			Total
	Dairy	Mixed	Suckler	Unknown [a]	Dairy	Mixed	Suckler	
Eastern	669	71	397	54	553	338	3896	5978
Mid and West	5498	427	828	146	4666	1531	10020	23116
Northern	2893	407	984	110	2472	1173	8354	16393
Scotland	1190	242	1195	23	1736	912	13987	19285
Southeast	1378	139	495	77	830	459	4401	7779
Southwest	6677	580	1079	132	3888	1617	10688	24661
Wales	2886	355	1074	36	3240	1704	11550	20845
Total	21191	2221	6052	578	17385	7734	62896	118057

[a] Holdings not present on MAFF or SOAEFD census rolls where no type description was given in the BSE database.

Table 3.5: Cox proportional hazards model showing the effect of region, holding size and holding type on the monthly hazard of experiencing a BSE index case.

Explanatory variable	Holdings	BSE-positive	Coefficient (SE)	P	Hazard	95% CI
Region comparison				< 0.01 [a]		
Eastern	5924 [b]	1137 [b]	0.7981 (0.0359)		2.2 [c]	2.1 – 2.4
Mid and West	22970	6753	0.583 (0.0241)		1.8	1.7 – 1.9
Northern	16283	4284	0.5632 (0.0254)		1.8	1.7 – 1.8
Scotland	16635	2627			1.0	
Southeast	7702	2012	0.8865 (0.0302)		2.4	2.3 – 2.6
Southwest	24529	8336	0.8746 (0.0234)		2.4	2.3 – 2.5
Wales	20809	4315	0.5127 (0.0256)		1.7	1.6 – 1.8
Holding size comparison				< 0.01		
1 – 6	31469	644	-0.8344 (0.0468)		0.43	0.40 – 0.48
7 – 21	25142	2021			1.0	
22 – 53	29145	8479	1.061 (0.0257)		2.9	2.7 – 3.0
> 53	29702	18320	1.7758 (0.0254)		5.9	5.6 – 6.2
Holding type comparison				< 0.01		
Dairy	38576	21191	1.117 (0.0168)		3.1	3.0 – 3.2
Mixed	9955	2221	0.5462 (0.0253)		1.7	1.6 – 1.8
Beef suckler	62896	6052			1.0	

Likelihood ratio test statistic 35570; df 11; P < 0.01.

[a] The significance of inclusion of the six region variables in the model.

[b] Cases with missing values have been excluded, so counts vary slightly from those shown in Table 3.4.

[c] Interpretation: compared with the reference category (holdings located in Scotland), after adjusting for the effect of holding size and holding type, holdings located in the Eastern region of England had 2.2 (95% CI 2.1 – 2.4) times the monthly hazard of having a BSE index case.

SE: standard error.

CI: confidence interval.

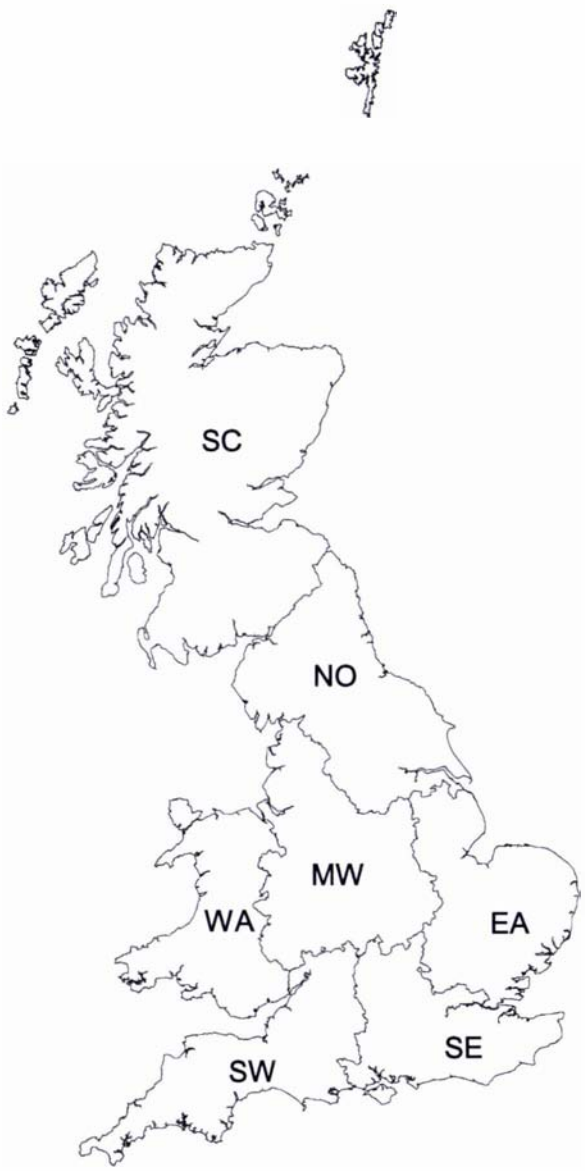


Figure 3.1: Map of Great Britain showing location of regions described in this study. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

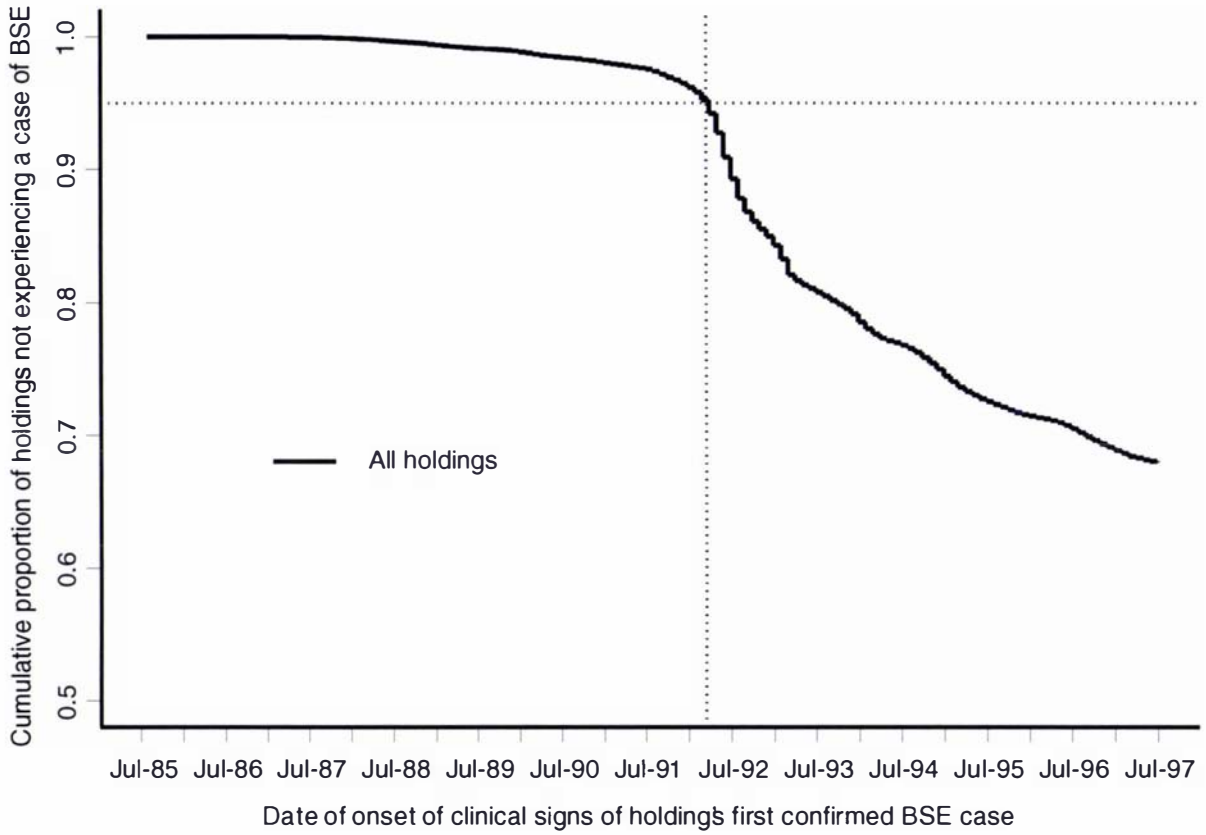


Figure 3.2: Kaplan-Meier survival curve showing the cumulative proportion of cattle holdings that had not experienced a BSE index case as a function of calendar time. The reference line indicates when 5% of all cattle holdings in Great Britain were confirmed with BSE (February 1992). This survival curve has been calculated using the population of cattle holdings estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997.

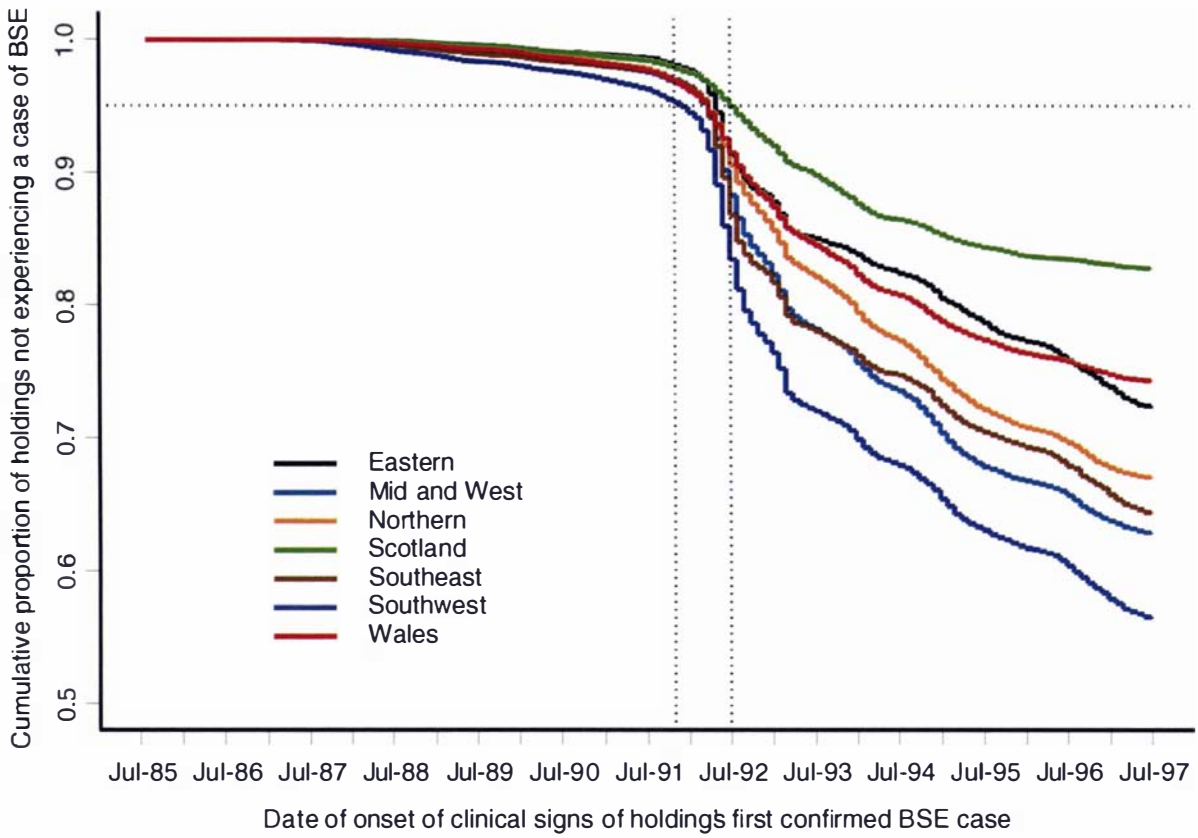


Figure 3.3: Kaplan-Meier survival curve showing the cumulative proportion of cattle holdings that had not experienced a BSE index case as a function of calendar time, stratified by region. Dates when 5% of a region's holdings were confirmed with BSE (in order) were: Southwest (November 1991, indicated by the left reference line), Mid and West (February 1992), Northern (February 1992), Southeast (February 1992), Wales (February 1992), Eastern (March 1992), and Scotland (June 1992, indicated by the right vertical reference line). These survival curves have been calculated using the population of cattle holdings estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997.

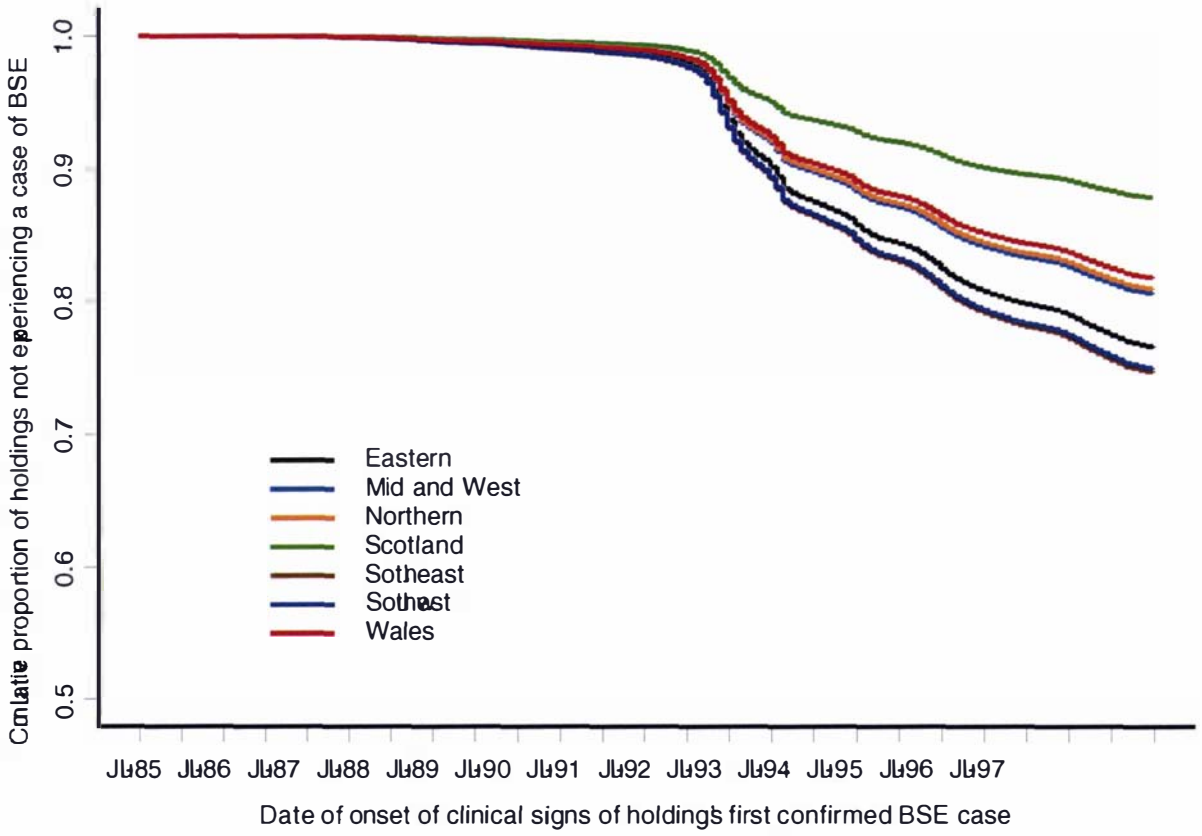


Figure 3.4: Kaplan-Meier survival curve showing the cumulative proportion of cattle holdings that had not experienced a BSE index case as a function of calendar time, stratified by region. These survival curves have been calculated using the population of cattle holdings estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of mean holding size and holding type, on the basis of the Cox proportional hazards regression model shown in Table 3.5.

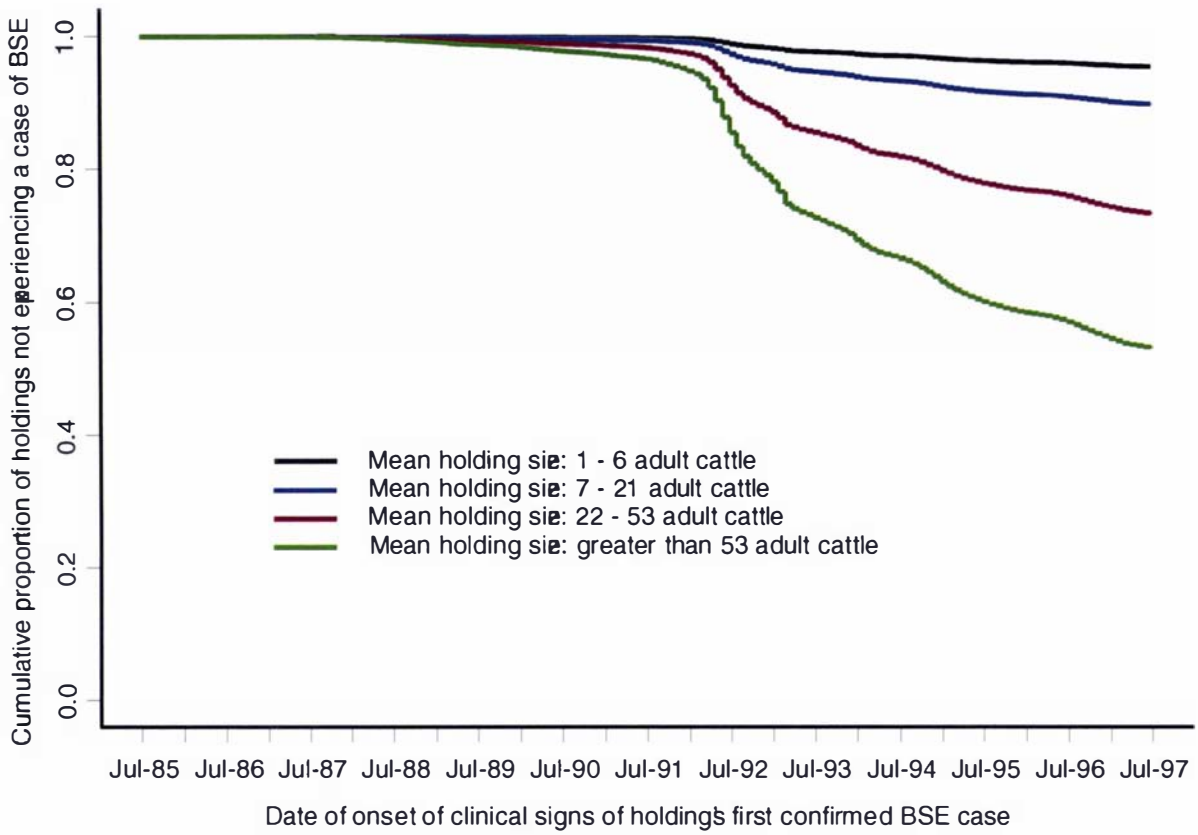


Figure 3.5: Kaplan-Meier survival curve showing the cumulative proportion of cattle holdings that had not experienced a BSE index case as a function of calendar time, stratified by holding size classification. These survival curves have been calculated using the population of cattle holdings estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of holding location and holding type, on the basis of the Cox proportional hazards regression model shown in Table 3.5.

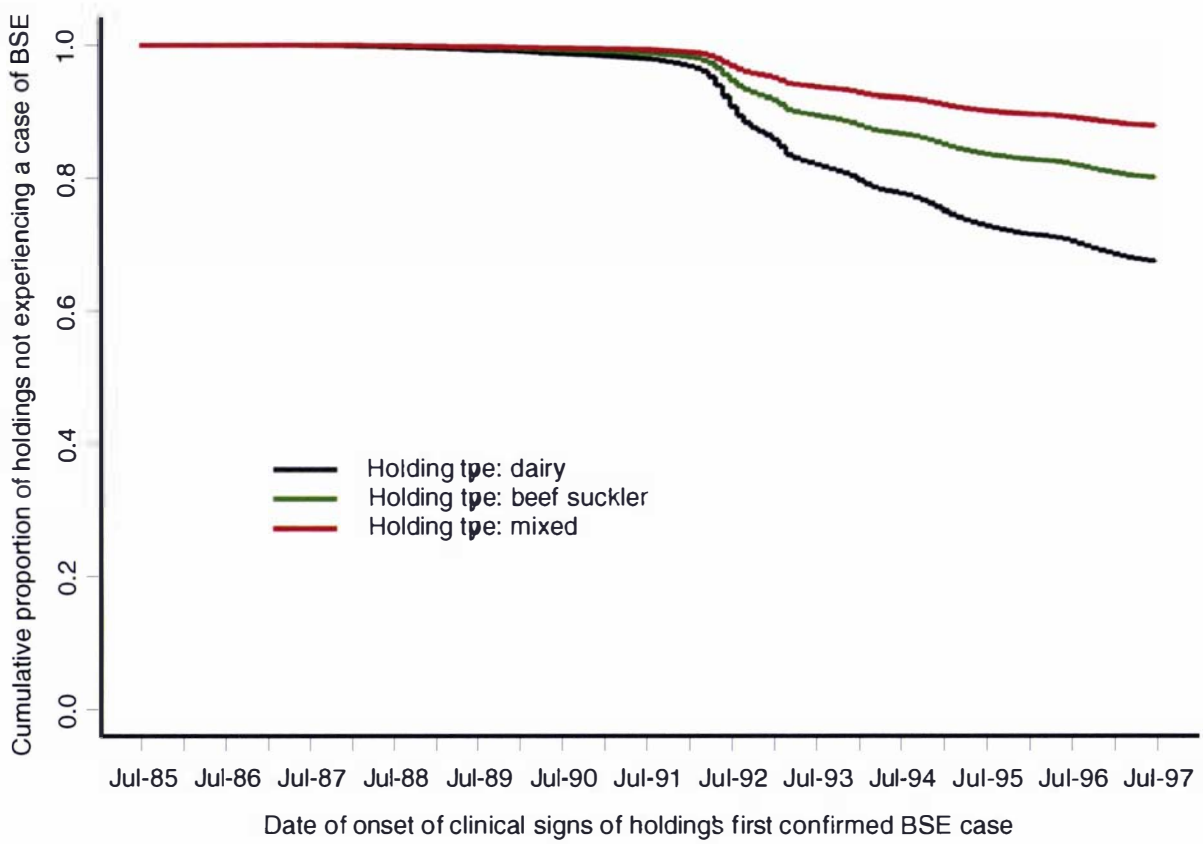


Figure 3.6: Kaplan-Meier survival curve showing the cumulative proportion of cattle holdings that had not experienced a BSE index case as a function of calendar time, stratified by holding type classification. These survival curves have been calculated using the population of cattle holdings estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of holding location and mean holding size, on the basis of the Cox proportional hazards regression model shown in Table 3.5.

3.4 Discussion

The BSE epidemic in Great Britain has been the subject of continuous epidemiological analysis to identify and improve the understanding of the major risk factors for disease and to assess the effects of the various control measures instituted (see for example Wilesmith, Ryan, Hueston & Hoinville 1992, Wilesmith 1996). This is the first of a sequence of studies in which the accumulated data on the BSE epidemic is analysed in conjunction with demographic data from the national population at risk.

This study was based on 118,057 cattle holdings situated in England, Scotland and Wales. Of these 30,042 had at least one confirmed case of BSE diagnosed up to 30 June 1997. Relating numbers of BSE-affected holdings to the size of the total population of cattle holdings at risk allows risk factors for the presence of the infective agent in a holding (such as type and size) to be quantified and enables comparisons to be made regarding the temporal progression of the epidemic in each region. In this study the unit of interest was the holding (rather than the individual animal) and as such, holdings that experienced a single confirmed case received the same analytical weighting as those that had multiple cases. This simplifies consideration of when holdings were first exposed to the BSE agent, uncomplicated by factors which may have influenced whether or not they had further cases.

Analysis of the geographical distribution of the earliest recorded cases of BSE revealed a remarkably contemporaneous occurrence throughout Great Britain with no evidence of an obvious initial focus (Wilesmith 1991b). However, a marked geographical heterogeneity in incidence was observed in the early stages of the epidemic and this has persisted throughout (Wilesmith et al. 1988, Wilesmith 1998). The results of the present study indicate that the temporal evolution of the BSE epidemic differed markedly between regions with different degrees of epidemic propagation. Five percent of holdings in the Southwest had experienced a BSE index case by November 1991 (Figure 3.3) and by 30 June 1997 44% of this region's holdings were BSE-positive. The Southwest of England had the greatest proportion of holdings affected, of all regions. Collectively, the Southeast, Mid and West and Northern regions showed similar temporal patterns of onset throughout the epidemic, all reaching 5% of holdings affected by February 1992. Their rates of breakdown were similar to the Southwest throughout the epidemic, and by 30 June 1997 in the three regions 36%, 37% and 33% of holdings were BSE-positive, respectively.

Wales and the Eastern region of England shared similar temporal patterns of onset (Figure 3.3). In these regions 5% of holdings were BSE-positive by February 1992 and March 1992, respectively. By 30 June 1997 26% and 28% of holdings were BSE-positive, respectively. In Scotland 5% of holdings were BSE-positive by June 1992 and by 30 June 1997 17% of holdings were BSE-positive. The unadjusted Kaplan-Meier survival curves shown in Figure 3.3 show consistent differences between the groups of regions throughout the entire epidemic, reflecting clear differences in the capacity for epidemic propagation, and possibly differences in the initial seeding of infection. Interestingly, after controlling for the influence of size and type, holdings in the south of England (that is, those located in the Southwest, the Southeast and Eastern regions) had similar (and the highest) monthly hazards of becoming BSE-positive (Table 3.5 and Figure 3.4). After removing the confounding influence of holding type and holding size, these results show that BSE onset date was likely to occur sooner in holdings situated in the south of England. Further analysis is required to clarify whether this south-to-north gradient in holding infection hazard was the result of transmission and amplification mechanisms or more simply due to variation in the scale of infection challenge.

Following the introduction of the primary control measure, a ban on the feeding of ruminant-derived protein to ruminants in July 1988 (HMSO 1988a, HMSO 1988b), the epidemic was monitored closely to assess the effectiveness of this intervention. Beneficial effects were observed (Wilesmith & Ryan 1992, Wilesmith & Ryan 1993, Hoinville 1994). Subsequent analyses of the geographical occurrence of cases in animals born after the July 1988 ban revealed that the response to the ban was less-marked in the Eastern region of England. This was attributed to the greater risk of cross-contamination with ruminant-derived meat and bone meal in feed mills in this region. This was because of the relatively much greater quantities of pig and poultry feedstuffs manufactured, as this region contains high densities of pigs and poultry and these species could legitimately receive ruminant derived protein (Wilesmith 1996). These findings were confirmed by a survey of finished feedstuffs using an ELISA to detect species-specific proteins (MAFF, unpublished findings). It is interesting therefore to note the increase in the holding confirmation rate of BSE in the Eastern region of England in 1996 – 1997 (Figure 3.3) which reflects this problem of cross-contamination in feed mills. This finding also reflects the incomplete compliance to the specified bovine offal ban introduced in September 1990

(HMSO 1990) which was intended to prevent inclusion of meat and bone meal derived from high risk bovine tissues in pig and poultry feedstuffs. Although individual animal data (rather than holding-level data) provides more accurate insights into the effectiveness of control measures, Figures 3.2 and 3.3 illustrate a decline in the rate of infection among holdings with a decrease in the slope of the survival curves evident 5 years after the introduction of control measures.

The high proportion of BSE index cases that could not be returned to their natal herd in Scotland (38% of all BSE index cases for this region, Table 3.2) reflects the fact that a large proportion of BSE cases that occurred in Scotland were in purchased animals, many of which were of unknown origin (Wilesmith, Ryan, Hueston & Hoinville 1992). A proportion of these would have been born within Scotland but most would have been born further south. As a result, the proportion of BSE-positive holdings reported for Scotland is probably overestimated in this analysis.

The origin of the exposure of the British cattle population to a transmissible spongiform encephalopathy agent is not the subject of this paper, but the working hypothesis for subsequent evolution of the epidemic is that the scale of the observed clinical BSE epidemic was the result of the recycling of BSE infected tissues via meat and bone meal back to cattle (Wilesmith & Wells 1991). Using this hypothesis, the rate of progression of the regional epidemics and the ultimate cumulative proportion of holdings affected is thought to be determined by three key factors. The first is the relative risk of cattle (particularly those in the late stage of incubation) being rendered and the resultant meat and bone meal being incorporated into cattle feedstuffs. Within a region the relative contributions that cattle make to the production of meat and bone meal and to the size of the cattle population relative to other meat and bone meal-consuming species (pigs and poultry) are therefore important. The second factor is the effect of the rendering system in reducing the infectivity of the resulting meat and bone meal. Laboratory-based studies have indicated that none of the basic rendering systems in Europe are capable of disinfecting tissues during rendering (Taylor et al. 1995). However, the solvent extraction process, albeit only partially simulated in a laboratory study, has been found to be the most effective, of the methods examined, in reducing the titre of the BSE agent (Taylor et al. 1998). The original survey of rendering plants in Great Britain had previously identified a reduction in the use of hydrocarbon solvents to maximise the extraction of tallow (fat) as a possi-

ble precipitating factor for the sudden exposure of the cattle population to a transmissible spongiform encephalopathy agent (Wilesmith et al. 1991). This survey of rendering plants also provided some explanations for the variation in geographical risk. These were the continuation in the use of hydrocarbon solvent in Scotland, and geographical variation in the proportion of meat and bone meal which had been produced as a result of double treatment, in the reprocessing of greaves. The third factor influencing the rate of progression of the regional epidemics and the ultimate cumulative proportion of holdings affected is the degree of effectiveness of the control measures which were imposed.

Given these factors, the greatest propagation of infection among holdings in the south of the country can be explained in part by: (1) the high density of cattle in the region, relative to other species (infected tissues from cattle were more likely to be converted to meat and bone meal and fed back to cattle in the Southwest compared with other regions), (2) changes in the use of solvent extraction during the amplification phase of the epidemic, and (3) the relatively small proportion of meat and bone meal produced in the region which had been subject to double heat treatment. An alternative, or possible complimentary explanation is that the epidemic commenced in the south. The contemporaneous occurrence of BSE cases throughout Great Britain is not, at first sight, consistent with this hypothesis, but this will be examined in later papers in this series.

An advantage of a multivariate analysis of this type is that the contributions of the various interacting risk factors for disease can be distinguished and quantified. In agreement with previous analyses (Wilesmith 1998) the hazard of a holding being BSE-positive depended on holding type and holding size. Holdings with greater than 53 adult cattle had 5.9 (95% CI 5.6 – 6.2) times the monthly hazard of having a BSE index case compared with holdings with 7 to 21 adult cattle (Table 3.5 and Figure 3.5). Dairy holdings had 3.1 (95% CI 3.0 – 3.2) times the monthly hazard of having a BSE index case compared with beef suckler holdings (Table 3.5 and Figure 3.6). In the context of the current knowledge of the epidemiology and pathogenesis of BSE these findings are logical: holdings of larger size were more likely to use significant amounts of compound feeds and holdings of larger size had, on average, a greater chance of an infected animal surviving long enough for clinical signs to be expressed. Controlling for the effect of location of the holding and holding size, dairy holdings had a higher monthly hazard of having their first BSE case confirmed compared with beef suckler holdings (Table 3.5 and Figure 3.6). This reflects

the greater propensity for compound feeds to be used in dairy holdings which are more intensively managed, compared with other enterprise types.

Because census data were used to provide details of the population at risk throughout the epidemic, these analyses have quantified factors that influenced BSE-onset date for cattle holdings in Great Britain to 30 June 1997. These analyses demonstrate different rates of onset by region, holding type and holding size and show that the epidemic propagated most strongly in the south of England. These analyses show that the growth of the epidemic followed essentially the same pattern in each region of the country, with modest temporal lags between each. The Eastern region showed evidence of additional later recycling of infection, consistent with the known problems of cross-contamination, which resulted in a slower response to the imposition of control measures. The control measures imposed in 1988 and 1990 brought the expansion of the epidemic under control, albeit more slowly in regions in which controls took some time to take full effect. Further analyses in this series will consider incidence at the individual animal level, and examine specific issues in the early stages of the epidemic.

CHAPTER 4

Individual animal risk factors for bovine spongiform encephalopathy in Great Britain

4.1 Introduction

Bovine spongiform encephalopathy (BSE) is a progressive fatal disorder of cattle first identified in November 1986 by the Central Veterinary Laboratory, Weybridge (Wells et al. 1987). Epidemiological studies (Wilesmith, Ryan, Hueston & Hoinville 1992) indicate that primary vehicle of BSE transmission was meat and bone meal prepared from animal material containing the infective agent. Rendering of subclinically infected cattle from 1982 to 1988 and inclusion of infective tissues (central nervous system and lymphatic tissue) into meat and bone meal fed to cattle caused widespread infection in the British cattle population during this period. Since cattle typically received meat and bone meal as part of calfhood feeding practices, the risk of exposure to the BSE agent has been considered to be birth cohort-dependent (Wilesmith & Ryan 1992) with those born between 1982 and 1988 subject to increasing levels of exposure.

While counts of the numbers of confirmed BSE cases have been regularly reported (MAFF 1999), descriptions of the epidemic in terms of the number of confirmed cases per head of total and demographically segmented populations at risk have been limited (Wilesmith et al. 1991, MAFF 1999). The objectives of the present study were to use agricultural census data collected between 1986 and 1996 as a basis for generating birth dates and disposal dates for the British adult cattle population present throughout this period. Using this population of animals as an estimate of the adult British cattle population at risk between 1986 and 1997 we sought to: (1) describe the cumulative incidence of BSE from July 1986 to June 1997, (2) identify individual animal-associated risk factors

that influenced the age-at-onset of clinical signs in confirmed cases, and (3) quantify the effectiveness of the various measures applied to control BSE throughout the epidemic.

4.2 Materials and methods

Study population

The period of interest in this study was from 30 June 1986 to 30 June 1997 and the unit of interest was adult cattle present in Great Britain during this time. To create an age-structured population we used data recorded at the annual agricultural censuses conducted by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998) in England and Wales and the Scottish Office, Agriculture, Environment and Fisheries Department, Scotland (SOAEFD 1987 – 1997) between 30 June 1986 and 30 June 1996. Census data provided for each cattle holding a count of the number of adult dairy and (beef) suckler animals present on 30 June of the respective census year. For holdings in England and Wales, census data was collected for all holdings employing at least one labour unit and a random sample of holdings employing less than one labour unit. In Scotland census data was collected for all holdings. Under MAFF and SOAEFD definitions, adult cattle were those over 2 years of age.

The following method was used to estimate the birth and disposal dates of all adult cattle that existed in Great Britain throughout the study period. The first census recorded for each holding (Y) was taken and the count of adult cattle present at that date was divided into counts of cows within each age group (from 2 to 13 years). Based on the age categorisation allocated at the first census date, a birth date was assigned (assumed to be 1 July for the calculated birth year). Age structures were assumed to be the same for all holdings and were based on the survey results of Pasman et al. (1995). At the end of the first census year cattle were removed on the basis of age-specific disposal hazards described by Pasman et al. (1995). The number of adult cattle in the second census year ($Y + 1$) was noted and where the ratio of cattle numbers in year (Y) to ($Y + 1$) was equal to or less than 1.0 (that is, reported holding size was static or increasing) additional 2 year old animals were introduced into the holding to maintain or increase holding size. Where the ratio of (Y) to ($Y + 1$) was greater than 1.0 (that is, holding size was decreasing) the age-specific removal hazard for each age group was multiplied by this ratio to evenly

reduce numbers across age groups and additional 2 year old animals were introduced into the holding (if necessary) to create a holding size equal to that recorded in year $(Y + 1)$. This cycle of introducing and disposing of cattle was repeated for each sequential census year for each holding. At the final census it was assumed that the holding was disbanded and all animals were assigned a disposal date 12 months after the date of the last census (that is, when the next census was expected).

Where holdings had census data recorded intermittently it was assumed that the holding was disbanded 12 months after the last continuously recorded census and all cattle were disposed of at that time. For the next census recorded a 'new' holding of cattle was initiated and the disposal-replacement cycle was repeated. The assignment of holding type descriptor and calculation of mean holding size for MAFF and SOAEFD census holdings has been described in an earlier paper (Wilesmith et al. 2000).

The BSE database (Wilesmith, Ryan, Hueston & Hoinville 1992, Sanson & Ryan 1997) provided details of the outcome of interest for this study, the date of onset of clinical signs for each of the 167,366 confirmed case of BSE that occurred before 30 June 1997. In total 2,715 BSE cases were from holdings that were unable to be matched with data present on either the MAFF or SOAEFD census rolls. Presumably these were holdings employing less than one labour unit (for those located in England and Wales) or holdings where there was a discrepancy in the unique identifier recorded in the census database and the BSE database. For these holdings the BSE database provided details of holding size and holding type. Under the assumption that the first census date was 30 June 1986, the last census date was 30 June 1996 and holding size was static these unmatched holdings were subject to the same process of population generation, described above.

Similar to the approach taken in our earlier study (Wilesmith et al. 2000) BSE cases that were recorded as purchased were included with data for the natal holding, if it was known. The natal holding was known for 145,152 (87%) of confirmed BSE cases (Table 4.1). Counts of BSE-affected cattle for each birth cohort (1 July – 30 June) for each holding were made and deducted from the generated population of cattle. Where the natal holding was unknown for purchased cases, counts of BSE-affected cattle were deducted from the holding of diagnosis.

The BSE database provided the birth date (if known) and date of onset of clinical signs for each confirmed case of BSE. This information was added to the data set containing

the generated population of cattle. To verify the model's accuracy, the final data set was queried to count the total number of stock estimated to be present for each 12 month period from 1 July 1986. These counts were compared with the actual numbers of cattle recorded at each year's agricultural census. The sum of the total number of adult cattle present throughout the study period provided an estimate of the total number of cattle at risk for the calculation of the cumulative incidence of BSE throughout the study period.

For the survival analyses conducted in this study, BSE cases with unknown birth dates (3,936 cases in total, 2.3% of confirmed BSE cases) were excluded. Date of birth was estimated from date of onset of clinical signs and year of birth (Sanson & Ryan 1997) for 37,526 (22%) of confirmed BSE cases: these were included in the survival analyses.

Statistical analyses

Kaplan-Meier survival curves were generated to describe the distribution of the number of months between birth and the onset of clinical signs of BSE. Cattle that were BSE-free were right-censored at the time of disposal. The effect of each stratum of a proposed prognostic variable on the age-at-onset of clinical signs of BSE for confirmed cases was assessed using the LIFETEST procedure in SAS (The SAS System for Windows, Release 6.12, SAS Institute Inc., Cary, NC, USA). Variables considered as potential predictors of age-at-onset of BSE were region of the natal holding (Eastern, Mid and West, Northern, Scotland, Southeast, Southwest and Wales, as defined in Figure 4.1), holding size, holding type and birth cohort. Definitions of the categorical variables region, holding size and holding type are provided in Wilesmith et al. (2000). Definitions of the levels of the cohort variable are shown in Table 4.4. The homogeneity of the Kaplan-Meier survival curves produced for each stratum was tested using the log rank statistic. Prognostic variables that showed an association (that is, a difference in the Kaplan-Meier survival curves at $P < 0.20$) were selected for inclusion in the multivariate analysis.

Regression coefficients for the Cox proportional hazards model were determined using a forward-stepping approach using the PHREG procedure in SAS Release 6.12. The Efron approximation was used to handle tied event times (Efron 1977). To verify that the proportional hazards assumption of the Cox model were valid a plot of $log[-logS(t)]$ against time was constructed for each stratum of a proposed prognostic variable. Covariate adjusted survival curves, based on the results of the Cox proportional hazards model,

were computed using the modified estimated risk score approach described by Hosmer & Lemeshow (1999).

4.3 Results

Of the 167,366 cases of BSE confirmed in Great Britain up to 30 June 1997 146,248 cases (87%) originated from dairy holdings, 10,951 (6.5%) from beef suckler holdings and the remainder from holdings designated as mixed or unknown (Table 4.2).

Counts of the total number of confirmed BSE cases, the estimated total number of adult cattle at risk and the cumulative incidence of BSE from 1 July 1986 to 30 June 1997 (expressed as cases per 100 adult cattle at risk) are shown in Table 4.3. The cumulative incidence of BSE in Great Britain for the period 1 July 1986 to 30 June 1997 was 1.10 confirmed cases per 100 adult cattle at risk (95% CI 1.09 – 1.10). There were significant differences in the cumulative incidence of BSE between regions (χ^2 53098; df 6; $P < 0.01$) with the highest incidence recorded in the Southeast, Eastern and Southwest regions of England (2.06, 1.93, and 1.86 cases per 100 adult cattle at risk, respectively). Scotland experienced the lowest cumulative incidence for the study period, with 0.25 confirmed cases per 100 adult cattle at risk.

Figure 4.2 shows the cumulative proportion of cattle that were BSE-free as a function of age (in months) for the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997. Kaplan-Meier survival curves showing the cumulative proportion of cattle that were BSE-free as a function of age, stratified by region of the natal holding are shown in Figure 4.3. In agreement with Table 4.3, the Southeast, Eastern and Southwestern regions of England had the highest proportions of cattle affected.

Regression coefficients for the Cox proportional hazards regression on age-at-onset of clinical signs for confirmed BSE cases are shown in Table 4.4. Cattle born in 1985, 1986 and 1987 had 6.1 (95% CI 5.9 – 6.2), 13 (95% CI 12 – 13), and 22 (95% CI 22 – 23) times the monthly hazard of having clinical signs of BSE occur, compared with cattle born before 30 June 1985. Cattle born in the twelve months after July 1988 had 7.4 (95% CI 7.2 – 7.5) times the monthly hazard of BSE occurring, compared with cattle born before 30 June 1985. Cattle born in cohorts after 1989 experienced progressive reductions in the

monthly hazard of experiencing BSE.

Cattle with their natal holding in the Eastern region of England had 6.0 (95% CI 5.8 – 6.1) times the monthly hazard of BSE occurring, compared to cattle in the reference category, those with their natal holding in Scotland (Table 4.4).

Based on the regression analysis shown in Table 4.4, Kaplan-Meier survival curves showing the cumulative proportion of cattle BSE-free as a function of age, stratified by region and adjusted for the effect of mean holding size, holding type and birth cohort are shown in Figure 4.4. Kaplan-Meier survival curves, stratified by birth cohort and adjusted for the effect of region of the natal holding, mean holding size and holding type are shown in Figures 4.5 and 4.6. Plots of the monthly hazard of BSE relative to those cattle born prior to June 1985 for each of the birth cohort categories are shown in Figure 4.7.

Table 4.1: Counts of confirmed BSE cases where origin of the case was homebred, purchased or unknown up to 30 June 1997, stratified by region.

Region	Homebred	Purchased		Unknown	Total
		Natal known	Natal unknown		
Eastern	7825	1955	1084	92	10956
Mid and West	22965	6968	4505	214	34652
Northern	11270	3451	2384	116	17221
Scotland	3713	1565	1722	64	7064
Southeast	14478	3881	1700	167	20226
Southwest	43685	10664	7115	500	61964
Wales	9053	3679	2444	107	15283
Total	112989	32163	20954	1260	167366

Table 4.2: Counts of confirmed BSE cases to 30 June 1997 by natal holding type classification and region.

Region	Dairy	Mixed	Suckler	Unknown [a]	Total
Eastern	9404	470	953	129	10956
Mid and West	31190	1550	1407	505	34652
Northern	14050	1314	1668	189	17221
Scotland	4332	695	2001	36	7064
Southeast	17539	937	1139	611	20226
Southwest	56995	2552	2189	228	61964
Wales	12738	906	1594	45	15283
Total	146248	8424	10951	1743	167366

[a] Holdings with unknown holding type recorded in the BSE database.

Table 4.3: Cumulative incidence of BSE from 1 July 1986 to 30 June 1997, stratified by region.

Region	Confirmed BSE cases	At risk ($\times 10^6$) [a]	Incidence [b] (95% CI [c])
Eastern	10956	0.57	1.93 (1.90 – 1.97)
Mid and West	34652	3.23	1.07 (1.06 – 1.08)
Northern	17221	2.22	0.77 (0.76 – 0.79)
Scotland	7064	2.84	0.25 (0.24 – 0.25)
Southeast	20226	0.98	2.06 (2.03 – 2.09)
Southwest	61964	3.34	1.86 (1.84 – 1.87)
Wales	15283	2.04	0.75 (0.74 – 0.76)
Total	167366	15.2	1.10 (1.09 – 1.10)

[a] Sum of adult animals estimated to have been present in region from 1 July 1986 to 30 June 1997.
[b] Cumulative incidence, cases per 100 adult cattle at risk.
[c] Fisher's 95% confidence interval.

Table 4.4: Cox proportional hazards model showing the effect of region, holding size, holding type and birth cohort on the monthly hazard of experiencing BSE.

Explanatory variable	Cattle	BSE cases	Coefficient (SE)	P	Hazard	95% CI
Region comparison				< 0.01 [a]		
Eastern	566846 [b]	10827 [b]	1.784 (0.0156)		6.0 [c]	5.8 – 6.1
Mid and West	3230600	34147	0.8853(0.0136)		2.4	2.4 – 2.5
Northern	2223334	17032	0.8158 (0.0145)		2.3	2.2 – 2.3
Scotland	2843296	7028			1.0	
Southeast	982791	19615	1.611 (0.0143)		5.0	4.9 – 5.1
Southwest	3336289	61736	1.449 (0.0131)		4.3	4.1 – 4.4
Wales	2037502	15238	0.8010 (0.0149)		2.2	2.2 – 2.3
Holding size comparison				< 0.01		
1 – 6	17793	944	0.6704 (0.0382)		2.0	1.8 – 2.1
7 – 21	1110126	3195			1.0	
22 – 53	3638153	22371	0.2426 (0.0198)		1.3	1.2 – 1.3
> 53	10294396	139113	0.5892 (0.0191)		1.8	1.7 – 1.9
Holding type comparison				< 0.01		
Dairy	9056088	146248	1.589 (0.0110)		4.9	4.8 – 5.0
Mixed	1250747	8424	1.035 (0.0150)		2.8	2.7 – 2.9
Beef suckler	4913823	10951			1.0	
Birth cohort comparison				< 0.01		
Before 30 Jun 1985	4130134	24730			1.0	
1 Jul 1985 – 30 Jun 1986	1251755	17174	1.804 (0.0104)		6.1	5.9 – 6.2
1 Jul 1986 – 30 Jun 1987	1157610	32240	2.544 (0.0090)		13	12 – 13
1 Jul 1987 – 30 Jun 1988	1193212	55769	3.112 (0.0083)		22	22 – 23
1 Jul 1988 – 30 Jun 1989	1165536	17068	2.000 (0.0105)		7.4	7.2 – 7.5
1 Jul 1989 – 30 Jun 1990	1077143	10650	1.734 (0.0121)		5.7	5.5 – 5.8
1 Jul 1990 – 30 Jun 1991	1040915	4343	1.112 (0.0169)		3.0	3.0 – 3.1
After 1 Jul 1991	4204353	3649	1.338 (0.0185)		3.8	3.7 – 4.0

Likelihood ratio test statistic 32556; df 11; $P < 0.01$.

[a] The significance of inclusion of the six region variables in the model.

[b] Cases with missing values have been excluded, so numbers differ slightly from those in Table 4.3.

[c] Interpretation: compared with adult cattle from the reference category (those from holdings in Scotland), after adjusting for the effect of the type and size of holding and birth cohort, adult cattle from holdings in the Eastern region of England had 6.0 (95% CI 5.8 – 6.1) times the monthly hazard of developing BSE.

SE: standard error.

CI: confidence interval.

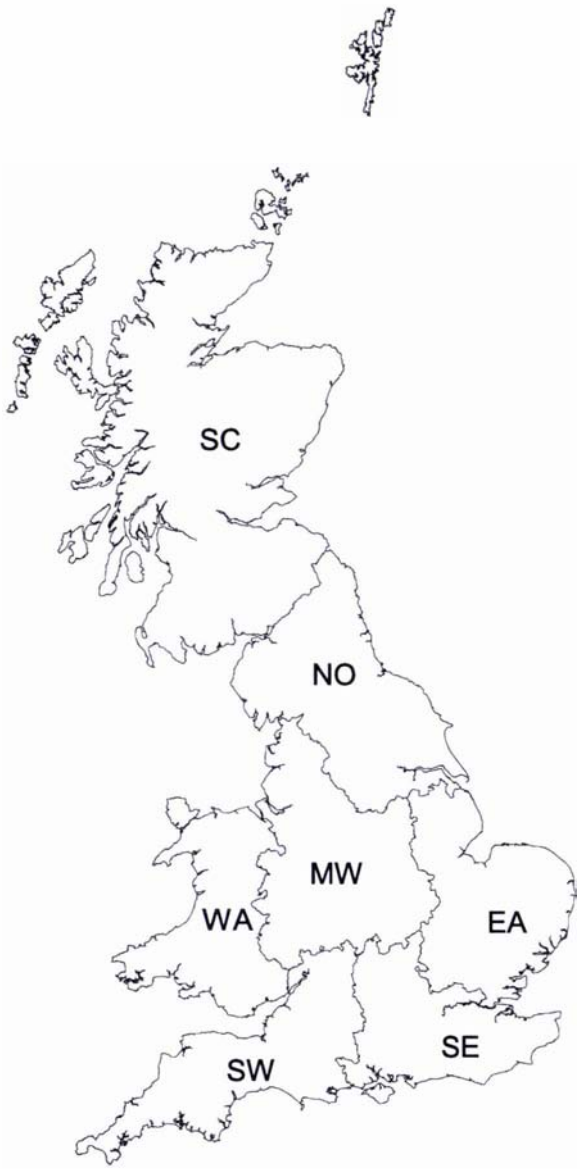


Figure 4.1: Map of Great Britain showing location of regions described in this study. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

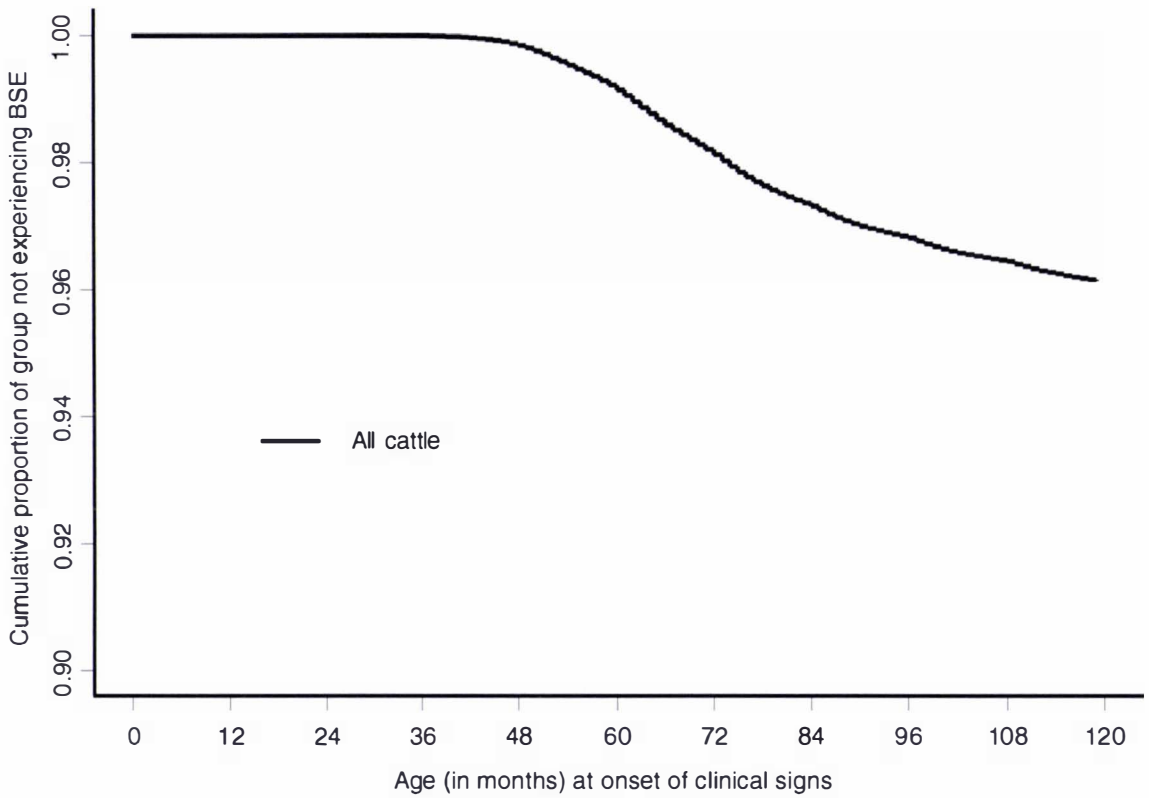


Figure 4.2: Kaplan-Meier survival curve showing the cumulative proportion of cattle BSE-free as a function of age (in months). This survival curve has been calculated using the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997.

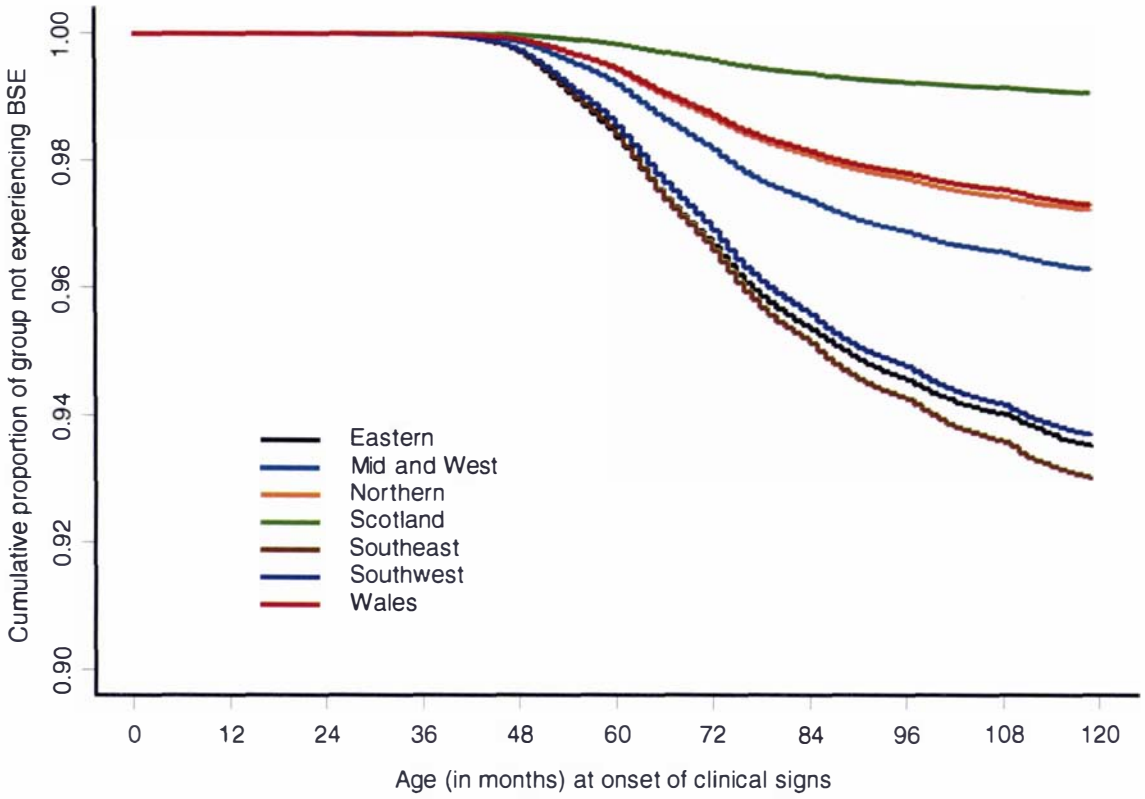


Figure 4.3: Kaplan-Meier survival curve showing the cumulative proportion of cattle BSE-free as a function of age (in months), stratified by region of the natal holding. These survival curves have been calculated using the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997.

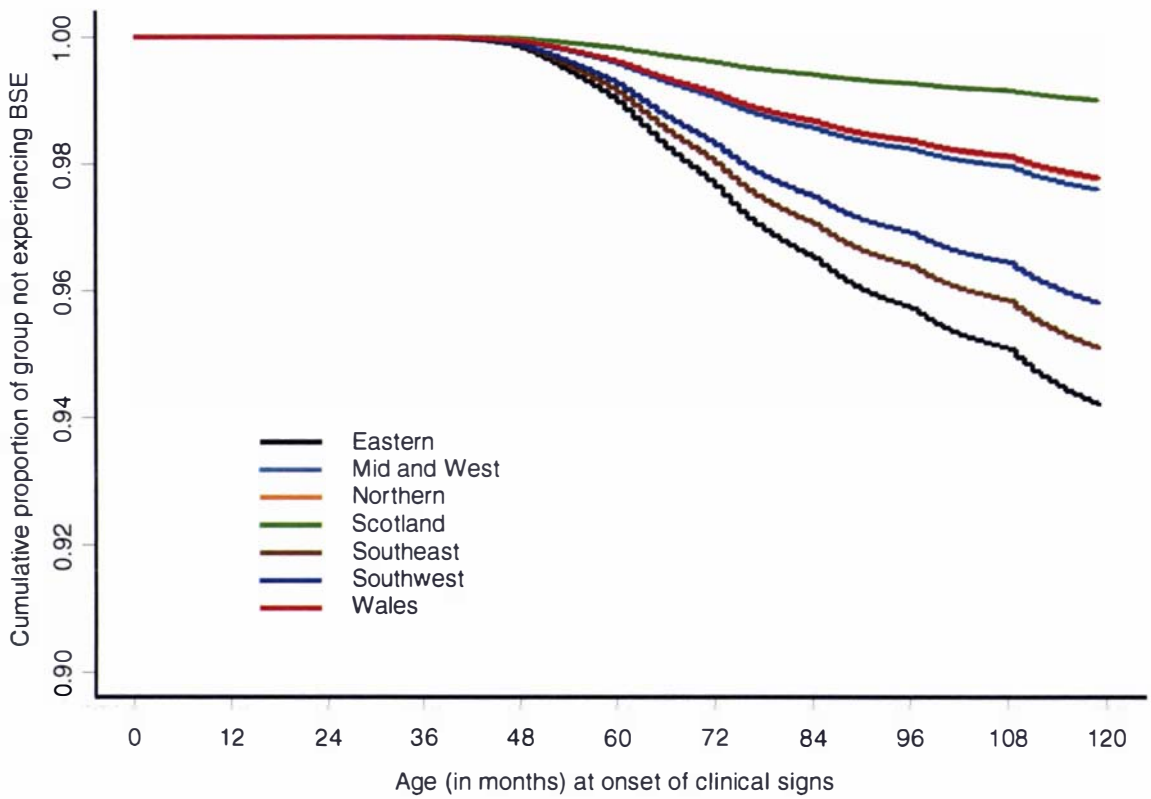


Figure 4.4: Kaplan-Meier survival curve showing the cumulative proportion of cattle BSE-free as a function of age (in months), stratified by region of the natal holding. These survival curves have been calculated using the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of mean holding size, holding type and birth cohort on the basis of the Cox proportional hazards regression model shown in Table 4.4.

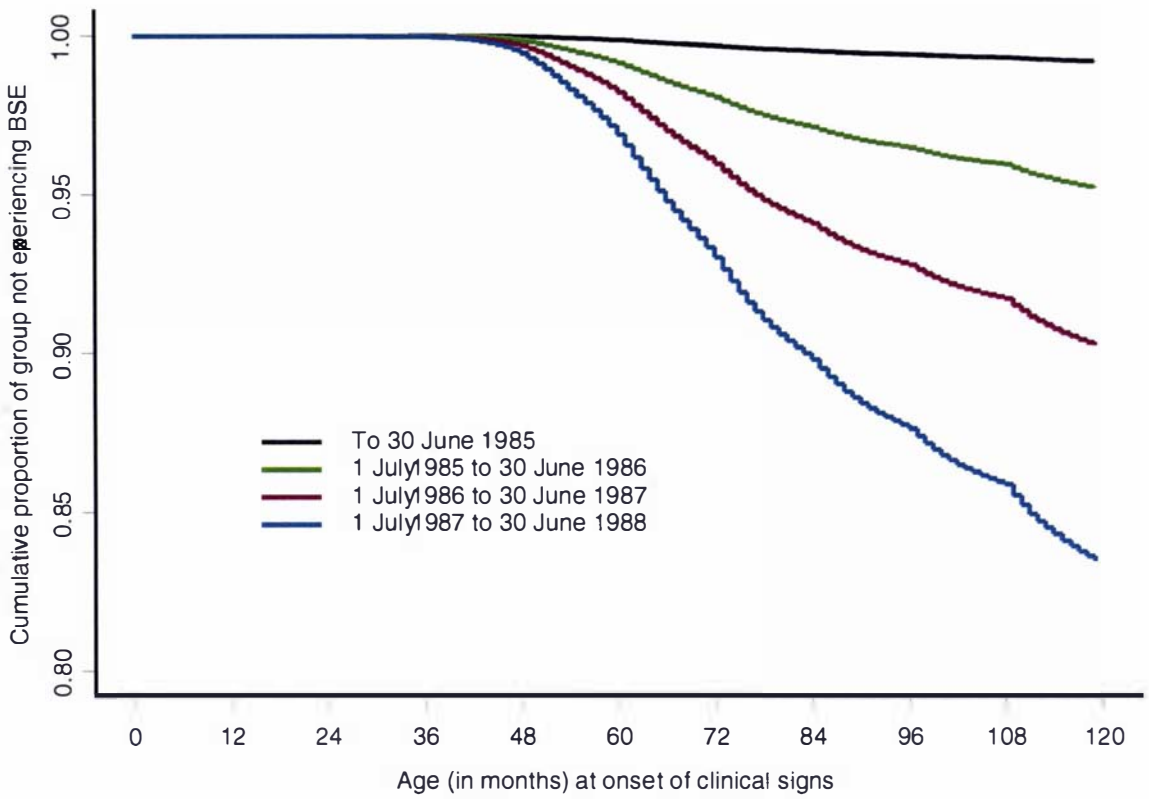


Figure 4.5: Kaplan-Meier survival curve showing the cumulative proportion of cattle BSE-free as a function of age (in months) for the 1985 and earlier, 1986, 1987 and 1988 birth cohorts. These survival curves have been calculated using the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of region of the natal holding, mean holding size and holding type on the basis of the Cox proportional hazards regression model shown in Table 4.4.

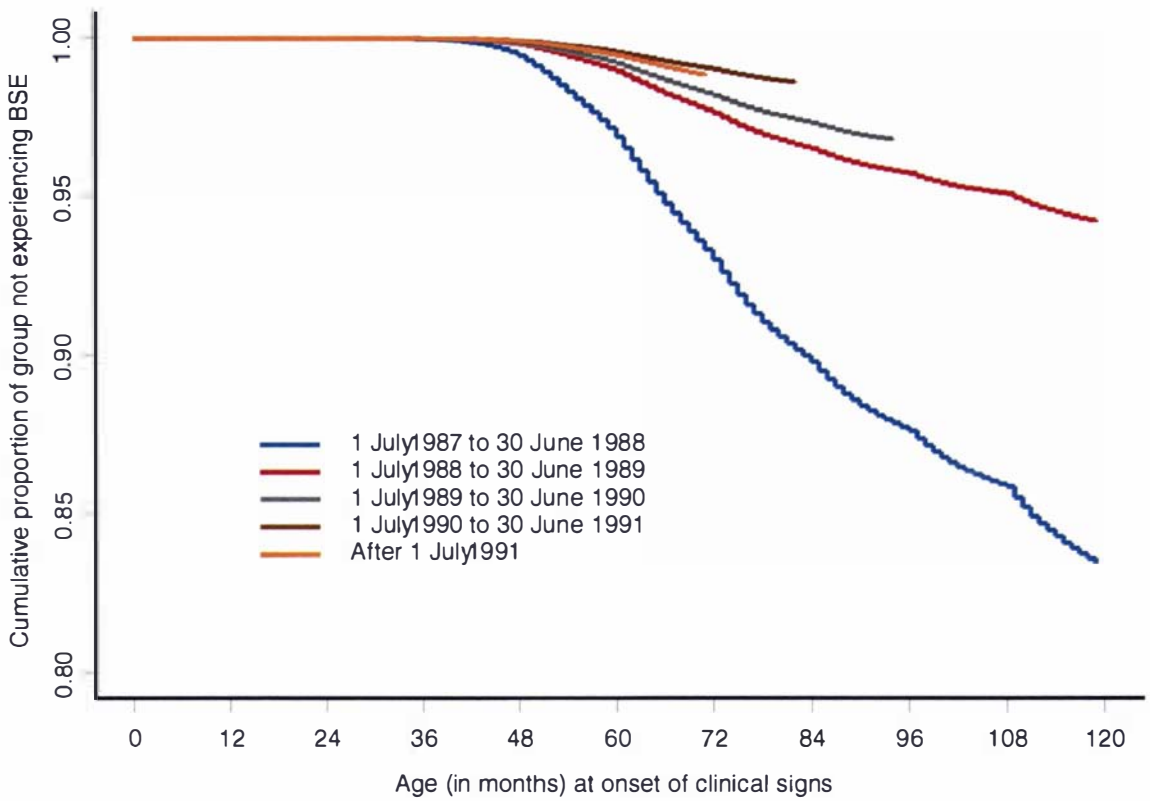


Figure 4.6: Kaplan-Meier survival curve showing the cumulative proportion of cattle BSE-free as a function of age (in months) for the 1988, 1989, 1990, and 1991 and greater birth cohorts. These survival curves have been calculated using the population of cattle estimated to have been present in Great Britain from 30 June 1986 to 30 June 1997 and have been adjusted to account for the effect of region of the natal holding, mean holding size and holding type on the basis of the Cox proportional hazards regression model shown in Table 4.4. The survival curve for the 1 July 1987 – 30 June 1988 cohort is shown, for reference.

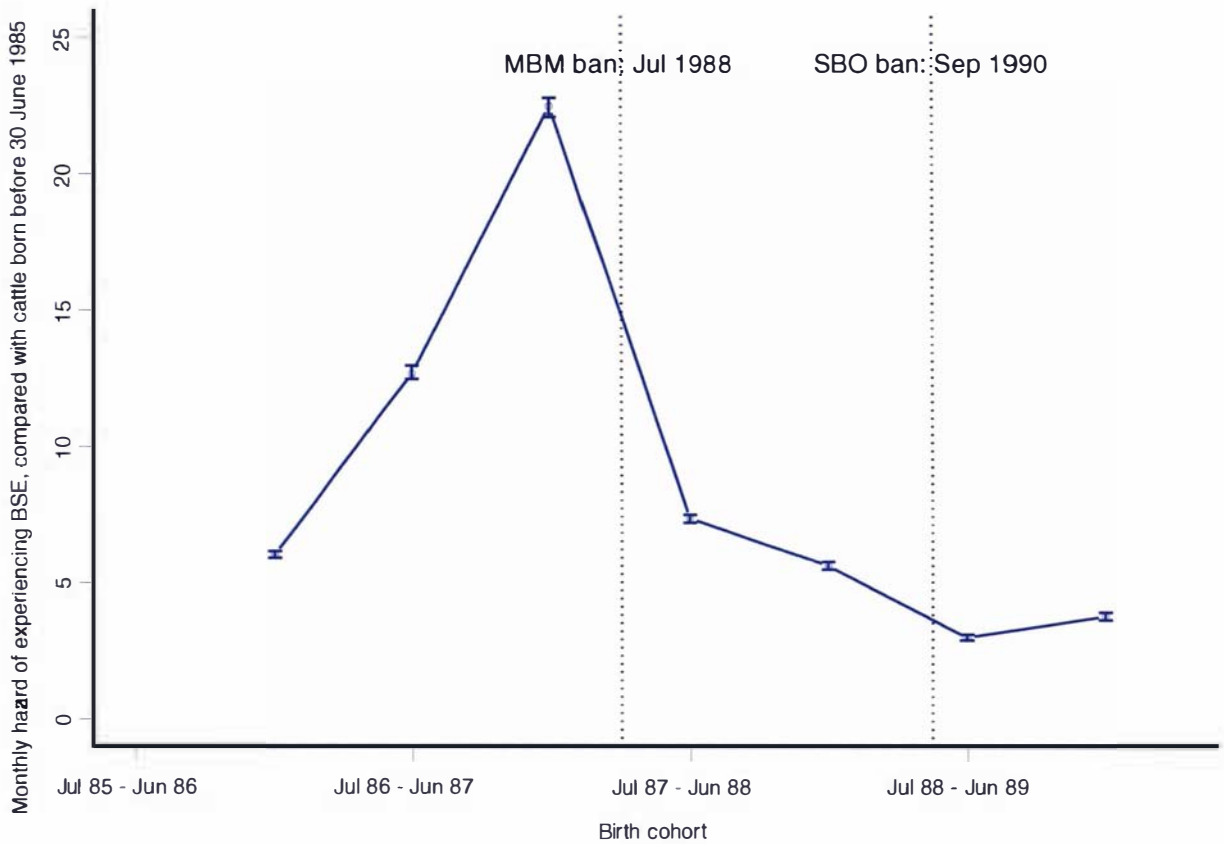


Figure 4.7: Line plot showing the monthly hazard of showing clinical signs of BSE, by birth cohort. Hazards are relative to the hazard of BSE computed for the cohort of cattle born before 30 June 1985. Error bars show the 95% confidence interval of the calculated hazard. The reference lines indicate the timing of legislation to control the disease. MBM: meat and bone meal; SBO: specified bovine offals.

4.4 Discussion

For any epidemiological investigation of a disease epidemic it is desirable to have accurate counts of the number of incident cases that occurred during the course of the epidemic as well as estimates of the size and nature of the population at risk. If these data are available, standardised estimates of incidence can be calculated which allows comparisons of the severity of the epidemic to be made across space and time. In this study routinely-recorded agricultural census data were used in conjunction with recent survey data of the age structure of British dairy herds to generate individual animal records for every beast alive between 1 July 1986 and 30 June 1997. Our estimated cattle numbers were within 9% of the actual numbers recorded at census (data not shown). In generating the cattle population we assumed that when animals were disposed from a holding they were either sold for slaughter or died. No specific provision for transfer to other holdings was allowed for. When holdings increased in size it was assumed that additional stock numbers were derived from 2 year old cattle; no provision was made for the purchase of older stock. The close match between the actual cattle numbers recorded at census and that estimated by our model suggests that these assumptions did not result in any major bias.

At the individual animal level the incidence of BSE in Great Britain from 1986 to 1997 showed a marked south-to-north gradient with the south of England (the Southeast, Eastern and Southwest regions in descending order) having the highest proportions of animals affected (Table 4.3, Figure 4.3). After controlling for the effect of mean holding size, holding type and birth cohort, this trend was still evident although the ranking of individual regions was altered slightly. Cattle in the Eastern region of England showed the highest monthly hazard of developing clinical signs of BSE (6.0, 95% CI 5.8 – 6.1) followed by the Southeast (5.0 95% CI 4.9 – 5.1) and the Southwest (4.3 95% CI 4.1 – 4.4) (Table 4.4, Figure 4.4). The effect of clustering of cases at the holding level has not been accounted for in these analyses, so the magnitude of the standard errors of the hazard ratios reported have been underestimated. Acknowledging the presence of this statistical artefact, we believe that the south-to-north trend in incidence is real and that differences between regions with similar hazard ratios (for example the East, the Southeast and Southwest) may not be as distinct as reported.

Explanations for the high hazard of BSE in the south of England is the subject of ongoing debate and include the following: (1) regional differences in the preparation of

meat and bone meal and inclusion of that meat and bone meal in concentrate cattle feeds, (2) regional differences in the cattle industry resulting in differential rates of amplification of infectious agent through the rendering process, and (3) regional differences in the effectiveness of control measures applied during the course of the epidemic. As discussed in our earlier paper in this series (Wilesmith et al. 2000) an alternative, or possibly complementary explanation is that the epidemic started in the south of the country. This hypothesis will be investigated in further analyses in this series.

The results of the Cox proportional hazards regression (Table 4.4) and the adjusted Kaplan-Meier survival curves showing the cumulative proportion of each birth cohort confirmed with BSE (Figures 4.5 and 4.6) quantify the effect of recycling of infective material within the cattle population up to July 1988. The hazard of time to onset of clinical signs for confirmed cases effectively doubled for each cohort born from 1985 to 1987, indicating the progressively increasing amounts of infective agent in the cattle food chain due to the recycling of the BSE agent in meat and bone meal (Wilesmith 1996). Relating the calculated hazards of disease for each birth cohort to the statutory BSE-control measures applied throughout the course of the epidemic allows the effectiveness of each control measure to be quantified. The monthly hazard of BSE for the 1988 cohort was 7.4 times (95% CI 7.2 – 7.5) that for cattle born before 30 June 1985. This was 0.33 times that calculated for the cohort born in the preceding year and is attributable to the July 1988 ban on the use of ruminant protein as a feed (HMSO 1988a, HMSO 1988b). The hazard of having BSE decreased further for the 1989 and 1990 birth cohorts (Table 4.4, Figures 4.6 and 4.7). The hazard of BSE in the 1989 cohort was 0.76 times that of the 1988 cohort and the hazard of BSE for the 1990 cohort was 0.54 that of the 1989 cohort. The decrease in the hazard observed for the 1990 cohort may be attributed to the effect of further control measures applied, notably the Specified Bovine Offal Ban introduced on 25 September 1990 (HMSO 1990). Because the study period ended on 30 June 1997 the records of animals born after 1 July 1991 are subject to varying amounts of right-censoring so the monthly hazard estimate for this group in Table 4.4 and Figure 4.7 is biased to an unknown degree and not directly comparable to the other cohorts.

The Cox proportional hazards regression model presented here (Table 4.4) is supportive of our earlier findings that BSE is principally a problem of cattle born on dairy holdings and cattle from larger holdings (Wilesmith et al. 2000). An interesting feature

in these analyses is that cattle from holdings of between 1 – 6 adult cattle had a higher monthly hazard of BSE compared to the reference category (those with 7 – 21 adult cattle). This result reflects the differential inclusion of BSE cases from small holdings (that is, those employing less than one labour unit) in the data set.

This and our previous study (Wilesmith et al. 2000) is, to the best of the authors' knowledge, the first time an entire animal disease epidemic has been described temporally at a national level. A novel approach in this study has been the use of routinely-collected census data to construct the population of cattle at risk throughout the epidemic and we believe, that until the Cattle Traceability System (MAFF 1999) is fully operational in Great Britain, this approach could be applied to epidemiological investigations into other animal diseases of public health and economic significance. These analyses demonstrate that the 1988 meat and bone meal feed ban was a highly effective means of controlling BSE with a 67% reduction in hazard of disease for those born in the first twelve months after its introduction. The Specified Bovine Offals ban of 1990 was further effective in controlling BSE, reducing the hazard of disease for those born in the first twelve months after its introduction by a further 46%. Had these control measures not been imposed, we conclude that the amount of infective BSE agent circulating may well have produced exposures sufficient to make control of BSE in Great Britain an almost impossible challenge.

CHAPTER 5

A descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in Great Britain to June 1997

5.1 Introduction

Bovine spongiform encephalopathy (BSE) is thought to have arisen in Great Britain as a result of events in the early 1980s that allowed the survival of the infective agent in meat and bone meal used as supplementary feed for cattle (Wilesmith 1994b). Once the infective agent was seeded into the cattle food chain the rendering of animals subclinically infected with BSE and the inclusion of the infective agent from these animals into meat and bone meal propagated the disease extensively throughout the country up until 1988, when control measures were introduced (Wilesmith et al. 1988, Wilesmith et al. 1991, Wilesmith et al. 2000).

While there is now little doubt that meat and bone meal has been the primary vehicle for transmission of the disease there are numerous hypotheses as to how the BSE agent originally entered the cattle food chain, and why the epidemic emerged during the 1980s. Perhaps the most widely accepted theory of the origin of the disease is that the BSE agent was derived from a strain of ovine scrapie (Wilesmith et al. 1988). This has been the working hypothesis of the origin of BSE to date. Alternative hypotheses, for example BSE having occurred in the cattle population for some time, but the change in rendering processes allowed the effective exposure of cattle resulting in the clinical epidemic, have been examined by Kimberlin & Wilesmith (1994) and Kimberlin (1996). Such examinations of alternative hypotheses have so far concentrated on the biological aspects

of transmissible spongiform encephalopathy agents, rather than seeking insights directly from the dynamic spatial and temporal aspects of the recorded epidemic.

Previous reports of the epidemiological features of the BSE epidemic have been based on case data in which the population at risk was either not considered explicitly, or was considered only at a highly aggregated level (Wilesmith et al. 1988, Wilesmith et al. 1991, Wilesmith & Ryan 1993, Hoinville 1994, Anderson et al. 1996, Ferguson et al. 1997). Because outlines of parishes within Great Britain (the smallest agricultural administrative unit) were not available and because of the requirement to protect the identity of BSE-affected holdings, to date there have been no detailed descriptive analyses of the epidemic in terms of its spatial or spatio-temporal pattern of occurrence. A better understanding of these aspects of the epidemic should allow the various theories of the origin to be more objectively debated and tested. This is essential if the true source of the disease is ever to be identified, and the possibility of further epidemics of BSE, independent of the British one, are to be effectively prevented. With this background the objectives of this study were to: (1) describe the spatial pattern of the bovine spongiform encephalopathy epidemic in Great Britain from June 1986 to June 1997 in terms of cattle holdings that experienced at least one confirmed case of BSE and (2) describe the spatial pattern of the BSE epidemic during the same period in terms of incident BSE cases.

5.2 Materials and methods

In the first section of this paper, referred to as the holding study, the unit of interest was every cattle holding that had details recorded on the agricultural censuses conducted annually by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998) in England and Wales and the Scottish Office, Agriculture, Environment and Fisheries Department, Scotland (SOAEFD 1987 – 1997), for the eleven censuses conducted between 30 June 1986 and 30 June 1996. The outcome of interest was whether or not a holding had at least one confirmed case of BSE (where the onset of clinical signs in the index case occurred prior to 30 June 1997). We term holdings that had at least one confirmed BSE case BSE-positive holdings. In the second section of this paper, referred to as the individual animal study, the unit of interest was the total number of adult cattle recorded at each annual agricultural census. The outcome of interest was the number of confirmed BSE cases that

occurred in the 12 month period following each census.

Description of study area

The area under investigation covered England, Wales and Scotland including the Isle of Wight, Scilly Isles, Western Isles, Orkney Islands and Shetland Islands. The Isle of Man and Channel Islands were excluded from this study. Geographical coordinates of the centroids of each of 12,297 English, Welsh and Scottish parishes were calculated within the Geographic Information System package ArcView for Windows Version 3.1 (Environmental Systems Research Institute, Redlands, California, USA). Parish perimeters were based on the 1974 boundary maps of Great Britain (Director General of Ordnance Survey 1974) with minor modifications made to reflect recent administrative changes. For parishes made up of multiple separate polygons (for example, parishes in the Western Isles of Scotland made up of several islands) the largest component polygon was selected for calculation of the parish centroid coordinates. For parishes made up of multiple, equally-sized polygons within a small distance of each other the centre of all component polygons was designated as the parish centroid. As part of the Ordnance Survey map each parish was uniquely characterised by a code made up of a two-digit county identifier and a three-digit parish identifier.

Holding study

For the holding study MAFF and SOAEFD census data provided a unique county-parish-holding (CPH) numeric identifier for each cattle holding. The county-parish component of each holding's CPH code was extracted and the geographic location of each holding was deemed to be at the centroid of the parish identified by the CPH identifier. If it is assumed that parishes are circular in shape and are, on average, 1813 hectares in size (Table 5.1) we estimate that this approach places cattle holdings to within 2.4 kilometres of their actual location.

A Gaussian kernel density surface of the location of cattle holdings recorded at any of the eleven agricultural censuses conducted between 30 June 1986 and 30 June 1996 was calculated using the Spatial Analyst Extension Version 1.1 in ArcView for Windows. A fixed bandwidth of 23 kilometres was used, calculated using the normal optimal method (Bowman & Azzalini 1997).

Details of those holdings that had experienced at least one confirmed case of BSE up to 30 June 1997 were extracted from the BSE database (Wilesmith, Ryan, Hueston & Hoinville 1992, Sanson & Ryan 1997). Since exposure to the BSE agent is thought to have occurred principally early in life as a result of the feeding of meat and bone meal supplement in calfhood (Wilesmith, Ryan, Hueston & Hoinville 1992), cases that were recorded as purchased were included with data for the natal holding, if it was known (Wilesmith et al. 2000). Where the natal holding was unknown for purchased cases, the case was included with the data for the holding of diagnosis. Gaussian kernel density surfaces of the cumulative number of BSE-positive holdings up until defined cut-points throughout the epidemic (30 June 1987, 30 June 1989, 30 June 1991, 30 June 1993, 30 June 1995 and 30 June 1997), were made using the same technique described above.

The ratio of the kernel density surface of BSE-positive holdings to that of the population of cattle holdings for each defined cut-point provided a relief map of the distribution of BSE-positive holdings per 100 holdings per square kilometre. These provided an estimate of the holding cumulative risk of experiencing a case of BSE throughout the epidemic, whilst correcting for the spatial distribution of the underlying holding population at risk (Bithell 1990, Lawson & Williams 1994, Bowman & Azzalini 1997).

Formal assessment of clustering of BSE-positive holdings as at 30 June 1997 was made using the spatial scan statistic implemented in SaTScan Version 2.1.3 (Kulldorff 1997). The number of BSE-positive holdings per parish was assumed to be Poisson distributed and the null hypothesis was that the risk of a holding being BSE-positive was the same across all areas of Great Britain. The spatial scan statistic imposes a circular window centred on each of the parish centroids. For each centroid, the radius of the window is set to vary continuously in size from zero to an upper limit of not greater than 50% of the total area under investigation. A large number of distinct circular windows are created, each with a different set of parishes within it and each window a possible candidate for delineating a cluster of BSE-positive holdings. For each window, the method tests the null hypothesis that all windows are equivalent against the alternative hypothesis that there is an elevated risk of BSE-positive holdings occurring within the current window when compared against those holdings outside of the window. The likelihood function is maximised over all windows, identifying those that constitute the most likely disease cluster; the likelihood ratio for this window is noted and constitutes the maximum likelihood ratio

test statistic. Its distribution under the null hypothesis and its corresponding P value was obtained by repeating the same analytic process on 999 random replications of the data set generated under the null hypothesis in a Monte Carlo simulation (Kulldorff 1997).

Individual animal study

In the individual animal study the number of adult cattle present within each holding at the date of each census was tabulated and the location of these animals was deemed to be at the centroid of the parish identified by the holding's CPH identifier. Gaussian kernel regression surfaces of the adult cattle population were made for each of the eleven census dates from June 1986 to June 1996. A fixed bandwidth of 30 kilometres was used, calculated using the normal optimal method (Bowman & Azzalini 1997).

Counts of the number of confirmed BSE cases diagnosed in each holding for the 12 months after each census date where extracted from the BSE database using the rule of including the data from purchased BSE cases with the natal holding, if it was known. Separate kernel regression surfaces of incident BSE cases were made for each of the eleven 12 monthly periods starting from 1 July 1986 using a fixed bandwidth of 30 kilometres.

The ratio of the kernel density surface for incident BSE cases to that of the kernel density surface for the population of cattle at risk for each census year provided a relief map of the number of confirmed BSE cases per 100 head of cattle per square kilometre. This provided an estimate of risk of BSE for each 12 month period, correcting for the spatial distribution of the cattle population at risk.

5.3 Results

Descriptive statistics of the details of the 12,297 parishes defined by the 1974 Ordnance Survey map of Great Britain are shown in Table 5.1. Descriptive statistics of the number of cattle holdings per parish are shown in Table 5.2 and descriptive statistics of the numbers of adult cattle per parish are shown in Table 5.3. Throughout the study the number of dairy holdings decreased steadily from 39,292 at the 1986 census to 28,497 at the 1996 census (data not shown). Adult dairy animals decreased from 2.8 million in 1986 to 2.3 million in 1996 (data not shown).

The data presented in these analyses are taken from 118,057 cattle holdings that were

located throughout Great Britain from June 1986 to June 1996. By 30 June 1997 30,042 of these holdings had experienced at least one confirmed case of BSE (Wilesmith et al. 2000). Throughout the same period 167,366 confirmed BSE cases were recorded from an estimated 15.2 million adult cattle at risk (Stevenson et al. 2000a).

Holding study

Figures 5.1a – 5.1f show the number of BSE-positive holdings per 100 holdings per square kilometre for the periods to 30 June 1987, 30 June 1989, 30 June 1991, 30 June 1993, 30 June 1995 and 30 June 1997, respectively. By 30 June 1997 the highest densities of BSE-positive holdings were in Avon, Somerset, Dorset and Wiltshire in the Southwest of England and Cheshire and Lancashire in the Mid and West region of England. The spatial scan statistic confirmed our visual assessment of clustering of BSE-positive holdings (Figure 5.2). By June 1997 the most likely cluster of BSE-positive holdings (that is, an area showing the greatest over-representation of BSE-affected holdings) was located in the eastern areas of the Southwest of England (log likelihood ratio statistic 385; $P < 0.01$). The second and third most likely clusters were centred around Cheshire (log likelihood ratio statistic 191; $P < 0.01$) and Cumbria (log likelihood ratio statistic 97; $P < 0.01$).

Individual animal study

Figures 5.3a – 5.3f show the number of confirmed BSE cases per 100 adult cattle per square kilometre for each of the 12 month periods to 30 June 1987, 30 June 1989, 30 June 1991, 30 June 1993, 30 June 1995 and 30 June 1997, respectively. In this analysis the highest densities of BSE cases occurred in the Southwest of England (Cornwall, Devon, Somerset, Dorset, Avon, Wiltshire and Gloucestershire) and in the south of Wales, principally in the county of Dyfed. In these counties the density of BSE was highest in the 24 month period between 1 July 1991 and 30 June 1993 with greater than 6.0 confirmed cases of BSE diagnosed per 100 cattle per square kilometre. For the 12 month period to 30 June 1997 the mean density of BSE had decreased to 0.3 confirmed cases per 100 cattle per square kilometre across all regions of Great Britain (Figure 5.3f).

Table 5.1: Descriptive statistics of British parish areas, stratified by region. Data based on the 1974 Ordnance survey map of Great Britain.

Region	Parishes	Parish size (hectares)				
	n	Mean	SD	Median	Q1, Q3	Total ($\times 10^6$)
Eastern	2321	1044	859	848	585, 1226	2.4
Mid and West	2704	970	915	729	471, 1183	2.6
Northern	1810	1553	1793	1019	578, 1825	2.8
Scotland	920	8512	12984	4174	2408, 8061	7.8
Southeast	1556	1212	852	993	641, 1571	1.5
Southwest	1931	1253	1061	981	603, 1593	2.4
Wales	1055	2007	2018	1343	683, 2569	2.0
Total	12297	1813	4272	959	585, 1691	22

SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.

Table 5.2: Descriptive statistics of the numbers of cattle holdings per parish, stratified by region. This table is based on MAFF and SOAEFD census data recorded between 30 June 1986 and 30 June 1996.

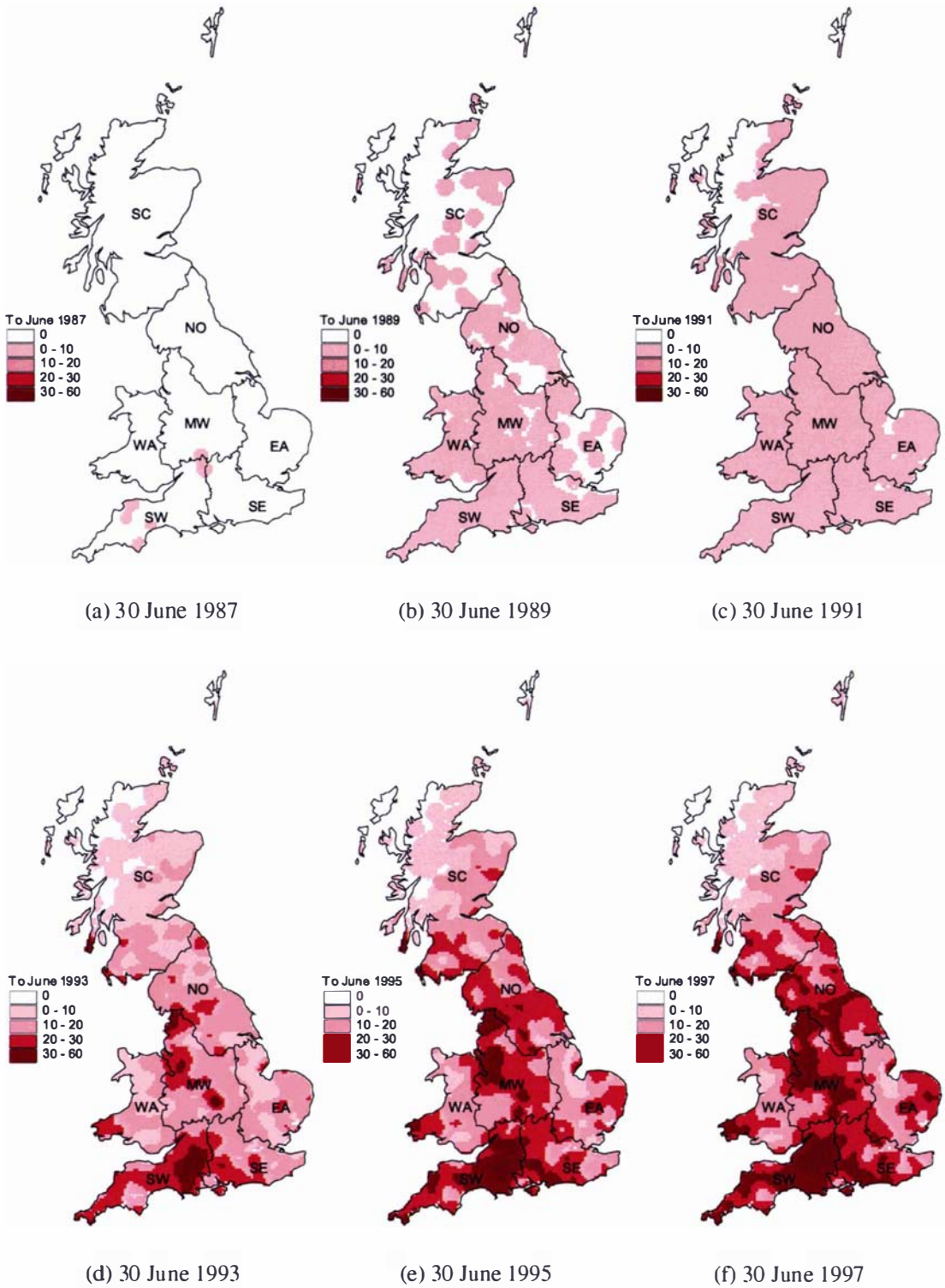
Region	Cattle holdings	Cattle holdings per parish			
	n [a]	Mean	SD	Median	Q1, Q3
Eastern	5978	3	3	2	1, 4
Mid and West	23116	9	10	6	3, 11
Northern	16393	10	12	6	3, 12
Scotland	19285	22	21	17	8, 29
Southeast	7779	6	6	4	2, 7
Southwest	24661	13	13	9	5, 17
Wales	20845	20	22	13	6, 26
Total	118057	10	14	6	3, 13

[a] The total number of holdings present for the agricultural censuses conducted between 30 June 1986 – 30 June 1996.
SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.

Table 5.3: Descriptive statistics of adult cattle numbers per parish, stratified by region. This table is based on MAFF and SOAEFD census data recorded between 30 June 1986 and 30 June 1996.

Region	Cattle	Cattle per parish			
	n [a]	Mean	SD	Median	Q1, Q3
Eastern	177495	96	115	56	18, 133
Mid and West	931796	356	461	215	87, 446
Northern	618756	362	483	210	75, 467
Scotland	842821	954	933	681	313, 1323
Southeast	300273	217	227	151	60, 299
Southwest	1023680	546	508	406	195, 726
Wales	597007	582	654	359	169, 738
Total	4491828	396	540	214	76, 495

[a] The mean number of adult cattle present in each region for the agricultural censuses conducted between 30 June 1986 and 30 June 1996.
SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.



(a) 30 June 1987 (b) 30 June 1989 (c) 30 June 1991

(d) 30 June 1993 (e) 30 June 1995 (f) 30 June 1997

Figure 5.1: Prevalence of BSE-positive holdings across Great Britain (expressed as the number of BSE-positive holdings per 100 holdings per square kilometre) for the periods to: (a) 30 June 1987, (b) 30 June 1989, (c) 30 June 1991, (d) 30 June 1993, (e) 30 June 1995, and (f) 30 June 1997. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

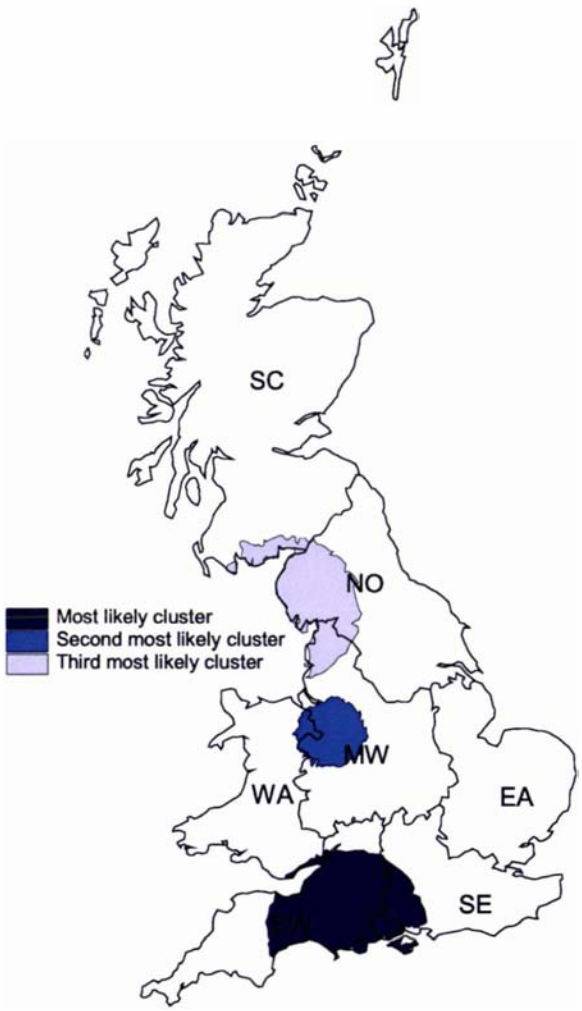
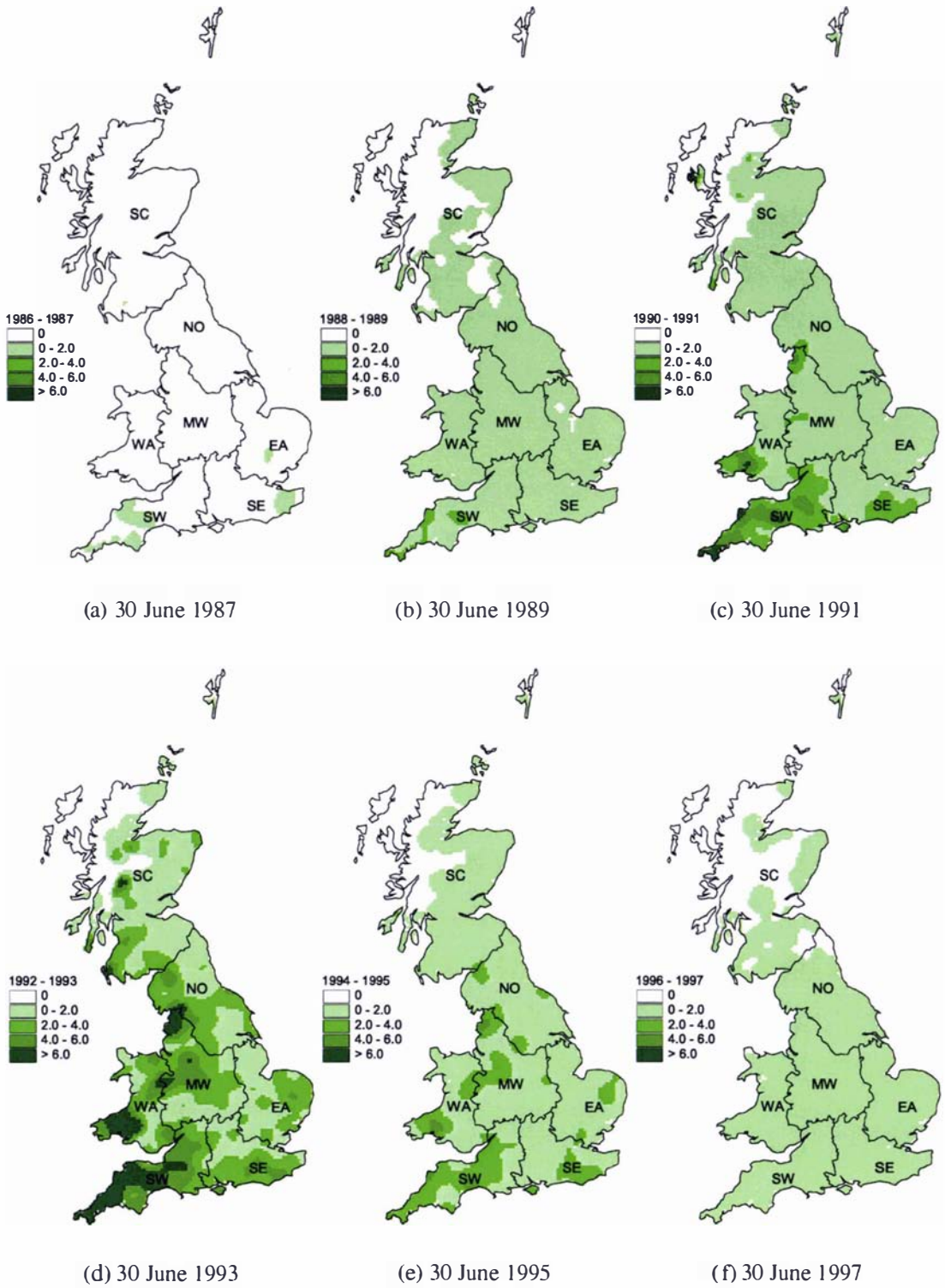


Figure 5.2: Location of the most likely, second most likely and third most likely high incidence clusters of BSE-positive holdings throughout Great Britain on June 30, 1997. Clusters were identified using the spatial scan statistic of Kulldorff (1997).



(a) 30 June 1987 (b) 30 June 1989 (c) 30 June 1991

(d) 30 June 1993 (e) 30 June 1995 (f) 30 June 1997

Figure 5.3: Incidence of BSE across Great Britain (expressed as confirmed BSE cases per 100 adult cattle per square kilometre) for the 12 month periods: (a) 1 July 1986 to 30 June 1987, (b) 1 July 1988 to 30 June 1989, (c) 1 July 1990 to 30 June 1991, (d) 1 July 1992 to 30 June 1993, (e) 1 July 1994 to 30 June 1995, (f) 1 July 1996 to 30 June 1997. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

5.4 Discussion

For privacy reasons we were obliged to avoid identifying the specific location of individual cattle holdings, an issue that has been avoided by using parish centroids and the kernel smoothing techniques described. While there has been some loss of detail by amalgamating data to the parish level this approach appears satisfactory, particularly at the whole-country level of resolution described in this study.

The size of the data sets used in these analyses posed a number of computational challenges and, as a result, some analytical compromises had to be made. Determination of smoothing parameters using robust techniques such as cross validatory methods (Stone 1974) were attempted but proved impracticable as a result of the need to process large matrices generated from the data set. The sensitivity of the bandwidths selected were tested and the overall pattern of our extraction surfaces were retained over a range of smoothing values (Lawson 1993). The application of edge correction techniques to our kernel density and regression surfaces would have been desirable, but was not computationally possible. Notwithstanding these minor limitations, we believe that the analyses presented here provide an accurate description of the spatial features of the BSE epidemic to 30 June 1997.

In order to make valid comparisons of disease incidence across space and time, it is necessary to relate the number of incident cases of the disease to the size of the population of animals at risk for each spatial and temporal window. In the individual animal study our temporal reference point was the date of onset of clinical signs for each confirmed BSE case and the sum of incident cases for each 12 month period was compared with the size of the cattle population recorded at census at the start of that period. With a disease such as BSE, where exposure to the causative agent generally occurs early in life, an alternative would have been to use the birth date of each confirmed case as the temporal reference point and to compare BSE incidence among birth cohorts. In this study we chose the former approach because: (1) accurate parish-level census data was only available for the duration of the study, that is from 1986 to 1996, (2) a number of confirmed BSE cases did not have accurate birth dates (Stevenson et al. 2000a) and (3) we wanted to describe the epidemic in terms of when and where clinical BSE cases occurred. The spatial distribution of BSE-affected birth cohorts will be addressed in a later paper in this series.

An analysis of the BSE epidemic identifying when holdings experienced their first

confirmed case of BSE identifies those holdings that were exposed to the BSE agent and avoids the 'noise' produced when within-holding incidence varies as a result of individual culling and replacement policies (Wilesmith et al. 2000). Using holding prevalence as the outcome of interest for a descriptive spatial analysis it is evident that while low numbers of holdings throughout Great Britain had a BSE index case in the early stages of the epidemic (Figure 5.1a – 5.1c), a relatively high density of BSE-positive holdings located in the eastern areas of the Southwest region of England developed early and was clearly identifiable by 30 June 1993 (Figure 5.1d). By 30 June 1997 the most likely high incidence cluster of BSE-positive holdings was evident at this location (Figure 5.1f and 5.2). By 30 June 1997 second and third most likely high incidence clusters of BSE-positive holdings were evident in two other areas of Great Britain: Cheshire and the south of Cumbria. These findings agree with our temporal analyses of the epidemic (Wilesmith et al. 2000) that larger-sized dairy holdings carried the highest risk of being BSE-positive and holdings in the south of the country were more at risk, compared with the north.

We propose that the spatial distribution of BSE-positive holdings across Great Britain, as at 30 June 1997, consisted of a first order (or trend) effect operating in a south-to-north direction and second order (or local) effects dependent on locally operating risks for BSE. Factors responsible for the first order effect are thought to include: (1) regional differences in the preparation of meat and bone meal and inclusion of that meat and bone meal in concentrate cattle feeds, and (2) regional differences in the cattle industry resulting in differential rates of amplification of infectious agent through the rendering process. Factors influencing the second order effects are thought to include: (1) local variations in holding type and size, and (2) local differences in farm practices with respect to the storage of concentrate feeds and use of concentrate feeds produced for other livestock species. Since in general for Great Britain meat and bone meal travels relatively short distances from its place of rendering to purchasing holdings we hypothesise that after July 1988, that is after the ban on the feeding of ruminant-derived protein to ruminants (HMSO 1988a, HMSO 1988b), areas with large numbers of pig and poultry units continued the process of recycling, due to cross-contamination of feed (Wilesmith 1996, Wilesmith et al. 2000).

At the individual animal level the highest density of confirmed BSE cases occurred in the greater part of the Southwest region of England and the Welsh county of Dyfed (Figure

5.3d). If Figures 5.1d and 5.3d are compared, it is evident that in Dyfed a relatively small number of holdings experienced the bulk of Welsh BSE cases whereas in the eastern areas of the Southwest of England high numbers of holdings in close spatial proximity were affected — for the most part experiencing low numbers of confirmed cases. The findings presented here provide additional insights to our previous study of individual animal-associated risk factors for BSE (Stevenson et al. 2000a). Whereas the overall cumulative incidence of BSE in Wales from 1986 to 1997 was at the low end of the range, many of the cases that occurred in this region were in close spatial proximity producing (at the height of the epidemic) some of the highest densities of confirmed BSE cases in Great Britain. This spatial pattern would suggest that a localised group of cattle holdings in Dyfed were exposed to high levels of BSE agent in their feed over an extended period.

In conclusion, the analyses presented in this paper are descriptive and provide a basis for further analytical spatial investigations into the BSE epidemic. Interpreting the findings presented here with the accompanying papers in this series (Wilesmith et al. 2000, Stevenson et al. 2000a), a consistent picture of the spatio-temporal evolution of the BSE epidemic in Great Britain is emerging. Although the first confirmed cases of BSE emerged simultaneously across Great Britain after November 1986 (Wilesmith 1991a), the epidemic concentrated itself in the south of England. From early 1992 onwards there was a sharp increase in the number of holdings experiencing an index case which we believe to be a direct consequence of the recycling of infectious material in the cattle food chain up until July 1988 (Wilesmith et al. 2000). With the bulk of BSE-positive holdings confirmed after 1992 (Wilesmith et al. 2000) we conclude that the spatial distribution of BSE-positive holdings across Great Britain at 30 June 1997 was largely determined by those factors that influenced the amount of recycled infectious material holdings were exposed to. Quantifying the effect of these factors as to whether or not specific locations were at high or low risk for BSE will provide confirmatory evidence of the characteristics of locations where a BSE epidemic is likely to propagate. For countries wishing to conduct surveillance for an emerging BSE epidemic or confirm the absence of BSE, as defined by passive surveillance, knowledge of these risk factors will ensure that surveillance is better-targeted.

CHAPTER 6

Area-level risks for BSE in Great Britain before and after the July 1988 meat and bone meal feed ban

6.1 Introduction

Descriptive spatial analyses of the bovine spongiform encephalopathy (BSE) epidemic in Great Britain have identified two characteristics of the geographical distribution of BSE-affected holdings and BSE-affected cattle identified from November 1986 to June 1997: (1) a first order effect, operating in a south-to-north direction with the south of the country experiencing a higher proportion of cattle confirmed with the disease, compared with the north and (2) second order effects thought to be explained in part by local variations in farming enterprise type and management practices (Anderson et al. 1996, Donnelly & Ferguson 2000, Stevenson et al. 2000b). In addition to these broad spatial trends, the latter phase of the epidemic has seen an increase in the proportional occurrence of BSE in the east of the country, thought to have arisen from cattle feed distributed in these areas being cross contaminated with high-protein concentrate feeds destined for the pig (and poultry) industry following the statutory control measures that banned the feeding of meat and bone meal to ruminants (Wilesmith 1996, Wilesmith 1997). Although cross contamination has been raised as a possible reason for the geographical shift in BSE risk that has been observed, to the best of our knowledge this hypothesis has not been tested against the accumulated epidemic data.

The first aim of this study was to describe the geographical distribution of BSE risk for two birth cohorts: those born before and those born after the July 1988 ban on feeding

meat and bone meal to ruminants. Our second aim was to assess the effect of putative risk factors on the spatial distribution of BSE for each of these cohorts. Identifying factors that influenced the distribution of cases during the different phases of the epidemic provides a better understanding of the mechanism of propagation of infection throughout the country. Additionally, factors thought to be important in influencing the incidence of disease throughout the eradication process can be quantified, allowing the effectiveness of control programs to be more critically evaluated.

6.2 Materials and Methods

The area of investigation covered England, Wales and Scotland including the Isle of Wight, Scilly Isles, Western Isles, Orkney Islands and Shetland Islands. Based on the average number of cattle recorded in each parish at the eleven agricultural censuses conducted by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998) in England and Wales and the Scottish Office, Agriculture, Environment and Fisheries Department, Scotland (SOAEFD 1987 – 1997) between 30 June 1986 and 30 June 1996 contiguous parishes were combined to define $i = 178$ areas (which we term 'districts'), each containing approximately 20,000 adult cattle (Figure 6.1). Aggregating the cattle population in this way enabled an ecological analysis of the BSE epidemic to be conducted at a satisfactory level of spatial resolution and the use of spatial units containing approximately equal numbers of the population at risk helped to stabilise the variance of the district-level disease rates reported.

The population of interest was the 15.2 million cattle estimated to have been present on British cattle holdings between 1 July 1986 and 30 June 1997. This was comprised of the cattle population present at the 30 June 1986 agricultural census and those animals recruited into the population from 1 July 1986 to 30 June 1997. The outcome of interest was the 175,511 confirmed cases of BSE where the onset of clinical signs occurred up to, and including, 15 May 2003. We evaluated the district-level risk of BSE for two birth cohorts: those animals born before the introduction of statutory control measures on 18 July 1988 (HMSO 1988a, HMSO 1988b) (termed the pre-control cohort), and those animals born between 18 July 1988 and 30 June 1997 (the post-control cohort). Details of the method for estimating birth dates of the cattle population present during this period

have been described previously (Stevenson et al. 2000a).

For each cohort and for each of the i districts, expected BSE counts, E_i (Carlin & Louis 2000, Lawson 2001a) were defined as:

$$E_i = n_i \left(\frac{\sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} n_i} \right) \tag{6.1}$$

where O_i was the observed number of BSE cases in the i^{th} district and n_i was the total number of cattle present in the i^{th} district. To describe changes in the spatial distribution of BSE incidence over time, standardised mortality ratios for BSE (that is, the ratio of the observed number of BSE cases in each district to the expected number of cases) were plotted for each cohort as choropleth maps. We considered it valid to compare the spatial distribution of standardised mortality ratios among the two cohorts on the basis that the age structure of the two cohort populations were similar (Breslow & Day 1987, Pickle & White 1995).

Because BSE is a non-contagious and relatively rare disease the observed number of cases in each district (O_i) was assumed to follow an independent Poisson distribution with the mean number of cases (μ_i) equal to the product of the expected number of cases (E_i) and an estimate of district-level relative risk of disease, e^{θ_i}:

$$O_i \sim Poisson(\mu_i) \tag{6.2}$$

$$O_i \sim Poisson(E_i e^{\theta_i}) \tag{6.3}$$

Bayesian ecological models were constructed to quantify the influence of hypothesised risk factors on district-level relative risk. Here, the logarithm of μ_i included the sum of a series of effects that modified relative risk. To account for the influence of spatial autocorrelation we used a convolution Gaussian prior (Besag 1989, Besag & Mollié 1989, Besag et al. 1991, Mollié 1996) where, in addition to m district-specific fixed-effects ($\beta_m x_{mi}$), unstructured (non-spatially correlated) U_i and structured (spatially correlated) heterogeneity terms S_i were included to estimate θ_i:

$$log\mu_i = logE_i + (\alpha + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + U_i + S_i) \tag{6.4}$$

Uninformed prior distributions (that is, normal distributions with a mean 0 and large variance) were assumed for the intercept α and regression coefficients $\beta_1 \ldots \beta_m$. Unstructured heterogeneity terms were parameterised as having a normal distribution with mean 0 and precision (inverse variance) τ. The structured heterogeneity terms were parameterised assuming a conditional intrinsic Gaussian autoregressive (CAR) structure with mean 0 and precision λ (Besag et al. 1991, Besag & Kooperberg 1995). The CAR structure models the log of the relative risk in district i, conditional on the risks in all other districts $i \neq j$, being normally distributed about the weighted mean of the log relative risks in the remaining districts, with variance inversely proportional to the sum of a spatial proximity matrix specified for the region under investigation. For the models presented we used a range of spatial proximity matrix specifications including those based on adjacency ($w_{ij} = 1$ if districts shared a common boundary and $w_{ij} = 0$ otherwise) and Euclidean distance ($w_{ij} = 1$ if the Euclidean distance between district centroids was less than a specified number of kilometres and $w_{ij} = 0$ otherwise). Distances of 25, 35, 50 and 100 kilometres were used to define the spatial weight matrix specifications based on Euclidean distance. To account for edge effects the ratio of the length of each district's coastline to its total perimeter was determined and the weights w_{ij} were multiplied by one less the coastline to perimeter ratio, for each district pair (Lawson 2001a). Correction for the influence of edge effects produced relative risk estimates that were closer to unity, compared with those where no edge effect correction was used. Edge-correction was therefore thought to provide a more conservative estimate of the magnitude of the relative risk estimates reported. A graphical representation of this mixed-effects parameterisation is shown in Figure 6.2.

Variables thought to influence district-level relative risk of BSE were: (1) the ratio of the mean number of dairy cattle in a district to the mean number of non-dairy cattle (termed dairy to non-dairy ratio), (2) the ratio of the mean number of adult pigs to the mean number of adult cattle in each district (termed pig to cattle ratio), (3) the ratio of the mean weight of mutton to the mean weight of beef processed in each district (termed mutton to beef ratio), and (4) the northing coordinate of the centroid of each district. Because the British pig industry is concentrated in the east of the country, it was hypothesised that contamination of cattle feed with high-protein concentrates destined for the pig industry accounted for the easterly shift in the post-control epidemic. Pig to cattle ratio provided

an estimate of the relative amount of pig feed consumed in a district and was therefore thought to be a suitable proxy variable to reflect the level of cross contamination, if it occurred. Northing coordinate of each district's centroid was included to account for the strong, south-to-north first order spatial trend in BSE risk.

Parish-level census data for cattle and pigs were available for the ten agricultural censuses conducted between 30 June 1987 and 30 June 1996 by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998). For the pre-control cohort 1987 and 1988 census data were used to estimate district-level explanatory variables. For the post-control cohort census data for 1989 to 1996 (inclusive) were used to estimate district-level explanatory variables. County-level slaughter counts for sheep and cattle for the same periods were allocated to each district on the basis of district area. The quantity of mutton produced in each district was estimated as the number of sheep slaughtered in that district multiplied by 30 kilograms, an approximation of dressed mutton carcass weight. The quantity of beef produced in each district was estimated as the number of cattle slaughtered in that district multiplied by 350 kilograms, an approximation of dressed beef carcass weight.

A three-stage approach was adopted to quantify model parameters. In the first stage, log transformed district-level standardised mortality ratios of BSE were plotted as a function of each of the district-level explanatory variable estimates. Relationships between the log transformed SMRs and explanatory variables were quantified using Pearson's product-moment correlation coefficient.

In the second stage, explanatory variables showing a relationship with BSE standardised mortality ratio at an alpha level of less than 0.10 were included in a fixed-effects model using a Markov chain Monte Carlo algorithm (Robert & Casella 1999) implemented in WinBUGS version 1.4.1 (Spiegelhalter et al. 2000). The presence of unexplained spatial autocorrelation in the data was informally assessed by plotting a correlogram of the residual terms from the fixed-effects model. Here, Moran's I statistic (I) (Moran 1950) for the residual terms at k spatial lags (W^k) was calculated and the 95% confidence interval for I_{W^k} determined by randomisation and plotted versus W^k as a correlogram. A Moran's I statistic (and its 95% confidence interval) greater than its expected value over one or more spatial lags was indicative of unexplained spatial autocorrelation and justified the use of a mixed-effects approach (Bailey & Gatrell 1995).

The third stage of the model-building process was to run the Markov chain Monte Carlo algorithm including terms for unstructured and structured heterogeneity. After Mollié (2000) we specified the average number of neighbours for the spatial proximity matrix as $\bar{m}$ and assumed that 95% of the logarithm of relative risks were in the interval $(-a, +a)$. Gamma hyperpriors for the unstructured (τ) and structured heterogeneity (λ) precision terms were estimated to have the following means:

$$2 \times \left(\frac{1.96}{a}\right)^2 \text{ for } \tau \tag{6.5}$$

and

$$\frac{2}{\bar{m}} \times \left(\frac{1.96}{a}\right)^2 \text{ for } \lambda \tag{6.6}$$

Variances of both gamma hyperpriors were taken to be large to reflect uncertainty about the values specified for the prior means. Model comparison was based on the Deviance Information Criterion (Spiegelhalter et al. 2002). For each of the spatial proximity matrix specifications considered, the best-supported model (that is, the model with the lowest Deviance Information Criterion) was used to estimate the regression coefficients reported.

For each of the Bayesian regression analyses we ran the Markov chain Monte Carlo sampler for 40,000 iterations and discarded the first 1,000 'burn-in' samples. Convergence was visually assessed by plotting cumulative path plots for each of the monitored parameters (Yu 1994, Yu & Mykland 1998, Robert & Casella 1999) and quantified using the Raftery and Lewis convergence diagnostic (Raftery & Lewis 1992b, Raftery & Lewis 1992a). Parallel chains were run using diverse initial values to ensure that convergence was achieved to the same distribution (Gelman 1996). Posterior sample sizes were determined by running sufficient iterations to ensure that the Monte Carlo standard error of the mean was at least one order of magnitude smaller than the posterior standard deviation for each parameter of interest. In each of the final models 400,000 iterations of the Gibbs sampler were used, with the results of every 100th iteration stored to estimate posterior parameter estimates.

Sensitivity of model estimates to changes in the specification of the spatial proximity matrix was assessed (Griffith & Layne 1999). Influence support plots were constructed to

evaluate inference robustness to alternative proximity matrix specifications (Marshall & Spiegelhalter 2000). Acknowledging that regression coefficients may be sensitive to the areal boundary definitions used (producing the so-called modifiable areal unit problem Openshaw 1984) analyses were repeated where the data was re-aggregated to the county level. Insensitivity of the estimated regression coefficients to re-aggregation provided some reassurance of the robustness of the effect estimates reported.

6.3 Results

Choropleth maps of the standardised mortality ratios and the standard errors of the standardised mortality ratios for BSE for the pre- and post-control cohorts are shown in Figure 6.3. For the pre-control cohort BSE risk was highest in the Southwest, Southeast, and Eastern regions of England. For the post-control cohort there was a reduction in the standardised mortality ratio for BSE in the Southwest and an increase in standardised mortality ratio for BSE in the Southeast and Eastern regions of England. The standard error of the standardised mortality ratios for BSE were greater for the post control cohort, reflecting the smaller case numbers on which these estimates have been based.

Descriptive statistics of district-level counts of adult dairy cattle, adult non-dairy cattle, adult pigs, and number of cattle and sheep slaughtered ($\times$ 1000), stratified by period (1987 to 1988 and 1989 to 1997) are shown in Table 6.1. Choropleth maps of dairy to non-dairy ratio, pig to cattle ratio and mutton to beef ratio for pre- and post-control cohorts are shown in Figure 6.4. Scatterplots of log transformed district-level SMR of BSE as a function of each of these variables for the pre- and post-control cohorts are shown in Figure 6.5. Table 6.2 shows the Pearson product-moment correlation coefficient computed for log transformed district-level BSE risk and dairy to non-dairy ratio, pig to cattle ratio, mutton to beef ratio and northing coordinate of district centroid for the pre-control and post-control cohorts. In each case log-transformed BSE risk was associated with each variable that was significant at an alpha level of 0.10, justifying their inclusion in a multivariable model of BSE risk.

Table 6.3 shows the posterior means and 95% credible intervals of the regression coefficients estimated for the fixed-effects model for each cohort. Correlograms based on the residual terms from the fixed-effects models showed strong evidence of spatial

autocorrelation, indicated by the Moran's I statistic being significantly greater than zero up to the third order spatial lag for the pre-control cohort and the second order spatial lag for the post-control cohort (data not presented).

Table 6.4 shows the posterior means and 95% credible intervals of the regression coefficients estimated for the mixed-effects model for each cohort. Box and whisker plots showing the 95% credible interval of the relative risk estimates for the mixed-effects models are shown in Figure 6.6.

Unit increases in dairy to non-dairy ratio increased district-level BSE relative risk by a factor of 1.02 and 1.03 for the pre- and post-control cohorts, respectively (pre-control cohort 95% credible interval 1.01 – 1.03, post-control cohort 95% credible interval 1.01 – 1.04). Pig to cattle ratio was risk neutral for the pre-control cohort (relative risk 1.01, 95% credible interval 0.99 – 1.02) and positively influenced area-level BSE risk for the post-control cohort (relative risk 1.05, 95% credible interval 1.02 – 1.08). Mutton to beef ratio was risk neutral for both cohorts (relative risk for the pre-control cohort 0.99, 95% credible interval 0.92 – 1.05, relative risk for the post-control cohort 0.97, 95% credible interval 0.93 – 1.01). The structured heterogeneity terms for each district showed greater variation than the unstructured heterogeneity terms, particularly in the model of BSE risk for the post-control cohort (Table 6.4 and Figure 6.7).

Influence support plots for each model are shown in Figure 6.8. In Figure 6.8 the horizontal axis shows, for each of the alternative spatial proximity matrix specifications considered, the loss of support relative to the reference model (the model with the lowest Deviance Information Criterion). Cross-hatches on each of the vertical lines identify individual districts where the estimated relative risk of BSE differed significantly from that determined for the reference model. Figure 6.8 provides evidence that changes to the spatial proximity matrix specification did not alter the majority of district-level relative risk estimates. Those districts where the estimated relative risks showed significant change as a result of re-specification of the proximity matrix were islands located off the coast of Scotland where the presence or absence of neighbouring districts was altered depending on the Euclidean distance used.

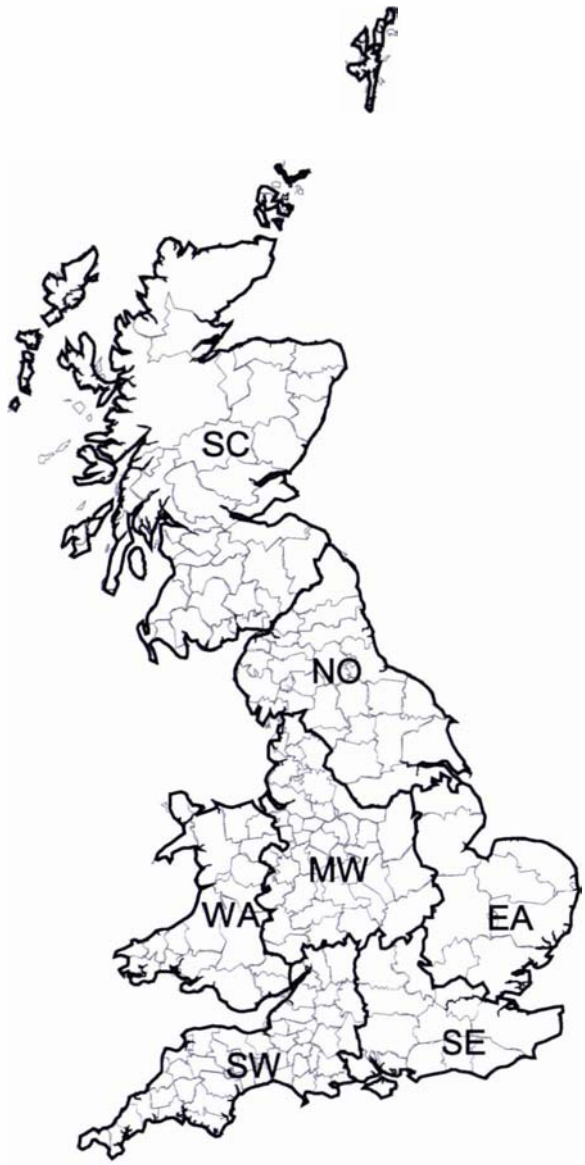


Figure 6.1: Map showing 178 districts of Great Britain. Districts have been defined, on the basis of MAFF census data collected from 30 June 1986 to 30 June 1996, as areas containing approximately 20,000 adult cattle. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

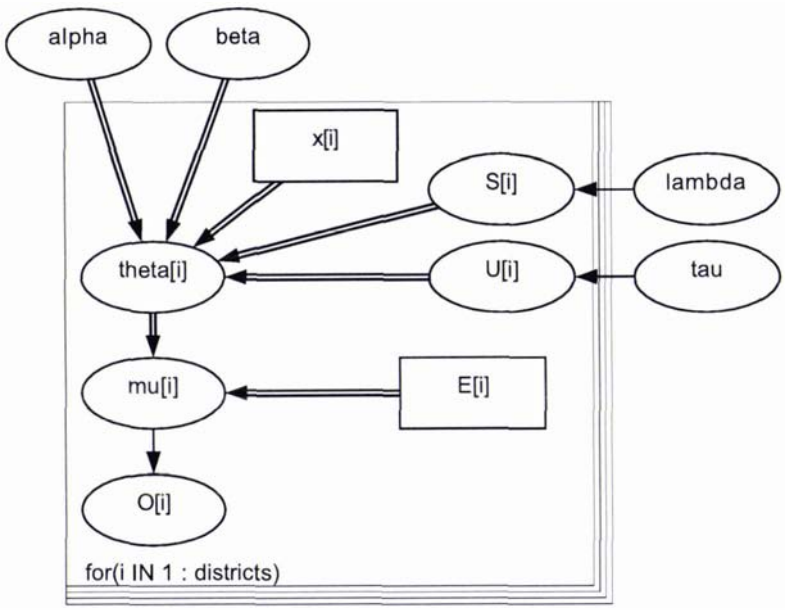
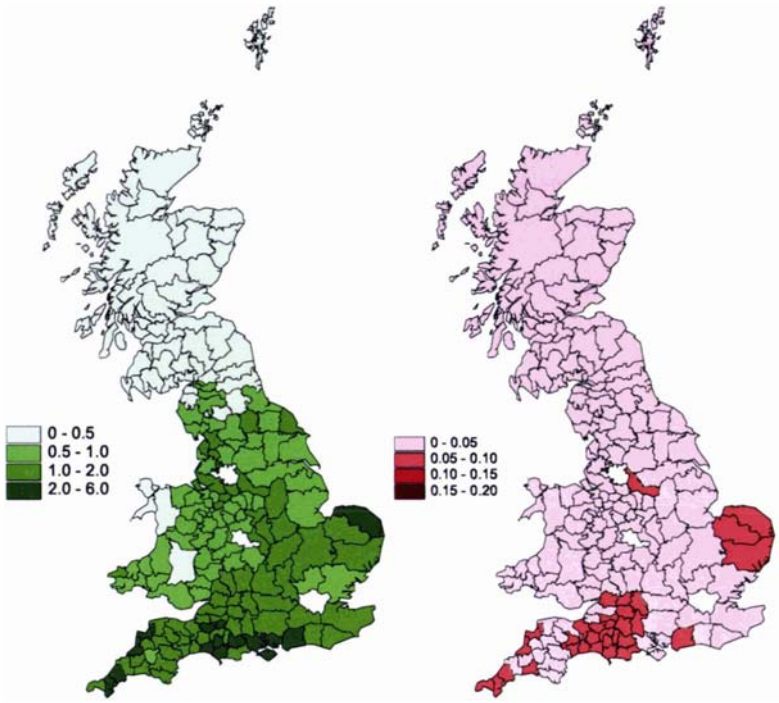
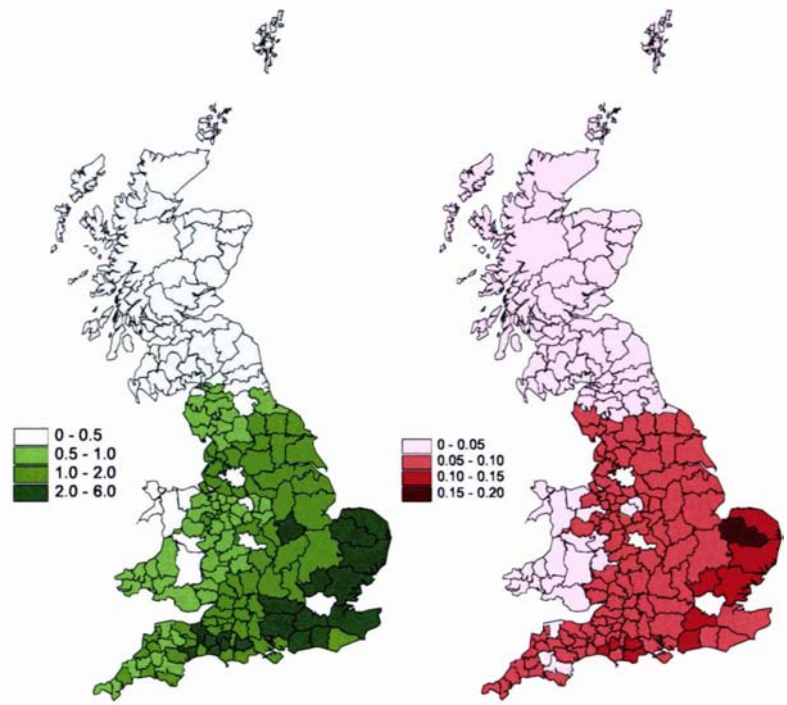


Figure 6.2: Directed acyclic graph showing the parameters in a mixed-effects model of district-level BSE risk. In the above plot the nodes of the graph represent the data and parameters of the model. Single rectangles denote observed variables, ovals denote unobserved parameters. Node labels follow the terminology used in Equation 6.4. Arrows are drawn to indicate the conditional dependence assumptions of the model; for example, `mu[i]` is is dependent on `theta[i]` and `E[i]`. The links in a graphical model either represent stochastic dependence (solid arrows) or a logical function (double arrows).



(a) SMR: pre-control cohort

(b) Standard error of SMR: pre-control cohort



(c) SMR: post-control cohort

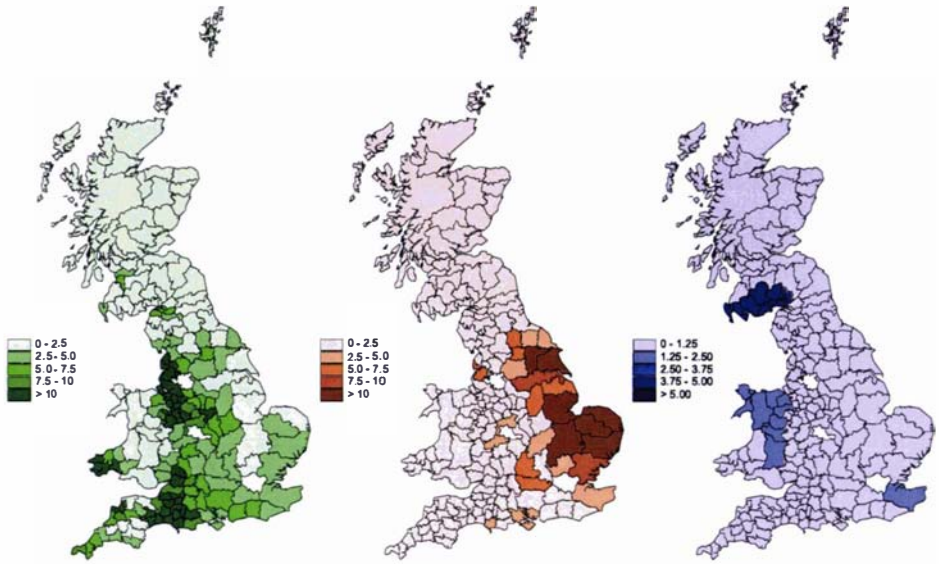
(d) Standard error of SMR: post-control cohort

Figure 6.3: Choropleth maps of district-level standardised mortality ratios (SMRs) and standard errors of district-level SMRs for BSE in British cattle 1986 – 1997: (a) and (b) the pre-control cohort, (c) and (d) the post-control cohort. Pre-control cohort: cattle born before 18 July 1988; post-control cohort: cattle born between 18 July 1988 and 30 June 1997.

Table 6.1: Descriptive statistics of district-level counts of adult dairy cattle, adult non-dairy cattle, adult pigs, and number of cattle and sheep slaughtered ($\times$ 1000), stratified by period (1987 to 1988 and 1989 to 1997).

Variable	1987 – 1988					1989 – 1997				
	n	Mean	SD	Median	Q1, Q3	n	Mean	SD	Median	Q1, Q3
Adult dairy cattle	2600	14	6	16	11, 18	2300	13	6	14	10, 17
Adult non-dairy cattle	1100	6	6	4	2, 8	1400	8	6	6	3, 10
Adult pigs	6700	38	79	15	4, 35	6400	39	81	14	4, 33
Cattle slaughtered	2800	16	19	11	7, 17	2400	13	15	9	5, 16
Sheep slaughtered	11600	65	60	45	29, 87	15600	88	82	80	41, 106

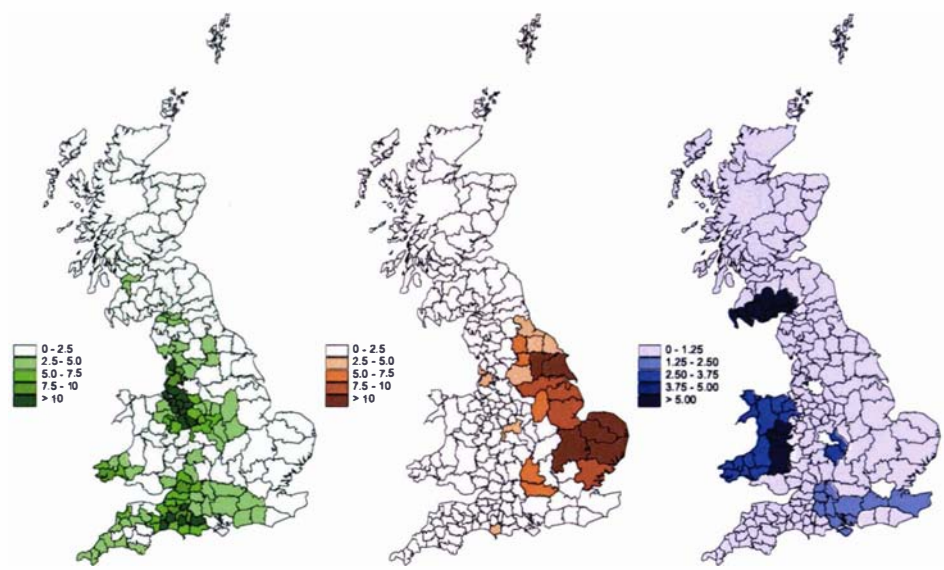
SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.



(a) Ratio dairy:non-dairy pre-control cohort

(b) Ratio pigs:cattle pre-control cohort

(c) Ratio mutton:beef pre-control cohort



(d) Ratio dairy:non-dairy post-control cohort

(e) Ratio pigs:cattle post-control cohort

(f) Ratio mutton:beef post-control cohort

Figure 6.4: Choropleth maps of district-level dairy:non-dairy ratio, pig:cattle ratio, and mutton: beef ratio for: (a), (b), and (c) the pre-control cohort, and (d), (e), and (f) the post-control cohort.

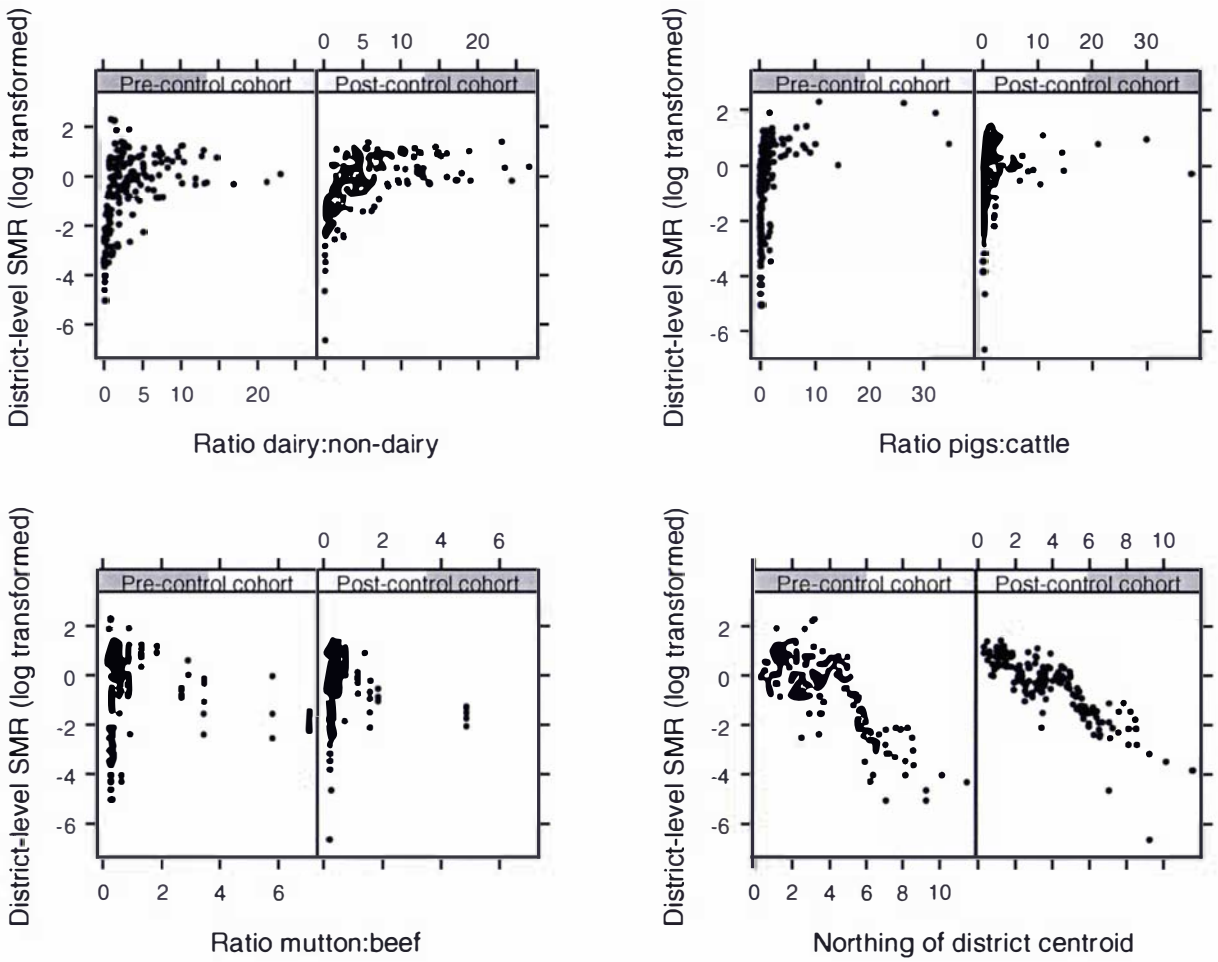


Figure 6.5: Scatterplots of log-transformed (base 2) district-level standardised mortality ratios (SMRs) for BSE versus dairy to non-dairy cattle ratio, pig to cattle ratio, mutton to beef ratio and northing coordinate of each district's centroid. In the above plots, unit increases in the transformed SMR value correspond to a doubling of BSE risk.

Table 6.2: Pearson's product-moment correlation coefficients for log-transformed district-level standardised mortality ratios for BSE and district-level estimates of dairy:non-dairy ratio, pig:cattle ratio, mutton:beef ratio and northing of district coordinate of each district's centroid.

Explanatory variable	Pearson's r	t	df	P
Pre-control cohort				
Ratio dairy:non-dairy	0.5247	8.18	176	< 0.01
Ratio pigs:cattle	0.1882	2.54	176	0.012
Ratio mutton:beef	-0.1634	-2.20	176	0.029
Northing [a]	-0.8651	-22.88	176	< 0.01
Post-control cohort				
Ratio dairy:non-dairy	0.4158	6.07	176	< 0.01
Ratio pigs:cattle	0.3278	4.60	176	< 0.01
Ratio mutton:beef	-0.1558	-2.09	176	0.038
Northing [a]	-0.7807	-16.57	176	< 0.01

[a] 100 kilometre increments.

Table 6.3: Posterior means and posterior standard deviations of the regression coefficients in the fixed-effects models of factors influencing district-level relative risk of BSE.

Explanatory variable	Posterior mean	SD	MC error	RR	95% CI of RR
Pre-control cohort					
Intercept	0.6629	0.0079	< 0.01		
Ratio dairy:non-dairy	0.0231	0.0005	< 0.01	1.02	1.02 – 1.02
Ratio pigs:cattle	0.0218	0.0006	< 0.01	1.02	1.02 – 1.02
Ratio mutton:beef	-0.1486	0.0051	< 0.01	0.86	0.85 – 0.87
Northing [a]	-0.2557	0.0017	< 0.01	0.77 [b]	0.77 – 0.78
Post-control cohort					
Intercept	0.7188	0.0128	< 0.01		
Ratio dairy:non-dairy	0.0208	0.0013	< 0.01	1.02	1.02 – 1.02
Ratio pigs:cattle	0.0496	0.0006	< 0.01	1.05	1.05 – 1.05
Ratio mutton:beef	-0.1332	0.0045	< 0.01	0.88	0.87 – 0.88
Northing [a]	-0.2551	0.0029	< 0.01	0.77	0.77 – 0.78

Pre-control Deviance Information Criterion: 13847.
Post-control Deviance Information Criterion: 12254.
[a] 100 kilometre increments.
[b] Interpretation: for every 100 kilometre increase in the northing coordinate of a district's centroid, district-level BSE risk was reduced by a factor of 0.77 (95% credible interval 0.77 – 0.76).
SD: Standard deviation.
MC error: Monte Carlo error.
RR: Relative risk.
CI: Bayesian credible interval.

Table 6.4: Posterior means and posterior standard deviations of the regression coefficients in the mixed-effects models of factors influencing district-level relative risk of BSE.

Explanatory variable	Posterior mean	SD	MC error	RR	95% CI of RR
Pre-control cohort [a]					
Intercept	0.7522	0.3523	0.04		
Ratio dairy:non-dairy	0.0208	0.0048	< 0.01	1.02	1.01 – 1.03
Ratio pigs:cattle	0.0073	0.0076	< 0.01	1.01	0.99 – 1.02
Ratio mutton:beef	-0.0140	0.0334	< 0.01	0.99	0.92 – 1.05
Northing [b]	-0.3085	0.0948	0.01	0.74 [c]	0.61 – 0.89
Structured heterogeneity [d]	0.4544	0.0830	0.01		
Unstructured heterogeneity [d]	0.0569	0.0228	< 0.01		
Post-control cohort [e]					
Intercept	0.6948	0.1106	< 0.01		
Ratio dairy:non-dairy	0.0281	0.0075	< 0.01	1.03	1.01 – 1.04
Ratio pigs:cattle	0.0513	0.0149	< 0.01	1.05	1.02 – 1.08
Ratio mutton:beef	-0.0286	0.0212	< 0.01	0.97	0.93 – 1.01
Northing [b]	-0.3349	0.0248	< 0.01	0.72	0.68 – 0.75
Structured heterogeneity [d]	0.5428	0.0192	< 0.01		
Unstructured heterogeneity [d]	0.0531	0.0080	< 0.01		

Pre-control Deviance Information Criterion: 1801.
Post-control Deviance Information Criterion: 1593.
[a] Structured heterogeneity terms based on a spatial weight matrix specification where districts are defined to be neighbours if they shared a common border.
[b] 100 kilometre increments.
[c] Interpretation: for every 100 kilometre increase in the northing coordinate of a district's centroid, district-level BSE risk was reduced by a factor of 0.74 (95% credible interval 0.61 – 0.89)
[d] Variance of heterogeneity term.
[e] Structured heterogeneity terms based on a spatial weight matrix specification where districts are defined to be neighbours if the Euclidean distance between district centroids was < 100 kilometres.
SD: Standard deviation.
MC error: Monte Carlo error.
RR: Relative risk.
CI: Bayesian credible interval.

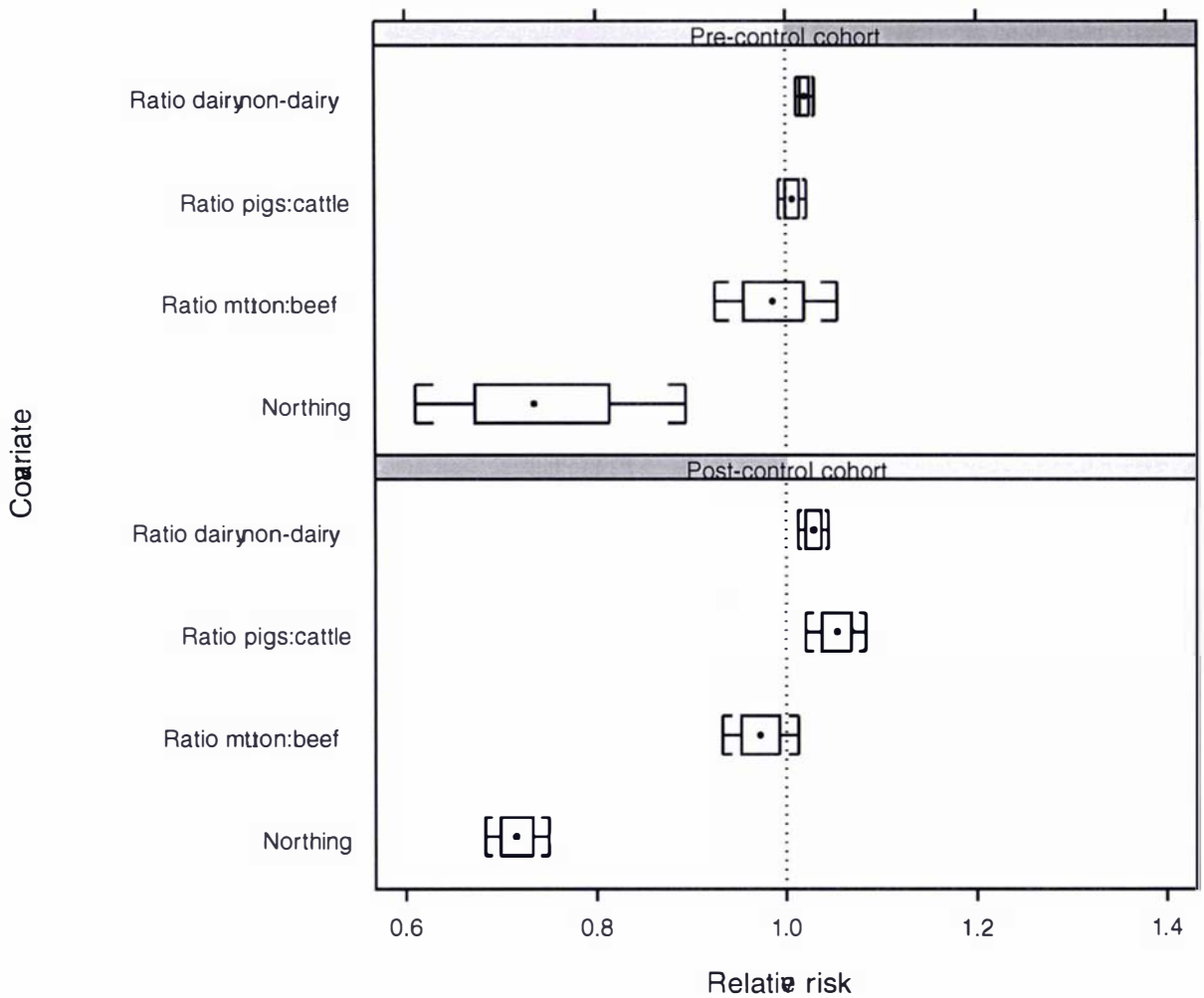
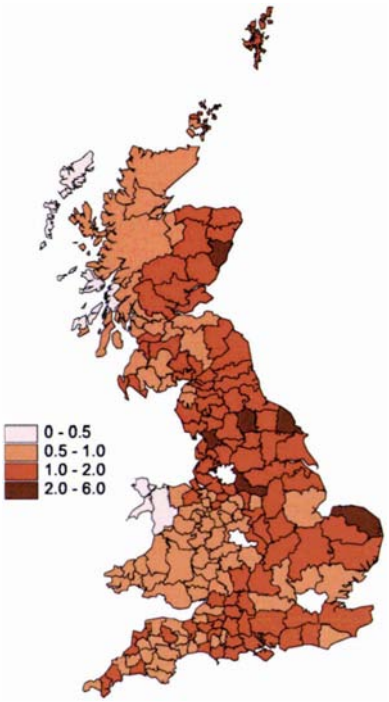
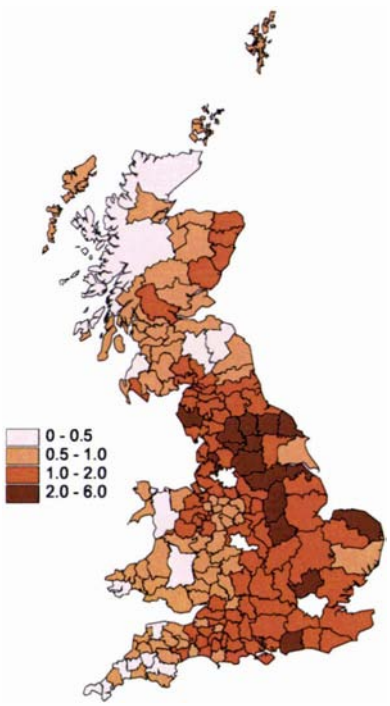


Figure 6.6: Box and whisker plots showing the 95% credible interval of the relative risk estimates for the mixed-effects models shown in Tables 6.4.

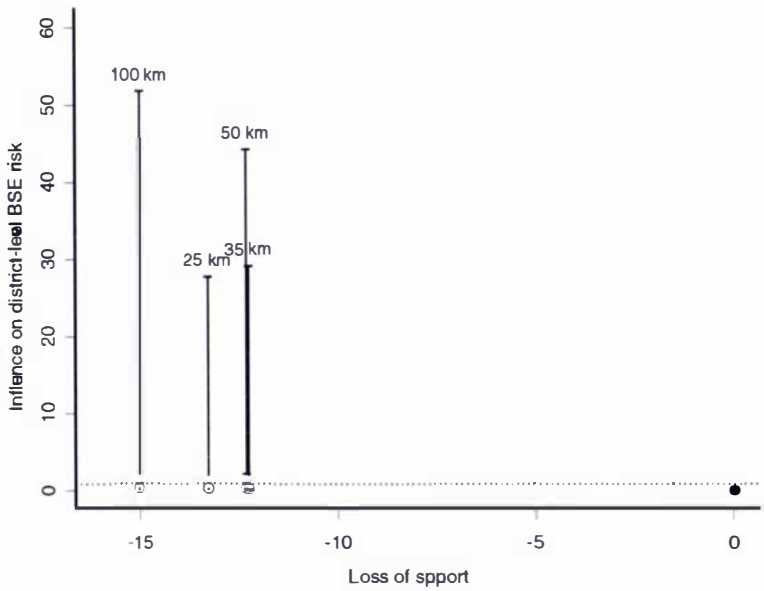


(a) Pre-control cohort

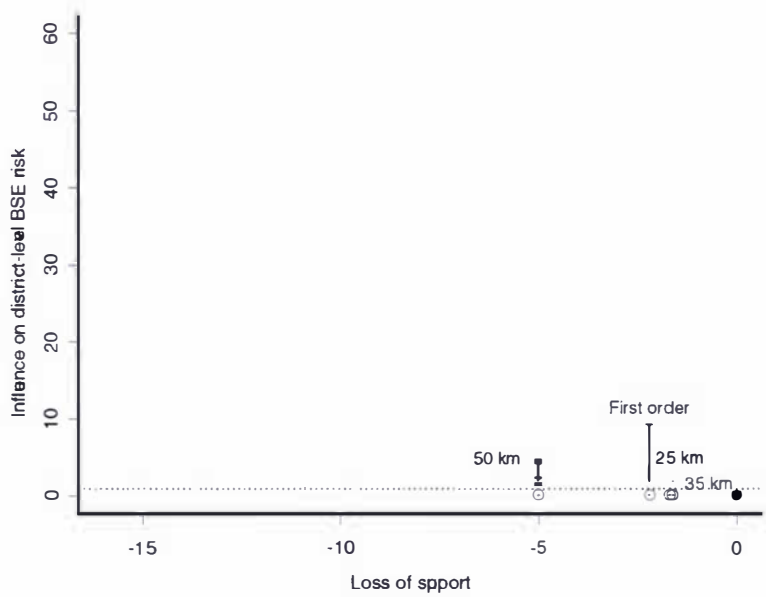


(b) Post-control cohort

Figure 6.7: Estimates of district-level relative risk of BSE attributable to the structured heterogeneity terms from the mixed-effects models shown in Table 6.4: (a) the pre-control cohort and (b) the post-control cohort.



(a) Pre-control cohort.



(b) Post-control cohort

Figure 6.8: Influence support plots for each of the alternative spatial weight matrix specifications for the mixed-effects models for: (a) the pre-control cohort and (b) the post-control cohort. Open circles (○) show the mean Kullback-Leibler divergence value for each of the alternative models relative to the reference model (•). Horizontal bars identify individual districts where the estimated relative risk of BSE differed significantly from that determined for the reference model. The horizontal reference lines show $q(d) = 0.95$, the Kullback-Leibler divergence value associated with a significant change in district-level relative risk estimate.

6.4 Discussion

This is an example of a small area study, where ecological regression methods have been used to relate counts of disease cases in geographical units to aggregated explanatory variable summaries, using area-based Gaussian random field models to accommodate local spatial correlation (Clayton & Kaldor 1987, Benardinelli et al. 1997). While this approach offers the advantage of using counts of disease cases and exposure estimates summarised at an area level it should be noted that the process of aggregation may potentially distort or mask the true exposure or response relationship for individuals within those areas, an effect known as cross-level or ecological bias (Piantadosi et al. 1988, Plummer & Clayton 1996). The value of analyses of this type therefore lies in the ability to specifically identify 'problem' areas within a region of study where the level of disease is not accounted-for by known confounders. Where disease eradication is an objective, these identified areas can then be targeted for further intervention — either in the form of more detailed investigational effort (at say, the individual farm level) or the application of additional control efforts.

The division of the 175,511 BSE cases included in this study into two cohorts was arbitrary and based on our requirement to compare risks for BSE before and after the introduction of the first statutory control measure for BSE, the July 1988 meat and bone meal ban (HMSO 1988a). Based on this criteria 129,825 and 45,686 BSE-affected cattle were included in the pre-control and post-control cohorts, respectively. Whereas testing the sensitivity of our conclusions to changing the date that defined birth cohorts would have been desirable, our ability to do so was limited by a number of factors. Firstly, it would seem logical to compare BSE risk for cohorts of cattle born before and after the date of enactment of one of the major legislated BSE control measures, that is either the July 1988 meat and bone meal feed ban (HMSO 1988a), the September 1990 specified offals feed ban (HMSO 1990) or the August 1996 ban on feeding meat and bone meal to all domestic species (HMSO 1996c). Given that the data in this analysis included BSE cases born up to and including 30 June 1997, a later cutpoint (say September 1990 or August 1996) would have resulted in truncation of the age distribution of the post control cohort and a subsequent decrease in the power to detect risks for this group. Secondly, if the cutpoint was shifted to a date much earlier than July 1988 it is possible that regional differences in case ascertainment might have introduced biases in the spatial distribution

of disease, the magnitude of which was unknown. Of the 129,825 cases included in the pre-July 1988 control cohort, 128,375 (99%) had onset of clinical signs dates after 18 July 1988 — a stage of the epidemic when BSE was widely known among the farming and veterinary community. The possibility of cases being missed prior to July 1988 exists, however there would need to have been considerable numbers of them (in the order of thousands) to substantially influence the spatial pattern shown by the 128,375 cases identified after 18 July 1988. This being the case, we reason that the bulk of cases comprising the pre-control cohort were diagnosed at a time when regional differences in case ascertainment would have been minimal, reducing the effect of the unknown level of bias caused by cases identified earlier in the epidemic.

For both cohorts dairy to non-dairy ratio positively influenced district-level BSE risk. These findings are consistent with those reported elsewhere (Wilesmith et al. 1988, Wilesmith 1991b, Wilesmith, Ryan, Hueston & Hoinville 1992, Stevenson et al. 2000a) and reflect the greater tendency of BSE to occur in areas with large numbers of dairy cattle relative to other cattle types, the former being more likely to receive compound feeds.

For cattle born after the ban on feeding ruminant-derived protein to ruminants the easterly shift in BSE risk (Figure 6.3c) implies that control measures were less-successfully applied in the Eastern counties of England compared with other regions of Great Britain (Wilesmith 1996, Wilesmith 1997). To account for the change in the spatial distribution of disease risk over time, pig to cattle ratio was included in each model as a proxy to represent the amount of pig feed produced in each district. In this case it was our hypothesis that high-protein concentrate feeds produced in areas with relatively large numbers of pigs (predominantly the east of England, Figures 6.4 b and e) contaminated cattle feed manufactured and distributed in those areas, resulting in higher relative risks of BSE. Of interest in this case was to determine if the easterly shift in BSE risk was explained entirely by higher district pig to cattle ratios, or by other, unmeasured factors.

Unit increases in pig to cattle ratio did not influence BSE risk for the pre-control cohort (relative risk 1.01, 95% credible interval 0.99 – 1.02) and positively influenced BSE risk for the post-control cohort (relative risk 1.05, 95% credible interval 1.02 – 1.08, Table 6.4) consistent with the stronger effect of cross-contamination as a determinant of disease after the imposition of the July 1988 meat and bone meal ban. Choropleth maps of district-level relative risk attributable to the structured heterogeneity terms from the

mixed-effects model identified areas in the Eastern, Northern and Southeast of England where there were spatial aggregations of BSE risk not explained by the fixed effects of dairy to non-dairy ratio, pig to cattle ratio, mutton to beef ratio or district northing (Figure 6.7). Thus, the observed easterly shift in BSE risk for the post control cohort was explained by a combination of the modelled fixed-effects and unmeasured, spatially aggregated influences. We can only speculate on the reasons for these localised influences, suffice to say that they would be consistent with the distribution areas of feed mills and/or compounders who failed to fully comply with the directives of the legislated control measures.

A notable feature of the British BSE epidemic has been the distinct south-to-north gradient in disease risk, an effect that is thought to be a result of: (1) regional differences in the processing of meat and bone meal by rendering plants and the reprocessing of greaves, (2) regional differences in the use of meat and bone meal by cattle feed compounders, (3) regional differences in transmission and amplification mechanisms, and (4) variation in the scale of original infection challenge (Wilesmith et al. 1991, Anderson et al. 1996, Wilesmith et al. 2000). While inclusion of explanatory variables to capture the effect of each of these influences would have been desirable, appropriate data was not available necessitating the use of northing coordinate as a proxy variable to account for these effects. The influence of northing coordinate on BSE risk was similar for the pre- and post-control cohort suggesting that the protective effect of a northerly location remained throughout the epidemic and was not influenced to any great degree by the control measures applied. Thus the broad spatial gradient established early in the recorded epidemic was maintained despite the various control measures which were introduced, but local influences interacted with the highly beneficial effects of the 1988 controls to dilute their impact in some areas but allow them to achieve their full effect in others. This has implications for other affected countries where failure to achieve adequate compliance with control measures may allow extended local spread to occur.

Whereas the other variables examined in this paper are concerned with processes which directly influence BSE exposure at the district level, the influence of sheep is more difficult to evaluate and results are also open to differing interpretations, particularly in light of recent debate about the possibility of persistent infection in the sheep population (Ferguson et al. 2002, Kao et al. 2002, Krebs et al. 2002). In this study our focus was to

test the hypothesis that sheep-derived material was a source of the BSE agent either early in the epidemic or over its duration. We examined district-level sheep density and the ratio of sheep to cattle population counts as possible risk factors for BSE and found no significant effect of either (results not presented). While acknowledging the limitations in the detail at which we could assess the issue, we then used county-level slaughter data for sheep and cattle to provide estimates of the weight in kilograms of carcass meat produced for each species. We used this ratio to represent the relative amounts of material entering the rendering system from each of these two species. Mutton to beef ratio was protective for both cohorts in the fixed-effects model, an effect that was neutral after accounting for structured and unstructured heterogeneity. If the BSE agent were present in sheep to a significant extent, either as a natural occurrence early in the epidemic or as a consequence of its inclusion in concentrate feed supplied to sheep during the period when this was around its peak in meat and bone meal, then it might be expected that mutton to beef ratio would present as a positive risk factor in one or both periods. The data do not support the existence of a risk-enhancing influence, thus giving some reassurance that sheep were not substantially contaminating rendered material with the BSE agent during either phase of the epidemic. The simplest interpretation of the results for both periods is that rendered sheep material diluted infected bovine material, hence mildly reducing area-level BSE risk.

6.5 Conclusion

For both cohorts examined in this paper district-level risk of BSE was increased by a more southerly location and greater numbers of dairy cattle, relative to beef cattle. For the cohort of cattle born after the July 1988 ban on feeding meat and bone meal district-level BSE risk was additionally associated with greater numbers of pigs, relative to cattle. These findings support the role of low level cross-contamination of cattle feed by pig feed as an influence on BSE incidence as the epidemic evolved. Prior to the 1988 meat and bone meal ban unexplained BSE risk was relatively uniformly distributed across the country whereas after the ban there were spatially aggregated areas of unexplained risk in the northern and eastern regions of England suggesting that local influences allowed BSE control measures to be less-successfully applied in these areas, compared with the rest of

the country. We conclude that spatially localised influences were operating in divergent ways during the two phases of the epidemic.

CHAPTER 7

The early evolution of the bovine spongiform encephalopathy epidemic in Great Britain

7.1 Introduction

It is now generally accepted that the bovine spongiform encephalopathy (BSE) epidemic in Great Britain was amplified by recycling of prion-infected material derived from clinical and subclinical cases to the bovine population via meat and bone meal in concentrate feeds (Wilesmith et al. 1988, Wilesmith, Ryan & Hueston 1992, Wilesmith, Ryan, Hueston & Hoinville 1992, Anderson et al. 1996, Hörnlimann et al. 1996, Donnelly, Ferguson, Ghani, Woolhouse, Watt & Anderson 1997, Donnelly, Ghani & Wilesmith 1997, Wilesmith & Ryan 1997). In contrast, it remains less clear how the epidemic started and why it started, of all times, during the mid 1980s. Initial research into the epidemiology of BSE attributed the origin of infection to a scrapie-like agent contaminating meat and bone meal, and it was hypothesised that changes in rendering practices had allowed the agent to retain its infectivity through the rendering process (Wilesmith et al. 1988, Wilesmith et al. 1991). Analyses of the epidemic conducted to date have inferred that effective exposure commenced after 1981 (Wilesmith et al. 1991) but the precise spatial and temporal distribution of the earliest recorded exposures have not been described in detail. As part of ongoing analyses of the BSE epidemic reported in this series of studies this paper focuses on the early phase of the British epidemic. Specifically we sought to: (1) estimate an incubation period for the disease, (2) estimate the date on which holdings were first exposed to the BSE agent at levels sufficient to produce the recorded epidemic, and (3) describe the spatial and temporal features of these earliest-recorded exposures.

7.2 Materials and methods

Analyses were undertaken in three stages (Figure 7.1). Firstly, individual BSE case details and cattle census data were used to estimate an incubation period for the disease — the period between exposure to the causative agent and onset of clinical signs (Rothman & Greenland 1998). We then used this incubation period to determine the date of exposure for all cases and hence the date of first exposure for the population of cattle holdings present in Great Britain at the 1986 agricultural census. In the final stage we described and quantified factors associated with the spatial and temporal pattern of holding-level exposure to 1 July 1983.

7.2.1 Estimation of incubation period

Details were derived from the BSE case database (Wilesmith, Ryan, Hueston & Hoinville 1992, Sanson & Ryan 1997) for all confirmed cases where the onset of clinical signs was before 1 July 1997. Case details included the identifier of the natal holding, date of birth, date of onset of clinical signs and indicator variables denoting the reliability of the recorded natal holding and date of birth. Date of birth was accurately known for 125,904 of the 167,366 cases in the data set. For BSE cases whose date of birth was estimated (n = 37,526) it was acknowledged that the recorded ages-at-onset were biased towards integer numbers of years. To reduce this bias a stochastic re sampling procedure was used to randomise the month of the estimated date of birth (Ferguson et al. 1997, Donnelly & Ferguson 2000). Age-at-onset of disease was calculated as the date of onset of clinical signs less the (known or estimated) date of birth for the 163,430 cases with valid values: age-at-onset was deemed to be missing for the 3,936 cases with no birth date recorded. In contrast to approaches taken in our previous analyses of the epidemic we regarded details of natal holding to be missing for 22,214 cases with uncertain data, even if an estimated natal holding identifier (typically the holding of diagnosis) was provided.

Multiple imputation techniques (Little & Rubin 1987) were used to estimate missing values of age-at-onset (as a continuous variable) and region of the natal holding (as a categorical variable of seven levels: Eastern, Mid and West, Northern, Scotland, Southeast, Southwest or Wales). Here we fitted a conditional Gaussian model using the expectation-minimisation and data augmentation algorithms implemented in the missing

library (Schimert et al. 2001) in S-PLUS (S-PLUS 6.1 for Windows, Insightful Corporation, Seattle, WA, USA). In the first instance the expectation-minimisation algorithm was used to obtain the maximum likelihood estimate and the value of the maximised log-likelihood for each of the missing values. These estimates were then used as starting values for the data augmentation algorithm. For the data augmentation algorithm a (flat) Jeffreys prior was used for all hyperparameters and a single Markov chain was run for ten times the number of iterations required for the expectation-minimisation algorithm to converge. Convergence of the simulated values of the data augmentation algorithm was assessed using time series and autocorrelation function plots. Having confirmed adequate convergence we generated five imputations of the missing data starting from five over dispersed values. Each of the five imputed data sets was analysed using the methods described below. The five sets of results were consolidated to provide point estimates and their uncertainty.

To estimate an incubation period for BSE, details of confirmed cases with either known, estimated or imputed dates of birth between 1 July 1987 and 1 July 1988 were selected and merged with details of the 1987-born population of unaffected cattle used in our earlier analyses of the epidemic (Stevenson et al. 2000a). The 1987 cohort was selected because these animals were born when contamination of meat and bone meal was greatest. Consequently, this cohort would have experienced the highest risk of successful exposure early in life and, being one of the first cohorts to reach adulthood after the July 1988 meat and bone meal ban (HMSO 1988a, HMSO 1988b), should have experienced low levels of successful exposure as adults. After Rothman (1981), it was assumed that an arbitrary time window after early-life exposure for each individual in this cohort was the only time during which an aetiologically relevant case of disease might occur: only those cases that occurred during this time interval counted in the numerator of the rate for exposure, and the sum of the number of animal years over the relevant time window constituted the cattle-time experience for the denominator of the rate. The analysis was repeated with changes in the assumptions about the size of the incubation period — the maximum value for the incidence density among exposed subjects obtained in repeated analyses provided the most appropriate point estimate of incubation period for animals exposed early in life.

For each confirmed BSE case the date of exposure to the causative agent was estimated

to be the date of onset of clinical signs less the estimated incubation period. Where the estimated date of exposure occurred before the recorded or estimated birth date, exposure date was taken as birth date plus a number drawn from a Poisson distribution with a mean of 90 days (assuming that these cases were exposed in early life, we approximated the number of days of age when concentrates containing meat and bone meal would have first been fed in significant amounts).

7.2.2 Estimation of the time and place of index exposure

The study population was all British cattle holdings recorded at the 1986 agricultural census conducted by the Ministry of Agriculture, Fisheries and Food (MAFF 1986 – 1998) in England and Wales and by the Scottish Office, Agriculture, Environment and Fisheries Department (SOAEFD 1987 – 1997) in Scotland. Census data for each holding included a unique identifier and a record of the number of adult dairy and beef suckler animals present on 30 June 1986. Holdings on which the number of dairy cattle recorded exceeded 80% of total stock numbers were classified as being of dairy type; holdings not meeting this criterion were classified as non-dairy.

Estimated BSE exposure dates were assigned to cattle holdings and the earliest of these was termed the index holding exposure date. To account for spatial variation in the accuracy of the recorded details of the natal holding of index exposure cases (that is, the BSE cases that defined index holding exposures) the location of index exposure was taken to be the natal holding of the index exposure case only if it was reliably known (21,684 of 30,042 exposed holdings). Those index exposures where the natal holding of the index case was estimated were regarded as missing values and in this case a categorical variable representing one of the seven regions of Great Britain (Eastern, Mid and West, Northern, Scotland, Southeast, Southwest, and Wales) was estimated by fitting a conditional Gaussian model using the data augmentation algorithm, described earlier. The precise location of the holding was then defined by randomly assigning a parish identifier from the list of all parishes within the imputed region.

On account of the observed geographical variation in the certainty of recorded BSE case details (date of birth and certainty of knowledge of the location of the natal holding) it was of interest to assess how well the multiple imputation procedure dealt with this type and level of bias in the data. A subset of the case holding database was created,

comprised of complete records only (n = 14,796). Individual records were then deleted from this data set in an effort to emulate the pattern of missing data in the 'actual' case data set. For example, if date of first of exposure was not reliably known for 9% of Eastern holdings in the actual data then 9% of first exposure dates from Eastern holdings in the 'complete' data set were randomly deleted. We then imputed the missing data for this derived data set using the expectation-minimisation and data augmentation algorithm. Descriptive statistics of the imputed data set were compared with descriptive statistics for the complete data set.

7.2.3 Spatial and temporal patterns of exposure

Kaplan-Meier survival curves were generated to describe the temporal pattern of initial BSE exposure among cattle holdings in Great Britain by 1 July 1983. Holdings that had no BSE exposure recorded were right-censored on 30 June 1983.

For our spatial descriptions of BSE exposure, holdings were deemed to be located at the centroid of their respective parishes. Two Gaussian kernel density surfaces were computed: the first comprised of holdings estimated to have been exposed to BSE by 1 July 1983 and the second for the population of cattle holdings present at the June 1986 census. The ratio of the density surface of BSE-exposed holdings to the population of cattle holdings at risk provided a relief map of the cumulative incidence of BSE-exposed holdings throughout Great Britain by 1 July 1983. Smoothing parameters used were the same for cases and the population of cattle holdings at risk and were determined by normal optimal methods (Bowman & Azzalini 1997).

To quantify factors associated with the spatial distribution of BSE exposure among British cattle holdings a logistic regression approach was used. Here, the response variable was a binary indicator of whether or not each of the i = 93,008 cattle holdings had been exposed to BSE by 1 July 1983. The probability of exposure was estimated to be a function of m vectors of holding-specific fixed-effects. To account for the presence of spatial autocorrelation in the data we used a Bayesian approach, adopting a convolution Gaussian prior (Besag 1989, Besag & Mollié 1989, Besag et al. 1991, Mollié 1996) where, in addition to the m vectors of fixed-effects, holding membership of one of $j = 178$ districts were included as unstructured heterogeneity terms U_j and structured (spatially correlated) heterogeneity terms S_j. Districts were arbitrarily defined as regions containing,

on the basis of census data, approximately 20,000 cattle (Chapter 6, Figure 7.2). In this context the probability of holding-level BSE exposure $p(h_{ij})$ was parameterised as:

$$log[p(h_{ij})/\{1-p(h_{ij})\}] = \alpha + \beta_1 x_{1ij} + \ldots + \beta_m x_{mij} + U_j + S_j \qquad (7.1)$$

Non-informative prior distributions were assumed for the fixed-effects regression coefficients β. Priors for the unstructured heterogeneity terms U_j were specified as being normally distributed with mean 0 and precision (inverse variance) τ. For the structured heterogeneity terms S_j we adopted a conditional intrinsic Gaussian autoregressive prior with mean 0 and precision λ (Besag et al. 1991, Besag & Kooperberg 1995). Here, $S_j \mid S_{(k \neq j)}$ has density proportional to $exp[-\frac{\lambda}{2}(a_j S_j - \sum_{k \neq j} w_{jk} S_k)^2]$. In this specification $w_{jk} \geq 0$ is a spatial proximity matrix reflecting the influence of S_k on the expectation of S_j and $a_j > 0$ is a sample size associated with district j. To account for edge effects the ratio of the length of each district's coastline to its total perimeter was determined and the weights w_{ij} were multiplied by one less the coastline to perimeter ratio, for each district pair (Lawson 2001a). In the model reported we used a spatial weight matrix specification based on adjacency (w_{jk} = 1 if districts shared a common boundary and w_{jk} = 0 otherwise).

The following risk factors were controlled-for in the model: (1) herd size (coded as three categories, based on terciles: 1 – 27 cattle, 27 – 53 cattle and > 53 cattle), and (2) enterprise type (dairy and non-dairy). We first constructed a model including only the fixed-effects using a Markov chain Monte Carlo algorithm (Gilks et al. 1996a, Gilks et al. 1996b) implemented in WinBUGS version 1.4 (Spiegelhalter et al. 2000). Residuals for holdings in each of the $j = 178$ districts were summed to provide district-level error terms from the fixed-effects model. Moran's I statistic (Moran 1950) at each of n spatial lags (W^n) were calculated and the 95% confidence interval for I_{W^n} was plotted as a function of W^n. Significant spatial autocorrelation in the district-level error terms justified the use of a mixed-effects model (Bailey & Gatrell 1995).

For the mixed-effects model, gamma hyperpriors for the unstructured and structured heterogeneity precision terms were estimated using the method described by Mollié (2000). Variances of both gamma hyperpriors were taken to be large to reflect uncertainty about the values specified for the prior means. Overall model fit was quantified using the Deviance Information Criterion (Spiegelhalter et al. 2002).

For the Bayesian regression analyses we ran the Markov chain Monte Carlo sampler for 40,000 iterations and discarded the first 1,000 'burn-in' samples. Convergence was visually assessed by plotting cumulative path plots for each of the monitored parameters (Yu 1994, Yu & Mykland 1998, Robert & Casella 1999) and quantified using the Raftery and Lewis convergence diagnostic (Raftery & Lewis 1992a, Raftery & Lewis 1992b). Parallel chains were run using diverse initial values to ensure that convergence was achieved to the same distribution (Gelman 1996). Posterior sample sizes were determined by running sufficient iterations to ensure that the Monte Carlo standard error of the mean was at least one order of magnitude smaller than the posterior standard deviation for each parameter of interest. In each of the final models 100,000 iterations of the Gibbs sampler were used, with the results of every 100th iteration stored to estimate posterior parameter estimates.

7.3 Results

Descriptive statistics of age-at-onset, stratified by birth cohort are shown in Table 7.1. Box and whisker plots of the same data are shown in Figure 7.3. Figure 7.4 shows box and whisker plots of age-at-onset, stratified by birth cohort and conditioned on region of the natal holding. Mean age-at-onset of clinical signs for the 167,366 BSE cases where the onset of clinical signs was before 1 July 1997 was 67 months (Table 7.1). Both the mean and the standard deviation of age-at-onset of clinical signs reduced steadily with each later-born birth cohort. Mean age-at-onset for those cases born before 1 July 1985 was 76 months (SD 20 months): for cases born after 1 July 1991 mean age-at-onset had decreased to 54 months (SD 8 months). For the cohort of cattle born before 1 July 1982, cattle born in the Southeast or Southwest of England had earlier ages-at-onset than cattle born in other regions (Wilcoxon rank sum Z statistic 3.9185; $P < 0.01$).

Of the 93,008 holdings included in the study population 1,245 were estimated to have been exposed to the BSE agent by 1 July 1983 with the earliest known date of exposure occurring in March 1981. Date of birth of the index exposure case was reliably known for 999 of the 1,245 case holdings (80%). Holdings in the south of the country (the Southeast, Eastern and Southwest) had the highest proportion of holdings where the date of birth of the index exposure case was reliably known (90%, 84%, and 81%, respectively, Table

7.2). The natal holding of the index exposure case was reliably known for 1,006 (81%) of the 1,245 case holdings. The Southeast, Eastern, and Southwest regions had the highest proportion of case holdings where the holding of origin of the index exposure case was reliably known (91%, 90%, and 83%, respectively, Table 7.3). Scotland and Wales had the lowest proportion of case holdings where the holding of origin of the index exposure case was reliably known (69% and 56%, respectively).

Figure 7.5 shows, for the 1987 birth cohort, the incidence density of BSE for incubation periods ranging from 24 to 144 months. The highest incidence density occurred at 72 months (6 years) and, in accordance with the method described by Rothman (1981), this was concluded to be the best point estimate of BSE incubation period and is used for this purpose for these analyses.

Figure 7.6 shows, for the variables region and days to index exposure, the distribution of holding counts that were deleted from the data set comprised of complete holding records and the distribution of holding counts for the imputed records. For counts of holdings per region the imputed counts were, on average, no greater than 10% different from the actual value.

Kaplan-Meier survival curves showing the cumulative proportion of holdings that had not experienced a BSE exposure by 1 July 1983, stratified by region are shown in Figure 7.7. Holdings in the Southeast and Southwest of England were exposed to BSE earlier, compared with holdings in the other regions of Great Britain (log rank test statistic 586; df 1; $P < 0.01$). By 1 July 1983 the Southeast and Southwest had the highest proportions of holdings exposed to BSE, of all the regions of Great Britain. Figure 7.8a shows the density of holdings estimated to have been exposed to BSE by 1 July 1983. The estimated number of holding exposures per 100 hectares, stratified by region are shown in Table 7.4. The Southwest of England had the highest density of exposures of all the regions of Great Britain (0.025 exposures per 100 hectares).

Table 7.5 shows the posterior means and 95% Bayesian credible intervals of the odds of holding-level BSE exposure occurring before 1 July 1983. Compared with medium-sized holdings (those with 27 – 53 adult cattle), holdings with greater than 53 adult cattle had 2.1 (95% credible interval 1.9 – 2.5) times the odds of exposure. Compared with non-dairy enterprise types, dairy holdings had 11 (95% credible interval 9.3 – 13) times the odds of exposure. District-level spatially structured heterogeneity terms showed greater

variability than the unstructured heterogeneity terms (variance 0.372 and 0.007, respectively) (Table 7.5 and Figure 7.9).

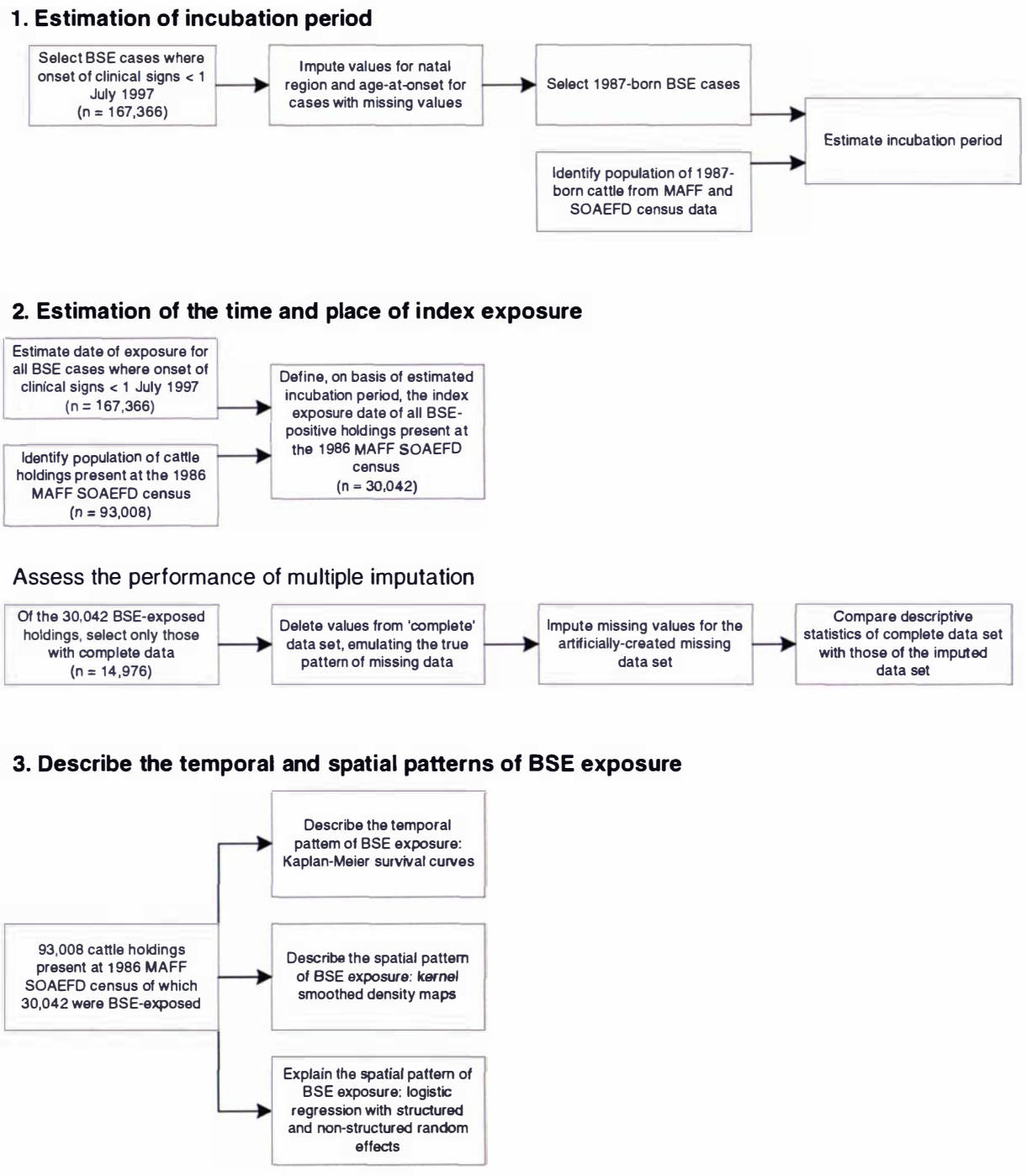


Figure 7.1: Diagramatic representation of the sequence of analyses conducted in this study.

Figure 7.2: Map showing 178 arbitrarily-defined districts of Great Britain. Districst have been defined, on the basis of MAFF census data collected from 1986 to 1997, as areas containing approximately 20,000 adult cattle. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

Table 7.1: Descriptive statistics of age (in months) at onset of clinical signs of BSE where the date of onset of clinical signs was before 1 July 1997, stratified by birth cohort. Age-at-onset has been imputed for those cases with no recorded birth date (n = 3,936).

Cohort	n	Mean	SD	Median	Q1, Q3
Before 1 July 1985	23695	76	20	72	62,85
1 July 1985 to 30 June 1986	19731	70	16	68	58,79
1 July 1986 to 30 June 1987	35344	68	15	66	57,76
1 July 1987 to 30 June 1988	49620	64	14	62	54,72
1 July 1988 to 30 June 1989	19704	64	13	63	55,72
1 July 1989 to 30 June 1990	11026	62	12	62	54,70
1 July 1990 to 30 June 1991	4519	60	10	60	53,66
After 1 July 1991	3726	54	8	54	49,60
Total	167366	67	16	65	56,75

SD: Standard deviation.
Q1: 25th percentile.
Q3: 75th percentile.

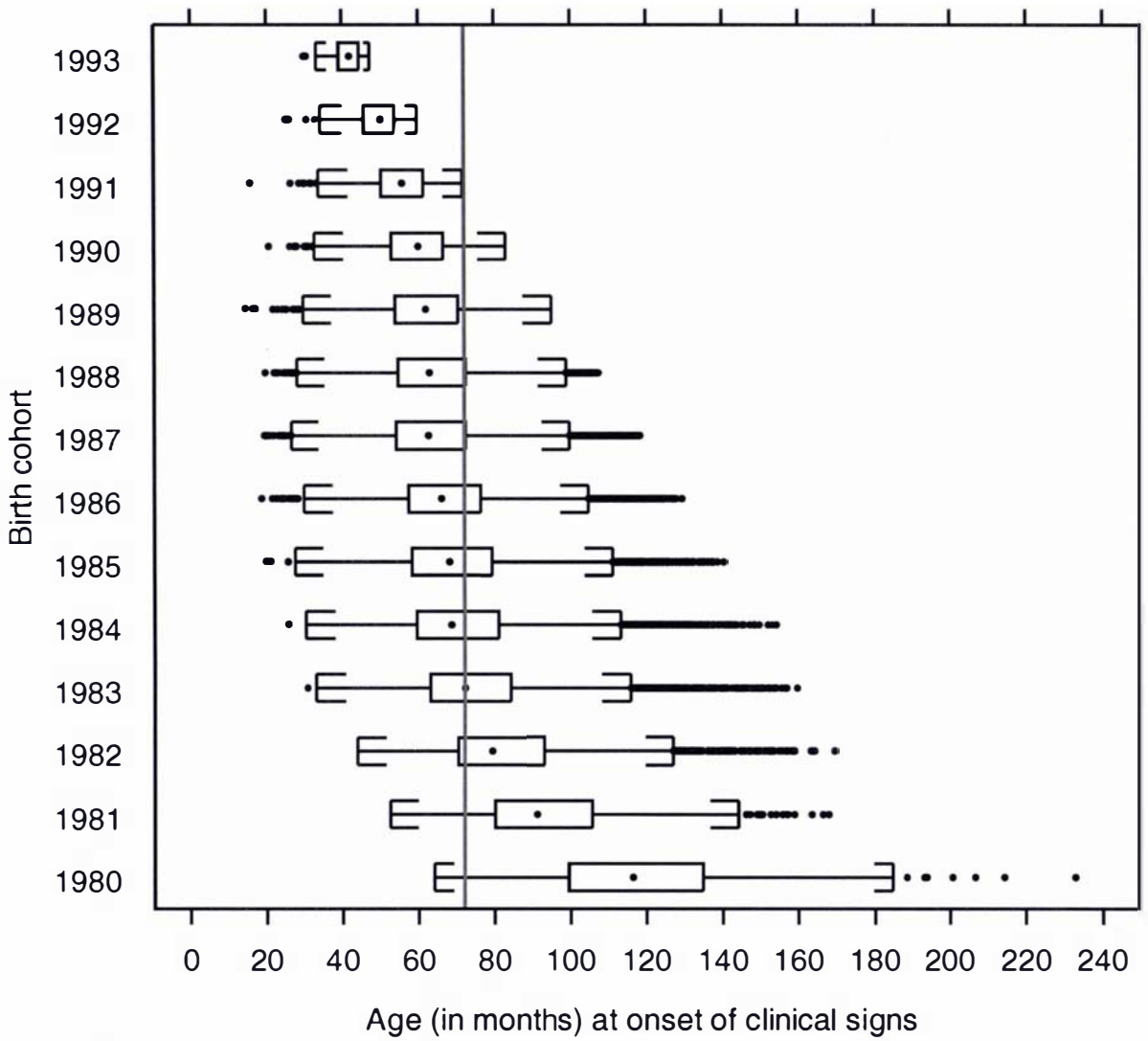


Figure 7.3: Box and whisker plot of age-at-onset of clinical signs of BSE for confirmed cases where the onset of clinical signs was before 1 July 1997, stratified by birth cohort. Enclosed boxes identify the range of the first and third quartile of age-at-onset. Staples identify 1.5 times the inter quartile range. Points beyond the staples identify outliers. Each birth cohort includes all cases born from 1 July of the identified year to 30 June of the following year. The cohort labelled '1980' includes all BSE cases born up to 1 July 1981. The cohort labelled '1993' includes all cases born after, and including, 1 July 1993. The vertical reference line is positioned at age-at-onset of 72 months.

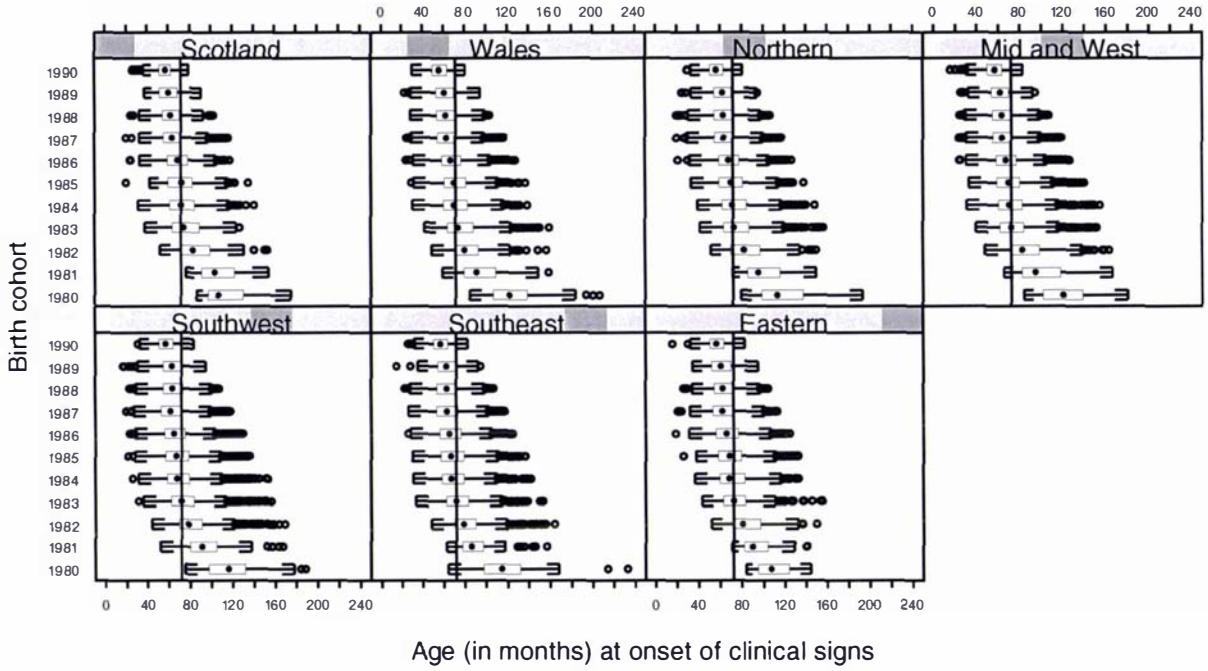


Figure 7.4: Box and whisker plot of age-at-onset of clinical signs of BSE for confirmed cases where the onset of clinical signs was before 1 July 1997, stratified by birth cohort and conditioned by region of natal holding. Enclosed boxes identify the range of the first and third quartile of age-at-onset. Staples identify 1.5 times the inter quartile range. Points beyond the staples identify outliers. Each birth cohort includes all cases born from 1 July of the identified year to 30 June of the following year. The cohort labelled '1980' includes all BSE cases born up to 1 July 1981. The cohort labelled '1990' includes all cases born after, and including, 1 July 1990. The vertical reference line is positioned at age-at-onset of 72 months.

Table 7.2: Counts of holdings estimated to have been exposed to BSE by 1 July 1983, stratified by region and certainty of birth date of the index exposure case. Percentages in brackets refer to the number in each group as a proportion of total exposed holdings in the same region.

Region	Birth date		Total
	Known	Unknown	
Eastern	49 (84%)	9 (16%)	58 (100%)
Mid and West	114 (79%)	31 (21%)	145 (100%)
Northern	64 (80%)	16 (20%)	80 (100%)
Scotland	36 (73%)	13 (27%)	49 (100%)
South east	162 (90%)	19 (10%)	181 (100%)
South west	495 (81%)	114 (19%)	609 (100%)
Wales	79 (64%)	44 (36%)	123 (100%)
Total	999 (80%)	246 (20%)	1245 (100%)

Table 7.3: Counts of holdings estimated to have been exposed to BSE by 1 July 1983, stratified by region and certainty of natal holding of index exposure case. Percentages in brackets refer to the number in each group as a proportion of total exposed holdings in the same region.

Region	Natal holding		Total
	Known	Unknown	
Eastern	52 (90%)	6 (10%)	58 (100%)
Mid and West	117 (81%)	27 (19%)	144 (100%)
Northern	61 (76%)	19 (24%)	80 (100%)
Scotland	34 (69%)	15 (31%)	49 (100%)
South east	165 (91%)	16 (9%)	181 (100%)
South west	508 (83%)	102 (17%)	610 (100%)
Wales	69 (56%)	54 (44%)	123 (100%)
Total	1006 (81%)	239 (19%)	1245 (100%)

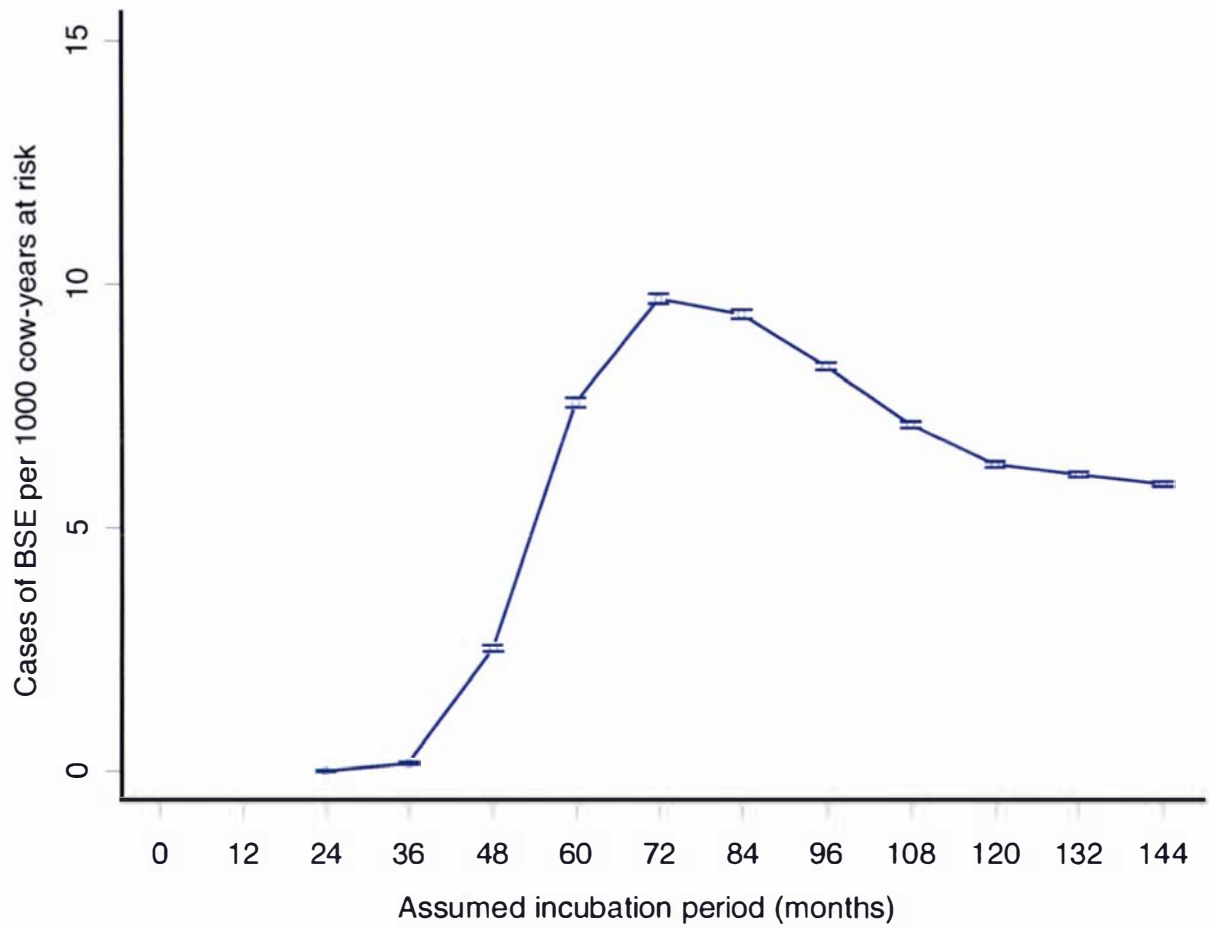
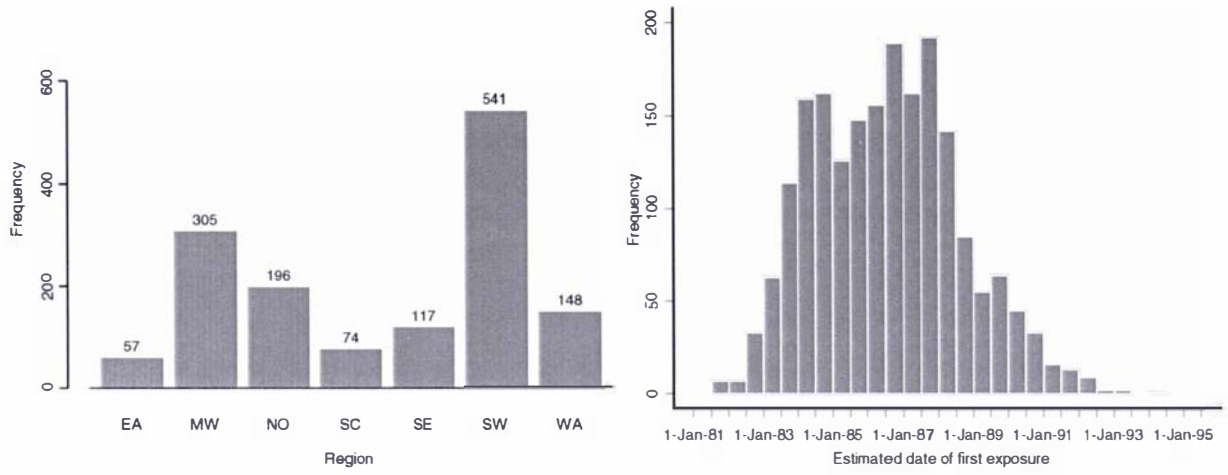
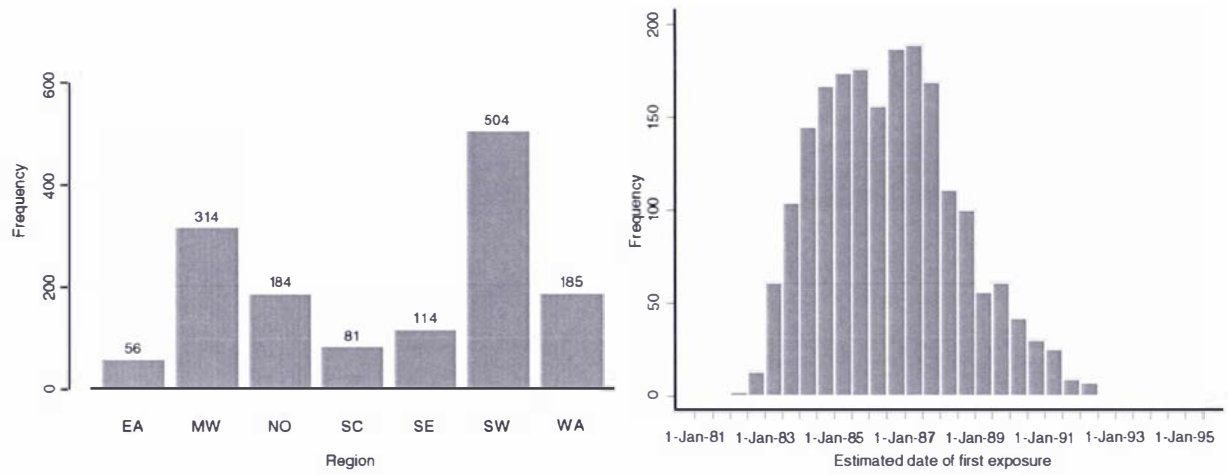


Figure 7.5: Line plot of incidence density of BSE (expressed as cases of BSE per 1000 cow-years at risk) for the 1987-born cohort at assumed incubation periods ranging from 24 to 144 months.



(a) Region counts: complete data

(b) Date of exposure: complete data



(c) Region counts: imputed data

(d) Date of exposure: imputed data

Figure 7.6: Results of analyses conducted to assess the performance of the multiple imputation procedure. Plots (a) and (b) show frequency histograms of holdings by region and date of first holding exposure for data deleted from a complete-record subset of holding data. Plots (c) and (d) show the values imputed for the same variables. Key for regions: EA Eastern; MW Mid and West; NO Northern; SC Scotland; SE Southeast; SW Southwest; WA Wales.

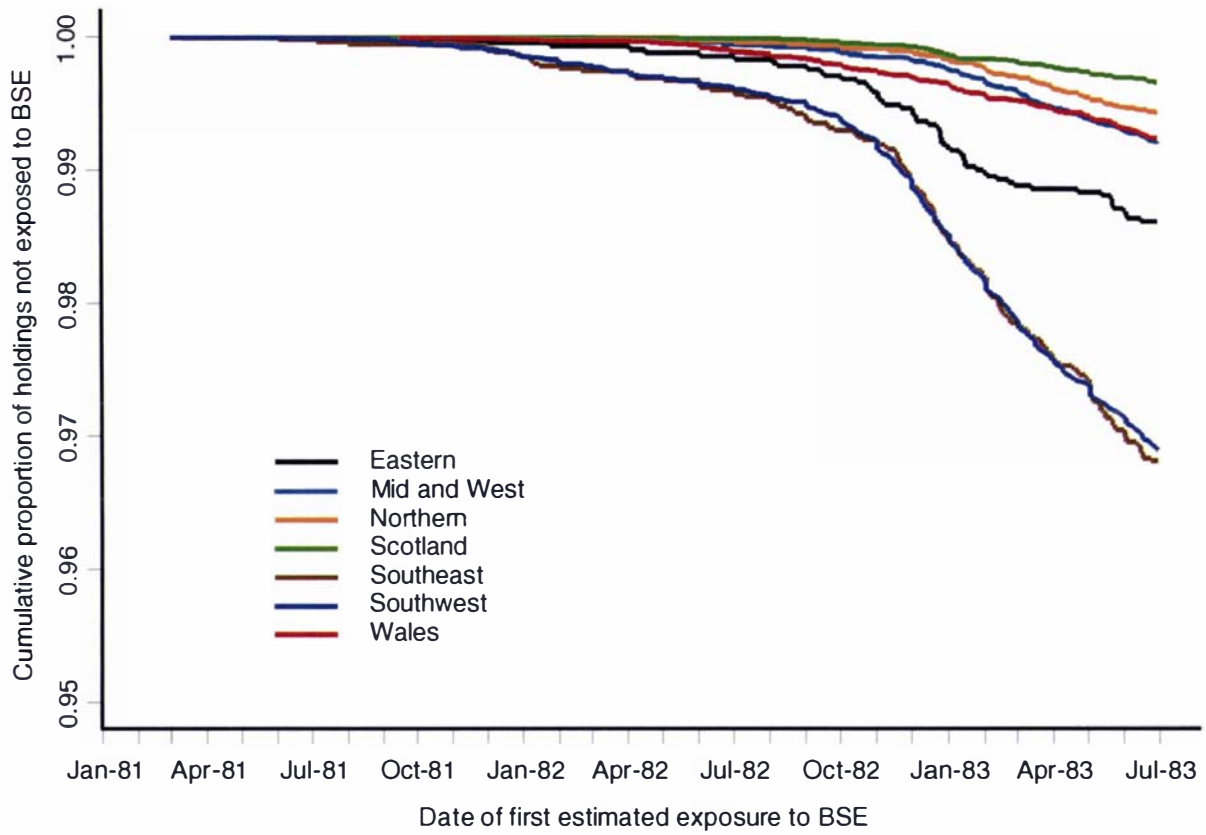
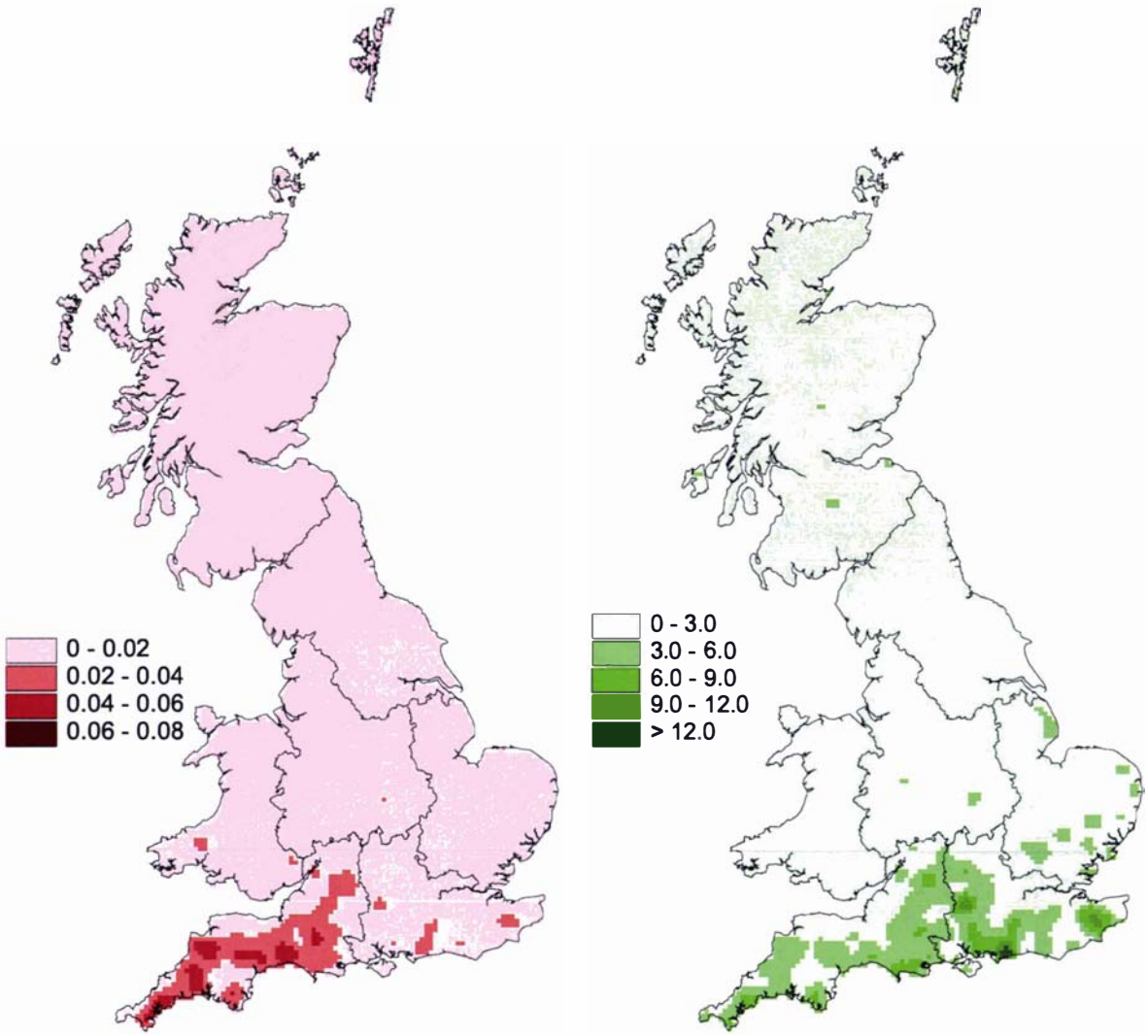


Figure 7.7: Kaplan-Meier survival curves showing the cumulative proportion of holdings that had not experienced a BSE exposure by 1 July 1983, stratified by region. These survival curves have been calculated using the population of cattle holdings in Great Britain on 30 June 1996.



(a) Exposure density

(b) Cumulative incidence

Figure 7.8: Image plots showing: (a) the kernel density of holdings estimated to have been exposed to BSE by 1 July 1983 (expressed as exposed holdings per square kilometres), and (b) the cumulative incidence of BSE-exposed holdings by 1 July 1983 (expressed as exposed holdings per 100 holdings per square kilometre).

Table 7.4: Counts of holdings estimated to have been exposed to BSE by 1 July 1983, estimates of the area (in hectares) of the seven regions of Great Britain, and density of BSE-exposed holdings (expressed as the number of exposures per 100 hectares).

Region	Exposed holdings	Area (hectares $\times 10^6$)	Exposures per 100 hectares
Eastern	58	2.4	0.002
Mid and West	144	2.6	0.006
Northern	80	2.8	0.003
Scotland	49	7.8	0.001
Southeast	181	1.5	0.012
Southwest	610	2.4	0.025
Wales	123	2.0	0.006
Total	1245	22	0.006

Table 7.5: Posterior means and standard deviations of the regression coefficients in the mixed-effects model of factors influencing district-level odds of BSE exposure to 30 June 1983. Structured heterogeneity terms have been based on an adjacency matrix where districts are considered to be neighbours if they share a common border.

Explanatory variable	Posterior mean	SD	MC error	OR	95% CI of OR
Intercept	-5.9046	0.0972	< 0.01		
Holding size					
1 – 27	-1.3912	0.1868	< 0.01	0.25 [a]	0.17 – 0.36
27 – 53	-				
> 53	0.7624	0.0715	< 0.01	2.1	1.9 – 2.5
Holding type					
Dairy	2.4020	0.0923	< 0.01	11	9.3 – 13
Non-dairy	-				
Structured heterogeneity [b]	0.3752	0.1216	< 0.01		
Unstructured heterogeneity [b]	0.0070	0.0083	< 0.01		

Deviance Information Criterion: 2214.
[a] Interpretation: compared with holdings with 27 – 53 cattle, the odds of BSE exposure for holdings with 1 – 27 cattle was reduced by a factor of 0.25 (95% Bayesian credible interval 0.17 – 0.36).
[b] Variance of heterogeneity term.
SD: Standard deviation.
MC error: Monte Carlo error.
OR: Odds ratio.
CI: Bayesian credible interval.

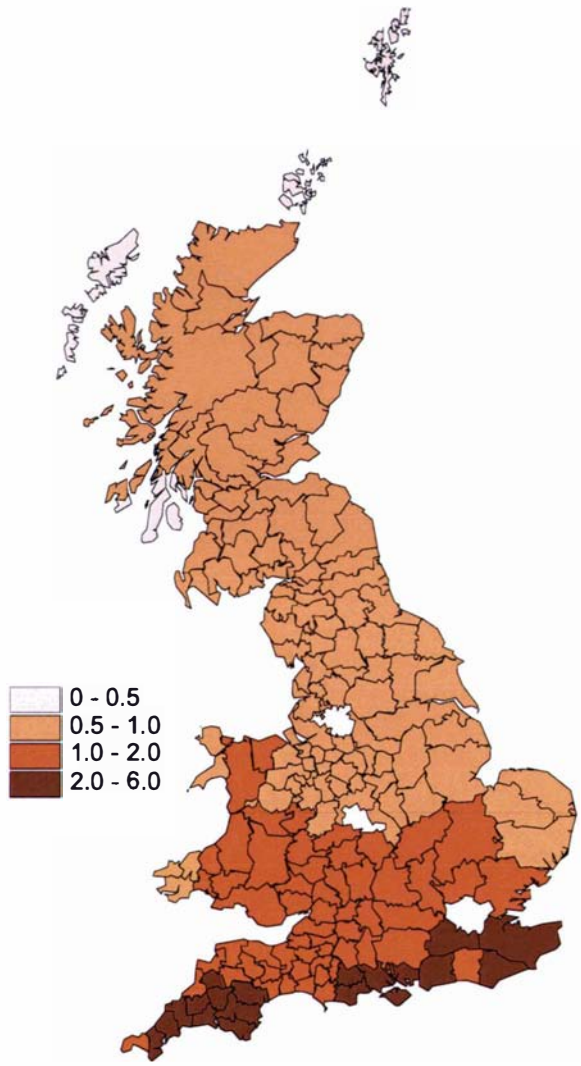


Figure 7.9: Estimates of district-level odds of BSE exposure attributable to the structured heterogeneity terms from the mixed-effects model shown in Table 7.5.

7.4 Discussion

The distribution of age-at-onset of BSE and its relationship with birth cohort (Table 7.1, Figure 7.3) presents two notable features. Firstly, older ages-at-onset for earlier-born cohorts is consistent with a relatively abrupt onset of effective exposure: cattle of all ages being exposed after a given date and the longer-lived subset of this group going on to experience BSE late in life. As the level of recycling of infective agent increased, the probability of exposure at an earlier age increased for later-born cohorts resulting in progressively earlier ages-at-onset of clinical signs. Secondly, the effect of the size of infective dose on incubation period needs clarification. Experimental studies have demonstrated an inverse relationship between dose of infective agent and length of incubation period (Sayers 1999, Sayers 2000). If these findings can be applied to field conditions one would expect the median age-at-onset to have decreased up until the 1987-born cohort (as recycling increased the amount of infective agent in cattle feed) then increase steadily as concentrations of infective agent in feed were reduced in response to the control measures applied. The further reduction in both the median and variability of age-at-onset after 1990 appears inconsistent with this explanation and may be a result of a combination of truncation of the case data set and selective culling in the late 1990s removing higher risk animals from the population before they could express clinical disease. Based on the distribution of ages-at-onset for cattle born before 1987 (Figure 7.3) the estimated 72 month incubation period appears consistent for cattle born in the early phase of the epidemic. Estimation of the incubation period of the disease for later-born cohorts would be of interest to determine if, and by how much, it may have changed during the course of the epidemic. In this way, this analysis could be enhanced by applying a distribution to the incubation period of the disease — a distribution which would have changed throughout the course of the epidemic in response to the level of infective agent circulating in the food chain.

In any attempt to draw conclusions about an epidemic's origin or source, care is required to evaluate the potential biases that might exist in the data available for analysis. The possibility of ascertainment bias exists in the early phase of the British BSE epidemic as a result of regional differences in the level of awareness of the disease among holding managers and veterinarians. In addition to under ascertainment there also exists the possibility that the details recorded for cases identified might also be biased, and the level of

this bias could also have varied regionally.

With respect to ascertainment bias, we note that the date of onset of clinical signs in 976 (78%) of the estimated 1,245 pre-1 July 1983 index exposures occurred after 1 January 1988. Given the relatively widespread publicity about the disease that commenced in early 1988 (the Veterinary Record published seven letters related to BSE in January and February 1988) we reason that the bulk of exposures reported in this study are derived from cases that were diagnosed when knowledge of the disease was relatively widespread among the veterinary community. Assuming that the tendency of holding managers to report cases to veterinarians did not vary geographically, we considered the exposures reported here to be a relatively unbiased sample of all BSE exposures that were occurring at the time. Although a possibly substantial number of cases were missed nationwide prior to 1988 (see for example Eddy 1995), because the timing and location of these will never be precisely known, our approach has been to place what is thought to be reasonably complete account of the early exposure pattern (offset from the true epidemic start date by an unknown period) in context with what has been learnt about the disease in terms of its ability to be recycled in the absence of control measures. In this way, more definitive conclusions about the origin of the disease in Great Britain may be reached.

Whereas the accuracy of details recorded for early cases was variable between areas (Tables 7.2 and Table 7.3) the descriptive statistics presented in Figure 7.6 provide evidence that the multiple imputation technique accounted for regional differences in the accuracy of recorded case details. Critical for this were two factors: (1) a relatively small proportion of missing values (in this case the level of 'missingness' was 20% for birth date and 17% for natal holding), and (2) the presence of additional variables in the data set to support the imputation algorithm (Schafer 1997).

Holdings in the Southeast and Southwest regions of Great Britain experienced a constant risk of additional holdings being exposed from December 1981 to October 1982 with an inflection point occurring in October 1982, after which time there was a substantially increased risk of additional holdings being exposed. In contrast, appreciable numbers of exposures in other regions commenced after mid 1982 (Figure 7.7). Given that exposure dates have been back-calculated from date of onset of clinical signs using a fixed incubation period of 72 months it is not surprising that date of first holding exposure follows a similar pattern to date of onset of clinical signs of holding index cases, described in an

earlier paper in this series (Wilesmith et al. 2000). In contrast to the pattern of date of onset of clinical signs of each holding's index case however, a notable feature of Figure 7.7 is the almost identical constant rate of exposure among holdings in the Southeast and Southwest regions of England, suggestive that the process of recycling (Wilesmith 1996) had established itself among cattle holdings in these areas of the country by December 1981.

By July 1983 the highest densities of cattle holdings that had been exposed to BSE were in the Southwest of England — areas with greater than 0.05 exposed holdings per 100 hectares were in the south of Cornwall, the north of Devon and the Somerset-Dorset border, Figure 7.8a. Although smaller numbers of holdings were exposed in West Sussex, lower numbers of holdings at risk in this county resulted in relatively high estimates of cumulative incidence (Figure 7.8b).

Plots of the district-level spatial heterogeneity terms from the mixed-effects model identify areas where there was a high probability of BSE exposure that was not accounted-for by the established risks of holding size or enterprise type. Figure 7.9 shows a relatively uniform distribution of unexplained, spatially-correlated risk across the south of England, highest in the counties of Cornwall, Devon, Hampshire, Sussex and Kent. This implies that, for the period investigated, the meat and bone meal supplied to these areas contained relatively large amounts of infective agent, sufficient to result in an excess of exposure in these areas that was unaccounted-for by either holding size or type.

It is unlikely that any single analysis of the accumulated epidemic data will ever provide definitive answers about the origin of bovine spongiform encephalopathy in Great Britain. Rather, careful assessment of the early phase of the epidemic in the context of lessons learnt about the epidemiological behaviour of the disease can help to provide further insight. On the basis of the analyses presented here and in earlier studies in this series (Wilesmith et al. 2000, Stevenson et al. 2000a, Stevenson et al. 2000b), the following statements can be made:

1. The wide distribution of age-at-onset for early-born cohorts (Figure 7.3) is indicative of an abrupt onset of effective exposure around 1981. As infection became established in each region, greater concentrations of BSE agent in the food chain resulted in most infections occurring in early life and reductions in ages-at-onset for later-born cohorts.

2. Given the accumulated knowledge of how BSE recycled via the cattle food chain

it appears that by July 1983 this process was established among cattle holdings in the Southeast and Southwest of England. We base this conclusion on the relatively high rate of holding-level exposure evident from July 1981 to July 1983 (Figure 7.7) and the excess BSE exposure risk that was not accounted-for by the established risk factors of holding type and size (Table 7.5 and Figure 7.9). Ages-at-onset for the pre-1982 birth cohorts in the Southeast and Southwest were less than those recorded for similar cohorts in other regions, indicative of early life exposures occurring in these regions and indirectly supportive of established recycling in these regions before July 1983. These findings support the conjecture that BSE had been present in Great Britain for some years prior to 1983 undergoing undetected iterations of recycling in the cattle food chain (Anonymous 2000, Review Committee 2001).

3. The simplest explanation for these findings is that contamination of the cattle feed supply commenced at an unknown date during the 1970s in the south of Great Britain, most likely the Southwest. An alternative explanation is that the initial infection seed was multifocal (that is, as an extended common source). In this case we reason that the disease was not able to establish itself in areas of the country where the demographics of the cattle population and the structure of the local rendering industry provided less favourable conditions for recycling. On this hypothesis the disease became established in areas of the country that provided favourable recycling conditions (that is, the Southeast and Southwest) enabling a critical level of agent in the cattle population to be reached, and infection subsequently 'spilling over' into other areas.

Since it would appear that early cases went undiagnosed for several years and hence details of the early phase of the epidemic can only be guessed at, it is unlikely the relative merits of these two explanations can be resolved by analysis. Modelling the epidemic offers one means for identifying which of these two scenarios is more likely.

CHAPTER 8

Spatio-temporal epidemiology of foot-and-mouth disease in two areas of Great Britain in 2001

8.1 Introduction

The 2001 foot-and-mouth disease (FMD) epidemic in Great Britain has been a timely reminder of the value of geo-referenced farm data in the management of animal disease outbreaks. Where many personnel are deployed to carry out control activities, accurate and up-to-date maps showing farm locations and herd disease status are key (Morris et al. 2002). In addition, geo-referenced data are essential — firstly for monitoring progress (Sanson et al. 1991) and secondly for predictive modelling of alternative control strategies (Howard & Donnelly 2000, Ferguson et al. 2001, Kao 2001, Keeling et al. 2001, Morris et al. 2001).

In the absence of vaccination, strategies for dealing with a FMD epidemic involve four main elements: (1) rapid slaughter of all stock on premises identified as infected, (2) pre-emptive slaughtering of stock on premises on the basis that they had been exposed to infection (termed 'pre-emptive' culling in this paper), (3) imposition of bans on the movement of stock and/or people within defined infected areas, and (4) surveillance to detect infected premises (Sanson 1993, Radostits et al. 2000). The removal of stock on infected premises and those potentially exposed in the immediate vicinity is to limit local spread, and restrictions on animal and animal-product movement is to reduce the spread of virus across larger distances.

The magnitude of FMD epidemics vary in response to the scale of the original infec-

tion challenge, the geographical distribution of the animal population at risk and the effectiveness of control efforts during eradication. In dynamic outbreak situations (where, as a result of pre-emptive culling, the animal population-at-risk is constantly changing and control measures vary in their effectiveness and intensity of application), the spatial and temporal components of disease risk can change markedly throughout an epidemic's course. Within the epidemic of FMD in Great Britain in 2001, there were apparent differences between geographical areas in response to control measures applied. In this paper we considered the two counties which experienced the greatest incidence of disease (Cumbria and Devon). Our first aim was to describe the spatial and temporal features of the incidence of FMD among farm holdings in these areas. The second aim was to describe the spatio-temporal interaction of infection risk and to describe changes in the components of this interaction that occurred over the course of the epidemic. Better appreciation of these aspects of an epidemic means that decisions concerning practical issues (such as pre-emptive culling radii, surveillance periods, the resources required and the appropriate effort to trace the spread of infection from infected premises) can be made with greater objectivity and modified appropriately for individual outbreaks.

8.2 Materials and methods

The areas investigated are shown in Figure 8.1. Both were 2,500 km^2 (50 km $\times$ 50 km): the first in the west of the county of Cumbria and the second in the north of the county of Devon. These areas were selected because they were regions where the behaviour of the disease (that is, the spread and ability to be controlled) showed marked differences (Gibbens et al. 2001). The period of interest was from 20 February 2001 to 30 September 2001.

The population of interest included all farm holdings containing at least one of the five FMD-susceptible domestic species (cattle, sheep, goats, deer or pigs) that were located within each of the defined areas. Holding-associated data were retrieved from the agricultural-census data collected by the Ministry of Agriculture, Fisheries and Food (MAFF 2001) for the 12 months to June 2000 and subsequently revised throughout the course of the 2001 FMD epidemic. Data recorded for each holding included the county-parish-holding identifier, easting and northing coordinates of the main farm building and

the number of adults present of each of the FMD-susceptible species. On the basis of adult-stock counts, a holding-level type classification was calculated as follows. Holdings where 80% of the total animal population on the day of census were dairy or beef suckler animals were designated as 'cattle' enterprise types. Holdings where 80% of the total animal population were a single species were designated as enterprises of that type. Holdings unable to be classified by this method were designated as 'mixed'.

Cases were holdings located within the boundaries of each defined study area that were declared as FMD-infected premises under the Foot-and-Mouth Disease Order 1983 (HMSO 1983). These were holdings where the clinical signs of FMD were observed by a Department for Environment, Food and Rural Affairs (DEFRA) investigating officer or holdings where there was laboratory confirmation of infection within the herd or flock (typically the case for holdings depopulated as direct contacts or holdings that were culled on suspicion of infection). Details of case holdings (county-parish-holding identifier, estimated infection date, confirmation date and cull date) were retrieved from the Disease Control System Database (Gibbens et al. 2001) and merged with the agricultural-census database.

We used three independent methods to analyse these data. Firstly, the temporal evolution of the epidemic was described using survival analyses. Secondly, the spatial distribution of infected premises was described using kernel density estimation methods. Thirdly, the spatio-temporal interaction among cases was described using the space-time K-function.

For the survival analyses, the outcome was the estimated FMD infection date for case holdings. Estimated infection date was recorded at the time of diagnosis, and was determined either on the basis of the history provided by the holding manager or on the estimated age of lesions observed at the time of examination. For the later case, estimated infection date was the date of examination minus the age (in days) of the oldest lesions identified minus an additional 5 days to represent an incubation period for the disease (Gibbens & Wilesmith 2002). Holdings that were culled pre-emptively or culled on suspicion of infection but not confirmed as infected were right-censored on the date of cull. Holdings that remained free of infection throughout the period of interest were right-censored at the end of the observation period on 30 September 2001. The weekly hazard of FMD (representing the weekly probability of becoming FMD infected, given

that a holding remained free of infection to at least the specified point in time) was computed using the LOCFIT library (Loader 1999) implemented in the R statistical package (Ihaka & Gentleman 1996). We compared the hazard of FMD for two groups: cattle holdings versus sheep, goat, deer, pig and mixed holdings (considered as a single group under the title of 'other'). Instantaneous hazard functions of this type allowed high-risk and low-risk periods for infection to be readily identified.

To describe the spatial pattern of infection among holdings in each area, four time periods were defined starting from 20 February 2001. These periods were thought broadly to represent the natural phases of the epidemic: the first (20 February to 28 March 2001) was the period of rapid spread of the disease; the second (29 March to 23 May 2001) a period in which there was a sharp reduction in incidence; the third (24 May to 18 July 2001) a period of relatively constant incidence, and the fourth (19 July to 30 September 2001) the period when eradication was achieved. For each period, two density surfaces representing the number of holdings per km^2 were constructed using a Gaussian-kernel smoothing function: the first (numerator) was based on all holdings identified as FMD-positive during the period and the second (denominator) on holdings that were present at the start of the period and considered at risk. The ratio of the density surface of FMD-positive holdings to the density surface of the population of holdings at risk at the start of each period provided a relief map of the distribution of the proportion of holdings per km^2 which became FMD-affected (Bithell 1990, Lawson & Williams 1994, Bowman & Azzalini 1997). These plots (termed 'extraction maps' by Lawson and Williams, 1993) provided an objective means for identifying areas where there were relatively high rates of infection. Bandwidth parameters for the kernel functions (used to control for the degree of smoothing of the estimated density surface) were calculated by cross validation (Bowman & Azzalini 1997) and were based on the holding population at risk at the start of each time interval. To account for edge effects (Lawson 2001a), details of holdings located within a 5 km guard area around the periphery of each study area were included in each data set. Kernel density surfaces were computed on the basis of holdings located in both the study area and the guard area; only the density estimates for each defined 2,500 km^2 area are reported.

The spatio-temporal interaction of infection risk was described using the space-time K-function (Diggle et al. 1995) implemented in the SPLANCS library (Rowlingson &

Diggle 1993, Bivand & Gebhardt 2000) in R. Here, the respective data sets were restricted to include case holdings only. The space-time K-function $\hat{K}(s,t)$ was calculated as the cumulative number of cases that were expected within distance s and time interval t of an arbitrarily-selected case divided by the intensity λ (the expected number of events per unit space and per unit time). Letting $\hat{K}_S(s)$ define the K-function in space and $\hat{K}_T(t)$ define the K-function in time, the K-function difference $\hat{D}(s,t)$ was computed as:

$$\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}_S(s)\hat{K}_T(t) \tag{8.1}$$

which estimates the cumulative number of cases expected within distance s and time interval t of an arbitrarily-selected case that were attributable to the interaction between space and time, scaled by λ. To facilitate comparison among time periods, we computed $D_0(s,t)$:

$$\hat{D}_0(s,t) = \frac{\hat{D}(s,t)}{\hat{K}_S(s)\hat{K}_T(t)} \tag{8.2}$$

which estimates, for given distance and time separations, the proportional increase in cases attributable to space-time interaction (Diggle et al. 1995). Separate space-time K-function analyses were conducted for each of the four time periods described using maximum distance and time separations of 10 km and 21 days, respectively. The use of a small maximum distance separation (relative to the overall dimensions of the study area) reduced the likely influence of first-order, trend effects on each K-function that was computed. Space-time K-functions were computed using details of holdings within each 2,500 km^2 study area: the K-function implementation within SPLANCS providing a correction term for edge effects. Crude estimates of the confidence limits for $\hat{D}_0(s,t)$ were determined by estimating the lower and upper limits of $\hat{D}(s,t) \times \lambda$, assuming that $\hat{D}(s,t) \times \lambda$ (being an estimate of count data) followed a Poisson distribution.

A formal test for the presence of space-time interaction was performed by conducting m Monte Carlo simulations in which each of the n case events were labelled with the observed n time 'markers'. A total of m estimates of $\hat{D}(s,t)$ were obtained and for each simulation, the sum of $\hat{D}(s,t)$ over all s and t were obtained. The sum of $\hat{D}(s,t)$ for the observed data then was ranked among the empirical frequency of m sums. If the observed sum ranked kth largest (or smallest) the one-sided attained significance level was k/m.

8.3 Results

Tables 8.1 and 8.2 provide counts of holdings present in each area at the start of the study period stratified by holding type and FMD status on 30 September 2001. The density of holdings in the Cumbria study area was 0.6 km^2, approximately half that of Devon (1.3 holdings per km^2). In Cumbria, there was one holding pre-emptively culled for each holding diagnosed FMD-positive. In Devon, 3.2 holdings were culled pre-emptively for each holding diagnosed FMD-positive.

In Cumbria the hazard of FMD peaked in the week commencing 8 March 2001 (Figure 8.2a). During this period, the infection hazard was greater for cattle holdings (6.0%) than for other holding types (2.5%). Infection hazard rapidly declined from 28 March as additional control measures were applied. From April through to August, the hazard of FMD for cattle holdings remained static at 0.8% — whereas for other holding types there was a steady rise in hazard (reaching 1.4% by the end of July). In Devon the peak infection hazard (0.7%) occurred in the week commencing 15 March followed by a rapid decline for both holding classifications.

There were marked changes in the spatial pattern of incident holdings in Cumbria for the four time periods. From 20 February to 28 March the highest density of incident holdings was in an area 570 km^2 north west of (and including) the town of Penrith (Figure 8.3a). The next period was characterised by incident holdings appearing further to the north west and south east of this primary focus — indicative of the infection moving into a population of susceptible holdings (Figure 8.3b). In the third period, a secondary focus of incident holdings of 500 km^2 developed within an area bounded by the towns of Windermere, Penrith, and Brough (Figure 8.3c). In Devon a single, relatively high-density focus of incident holdings was present in the first period (Figure 8.4a). In the second period there were no high density zones identifiable in Devon: incident holdings during this time were randomly distributed. The widely-dispersed pattern of FMD-affected holdings in Figure 8.4b and the absence of new infections after 31 May 2001 reflects the effectiveness of the control measures applied in this area.

In the surface plots of $\hat{D}_0(s,t)$ for Cumbria (Figs. 8.5a – 8.5d) surface values where $\hat{D}_0(s,t)$ exceeds 1.0 are marked (showing the distance and time separation from an arbitrarily-selected case where there was a 100% increase in FMD case numbers attributable to space-time interaction). A Monte Carlo test for space-time interaction was based on val-

ues of (s, t) in a 30 $\times$ 30 grid running in each coordinate direction. The observed values of the test statistic for each period were 0.852, 2.42, 3.34, and 3.89 $\times$ 10^6, whereas the values from 99 Monte Carlo permutations of the times ranged from -0.433 to 0.815, -1.31 to 1.30, -0.714 to 1.51, and -1.02 to 1.67 $\times$ 10^6 — corresponding to a one-sided P $<$ 0.01 in each instance. For the Devon study area (Figure 8.6a and 8.6b) the test statistics for the two periods were 0.935 and 0.124 $\times$ 10^6. The values from 99 Monte Carlo permutations of the times ranged from -0.817 to 1.62 and -0.58 to 0.56 $\times$ 10^6, corresponding to a one-sided P of 0.04 and 0.39 for the earlier and later periods, respectively.

The median days between estimated infection and confirmation date in the Cumbria study area was 6 (25th quartile 6 days, 75th quartile 7 days) (Figure 8.7); median days between confirmation and cull date was 2 (25th quartile 1 day, 75th quartile 2 days) (Figure 8.8). For the Devon study area, the median days between estimated infection and confirmation date was 7 (25th quartile 6 days, 75th quartile 7 days). Median days between confirmation and cull date was 2 (25th quartile 1 day, 75th quartile 2 day).

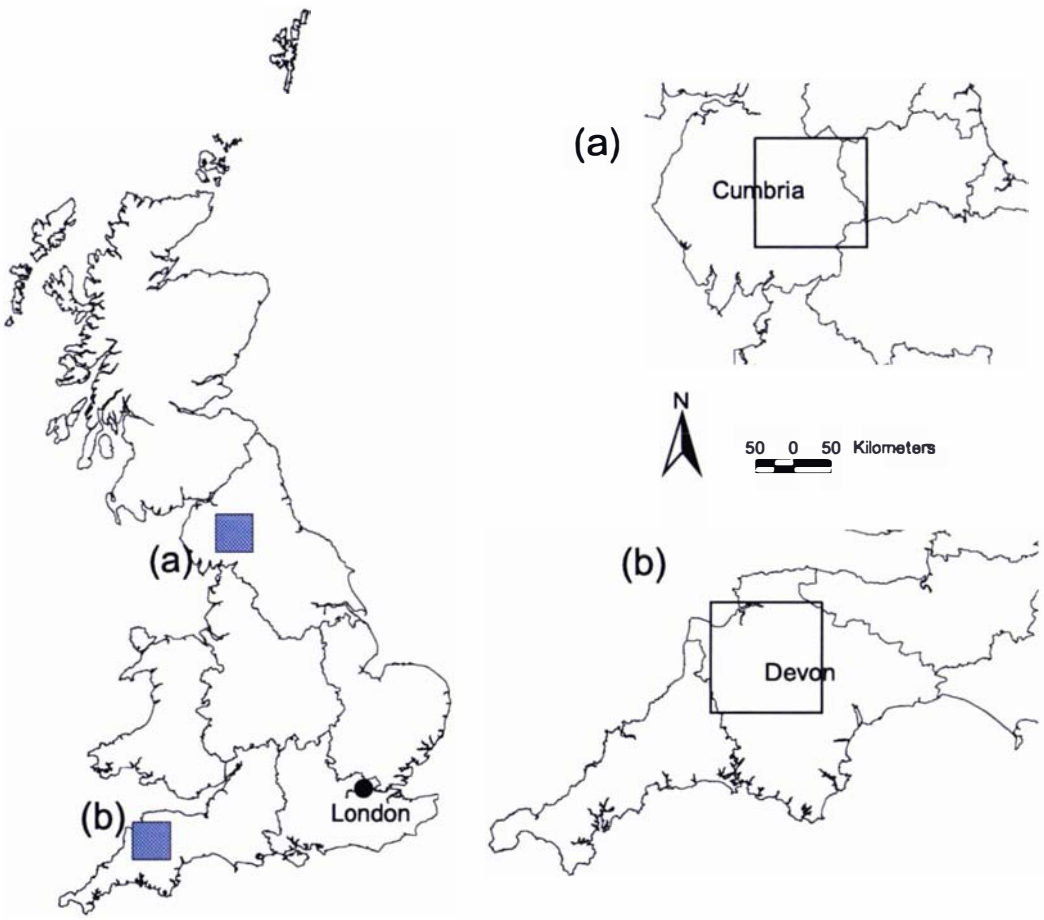


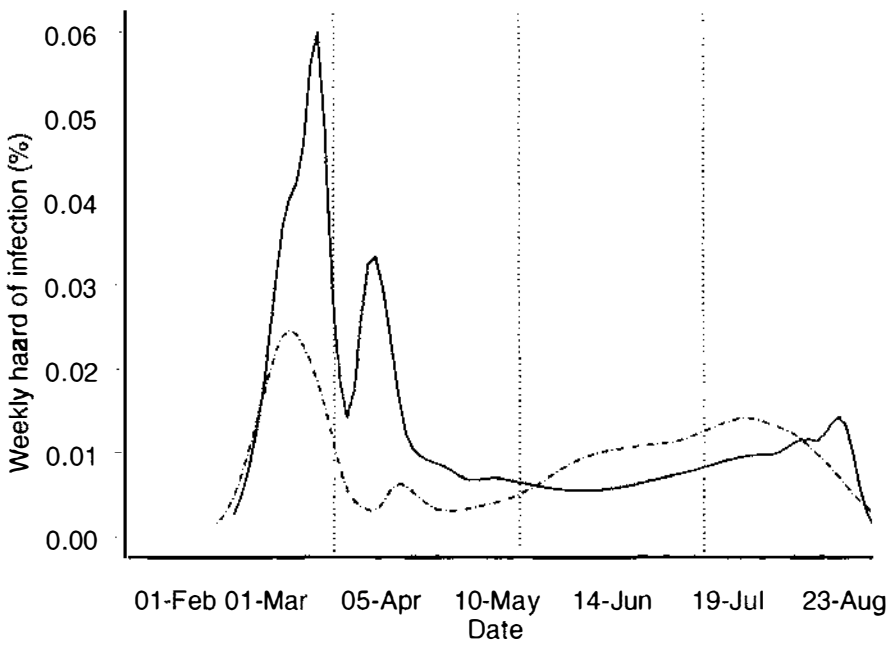
Figure 8.1: Map of Great Britain showing the location of the study areas described in this study: (a) Cumbria, (b) Devon.

Table 8.1: Counts of holdings present in the Cumbria study area (Great Britain) on 24 February 2001 stratified by holding type classification and foot-and-mouth disease status on 30 September 2001.

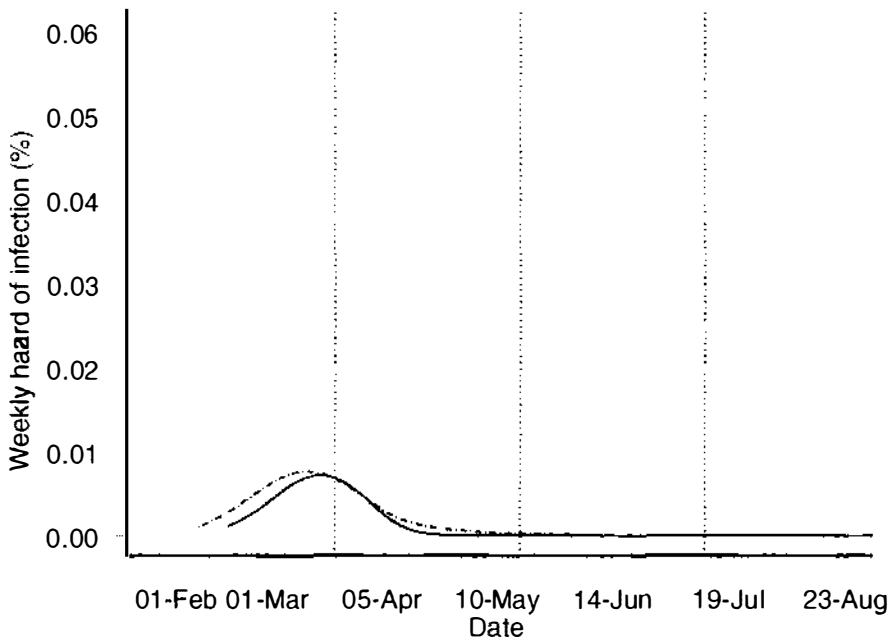
Holding type	FMD-positive	Pre-emptively slaughtered	Slaughtered on suspicion	FMD-negative	Total
Cattle	76	36	0	148	260
Other					
Deer	0	1	0	0	1
Goat	0	1	0	11	12
Mixed	11	1	0	2	14
Pig	3	2	0	6	11
Sheep	256	288	5	708	1257

Table 8.2: Counts of holdings present in the Devon study area (Great Britain) on 24 February 2001 stratified by holding type classification and foot-and-mouth disease status on 30 September 2001.

Holding type	FMD-positive	Pre-emptively slaughtered	Slaughtered on suspicion	FMD-negative	Total
Cattle	45	162	2	1024	1233
Other					
Deer	0	0	1	10	11
Goat	1	7	0	81	89
Mixed	9	4	0	34	47
Pig	2	15	0	56	73
Sheep	85	267	14	1475	1841

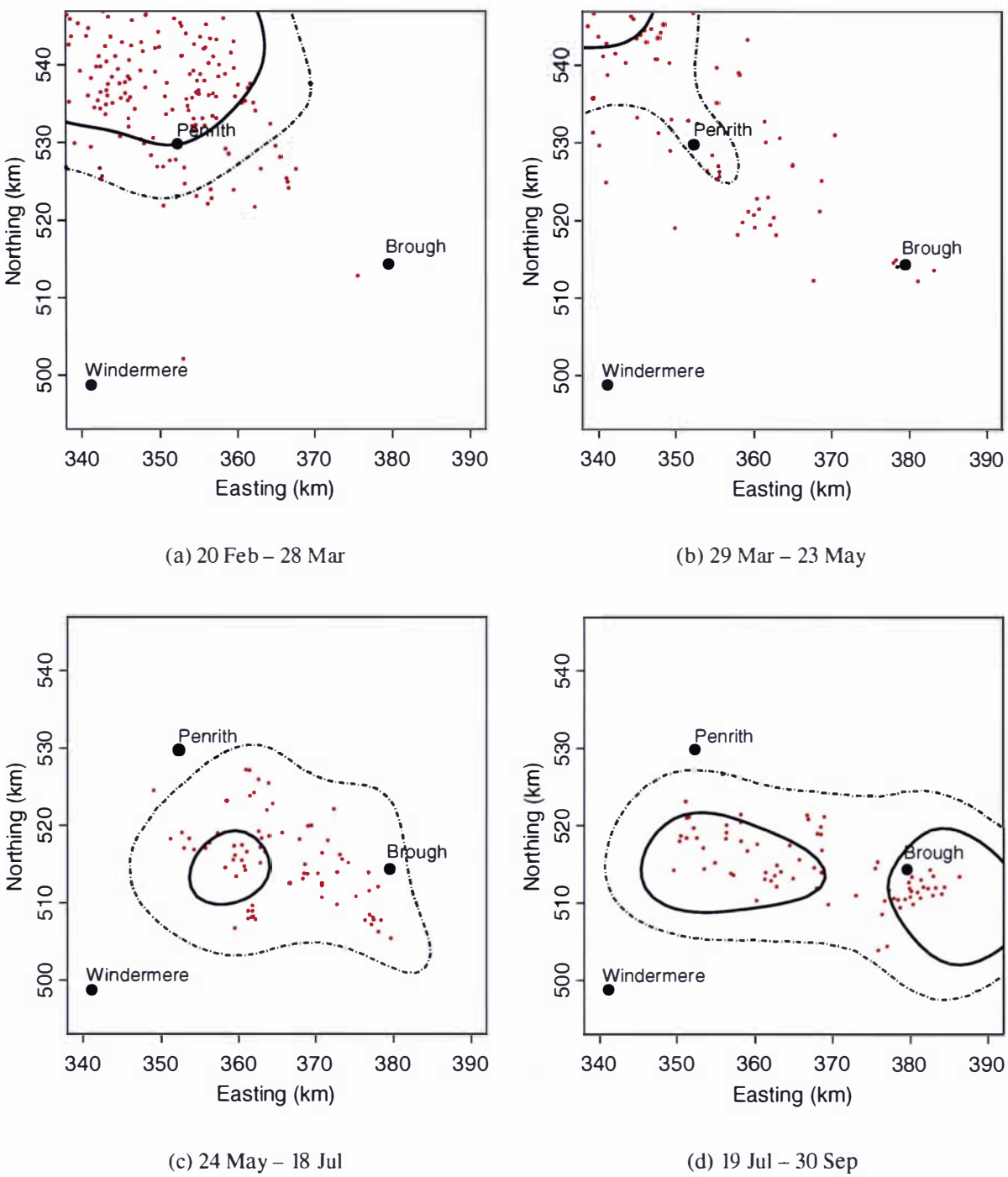


(a) Cumbria



(b) Devon

Figure 8.2: Weekly hazard of foot-and-mouth disease infection for cattle holdings (solid lines) and 'other' holdings (dashed lines) in Cumbria and Devon (Great Britain). Vertical lines mark the four time periods described (20 February – 28 March 2001, 29 March – 23 May 2001, 24 May – 18 July 2001, and 19 July – 30 September 2001, respectively).



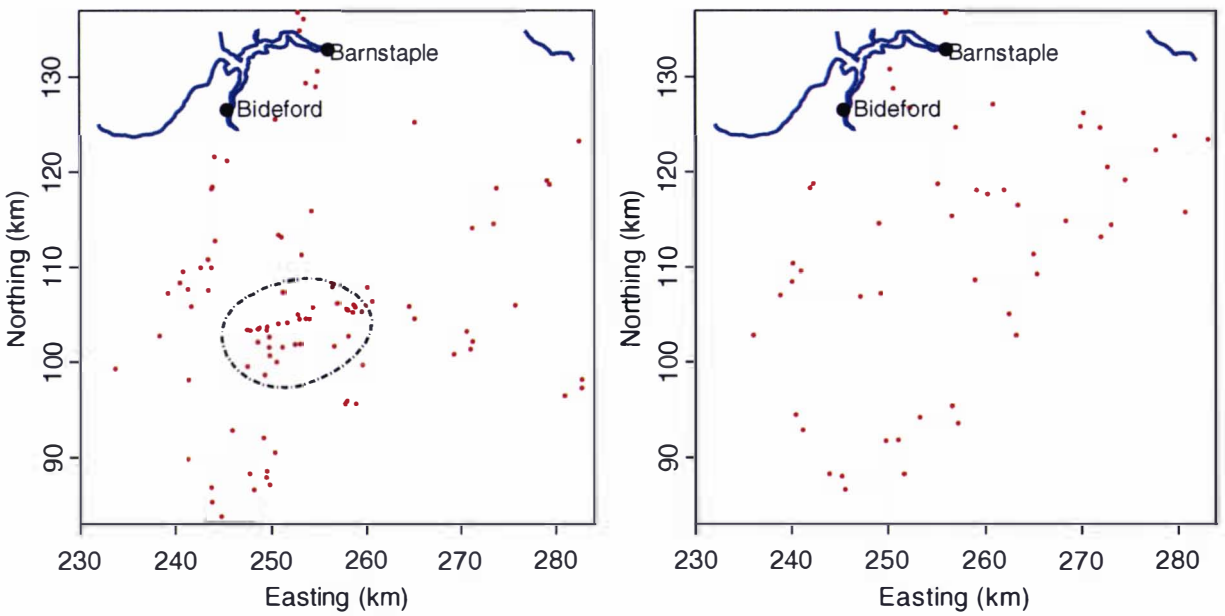
(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

(c) 24 May – 18 Jul

(d) 19 Jul – 30 Sep

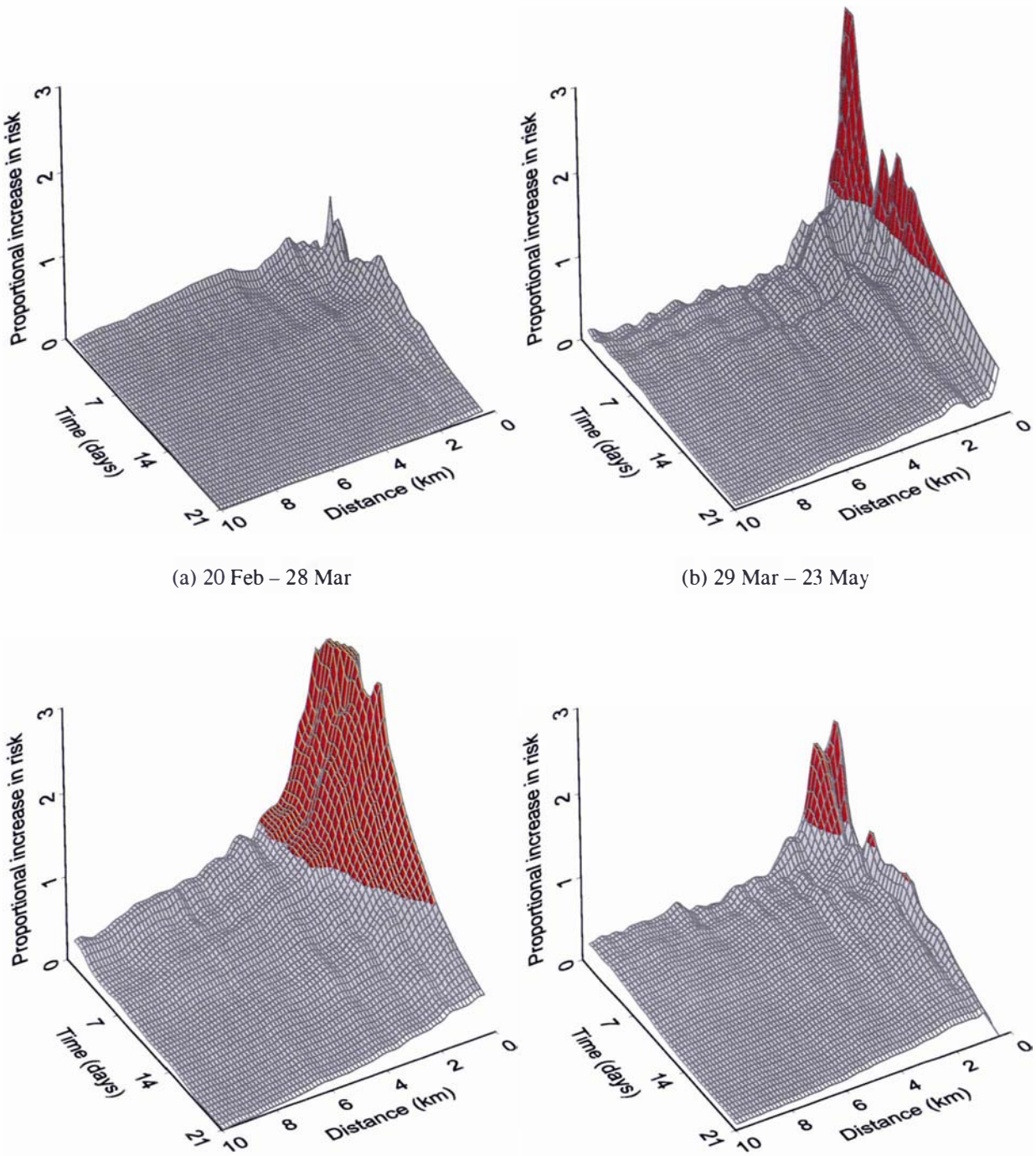
Figure 8.3: Contour plots showing locations in Cumbria (Great Britain) where > 10% (dashed lines) and > 20% (solid lines) of holdings were diagnosed with foot-and-mouth disease. Point locations of incident holdings in each period have been superimposed, for reference.



(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

Figure 8.4: Contour plots showing locations in Devon (Great Britain) where $> 10\%$ (dashed lines) of holdings were diagnosed with foot-and-mouth disease. Point locations of incident holdings in each period have been superimposed, for reference.



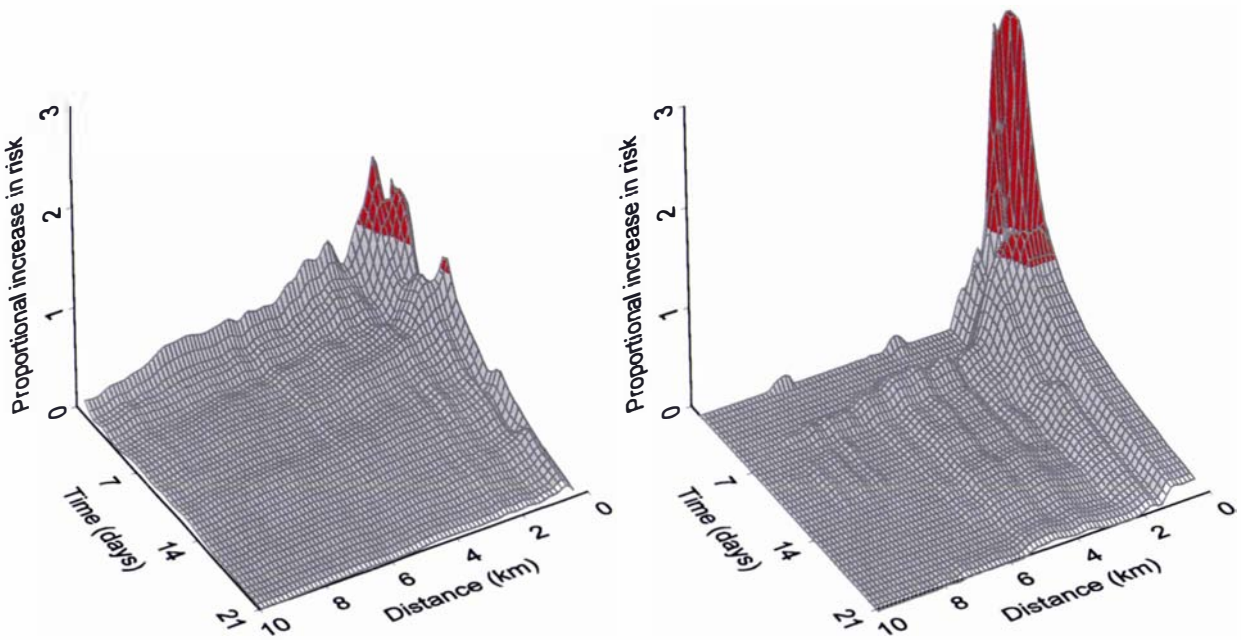
(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

(c) 24 May – 18 Jul

(d) 19 Jul – 30 Sep

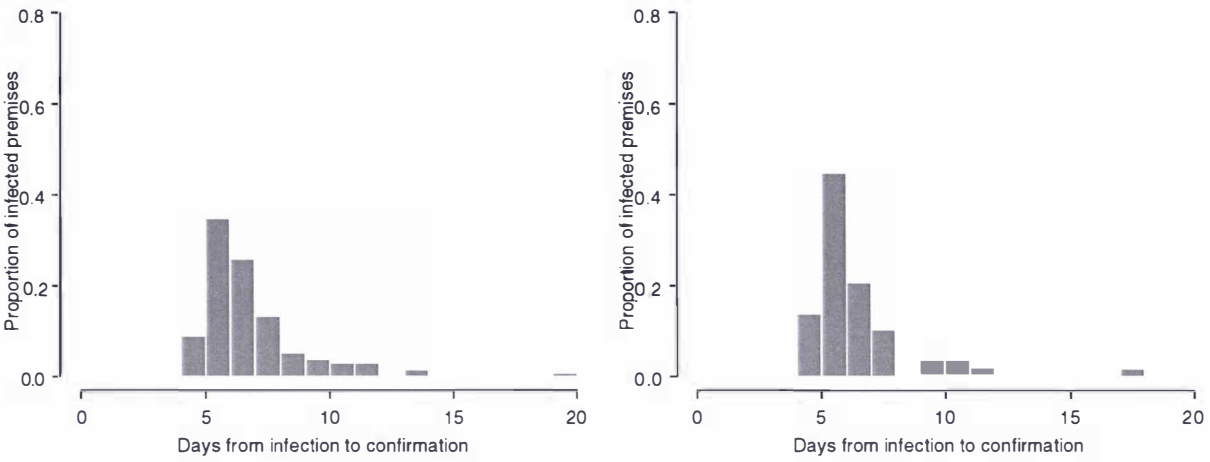
Figure 8.5: Spatio-temporal interaction of foot-and-mouth disease risk among infected premises in Cumbria (Great Britain). On each surface, the dark-shaded area shows the distance-time separations where the proportional increase in risk attributable to space-time interaction was greater than one.



(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

Figure 8.6: Spatio-temporal interaction of foot-and-mouth disease risk among infected premises in Devon (Great Britain). On each surface, the dark-shaded area shows the distance-time separations where the proportional increase in risk attributable to space-time interaction was greater than one.



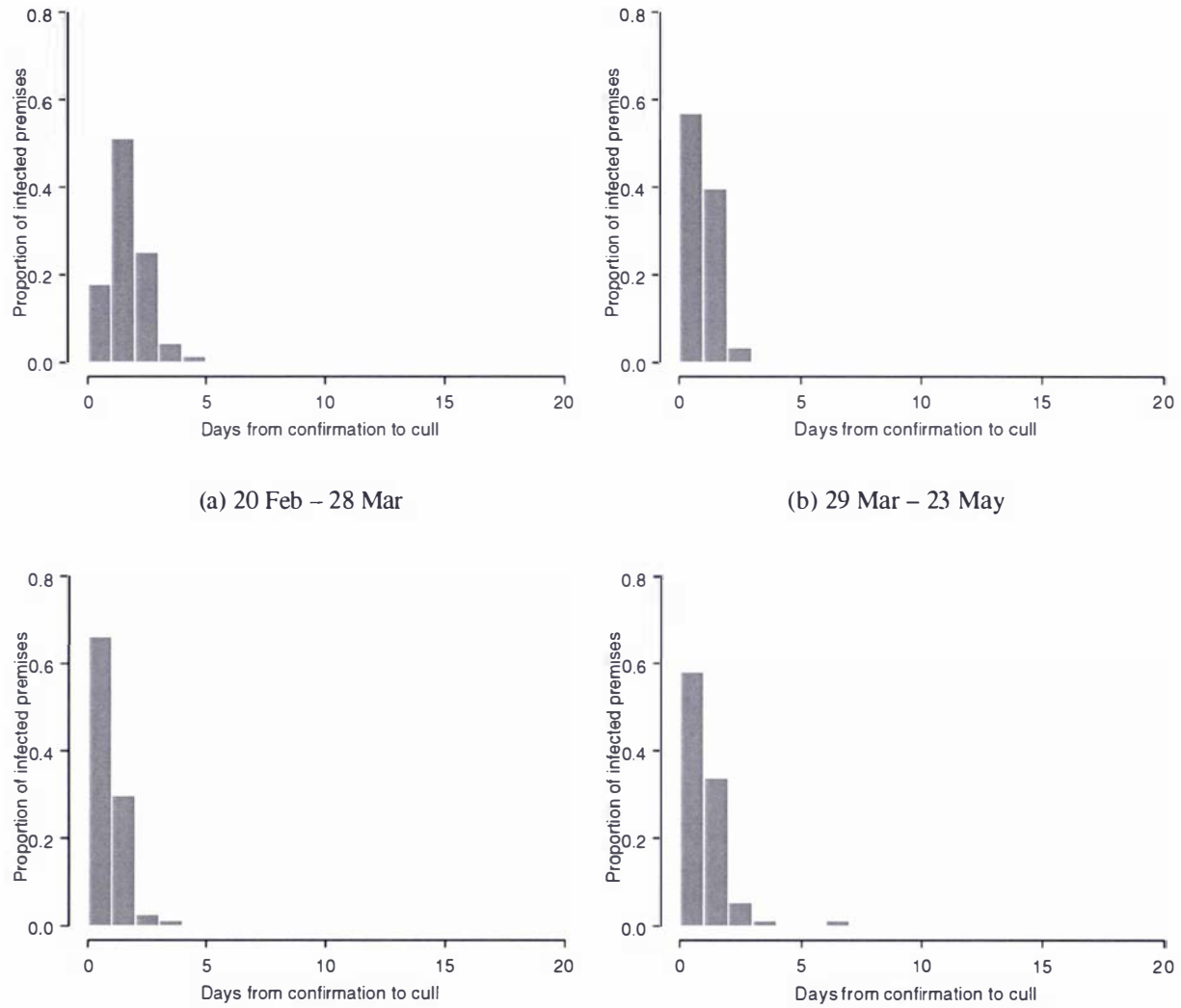
(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

(c) 24 May – 18 Jul

(d) 19 Jul – 30 Sep

Figure 8.7: Frequency distribution of the interval (in days) between foot-and-mouth disease infection date and confirmation date in Cumbria (Great Britain).



(a) 20 Feb – 28 Mar

(b) 29 Mar – 23 May

(c) 24 May – 18 Jul

(d) 19 Jul – 30 Sep

Figure 8.8: Frequency distribution of the interval (in days) between foot-and-mouth disease confirmation date and cull date in Cumbria (Great Britain).

8.4 Discussion

The observed start of the 2001 FMD epidemic in Great Britain was 19 February when the disease was identified in a group of cull sows sent for slaughter to an abattoir located in the south east of the country. By the end of the epidemic on 30 September 2001, 2,026 holdings throughout Great Britain had been confirmed as infected and had been culled and an additional 8,481 holdings had been culled pre-emptively as part of the control measures introduced after 24 February 2001, making this the largest and most-expensive FMD epidemic in a temperate country in recent years. Our analyses relate to two areas of Great Britain where there were relatively many farm holdings affected.

Several factors explain the striking differences in the behaviour of infection in the two areas investigated. In Cumbria in the early phase of the epidemic, many cattle holdings were infected — probably resulting in a large environmental viral load (Sellers 1971). Although control activities resulted in a rapid decrease in infection hazard after 28 March, the epidemic remained incompletely controlled. Persistence of the disease — accompanied by the onset of seasonal farming activities such as contract hay and silage making — facilitated medium-to-long distance spread of virus and is thought to be at least partly responsible for the steady increase in infection hazard that occurred from 19 April to 15 August (Figure 8.2a). Although not evident from these analyses (where farm holdings have been represented by a single point in space), holdings made up of disaggregated land parcels and the unavoidable movement of animals and animal products between these parcels also would have complicated control efforts (Ferguson et al. 2001, Kao 2001). In contrast, in Devon, we speculate that a lower initial environmental viral load and an already-mobilised control effort resulted in relatively quick eradication (Figure 8.2b).

Figure 8.5a shows that in the Cumbria study area for the period from 20 February to 28 March 2000 there was (in comparison with other periods) relatively low spatio-temporal interaction of infection. Given the relatively high infection hazard evident at this time (Figure 8.2a) this would suggest heavy and recent 'seeding' of virus throughout this area. As the epidemic progressed, spatio-temporal interaction became more prominent — evidenced by $\hat{D}_0(s,t)$ values of > 1 at relatively small distance-time separations (Figs. 8.5b – 8.5d). In Devon, spatio-temporal interaction of infection risk was evident in both the 20 February to 28 March and 29 March to 23 May periods (Figure 8.6a and 8.6b) — though compared with the equivalent series for Cumbria, the space-time separations

where $\hat{D}_0(s,t) > 1$ were relatively small.

The distance component of the $\hat{D}_0(s,t) = 1$ risk contour identifies the extent of spatio-temporal interaction of infection risk (colloquially termed 'contagiousness') in space and the temporal component of the $\hat{D}_0(s,t) = 1$ risk contour identifies the extent of contagiousness in time. Figs. 8.5 and 8.6 show that these two components are not static throughout an epidemic; increases in the maximum distance component of the $\hat{D}_0(s,t) = 1$ contour in sequential analyses is indicative of increases in local-spread distances — implying a need to re-assess the effectiveness of control measures aimed at reducing local spread of the infection. Increases in the maximum temporal position of the $\hat{D}_0(s,t) = 1$ risk contour suggest that infected holdings are remaining infectious for longer — implying a need to enhance detection, the speed of depopulation and/or cleaning and disinfection procedures. Figure 8.5b shows that for holdings infected between 29 March and 23 May 2001 in Cumbria, the $\hat{D}_0(s,t) = 1$ risk contour extended no greater than 1 km (95% CI 0 km, 3 km). For holdings infected between 24 May and 18 July, the shape of the contour differed: the $\hat{D}_0(s,t) = 1$ risk contour persisted at 3 km (95% CI 0 km, 5 km) for 7 days after infection date, then declined to 0 km by 14 days (Figure 8.5c). For the period 19 July to 30 September 2001, the $\hat{D}_0(s,t) = 1$ risk contour (Figure 8.5d) declined from 2 km (95% CI 0 km, 5 km) at $t = 0$ days to 0 km at 7 days. Since the process of culling, disposal and disinfection of infected premises was well-established in Cumbria early in the epidemic (from 29 March 2001 to the end of the epidemic > 50% of infected premises were culled within 1 day of confirmation; Figure 8.8b) we therefore conclude that the major reason for the increased time infected premises posed a risk to others during May to July was delayed detection of disease. The movement of the disease into 'other' (predominantly sheep) holding types from May to July — where the identification of clinical signs was more difficult (Kitching & Mackay 1995) probably influenced this pattern. This is important because the epidemic in the studied area Cumbria was predominantly a cattle-based outbreak with infection spilling over into sheep flocks.

Although this has been a retrospective analysis of two purposively-selected areas we propose that the analytical approach described could be applied on a routine basis during future FMD epidemics to describe and better understand the temporal, spatial and spatio-temporal features of infection risk. Instantaneous hazard functions provide a description of the probability of infection per unit time, accounting for temporal changes in the size of

the holding population at risk. Extraction maps account for the spatial distribution of the farm holding population at risk and provide a means for defining locations where there are high proportions of infected premises per unit area. Whereas these techniques consider time and space independently, the ability to describe the risk of infection attributable to spatio-temporal interaction provides further insight — identifying the extent of 'contagiousness' firstly in space (and therefore providing an objective means for defining suitable pre-emptive culling distances) and secondly in time (indicating how quickly infection risk is 'extinguished' after a holding becomes infected). The space-time K-function appears well-suited for this purpose, and in the example provided in these analyses we have shown that the level of spatio-temporal interaction itself can change over time, indicative of important changes in the epidemiological behaviour of the disease.

8.5 Conclusion

Our analyses showed that in Cumbria (where many cattle holdings initially were infected and eradication took several months) the distance over which contagiousness operated and the temporal duration of contagiousness from source farms changed throughout the course of the epidemic. In contrast, in Devon (where the epidemic was controlled quickly), there was less evidence of space-time interaction and the relevant spatial and temporal effects were much shorter.

CHAPTER 9

The predictive accuracy of a stochastic spatial model of foot-and-mouth disease

9.1 Introduction

Foot-and-mouth disease (FMD) is one of the most contagious pathogens known to microbiology and is the single most important constraint to international trade in livestock and animal products (Donaldson & Alexandersen 2002). In countries where the disease is not endemic FMD outbreaks can have a severe effect on a national economy: not only are losses associated with decreased animal production but also due to the cost of eradication, the imposition of local and international trade restrictions, negative effects on industries that service the agricultural sector, community disruption accompanying restrictions on movement, and reductions in domestic and international tourism (Thompson et al. 2002).

Given the unprecedented expense of the 2001 British FMD epidemic and the possibility of further outbreaks occurring at unpredictable intervals in the future, a number of independent inquiries have made recommendations concerning the ways in which infectious disease epidemics in Great Britain might be better handled (Anderson 2002, Curry 2002, Follet 2002). In addition, alternative approaches to FMD control and eradication have been proposed and discussed (Leforban 2002, Keeling et al. 2003). Recognising the potential logistic problems associated with FMD vaccination strategies applied in either prophylactic or reactive forms, Keeling et al. (2003) have proposed a novel strategy for FMD control and eradication which they term a 'predictive' vaccination approach. This strategy is based on the ability to identify farms that are likely to acquire the disease as second generation infections (that is, two infection cycles on from currently infected farms) and to vaccinate stock on these farms early enough so that animals are immune

by the time they are challenged by virus. When used in combination with pre-emptive slaughtering, Keeling et al. (2003) reason that this approach could markedly reduce the long tail that has characterised the two most recent, large scale British epidemics.

The 2001 British FMD epidemic was uniquely well documented via a spatial farm database and associated case reporting and tracing. This has allowed the development of a range of models which have attempted to quantify the effectiveness of a range of proposed control strategies (Howard & Donnelly 2000, Ferguson et al. 2001, Kao 2001, Morris et al. 2001). Whereas these models have provided valuable insight into factors that influence the spread of disease at the national level, to the best of our knowledge there have been no reports of their predictive accuracy sufficient to assess their suitability for meeting the requirements of a predictive vaccination strategy or other policy initiatives. Throughout the 2001 British FMD epidemic the stochastic spatial model InterSpread (Sanson et al. 1999) was run on a regular basis to assess the effectiveness of proposed control strategies and to provide on-going forecasts of eradication progress at national and local levels (DEFRA 2001, Morris et al. 2001). With a complete record of the epidemic available, the aims of this study were firstly to quantify the spatial and temporal accuracy of InterSpread's predictions throughout the epidemic at both the national and sub-regional level and secondly to identify ways in which the predictive accuracy of the model might be improved. Greater confidence in model predictions would seem a necessary prerequisite if novel vaccination strategies were to be used as part of control measures in the event of future FMD outbreaks.

9.2 Materials and methods

InterSpread is a stochastic spatial model of FMD, developed as part of the EpiMAN information system for emergency disease control (Sanson 1993). As its base the model uses the coordinates of all farms within a defined area (typically an entire country) and the numbers and types of each of the five FMD-susceptible species present at each farm location, as well as other relevant data such as the Cartesian coordinates of animal markets and boundaries of defined control zones. The model can be initiated from the index case of an outbreak or on the basis of the recorded history of the sequence of farms already confirmed with the disease to a given date, termed the simulation start date. For each farm

confirmed with the disease the recorded or estimated date of infection is used to start the simulated spread of disease from that farm. In addition to data related to the population of farms at risk and the distribution of infection at the start of simulation, InterSpread uses a series of epidemiological parameters to govern pathogen behaviour and a series of control parameters to define details of the control measures applied throughout the course of an outbreak. Once simulation commences, farm-to-farm spread of infection continues until the specified control measures extinguish the epidemic. A detailed description of the parameters used by InterSpread (with specific reference to the 2001 British epidemic) has been provided by Morris et al. (2001).

For the InterSpread simulations conducted throughout the 2001 British epidemic the farm population at risk data was derived from the 2000 agricultural census conducted by the Ministry of Agriculture, Fisheries and Food (MAFF 2001) in England and Wales and the Scottish Executive Rural Affairs Department (SERAD 2001) in Scotland. Census data for each farm holding included a unique identifier, the easting and northing coordinate of the main farm building and the numbers of adult pigs, cattle, sheep, goats and deer present at the date of census (30 June 2000).

InterSpread simulations were run twice to three times weekly throughout the 2001 British epidemic, with a total of 90 individual national simulations completed between 24 February and 1 October 2001 plus a range of *ad hoc* local and regional simulations conducted at irregular intervals. A contour plot showing the predicted density of infected premises from a typical InterSpread simulation conducted during the 2001 British epidemic is shown in Figure 9.1.

For each simulation details of all known confirmed case holdings to 2 weeks before simulation start date were derived from the Disease Control System Database (Gibbens et al. 2001) and used to initiate the model wherein the epidemic was simulated for a period of 300 days or until eradication of the disease occurred, whichever occurred first. On account of the pre-emptive slaughter policy adopted to control the disease and on-going improvements to the accuracy of recorded farm holding locations, the farm population at risk database was revised before each model run. In this study we present in detail the results of InterSpread predictions from three simulation start dates: 7 April, 10 June and 3 September 2001, thought to be representative of the early, middle and late phases of the epidemic, respectively. The epidemiological and control parameters used for each of

these simulations are summarised in Table 9.1, Table 9.2, and Figure 9.2.

The predictive accuracy of InterSpread was evaluated at the national and sub-regional level. National level analyses included all holdings within the farm population at risk database. For the sub-regional analyses we selected a 2,500 km^2 (50 km $\times$ 50 km) area within the county of Cumbria (Figure 9.3), described in detail in Chapter 8. For each simulation start date InterSpread was run for 99 iterations and predictive accuracy in the 28-day period after simulation start date evaluated.

Frequency histograms showing the distribution of the 99 predicted number of infected holdings in the 28 days after simulation start date were compared with the actual number of holdings diagnosed in the same time frame. A comparison of the actual number of diagnosed holdings with the distribution of predicted infected holdings was conducted as follows. The null hypothesis was that the actual number of diagnosed holdings was consistent with the distribution of predicted infected holding numbers. Where the actual number ranked k*th* among the 99 predictions we rejected the null hypothesis if $k/99$ was less than 0.05 or greater than 0.95. Using this method, the two-sided significance level was $k/99$ where $k < 50$ and $1 - (k/99)$ where $k \geq 50$.

The spatial accuracy of predictions was determined by dividing each area (that is, the whole of Great Britain for the national-level analysis and the 2,500 km^2 area in Cumbria for the sub-regional analysis) into a regular grid of square cells. Holding locations plotted as points were superimposed on this grid (Figure 9.4a) and a cell was identified as positive if there was at least one infected holding diagnosed within it during the specified period following simulation start date (Figure 9.4b). This process was repeated using the location details of holdings predicted to be infected by InterSpread in the same time frame (Figure 9.4c and Figure 9.4d). Each cell was then considered in turn and classified as follows: (1) those where FMD was predicted and FMD occurred (correctly predicted positives), (2) those where FMD was predicted and FMD did not occur (false positives), (3) those where FMD was not predicted and FMD occurred (false negatives), and (4) those where FMD was not predicted and FMD did not occur (correctly predicted negatives). On the basis of counts of cells within each category a two-by-two table was constructed and predictive accuracy quantified in terms of sensitivity, specificity, and positive and negative predictive value (Rothman & Greenland 1998).

9.3 Results

Table 9.3 shows the predicted number of infected premises and the actual number of diagnosed premises within each analysed region for the 28-day period after the April, June and September simulation start dates. Frequency histograms showing the distribution of the predicted number of FMD-infected premises for the 28-day period after the April, June and September simulation start dates are shown in Figures 9.5 and 9.6.

Table 9.4 shows the sensitivity, specificity, positive predictive value and negative predictive value of InterSpread predictions made at the April, June, and September simulation start dates. Means and 95% confidence intervals of sensitivity, specificity and positive and negative predictive value are reported for those farm holdings predicted to become infected within a 0 to 14- and 14 to 28-day period after each simulation start date. In this table inferential cell sizes were 400 km^2 and 4 km^2 for the national-level and Cumbrian analyses, respectively. For the national-level simulations the positive predictive value of InterSpread's spatial predictions (that is, the proportion of cells predicted to be disease positive that were actually disease positive) ranged from 55% to 75% at 0 to 14 days after simulation start date. Negative predictive values (the proportion of cells predicted to be disease negative that were actually disease negative) ranged from 99% to 100% for the same time frame. Positive predictive values were lower for the Cumbrian sub-regional analysis: 4% to 37% for the 0 to 14 day period after simulation start date.

Figures 9.7 and 9.8 show the effect of varying the inferential grid cell size on the positive and negative predictive value of predictions for the 0 to 14-day and 14 to 28-day period after simulation start. Increasing inferential grid cell size increased positive predictive values and decreased negative predictive values, the extent of the change varying among the three simulation start dates described.

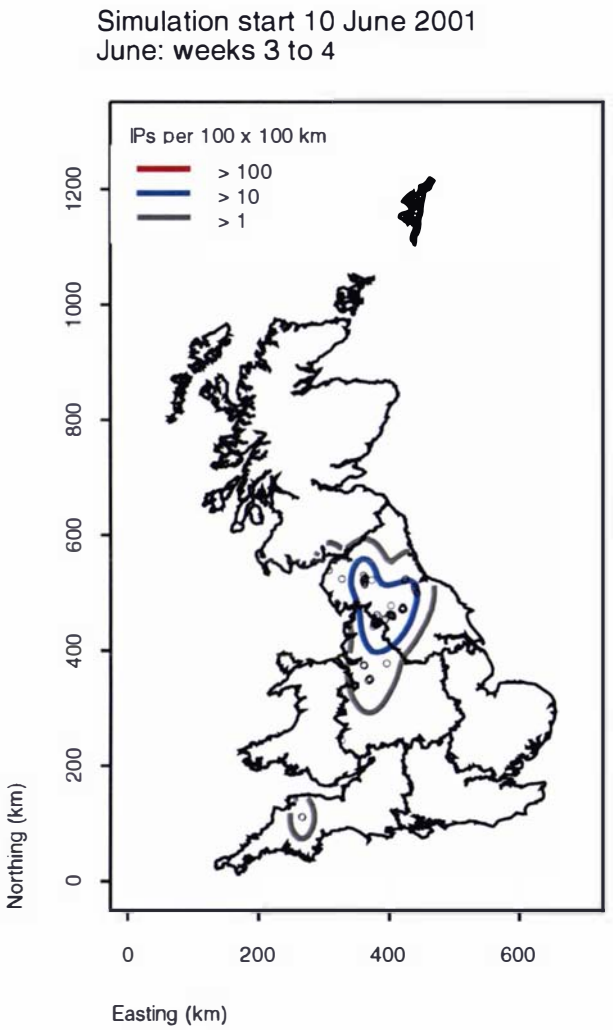


Figure 9.1: Map of Great Britain showing the predicted spatial distribution of foot-and-mouth disease infected premises across Great Britain in the 14-day period after a 10 June 2001 simulation start date. Contour lines identify areas where greater than 1, 10, and 100 infected premises per 100 km^2 were predicted by the model.

Table 9.1: Epidemiological parameters used by InterSpread for simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001. Numbers in brackets refer to item numbers within the InterSpread parameter file.

Details	April	June	September
Days to onset of clinical signs (1)	See Figure 9.2a		
Days to diagnosis [a] (2 - 4)	See Figure 9.2b		
Days before onset of signs that virus appears in milk (33)	4	4	4
Risk movements off farm per day [b] (5 – 7)	0.1, 0.5, 1	0.1, 0.5, 1	0.1, 0.5, 1
Risk movements to markets off farm per day [c] (8 – 9)	0.2, 0	0.2, 0	0, 0
Additional farm contacts generated by market movements (10)	1.5	1.5	1.5
Movement distance probabilities (11)	0 – 16 km: 0.90, 16 – 64 km: 0.10		
Probability of infection from risk movements [b] (12 – 14)	0.4, 0.15, 0.005	0.4, 0.15, 0.005	0.4, 0.15, 0.005
Days after onset of signs that transmission via movement will occur (47)	-2	-2	-2
Proportion of dairy farms with lactating dairy cattle (24)	1	1	1
Maximum length of tanker routes (25)	20000	20000	20000
Probability of farm being on tanker route (26)	0.5	0.5	0.5
Farms on tanker route in non-movement controlled area [d] (27 – 28)	8 (4)	8 (4)	8 (4)
Farms on tanker route inside infected area [d] (29 – 30)	8 (4)	8 (4)	8 (4)
Dairy tanker pickups per week (40)	7	7	7
Probability of infection from tanker contact [e] (44 – 45)	0.05, 0.01	0.05, 0.01	0.05, 0.01
Separate or combined local and airborne spread (46)	Combined	Combined	Combined
Local spread probabilities (15 – 16)	See Figure 9.2c		
Proportion of continuing local spread after farm on surveillance (49)	0.9	0.9	0.9
Use airborne spread? (20)	Yes	Yes	Yes
Airborne spread probabilities (46.1)	See Figure 9.2d		
Species multiplier for airborne spread [a] (17 – 19)	1, 1, 1	1, 1, 1	1, 1, 1
Airborne susceptibility modifier [a] (52 – 54)	1, 0.05, 0.001	1, 0.05, 0.001	1, 0.05, 0.001
Airborne spread using prevailing wind (21)	Yes	Yes	Yes
Probability of airborne spread on a given day (46.4)	0.7	0.7	0.7
Direction of prevailing wind (22)	90	90	90
Flexible or fixed wind direction (46.2)	Flexible	Flexible	Flexible
Wind direction (degrees, probability)	50 0.10; 130 0.50; 360 0.40		
Days to slaughter an infected farm (31)	1	1	1
Diagnosed farms that cause a resource problem (32)	2×10^6	2×10^6	2×10^6
Hours to trace movements [b] (36 – 38)	48, 120, 336	48, 120, 336	48, 120, 336
Days to identify an infected farm by back tracing (39)	2	2	2
Probability that a movement is not reported [b] (41 – 43)	0.05, 0.05, 0.1	0.05, 0.05, 0.1	0.05, 0.05, 0.1
Days for immunity after vaccination [f] (34 – 35)	7, 7	7, 7	7, 7

[a] Cattle, sheep, pigs.

[b] High, medium, low.

[c] High, medium.

[d] Mean (standard deviation).

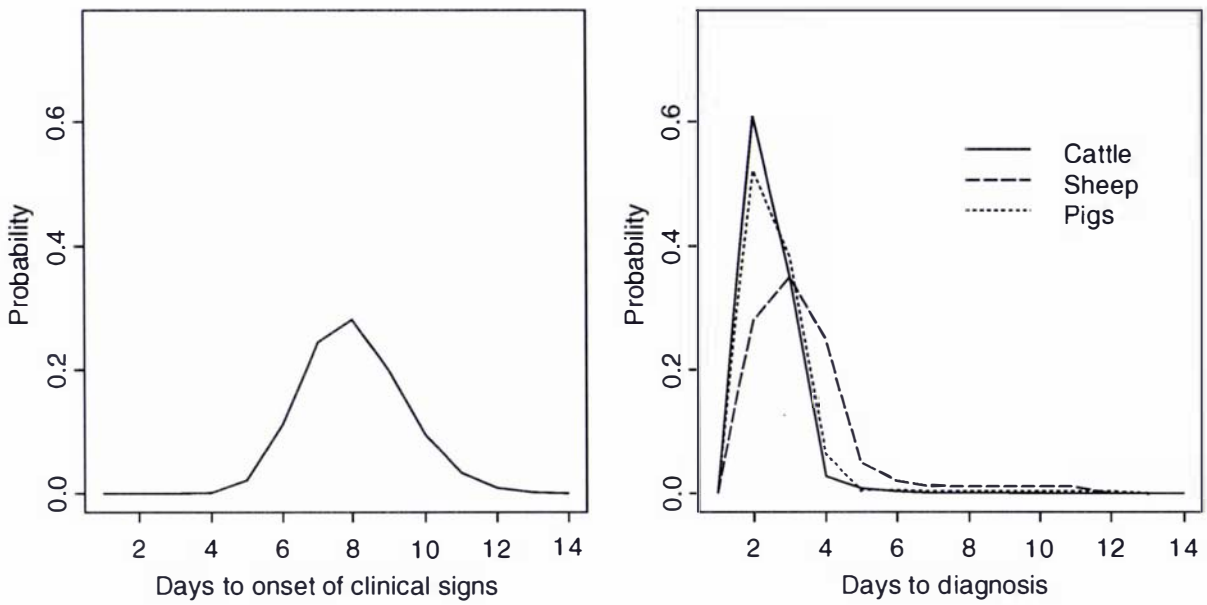
[e] Medium, low.

[f] Cattle, pigs.

Table 9.2: Control strategy procedural definitions and settings used by InterSpread for simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001.

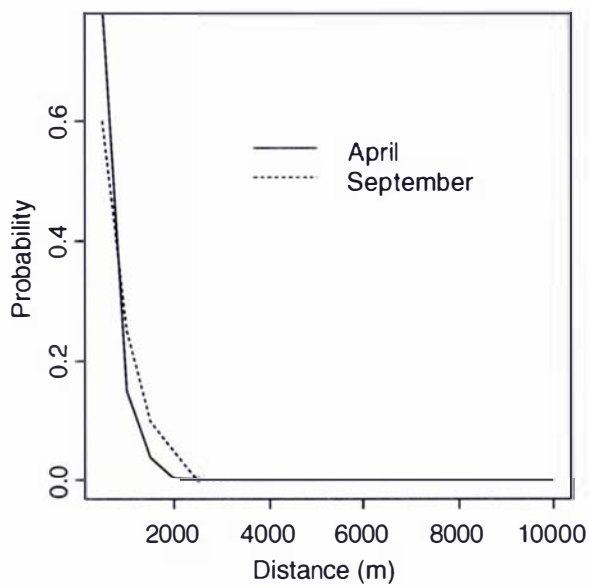
Details	April	June	September
Number of infected areas	1	1	1
Date infected areas activated	24 Feb 01	24 Feb 01	24 Feb 01
Coordinates of infected area	List	List	List
Number of pre-emptive slaughter risk policies	2	2	2
Date pre-emptive slaughter risk policy activated	24 Feb 01	24 Feb 01	24 Feb 01
Proportions to slaughter for each risk category [a]	0.05, 0, 0	0.05, 0, 0	0.05, 0, 0
Date pre-emptive slaughter risk zone activated	13 Mar 01	13 Mar 01	13 Mar 01
Proportions to slaughter for each risk category	0.05, 0, 0	0.05, 0, 0	0.05, 0, 0
Number of pre-emptive slaughter by area policies	0	0	0
Number of within-infected area movement control policies	1	1	2
Date within-infected area movement policies activated	24 Feb 01	24 Feb 01	24 Feb 01
Proportions of movements to all within-infected area [a]	0, 0.5, 0.9	0, 0.5, 0.9	0, 0.5, 0.9
Date within-infected area movement policy activated	-	-	1 May 01
Proportions of movements to allow within-infected area [a]	-	-	0.1, 0.5, 0.9
Number of off-surveillance farm movement policies	1	1	2
Date off-surveillance farm movement policy activated	22 Feb 01	22 Feb 01	22 Feb 01
Proportions of movements to allow off farms on surveillance [a]	0, 0.01, 0.5	0, 0.01, 0.5	0, 0.01, 0.5
Date off-surveillance farm movement policy activated	-	-	1 May 01
Proportions of movements to allow off farms on surveillance [a]	-	-	0.1, 0.01, 0.5
Number of off-vaccinated farm movement policies	0	0	0
Number of surveillance-level control policies	1	1	1
Date surveillance-level control policy activated	22 Feb 01	22 Feb 01	22 Feb 01
Proportion of farms on surveillance, based on risk [a]	1, 1, 0.5	1, 1, 0.5	1, 1, 0.5
Number of surveillance-length control policies	1	1	1
Date surveillance-length control policies activated	22 Feb 01	22 Feb 01	22 Feb 01
Length of surveillance period (days)	14	14	14
Number of patrol zone policies defined	1	1	1
Date patrol zone policy activated	22 Feb 01	22 Feb 01	22 Feb 01
Radius of patrol zone	3000	3000	3000
Number of radial zones defined	1	1	2
Inner and outer radius of zone (m)	0 – 1500	0 – 500	0 – 500
Date to start creating zones around diagnosed farms	20 Mar 01	20 Mar 01	20 Mar 01
Date to stop creating zones around diagnosed farms	31 Dec 01	31 Dec 01	30 Apr 01
Type of control for this zone	Slaughter	Slaughter	Slaughter
Resource for this zone	Unlimited	Unlimited	Unlimited
Number of days to process this zone	3	3	3
Species	C, S, P	C, S, P	C, S, P
Inner and outer radius of zone (m)	-	-	0 – 1100
Date to start creating zones around diagnosed farms	-	-	1 May 01
Date to stop creating zones around diagnosed farms	-	-	31 Dec 01
Type of control for this zone	-	-	Slaughter
Resource for this zone	-	-	Unlimited
Number of days to process this zone	-	-	3
Species	-	-	C,S,P

[a] High, medium, low.
C,S,P: Cattle, sheep, pigs

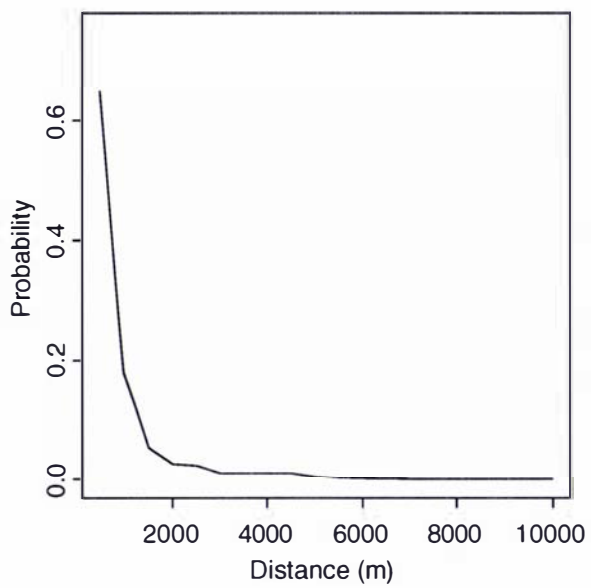


(a) Days to onset of clinical signs (item 1)

(b) Days to diagnosis (item 2)



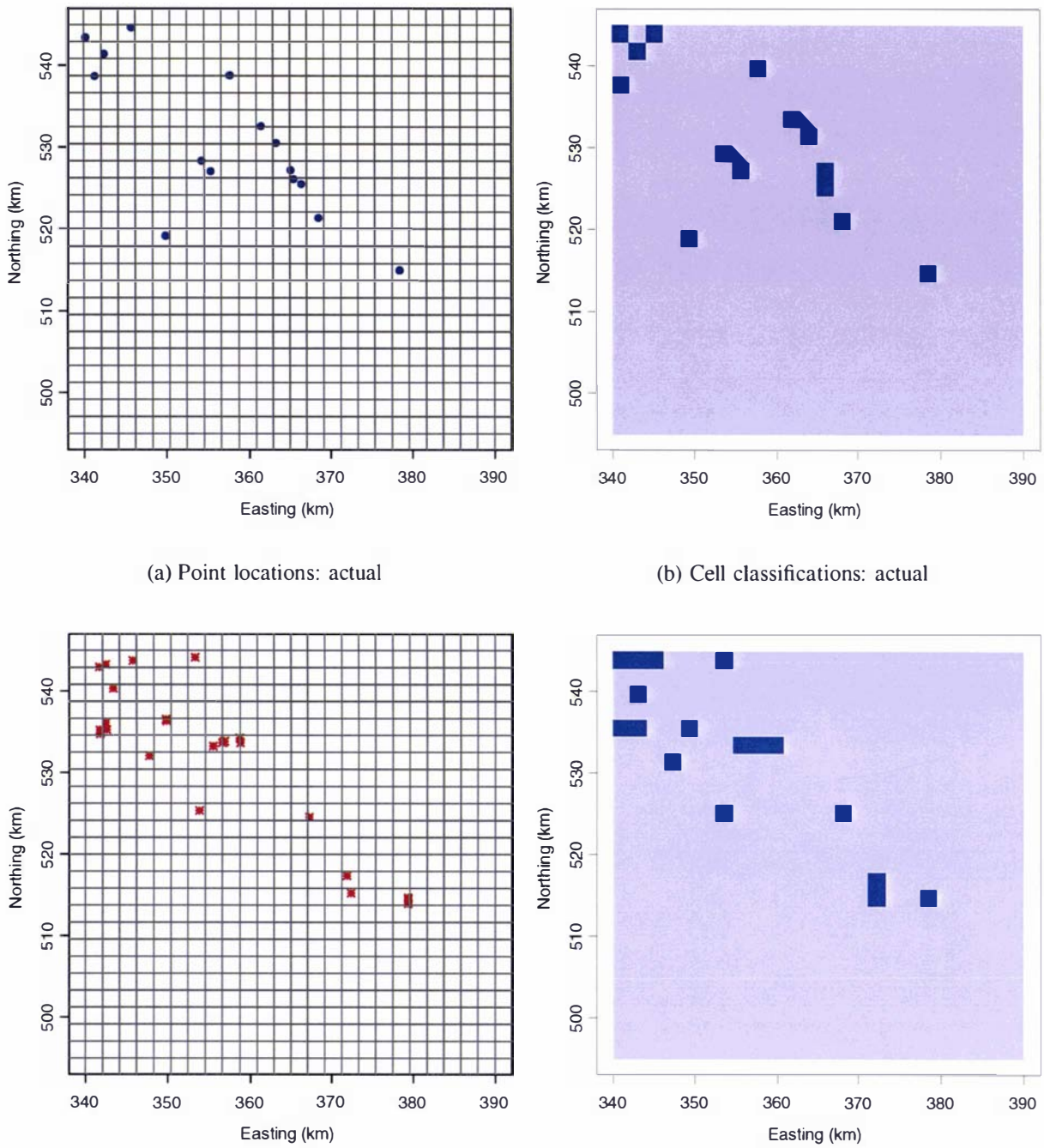
(c) Local spread probabilities (item 15 – 16)



(d) Airborne spread probabilities (item 46.1)

Figure 9.2: Selected epidemiological parameter settings used by InterSpread for simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001.

Figure 9.3: Map of Great Britain showing the location of the Cumbria study area described in this study.



(a) Point locations: actual

(b) Cell classifications: actual

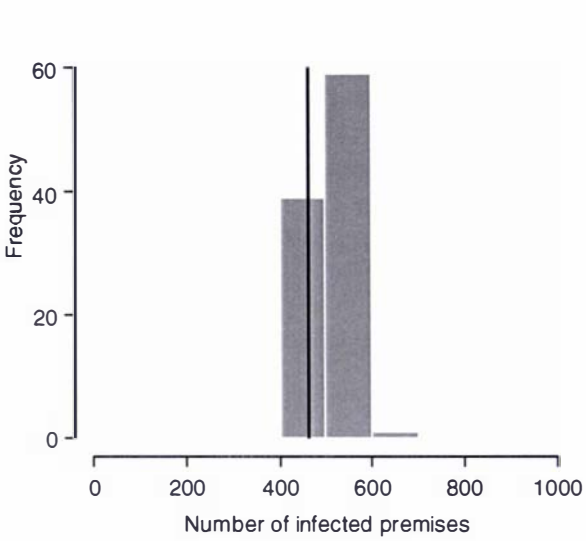
(c) Point locations: predicted

(d) Cell classifications: predicted

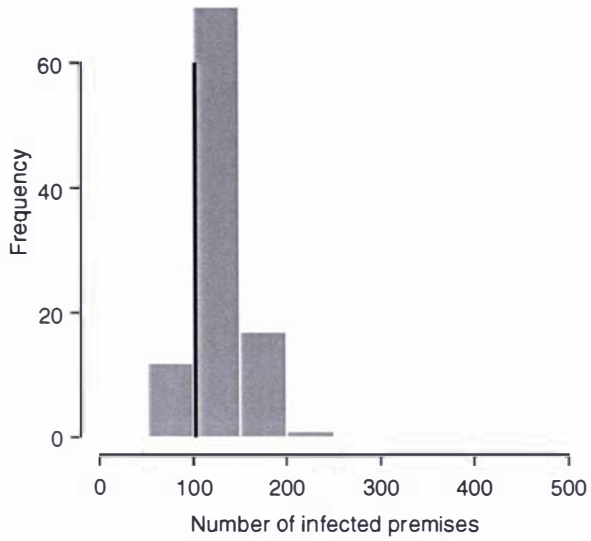
Figure 9.4: Explanation of the method used to assess the accuracy of InterSpread's predictions. Each study area was divided into a regular grid of square cells. Foot-and-mouth disease infected holdings diagnosed during each defined time interval after simulation start date were superimposed on this grid (a) and a cell was classified positive if there was at least one infected premises diagnosed within it cell during the specified period following simulation start date, (b). This process was repeated for Interspread-predicted infected premises for the same area and time frame, (c) and (d).

Table 9.3: Counts of premises predicted to become infected and counts of premises diagnosed with foot-and-mouth disease in the 28 days after InterSpread simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001. Predictions have been summarised over 99 simulations and the mean, minimum and maximum numbers reported.

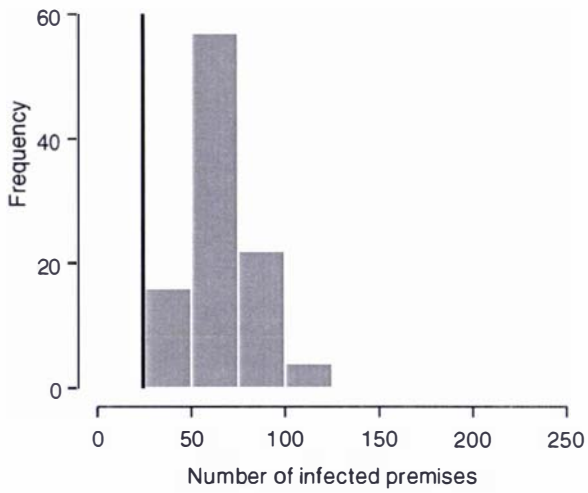
Region	Simulation		Predicted	Actual	P
	Start	End	Mean (range)		
Great Britain	7 Apr 2001	5 May 2001	509 (405 – 640)	462	0.18
	10 Jun 2001	8 Jul 2001	127 (74 – 225)	102	0.15
	3 Sep 2001	1 Oct 2001	68 (40 – 104)	24	0.01
Cumbria	7 Apr 2001	5 May 2001	39 (16 – 63)	33	0.33
	10 Jun 2001	8 Jul 2001	12 (2 – 27)	44	< 0.01
	3 Sep 2001	1 Oct 2001	32 (13 – 60)	13	0.01



(a) April

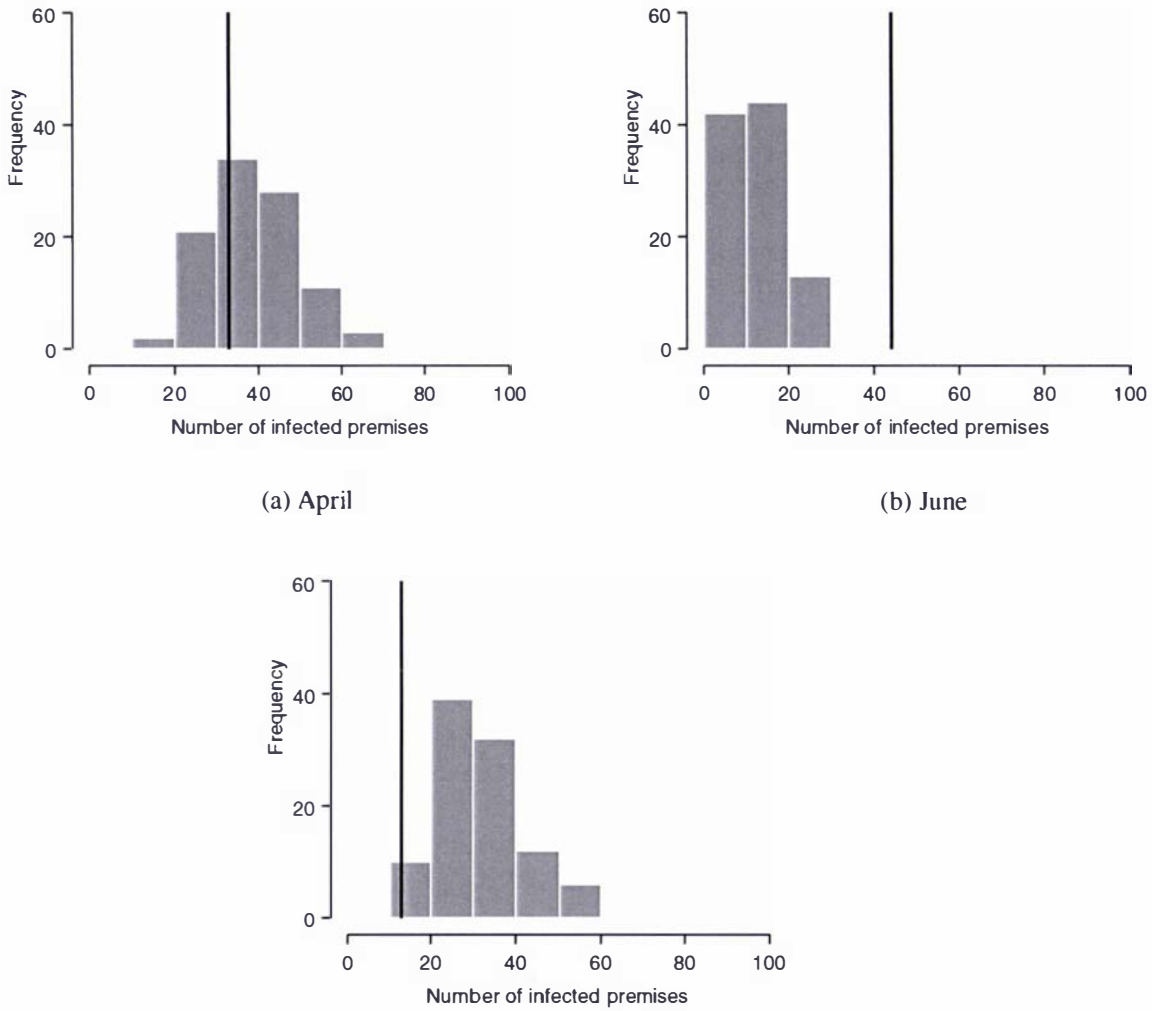


(b) June



(c) September

Figure 9.5: Frequency histograms showing the distribution of the predicted number of foot-and-mouth disease infected premises across Great Britain in the 28 days after simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001. The vertical line shows the actual number of diagnosed premises during the same time frame.



(a) April

(b) June

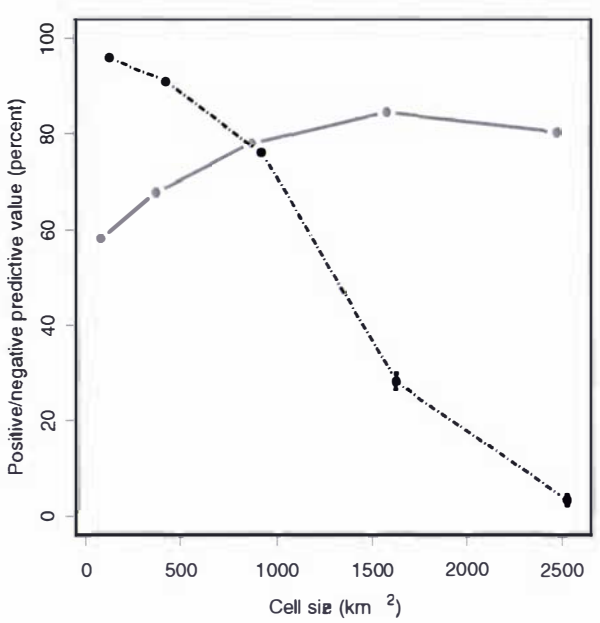
(c) September

Figure 9.6: Frequency histograms showing the distribution of the predicted number of foot-and-mouth disease infected premises across Cumbria in the 28 days after simulations starting on 7 April 2001, 10 June 2001 and 3 September 2001. The vertical line shows the actual number of diagnosed premises during the same time frame.

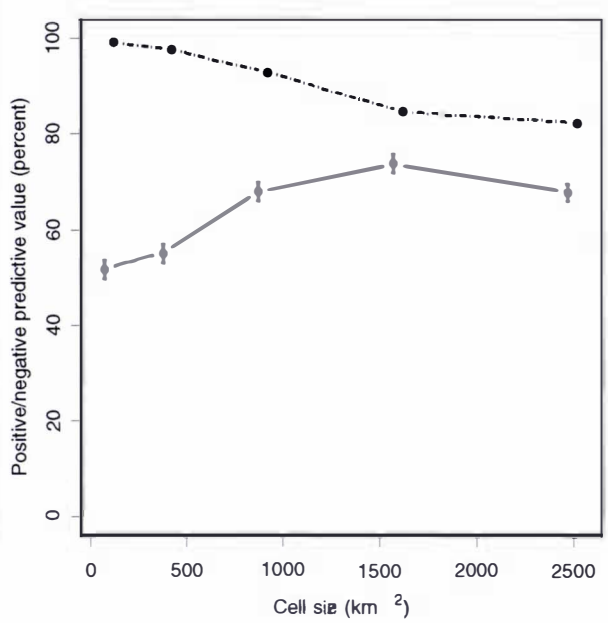
Table 9.4: Means (95% confidence intervals) of sensitivity, specificity, positive predictive value and negative predictive value of spatial predictions of foot-and-mouth disease infection made by InterSpread in April, June and September of the 2001 British epidemic. Predictions have been made for the intervals 0 to 14 and 14 to 28 days after simulation start date and summarised for 99 model simulations.

Region	Date	Days	Sensitivity	Specificity	PPV	NPV
Great Britain [a]	Apr	0 – 14	59 (58 – 59)	99 (99 – 99)	68 (67 – 69)	99 (99 – 99)
		14 – 28	36 (35 – 37)	99 (99 – 99)	44 (43 – 45)	99 (99 – 99)
	Jun	0 – 14	37 (36 – 38)	100 (100 – 100)	55 (53 – 57)	100 (99 – 100)
		14 – 28	29 (28 – 29)	100 (100 – 100)	45 (43 – 47)	100 (99 – 100)
	Sep	0 – 14	71 (69 – 73)	100 (100 – 100)	75 (72 – 78)	100 (100 – 100)
		14 – 28	38 (29 – 48)	100 (100 – 100)	8 (6 – 10)	100 (100 – 100)
Cumbria [b]	Apr	0 – 14	27 (26 – 29)	97 (97 – 97)	23 (22 – 24)	98 (98 – 98)
		14 – 28	3 (2 – 4)	99 (99 – 99)	6 (4 – 7)	98 (98 – 98)
	Jun	0 – 14	2 (1 – 2)	99 (99 – 99)	4 (2 – 5)	98 (98 – 98)
		14 – 28	-	-	-	-
	Sep	0 – 14	35 (34 – 36)	99 (99 – 99)	37 (35 – 39)	99 (99 – 99)
		14 – 28	-	-	-	-

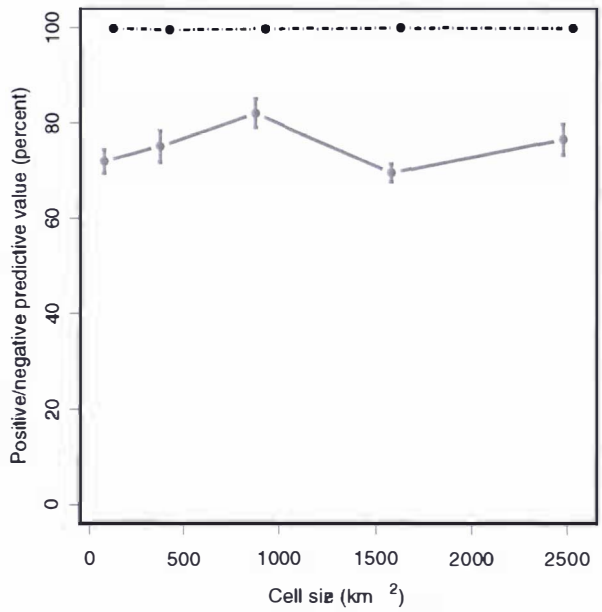
[a] Predictions based on a grid cell size of 400 km^2 (20 km × 20 km).
[b] Predictions based on a grid cell size of 4 km^2 (2 km × 2 km).
PPV: Positive predictive value.
NPV: Negative predictive value.



(a) April

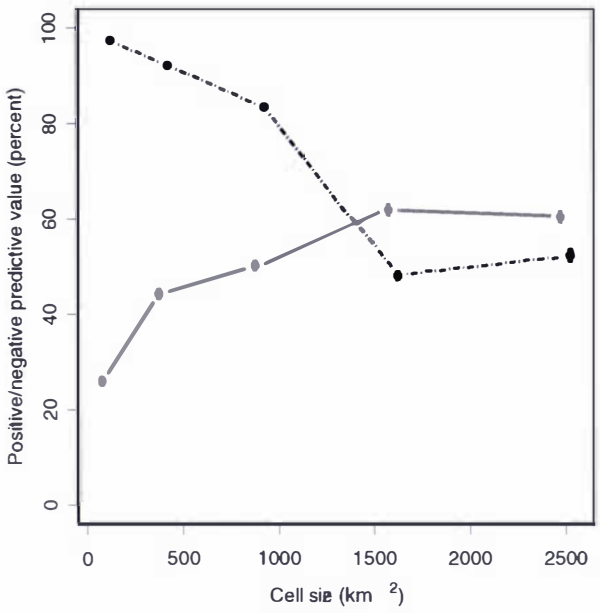


(b) June

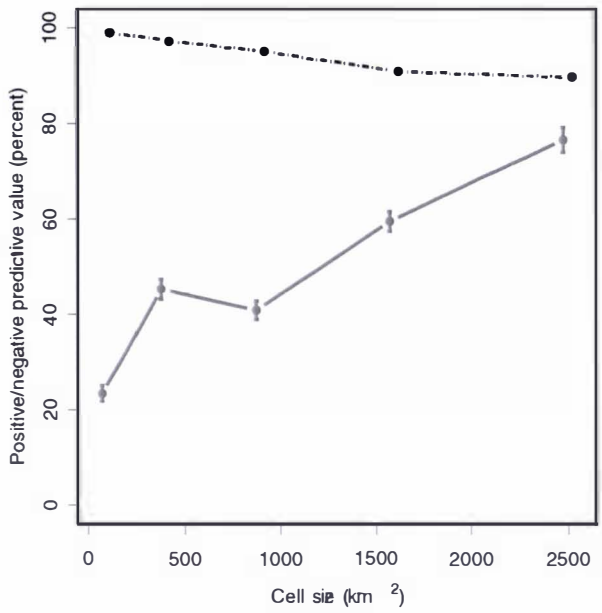


(c) September

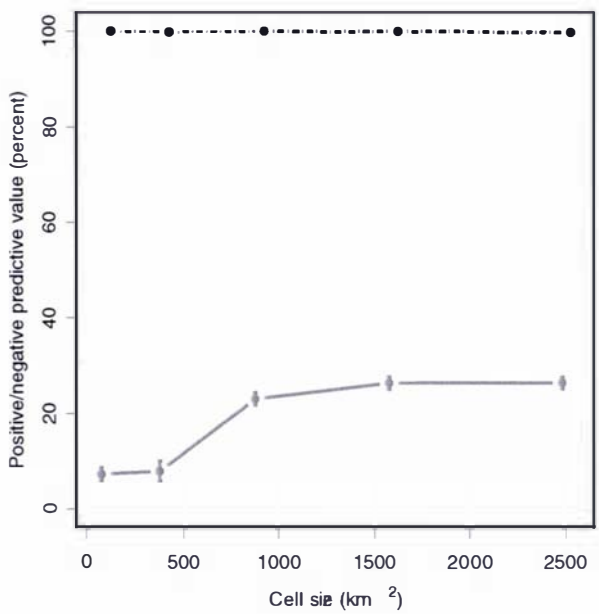
Figure 9.7: Spatial accuracy of InterSpread predictions of foot-and-mouth disease infected premises throughout Great Britain for the April, June and September 2001 simulation start dates. The line plots above show the mean and 95% confidence interval of the positive predictive value (solid lines) and negative predictive value (dashed lines) of predictions as a function of inferential grid cell size. Predictions have been made for the 0 to 14 day period after each simulation start date.



(a) April



(b) June



(c) September

Figure 9.8: Spatial accuracy of InterSpread predictions of foot-and-mouth disease infected premises throughout Great Britain for the April, June and September 2001 simulation start dates. The line plots above show the mean and 95% confidence interval of the positive predictive value (solid lines) and negative predictive value (dashed lines) of predictions as a function of inferential grid cell size. Predictions have been made for the 14 to 28 day period after each simulation start date.

9.4 Discussion

Predictive models play a potentially important role in managing large scale infectious disease outbreaks in animal populations by providing a means by which the effectiveness of various candidate control policies can be critically evaluated (Moutou & Durand 1994, Garner & Lack 1995, Nielen et al. 1996, Keeling et al. 2001). Models that use as their basis details of farm location offer the additional benefit of identifying the location farm holdings likely to become infected during the prediction period. This facility — if proven to be sufficiently accurate — allows resources to be more effectively targeted at identified high risk areas at a time when they are invariably at a premium. In the case of FMD, accurate spatial and temporal predictive ability opens up the possibility of successfully applying novel control and eradication strategies (such as predictive vaccination) to future outbreaks. The focus of this paper has been to quantify the predictive performance of the stochastic spatial model of FMD, InterSpread, with a view to evaluating its suitability as a tool to guide a predictive vaccination strategy for FMD control and eradication.

At the national level the number of incident infected premises predicted by InterSpread compared favourably with the numbers of premises actually diagnosed (Figure 9.5). At the sub-regional level there were stages of the epidemic when the actual number of holdings diagnosed differed significantly from the distribution of holdings predicted to become infected (Table 9.3 and Figure 9.6). These discrepancies, we believe, are to be expected when modelling the behaviour of a large-scale epidemic — mainly arising (in this case) from the effectiveness of control efforts varying both geographically and over time. These discrepancies were identified at the time of simulation and in an effort to improve the match between the observed numbers of incident holdings and the number predicted, a regular process of parameter adjustment was undertaken. In April, for example, InterSpread over-predicted incident case numbers at the national level (Table 9.3 and Figure 9.5a). On the basis of discussions with field epidemiology staff the size of pre-emptive slaughter radii around infected premises was adjusted to better-represent actual practice. These modifications resulted in predicted case numbers that more closely matched those actually diagnosed (data not shown). In response to under-prediction of incident infections in some areas of the country from early July onwards (Figure 9.6b) parameters were altered further to allow small numbers of high risk movements through infected areas and off farms that were on surveillance. Again, these changes were made

on the basis of discussing the current efficacy of control measures with field epidemiological staff and making biologically appropriate changes to model parameters where necessary. This process (identifying areas and time-frames where model predictions failed to reflect the observed epidemic and altering model parameter settings accordingly) was a necessary component of modelling FMD using a simulation-based approach and provided control personnel with additional insight into epidemic behaviour as it was under modification (via control measures) within a complex, heterogenous environment of farms, animal density and human activity.

Prediction specificities and negative predictive values were greater than prediction sensitivities and positive predictive values (Table 9.4). High negative predictive values mean that a high proportion of grid cells predicted to be disease negative will not have disease. Lower positive predictive values means that some grid cells predicted to be disease positive will be free of disease. Of all the possible characteristics of an 'imperfect test' these are perhaps the most desirable of a predictive model of FMD, tending towards over-prediction (too many false positives) rather than under-prediction (too many false negatives). InterSpread provided spatial predictions with useful positive and negative predictive values at 0 to 14 days after simulation start date. In April the positive and negative predictive values of national level predictions at relatively small inferential grid cell sizes of 100 km^2 (10 km $\times$ 10 km) were 58% and 96% respectively. At an inferential cell size of 400 km^2 (20 km $\times$ 20 km) positive and negative predictive values were 68% and 91% respectively. Predictive accuracy of the model varied: periods when predictive accuracy was lower (for example June, Figure 9.7b) were stages of the epidemic when several 'foci' of FMD infection were present throughout the country and we believe that at this time a single set of model parameters were not sufficient to adequately reflect the heterogenous behaviour of disease throughout the country. We conclude that InterSpread is able to predict FMD-infected locations within 0 to 14 days after simulation start sufficient to better-target surveillance activities and provide estimates of resource requirements for contingency planning. The spatial accuracy of InterSpread predictions might be further improved through the use of a series of sub-regional models, better-capturing the characteristics of individual outbreaks that typically emerge during extended large scale, multicentred epidemics.

Given that the time taken for immunity to develop after vaccination is at least 7 days

(Doel et al. 1994, Salt et al. 1998, Cox et al. 1999) and that a variable amount of time is required to conduct a vaccination program it would seem that a predictive vaccination approach to FMD control would require a model to accurately predict the location of infection from at least 14 days after simulation start date. The relatively poor predictive ability at 14 – 28 days after simulation start (Figure 9.8) indicates that InterSpread — as used in the 2001 British epidemic — was not suitable as a tool to guide a predictive vaccination strategy. As suggested, a series of sub-regional models might offer one means for improving predictive ability at greater periods of time after simulation start date sufficient for the needs of a predictive vaccination strategy for FMD control and eradication, and this would be a profitable area of future research.

CHAPTER 10

General discussion

This thesis applies various analytical techniques to these two quite different epidemics: a rapidly spreading highly contagious disease for which urgent decisions are essential (FMD), and a feed-borne non-contagious disease with an exceptionally long incubation period (BSE). The BSE epidemic, in particular, presented major investigational challenges because its recent emergence meant that its epidemiological features were not yet fully clear.

Descriptions of the temporal pattern of BSE onset among holdings (Chapter 3) and among individual cattle (Chapter 4), by accounting for the population of cattle holdings and individual animals at risk, have allowed the incidence of disease to be compared among regions, holding types and sizes and among birth cohorts. These analyses quantified the success of the primary control measures in Great Britain (the July 1988 meat and bone meal ban resulted in a 60% reduction in BSE risk for cattle born in the first 12 months after its introduction) and have better-characterised the behaviour of the disease in areas of the country where control measures were less-successfully applied. Mixed-effects spatial models of BSE risk (Chapter 6) for cattle born after July 1988 identified moderate increases in the risk of disease in areas where there were greater numbers of pigs relative to cattle, consistent with disease being associated with cross-contamination of cattle feed with high-protein pig and poultry feed. In addition, spatial aggregations of BSE risk not accounted-for by the explanatory variables included in the model were identified — better defining areas where control efforts had been less successfully applied. These analyses demonstrate that the 1988 meat and bone meal ban was a highly effective means of controlling BSE and that localised influences were important during the control phase of the epidemic. The risk factors for BSE that have been quantified in this thesis (in particular those related to holding type classification and holding size) provide useful

guidelines for those countries wishing to conduct targeted surveillance for BSE. The ability to identify areas where there is an excess of not accounted-for risk of the disease has important implications for those countries in the process of eradicating the disease where failure to achieve adequate compliance with control measures may allow extended local spread to occur.

Drawing on the findings from Chapters 3 to 6, Chapter 7 sought to more precisely describe the spatial and temporal features of the earliest known exposures to BSE. From these analyses it was concluded that there was evidence to support the hypothesis that recycling of infection was established in the Southeast and Southwest of the country by July 1983. The question of the epidemic starting as a single or multi focal source still remains unclear and given that details of the early phase of the epidemic can only be guessed at, simulation modelling of this phase of the epidemic should help to identify which of these two scenarios was the more likely.

The BSE epidemic is clearly in decline, and in the coming years attention will be focused on those cases that were born after the total ban of feeding mammalian protein to farm animals (HMSO 1996b, Wilesmith 2002). The analytical techniques applied in this thesis could legitimately find application to better define risk factors for these 'final' cases. Due to small numbers it is likely that a case-control approach would be the most appropriate with detailed covariate information available for all study subjects. Model-based approaches, if adopted, would need to consider holdings as individual point locations and cases be paired with spatially-matched controls.

Descriptive spatial, temporal and spatio-temporal analyses of the FMD epidemic demonstrated marked differences in the epidemiological behaviour of this disease in the two areas of Great Britain that were investigated in Chapter 8. The ability to describe the risk of infection attributable to spatio-temporal interaction provides insight into the epidemiology of FMD within infected areas — identifying the extent of 'contagiousness' firstly in space (and therefore providing an objective estimate of suitable pre-emptive slaughter distances) and secondly in time (indicating the effectiveness of disease detection, speed of de-population and/or cleaning and disinfection procedures). In Chapter 8 it was shown that the space-time K-function was well-suited for this purpose, and in the Cumbrian epidemic described in this chapter it was shown that the level of spatio-temporal interaction itself can change over time, indicative of important changes in the epidemiological be-

haviour of the disease. Greater utility from analyses of this type will be obtained when robust methods are available to allow the degree of space-time interaction of infection to be explicitly quantified at each location and time point within a study area. This will allow 'problem areas' to be promptly and definitively identified during the course of an epidemic, allowing more focussed allocation of resources.

The stochastic spatial model, InterSpread, provided credible spatial and temporal predictions of FMD spread throughout the 2001 British epidemic, with prediction sensitivities and specificities sufficient to direct surveillance activities and provide estimates of resource requirements for contingency planning (Chapter 9). On account of the marked differences in behaviour of the disease among affected areas of the country (Gibbens et al. 2001) it was not possible to accurately predict FMD-positive locations at fine spatial scales greater than 14 days after simulation start date. The spatial accuracy of InterSpread predictions might be further improved through the use of a series of sub-regional models, better-capturing the characteristics of individual outbreaks that typically emerge during extended large scale, multicentred epidemics.

10.1 Surveillance for endemic disease

The analyses presented in this thesis have demonstrated the value of spatially referenced data for the epidemiological analysis of animal disease. Given the more widespread availability of spatial data, the more widespread use of desktop GIS packages and statistical techniques better-suited for analysing spatial data, the challenge for veterinary epidemiologists in the years ahead is to conduct analyses in real time rather than as academic exercises well after periods of crisis. Not only would real time analyses be of use for understanding the epidemiology of emerging disease syndromes, but also for endemic diseases or variants of endemic diseases such as *Salmonella enterica* subspecies *enterica* serovar Typhimurium, DT104, *Salmonella enteritidis* and *Escherichia coli* O157:H7. This raises the subject of disease surveillance, and ideally spatial and temporal analyses should form a key component of active surveillance systems where emerging animal health concerns are detected by scanning data from providers on a regular basis (Scudamore 2002). In the design of such systems microbial contamination should not be the sole area of concern: chemical contamination of animal product has recently damaged consumer confidence in

Australia (Spence et al. 1998) and Belgium (Van Larebeke et al. 2002) and arguably these threats should receive equal attention.

10.2 Case data

In most countries once an animal is suspected or diagnosed by the holding manager or attending veterinarian as having an exotic or notifiable disease the case is handed over to a representative of the state veterinary service who then supervises the collection of necessary individual and farm-level case details (including, in most cases details of farm location). Hand written details are then transferred to a centralised database which, when supplemented with relevant laboratory details, forms the complete register of diagnosed cases. While this strategy works satisfactorily for 'unusual' disease conditions where there is little problem with case definition (and was the means by which case data was recorded for both epidemics described in this thesis (Sanson & Ryan 1997, Gibbens et al. 2001) the approach is inherently unsuitable for endemic diseases which do not reliably come to official attention.

To accomplish the objectives of active or 'scanning' surveillance systems (Scudamore 2002) and to exploit more fully the potential of spatial epidemiology to better understand endemic disease patterns there is a need to capture a wider variety of disease event information from individual farm holdings. Not only is an accurate clinical diagnosis required, but also the time of onset of clinical signs and associated case details. This goal has long been the aim of surveillance systems in human health such as the hospital episode system and the National Health Service Prescription Pricing Authority system in England and various practice-based schemes using the International Classification of Diseases coding system (WHO 1977, WHO 1992). Given the variety of databases maintained and the variety of different disease coding systems in use, none have been universally applied and can therefore be regarded as fully effective. Staines & Järup (2000) reviewed systems for recording human health event data for spatial epidemiological use and concluded that the major problems of these initiatives were associated with disease coding systems changing over time and a lack of universal adoption of any one particular system.

With regard to endemic disease recording systems for domestic species, national-level programs have been implemented in Norway (Helsetjenesten for Storfe 2003), Denmark,

and New Zealand (Livestock Improvement Corporation 2003). Whereas these systems are heroic in their aims and objectives and should be viewed as a necessary first step towards any useful active disease surveillance system, the second (and perhaps more difficult) phase of such projects should be to ensure accuracy of diagnosis, consistency of disease definitions across farm holdings and consistency of observation of the herd or flock. Stevenson (2000) intensively monitored disease incidence in six dairy herds in New South Wales, Australia over a 2 year period and concluded that regular contact between herd veterinarian and holding manager was necessary for disease events to be consistently and accurately recorded. While accurate estimates of disease incidence may represent the goal of comprehensive surveillance systems, it is unlikely that this objective will ever be met given the inevitable heterogeneity in diagnostic and data recording skill that will exist among holding managers within any given region. Acknowledging this, perhaps the benefit of endemic disease recording systems will come from their ability to allow *changes* in a 'noisy' pattern of disease incidence in time and space to be detected, indicative of epidemiologically important changes to a previously stable disease profile. This characteristic increases the need for authorities to develop experience in analysing data of this type, recognising that the amount of 'noise' will no doubt vary over time (associated with seasonal farming activities) and among regions (associated with farm enterprise types).

Given higher rates of on-farm computer use in recent years (NAHMS 1996, Hayes 1997), software for herd and flock management could provide an important means for capturing detailed farm level data which might then be provided to an integrated animal disease surveillance system. Veterinarians are uniquely placed to play an important role in this process by assisting herd managers to make best use of these tools, exploiting the analytical capability of these packages to identify and correct shortfalls in herd-level performance. If herd managers can appreciate the value of accurate, up-to-date information (enabling them to make timely, well-informed business decisions) the same high-quality data can then be provided to surveillance authorities with corresponding increases in the accuracy and usefulness of the system as a whole. Figure 10.1 provides an outline of how on-farm animal management software might play a role in a national endemic disease surveillance system. In such a system health-related data is obtained from a variety of sources (the farm holding itself, veterinarians, laboratory diagnoses, slaughterhouse monitoring and so on) and merged within a 'national animal health data

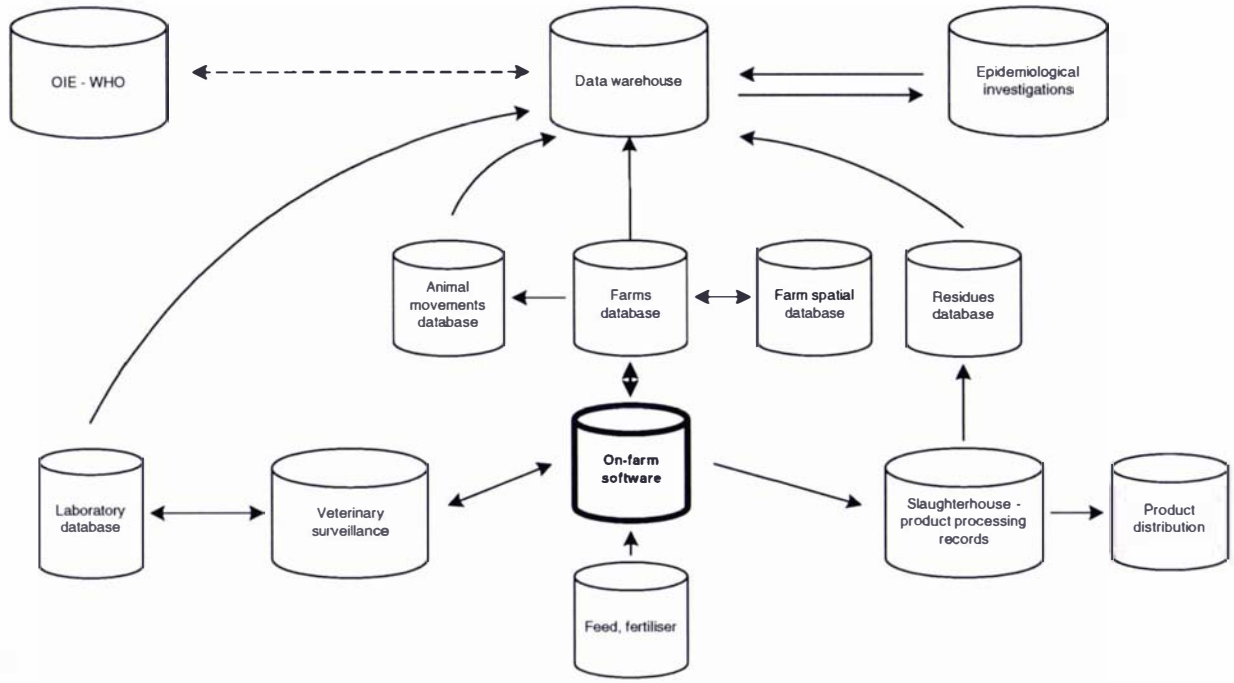


Figure 10.1: Schematic diagram showing how on-farm data, veterinary practice records, laboratory data, slaughterhouse processing records, details of residue assessments and animal movements might be integrated within an active surveillance data network.

warehouse'. This framework, ideally, should not be limited to a single species or production type: the simultaneous monitoring of so-called 'baskets' of disease conditions (Lawson & Williams 2000) across multiple species would provide an important means for increasing the sensitivity to detect true changes in disease patterns.

10.3 Population data

As a general premise, and as clearly demonstrated in this thesis, epidemiological applications are more informative when case details are interpreted in context of the population from which they arose (Rothman & Greenland 1998). When considering the spatial distribution of disease large numbers of cases diagnosed at a given location may truly represent an excess of disease or may simply be a function of the underlying population structure (Kelsall & Diggle 1995a, Kelsall & Diggle 1995b, Lawson 2001b). In Chapter 5 extraction mapping techniques were used to show the distribution of BSE throughout Great Britain after accounting for the spatial distribution of the cattle population at risk. In addition, census data was used in conjunction with age-specific removal hazards derived from British survey data (Pasman et al. 1995) to create a set of individual animal records

representing each animal estimated to have been present in Great Britain from 1986 to 1996. This data set was used in the analysis of individual-animal risks of BSE (Chapter 4) and for the comparison of BSE risk for animals born before and after the 1988 meat and bone meal feed ban (Chapter 6). Necessarily, this process was time consuming and involved making broad assumptions about the age structure of the British cattle population. Nonetheless, in the absence of more suitable alternatives it provided a useful estimate of the cattle population at risk and with regard to the British BSE epidemic allowed informative comparisons to be made among regions, holding types and sizes, and birth cohorts. With the implementation of the Cattle Passports Order in Great Britain (HMSO 1996a), it is anticipated that the level of detail of population at risk data will substantially improve in future years and data from these information sources will allow epidemiological investigations to be conducted in greater detail, allowing the influence of animal movement to be accounted for. For these schemes to be effective it will be necessary that accurate herd-level data is transferred regularly from the farm to responsible agencies, a process that should be facilitated by appropriately designed on-farm software.

10.4 Conclusion

The application of spatial and temporal analytical techniques to animal health data are of immense help in unravelling the pressing scientific and managerial questions faced by those involved in controlling diseases in animal populations. Descriptive temporal analyses, such as those described in Chapters 3, 4 and 8 allow temporal patterns of epidemic data to be characterised and are of particular use in situations where the population at risk is under constant change. Descriptive spatial analyses — such as kernel smoothing (Chapters 5 and 8) and choropleth mapping (Chapter 6) — are useful to describe the geographical features of disease and to investigate relationships between putative risks and disease. Mixed-effects spatial models provide additional insight by quantifying the magnitude of the effect of identified risk factors and by defining locations where disease is unaccounted-for by these risks (Chapter 6). This facility is of particular use for those involved in disease control since these locations can then be targeted for further investigation to elucidate additional mechanisms that might be involved in disease causation.

Perhaps the only certainty in the years ahead is that new and novel disease syndromes

will emerge from time to time, just as BSE did in the late 1980s. To ensure that food derived from animals remains 'safe' existing information systems need to be refined and new systems developed that allow rapid identification and analysis of emerging animal health problems. The studies presented in this thesis demonstrate that the application of temporal, spatial, and spatio-temporal analytical methods can enhance the understanding of the epidemiological features of diseases in animal populations. The challenge now is to make best use of existing data sources to conduct investigations in real time and to ensure that animal health surveillance data collection systems are progressively re-engineered so that the necessary information for spatial and temporal analyses are readily available when and where they are required.

Bibliography

Abernethy, D., Pfeiffer, D., Denny, G., Torrens, T., McCullough, S. & Graham, D. (2000), Evaluating airborne spread in a Newcastle disease epidemic in Northern Ireland, *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of 9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Altman, D., Machin, D., Bryant, T. & Gardner, M. (2001), *Statistics with Confidence*, British Medical Journal, London.

Anderson, I. (2002), *Foot and mouth disease 2001: Lessons to be learned inquiry*, HMSO, London.

Anderson, R., Donnelly, C., Ferguson, N., Woolhouse, M., Watt, C., Udy, H., MaWhinney, S., Dunstan, S., Southwood, T., Wilesmith, J., Ryan, J., Hoinville, L., Hillerton, J., Austin, A. & Wells, G. (1996), 'Transmission dynamics and epidemiology of BSE in British cattle', *Nature London* **382**(6594), 779–788.

Anonymous (2000), *The BSE inquiry. The Phillips Inquiry into BSE and variant CJD in the United Kingdom*, HMSO, London.

Anselin, L. (1995), 'Local indicators of spatial association — LISA', *Geographical Analysis* **27**(2), 93–114.

Anselin, L. (1996), The Moran scatterplot as an ESDA tool to assess local instability in spatial association, *in* M. Fischer, H. Scholten & D. Unwin, eds, 'Spatial analytical perspectives on GIS', Taylor and Francis, London, pp. 111–125.

Anselin, L. & Smirnov, O. (1998), 'The SpaceStat Extension for ArcView Version 2.0', The SpaceStat Project.

Antunes, J. & Waldman, E. (2002), 'Trends and spatial distribution of deaths of children aged 12 – 60 months in Sao Paulo, Brazil, 1980 – 98', *Bulletin of the World Health Organization* **80**(5), 391–398.

Arlinghaus, S. (1994), *Practical Handbook of Digital Mapping: Terms and Concepts*, CRC Press, New York.

Arlinghaus, S. (1995), *Practical Handbook of Spatial Statistics*, CRC Press, New York.

Baddeley, A. & Turner, R. (2000), 'Practical maximum pseudolikelihood for spatial point patterns (with discussion)', *Australian and New Zealand Journal of Statistics* **43**(3), 283–322.

Baddeley, A. & Turner, R. (2002), 'Spatstat — Analysis for spatial point patterns Version 1.2.1', http://www.maths.uwa.edu.au/ adrian/spatstat.html.

Bailey, T. (1994), A review of statistical spatial analysis in geographical information systems, *in* S. Fotheringham & P. Rogerson, eds, 'Spatial Analysis and GIS', Taylor and Francis, London, pp. 13–44.

Bailey, T. & Gatrell, A. (1995), *Interactive Spatial Data Analysis*, Longman Scientific and Technical, London.

Bartlett, M. (1964), 'The spectral analysis of two-dimensional point processes', *Biometrika* **51**, 299–311.

Bell, B. & Broemeling, L. (2000), 'A Bayesian analysis for spatial processes with application to disease mapping', *Statistics in Medicine* **19**(7), 957–974.

Bellido, J., Pierce, G. & Wang, J. (2001), 'Modelling intra-annual variation in abundance of squid *Loligo forbesi* in Scottish waters using generalised additive models', *Fisheries Research* **52**(1-2), 23–39.

Benardinelli, L., Clayton, D. & Montomoli, C. (1995), 'Bayesian estimates of disease maps: How important are priors?', *Statistics in Medicine* **14**, 2411–2431.

Benardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. & Ghislandi, M. (1995), 'Bayesian analysis of space-time variation in disease risk', *Statistics in Medicine* **14**, 2433–2443.

Benardinelli, L. & Montomoli, M. (1992), 'Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk', *Statistics in Medicine* **11**, 983–1007.

Benardinelli, L., Pascutto, C., Best, N. & Gilks, W. (1997), 'Disease mapping with errors in covariates', *Statistics in Medicine* **16**, 741–752.

Berke, O. (2001), 'Choropleth mapping of regional count data of *Echinococcus multilocularis* among red foxes in Lower Saxony, Germany', *Preventive Veterinary Medicine* **52**, 119–131.

Berman, M. & Turner, T. (1992), 'Approximating point process likelihoods with GLM', *Applied Statistics* **41**(1), 31–38.

Besag, J. (1989), 'Towards Bayesian image analysis', *Journal of Applied Statistics* **16**, 395–407.

Besag, J. & Kooperberg, C. (1995), 'On conditional and intrinsic autoregressions', *Biometrika* **82**, 733–746.

Besag, J. & Mollié, A. (1989), 'Bayesian mapping of mortality rates', *Bulletin of the International Statistical Institute* **53**, 127–128.

Besag, J. & Newell, J. (1991), 'The detection of clusters in rare diseases', *Journal of the Royal Statistical Society Series A* **154**, 143–155.

Besag, J., York, J. & Mollié, A. (1991), 'Bayesian image restoration, with two applications in spatial statistics', *Annals of the Institute of Statistical Mathematics* **43**, 1–21.

Best, N., Marshall, C. & Thomas, A. (2000), Bayesian inference using WinBUGS, *in* N. Best, ed., 'Bayesian inference using WinBUGS', Queensland University of Technology, Brisbane, pp. –.

Best, N., Waller, L., Thomas, A., Conlon, E. & Arnold, R. (1999), Bayesian models for spatially correlated disease and exposure data, *in* J. Bernardo, J. Berger, A. Dawid & A. Smith, eds, 'Bayesian Statistics 6: Proceedings of the Sixth Valencia Meeting on Bayesian Statistics', Oxford University Press, Oxford, pp. 131–56.

Bhopal, R., Diggle, P. & Rowlingson, B. (1992), 'Pinpointing clusters of apparently sporadic cases of Legionnaires' disease', *British Medical Journal* **304**, 1022–1027.

Bithell, J. (1990), 'An application of density estimation to geographical epidemiology', *Statistics in Medicine* **9**, 691–701.

Bivand, R. & Gebhardt, A. (2000), 'Implementing functions for spatial statistical analysis using the R language', *Journal of Geographical Systems* **2**, 307–317.

Böhning, D., Dietz, E. & Schlattmann, P. (2000), 'Space-time mixture modelling of public health data', *Statistics in Medicine* **19**, 2333–2344.

Böhning, D. & Schlattmann, P. (1992), 'Computer-assisted analysis of mixtures (CA-MAN): Statistical algorithms', *Biometrics* **48**, 283–303.

Borchers, D., Buckland, S., Priede, I. & Ahmadi, S. (1997), 'Improving the precision of the daily egg production method using generalised additive models', *Canadian Journal of Fish and Aquatic Sciences* **54**, 2727–2742.

Bowman, A. (1984), 'An alternative method of cross-validation for the smoothing of density estimates', *Biometrika* **711**, 353–360.

Bowman, A. & Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-PLUS Illustrations*, Oxford University Press, London.

Breslow, N. & Clayton, D. (1993), 'Approximate inference in generalised linear mixed models', *Journal of the American Statistical Association* **88**, 9–25.

Breslow, N. & Day, N. (1987), *The Design and Analysis of Cohort Studies Volume 2 IARC Scientific Publication No. 82*, International Agency for Research on Cancer, Lyon.

Campos, M., Valencia, L., Fortes, B., Braga, R. & Medronho, R. (2002), 'Spatial distribution of *Ascaris lumbricoides* infection', *Revista De Saude Publica* **36**(1), 69–74.

Carlin, B. & Louis, T. (2000), *Bayes and Empirical Bayes Methods for Data Analysis — Monographs on Statistics and Applied Probability 69*, Chapman and Hall, London.

Carpenter, T., Hird, D. & Snipes, K. (1996), 'A time-space investigation of the epidemiology of fowl cholera', *Preventive Veterinary Medicine* **28**(3), 159–163.

Carr, D., Wallin, J. & Carr, D. (2000), 'Two new templates for epidemiology applications: Linked micromap plots and conditioned choropleth maps', *Statistics in Medicine* **19**, 2521–2538.

Carrat, F. & Valleron, A.-J. (1992), 'Epidemiologic mapping using the 'kriging' method: Application to an influenza-like epidemic in France', *American Journal of Epidemiology* **135**(11), 1293–1300.

Chetwynd, A. & Diggle, P. (1998), 'On estimating the reduced second moment measure of a stationary spatial point process', *Australian and New Zealand Journal of Statistics* **40**, 11–15.

Chetwynd, A., Diggle, P., Marshall, A. & Parslow, R. (2000), 'Investigation of spatial clustering from matched and stratified case-control studies (submitted September 2001)', - **-**, –.

Christensen, O., Ribeiro Jr, P. & Diggle, P. (2002), 'geoRglm', *R News* **2**(2).

Cislaghi, C., Biggeri, A., Braga, A., Lagazio, C. & Marchi, M. (1995), 'Exploratory tools for disease mapping in geographical epidemiology', *Statistics in Medicine* **14**, 2363–2381.

Clayton, D. & Benardinelli, L. (1992), Bayesian methods for mapping disease risk, *in* P. Elliot, J. Cuzick, D. English & R. Stern, eds, 'Geographical and Environmental Epidemiology: Methods for Small-Area Studies', Oxford University Press, London, pp. 205–220.

Clayton, D. & Kaldor, J. (1987), 'Empirical Bayes estimates of age-standardized relative risks for use in disease mapping', *Biometrics* **43**, 671–681.

Cliff, A. & Ord, J. (1981), *Spatial Processes, Models and Applications*, Pion, London.

Collett, D. (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, London.

Conlon, E. & Louis, T. (1999), Addressing multiple goals in evaluating region-specific risk using Bayesian methods, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 31–47.

Cox, S., Barnett, P., Dani, P. & Salt, J. (1999), 'Emergency vaccination of sheep against foot and mouth disease: protection against disease and reduction in contact transmission', *Vaccine* **17**(1855-1868).

Cressie, N. (1992), 'Smoothing regional maps using empirical Bayes predictors', *Geographical Analysis* **24**, 75–95.

Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley and Sons Ltd, London.

Cressie, N. & Chan, N. (1989), 'Spatial modelling of regional variables', *Journal of the American Statistical Association* **84**, 393–401.

Cromley, E. & McLafferty, S. (2002), *GIS and Public Health*, The Guilford Press, New York.

Curry, D. (2002), *Farming and food: a sustainable future. Report of the Policy Commission on the Future of Farming and Food*, HMSO, London.

Cuzick, J. & Edwards, R. (1990), 'Spatial clustering for inhomogenous populations (with discussion)', *Journal of the Royal Statistical Society Series B* **52**, 73–104.

Cuzick, J. & Edwards, R. (1996), Cuzick-Edwards one-sample and inverse two-sampling statistics, *in* F. Alexander & P. Boyle, eds, 'Methods for Investigating Localized Clustering of Disease, IARC Scientific Publication 135', International Agency for Research on Cancer, Lyon, pp. 200–202.

DEFRA (2001), *Foot-and-mouth disease 2001 State Veterinary Service Epidemiology Report*, Department for Environment, Food and Rural Affairs, London.

Devine, O., Louis, T. & Halloran, M. (1994a), 'Empirical Bayes estimators for spatially correlated incidence rates', *Environmetrics* **5**, 381–398.

Devine, O., Louis, T. & Halloran, M. (1994b), 'Empirical Bayes methods for stabilising incidence rates before mapping', *Epidemiology* **5**, 622–630.

Diggle, P. (1983), *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.

Diggle, P. (2000), Overview of statistical methods for disease mapping and its relationship to cluster detection, *in* P. Elliott, J. Wakefield, N. Best & D. Briggs, eds, 'Spatial Epidemiology Methods and Applications', Oxford University Press, London, pp. 87–103.

Diggle, P. & Chetwynd, A. (1991), 'Second order analysis of spatial clustering for inhomogenous populations', *Biometrics* **47**, 1155–1163.

Diggle, P., Chetwynd, A., Häggkvist, R. & Morris, S. (1995), 'Second order analysis of space-time clustering', *Statistical Methods in Medical Research* **4**(2), 124–136.

Diggle, P. & Elliott, P. (1995), 'Disease risk near point sources: Statistical issues in the analysis of disease risk near point sources using individually or spatially aggregated data', *Journal of Epidemiology and Community Health* **49**, S20–S27.

Diggle, P. & Gatrell, A. (2001), Spatial statistics for environmental epidemiology, *in* A. Gattrell, ed., 'GIS Vet', University of Lancaster, Lancaster, pp. –.

Diggle, P., Gatrell, A. & Lovett, A. (1990), Modelling the prevalence of cancer of the larynx in part of Lancashire: A new methodology for spatial epidemiology, *in* R. Thomas, ed., 'Spatial Epidemiology', Pion Limited, London, pp. 35–47.

Diggle, P., Moyeed, R., Rowlingson, B. & Thomson, M. (2002), 'Childhood Malaria in the Gambia: A case study in model-based geostatistics', *Journal of the Royal Statistical Society Series C - Applied Statistics* **51**, 493–506.

Diggle, P. & Rowlingson, B. (1994), 'A conditional approach to point process modelling of elevated risk', *Journal of the Royal Statistical Society Series A* **157**, 433–440.

Diggle, P., Tawn, J. & Moyeed, R. (1998), 'Model-based geostatistics (with discussion)', *Applied Statistics* **47**, 299–350.

Director General of Ordnance Survey (1974), *Administrative area diagram showing the boundaries of the county, districts and communities or equivalent areas under the Local Government Act 1972*, Ordnance Survey, Southampton.

Doel, T., L, W. & Barnett, P. (1994), 'Emergency vaccination against foot-and-mouth disease: rate of development of immunity and its implications for the carrier state', *Vaccine* **12**(7), 592–600.

Doherr, M., Heim, D. & Vandevelde, M. (2000), Spatial clustering of Swiss bovine spongiform encephalopathy (BSE) cases, *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of 9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Doherr, M., Hett, A., Rüfenacht, J., Zurbriggen, A. & Heim, D. (2002), 'Geographical clustering of cases of bovine spongiform encephalopathy (BSE) born in Switzerland after the feed ban', *Veterinary Record* **151**, 467–472.

Donaldson, A. & Alexandersen, S. (2002), 'Predicting the spread of foot and mouth disease by airborne virus', *Revue Scientifique Et Technique De L'Office International Des Epizooties* **21**(3), 569–575.

Donnelly, C. & Ferguson, N. (2000), *Statistical Aspects of BSE and vCJD Models for Epidemics. Monographs on Statistics and Applied Probability*, Chapman and Hall/CRC Press, London.

Donnelly, C., Ferguson, N., Ghani, A., Woolhouse, M., Watt, C. & Anderson, R. (1997), 'The epidemiology of BSE in cattle herds in Great Britain. I. Epidemiological process, demography of cattle and approaches to control by culling', *Philosophical Transactions of the Royal Society of London — Series B: Biological Sciences* **352**, 781–801.

Donnelly, C., Ghani, A. & Wilesmith, J. (1997), 'Analysis of the bovine spongiform encephalopathy maternal cohort study: Evidence for direct maternal transmission', *Applied Statistics* **46**(3), 321–344.

Dorling, D. (1995), 'The visualization of local urban change across Britain', *Environment and Planning B — Planning and Design* **22**(3), 269–290.

Ebdon, D. (1985), *Statistics in Geography*, Basil Blackwell, New York.

Eddy, R. (1995), 'Origin of BSE', *Veterinary Record* **137**(25), 648.

Efron, B. (1977), 'The efficiency of Cox's likelihood function for censored data', *Journal of the American Statistical Association* **76**, 312–319.

Elliott, P., Wakefield, J., Best, N. & Briggs, D. (2000), *Spatial Epidemiology Methods and Applications*, Oxford University Press, London.

Ferguson, N., Donnelly, C. & Anderson, R. (2001), 'The foot-and-mouth disease epidemic in Great Britain: pattern of spread and impact of interventions', *Science* **292**, 1155–1160.

Ferguson, N., Donnelly, C., Woolhouse, M. & Anderson, R. (1997), 'The epidemiology of BSE in cattle herds in Great Britain. II. Model construction and analysis of transmission dynamics', *Philosophical Transactions of the Royal Society of London — Series B: Biological Sciences* **352**, 803–838.

Ferguson, N., Ghani, A., Donnelly, C., Hagenaars, T. & Anderson, R. (2002), 'Estimating the human health risk from possible BSE infection of the British sheep flock', *Nature* **415**, 420–424.

Foley, J., Foley, P. & Madigan, J. (2001), 'Spatial distribution of seropositivity to the causative agent of granulocytic ehrlichiosis in dogs in California', *American Journal of Veterinary Research* **62**(10), 1599–1605.

Follet, B. (2002), *Infectious diseases in livestock. Scientific questions relating to the transmission and control of epidemic outbreaks of infectious disease in livestock in Great Britain*, The Royal Society, London.

French, N., Berriatua, E., Wall, R., Smith, K. & Morgan, K. (1999), 'Sheep scab outbreaks in Great Britain between 1973 and 1992: Spatial and temporal patterns', *Veterinary Parasitology* **83**(3-4), 187–200.

French, N., Clough, H., Berriatua, E., McCarthy, H., Proudman, C. & Hillyer, M. (2000), The use of K-function analysis to detect and describe space-time clustering of animal diseases, *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of

9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Fuchs, K. & Deutz, A. (2002), 'Use of variograms to detect critical spatial distances for the Knox's test', *Preventive Veterinary Medicine* **54**(1), 37–45.

Garner, M. & Lack, M. (1995), 'Modelling the potential impact of exotic diseases on regional Australia', *Australian Veterinary Journal* **72**(3), 81–87.

Gatrell, A., Bailey, T., Diggle, P. & Rowlingson, B. (1996), 'Spatial point pattern analysis and its application to medical geography', *Transactions of the Institute of British Geographers* **21**(1), 256–274.

Geary, R. (1954), 'The contiguity ratio and statistical mapping', *The Incorporated Statistician* **5**, 115–145.

Gelman, A. (1996), Inference and monitoring convergence, *in* W. Gilks, S. Richardson & D. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman and Hall, London, pp. 131–143.

Gerbier, G., Bacro, J., Pouillot, R., Durand, B., Moutou, F. & Chadoeuf, J. (2002), 'A point pattern model of the spread of foot-and-mouth disease', *Preventive Veterinary Medicine* **56**(1), 33–49.

Ghosh, M., Natarajan, K., Stroud, T. & Carlin, B. (1998), 'Generalized linear models for small-area estimation', *Journal of the American Statistical Association* **93**, 273–282.

Gibbens, J., Sharpe, C., Wilesmith, J., Mansley, L., Michalopoulou, E., Ryan, J. & Hudson, M. (2001), 'Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months', *Veterinary Record* **149**, 729–743.

Gibbens, J. & Wilesmith, J. (2002), 'Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain', *Veterinary Record* **151**, 407–412.

Gilks, W., Richardson, S. & Spiegelhalter, D. (1996a), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Gilks, W., Richardson, S. & Spiegelhalter, D. (1996b), Introducing Markov chain Monte Carlo, *in* W. Gilks, S. Richardson & D. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman and Hall, London, pp. 1–19.

Glick, B. (1979), 'The spatial autocorrelation of cancer mortality', *Social Science and Medicine* **13**, 123–130.

Gregorio, D., Kulldorff, M., Barry, L. & Samociuk, H. (2002), 'Geographic differences in invasive and in situ breast cancer incidence according to precise geographic coordinates, Connecticut, 1991 – 95', *International Journal of Cancer* **100**(2), 194–198.

Gregorio, D., Kulldorff, M., Barry, L., Samocuik, H. & Zarfos, K. (2001), 'Geographical differences in primary therapy for early-stage breast cancer', *Annals of Surgical Oncology* **8**(10), 844–849.

Griffith, D. & Layne, L. (1999), *A Casebook for Spatial Statistical Data Analysis*, Oxford University Press, London.

Haining, R., Wise, S. & Ma, J. (1998), 'Exploratory spatial data analysis in a geographic information systems environment', *The Statistician* **47**(3), 457–469.

Hammond, R., McGrath, G. & Martin, S. (2001), 'Irish soil and land-use classifications as predictors of numbers of badgers and badger setts', *Preventive Veterinary Medicine* **51**(3-4), 137–148.

Hanchette, C. & Schwartz, G. (1992), 'Geographic patterns of prostate-cancer mortality — evidence for a protective effect of ultraviolet-radiation', *Cancer* **70**(12), 2861–2869.

Harper, W. & Furr, J. (1986), 'Geostatistical analysis of potentiometric data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas'.

Haslett, J., Bradley, R., Craig, P., Unwin, A. & Wills, G. (1991), 'Dynamic graphics for exploring spatial data with application to locating global and local anomalies', *The American Statistician* **45**(3), 234–242.

Hastie, T. & Tibshirani, R. (1990), *Generalised Additive Models*, Chapman and Hall, London.

Hayes, D. (1997), The Development of an Expert System for Diagnosing Reproductive Problems in Seasonal Dairy Herds, Master's thesis, Massey University.

Heisterkamp, S., Doornbos, G. & Gankema, M. (1993), 'Disease mapping using Empirical Bayes and Bayes methods on mortality statistics in the Netherlands', *Statistics in Medicine* **12**, 1895–1913.

Helsetjenesten for Storfe (2003), 'Norwegian Cattle Health Services (NCHS)', http://www.storfehelse.tine.no.

Hjalmars, U., Kulldorff, M., Gustafsson, G. & Nagarwalla, N. (1996), 'Childhood leukaemia in Sweden: Using GIS and a spatial scan statistic for cluster detection', *Statistics in Medicine* **15**, 707–715.

Hjalmars, U., Kulldorff, M., Wahlqvist, Y. & Lannering, B. (1999), 'Increased incidence rates but no space-time clustering of childhood astrocytoma in Sweden, 1973 – 1992 — A population-based study of pediatric brain tumors', *Cancer* **85**(9), 2077–2090.

HMSO (1983), *Foot-and-Mouth Disease Order. Statutory Instrument 1983, Number 1950*, HMSO, London.

HMSO (1988a), 'The Bovine Spongiform Encephalopathy Order. Statutory Instrument 1988, Number 1039', HMSO, London.

HMSO (1988b), 'The Bovine Spongiform Encephalopathy Amendment Order. Statutory Instrument 1988b, Number 1345', HMSO, London.

HMSO (1990), 'The Bovine Spongiform Encephalopathy (No 2) Amendment Order. Statutory Instrument 1990, Number 1930', HMSO, London.

HMSO (1996a), 'The Cattle Passports Order. Statutory Instrument 1996, Number 1686', HMSO, London.

HMSO (1996b), 'The Bovine Spongiform Encephalopathy (Amendment) Order. Statutory Instrument 1996, Number 962', HMSO, London.

HMSO (1996c), 'The Bovine Spongiform Encephalopathy Order. Statutory Instrument 1996, Number 2007', HMSO, London.

Hogg, R. (1979), 'Statistical robustness: One view of its use in applications today', *American Statistician* **33**, 108–116.

Hoinville, L. (1994), 'Decline in the incidence of BSE in cattle born after the introduction of the 'feed ban'', *Veterinary Record* **134**(11), 274–275.

Hörnlimann, B., Heim, D. & Griot, C. (1996), Evaluation of BSE risk factors among European countries, *in* C. Gibbs, ed., 'Bovine Spongiform Encephalopathy: The BSE Dilemma', Springer-Verlag, New York, pp. 384–394.

Hosmer, D. & Lemeshow, S. (1999), *Applied Survival Analysis Regression Modelling of Time to Event Data*, Jon Wiley and Sons Inc, London.

Howard, S. & Donnelly, C. (2000), 'The importance of immediate destruction in epidemics of foot-and-mouth disease', *Research in Veterinary Science* **69**, 189–196.

Hungerford, L. (1991), 'Use of spatial statistics to identify and test significance in geographic disease patterns', *Preventive Veterinary Medicine* **11**(3-4), 237–242.

Hungerford, L. & Smith, R. (1996), 'Spatial and temporal patterns of bovine anaplasmosis as reported by Illinois veterinarians', *Preventive Veterinary Medicine* **25**(3-4), 310–313.

Ihaka, R. & Gentleman, R. (1996), 'R: A language for data analysis and graphics', *Journal of Computational and Graphical Statistics* **5**, 299–314.

International Agency for Research on Cancer (1985), *Scottish Cancer Atlas, IARC Scientific Publication 72*, IARC, Lyon.

Jalvingh, A., Nielen, M., Maurice, H., Stegeman, A., Elbers, A. & Dijkhuizen, A. (1999), 'Spatial and stochastic simulation to evaluate the impact of events and control measures ion the 1997-1998 classical swine fever epidemic in The Netherlands. I. Description of simulation model', *Preventive Veterinary Medicine* **42**(3-4), 271–295.

Jarup, L., Best, N., Toledano, M., Wakefield, J. & Elliott, P. (2002), 'Geographical epidemiology of prostate cancer in Great Britain', *International Journal of Cancer* **97**(5), 695–699.

Jemal, A., Kulldorff, M., Devesa, S., Hayes, R. & Fraumeni, J. (2002), 'A geographic analysis of prostate cancer mortality in the United States, 1970 – 89', *International Journal of Cancer* **101**(2), 168–174.

Jerrett, M., Burnett, R., Kanaroglou, P., Eyles, J., Finkelstein, N. & Brook, J. (2001), 'A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada', *Environment and Planning A* **33**(6), 955–973.

Kafadar, K. (1996), 'Smoothing geographical data, particularly rates of disease', *Statistics in Medicine* **15**, 2539–2560.

Kao, R. (2001), 'Landscape fragmentation and foot-and-mouth disease transmission', *Veterinary Record* **148**, 746–747.

Kao, R., Gravenor, M., Baylis, M., Bostock, C., Chihota, C., Evans, J., Goldmann, W., Smith, A. & McLean, A. (2002), 'The potential size and duration of an epidemic of bovine spongiform encephalopathy in British sheep', *Science* **295**, 332–335.

Keeling, M., Woolhouse, M., May, R., Davies, G. & Grenfell, B. (2003), 'Modelling vaccination strategies against foot-and-mouth disease', *Nature* **421**(6919), 136–142.

Keeling, M., Woolhouse, M., Shaw, D., Matthews, L., Chase-Topping, M., Haydon, D., Cornell, S., Kappey, J., Wilesmith, J. & Grenfell, B. (2001), 'Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape', *Science* **294**(5543), 813–817.

Kelsall, J. & Diggle, P. (1995a), 'Kernel estimation of relative risk', *Bernoulli* **1**, 3–6.

Kelsall, J. & Diggle, P. (1995b), 'Non-parametric estimation of spatial variation in relative risk', *Statistics in Medicine* **14**, 2335–2342.

Kelsall, J. & Diggle, P. (1998), 'Spatial variation in risk: A nonparametric binary regression approach', *Applied Statistics* **47**, 559–573.

Kimberlin, R. (1996), Speculations on the origin of BSE and the epidemiology of CJD, *in* C. Gibbs, ed., 'Bovine spongiform encephalopathy. The BSE dilemma', Springer-Verlag, New York, pp. 155–175.

Kimberlin, R. & Wilesmith, J. (1994), Bovine spongiform encephalopathy; epidemiology, low dose exposure, and risks, *in* 'Slow Infections of the Central Nervous System', New York Academy of Sciences, New York, pp. 210–220.

Kitching, R. & Mackay, D. (1995), 'Foot-and-mouth disease', *State Veterinary Journal* **5**, 4–8.

Kleinschmidt, I., Sharp, B., Mueller, I. & Vounatsou, P. (2002), 'Rise in malaria incidence rates in South Africa: A small-area spatial analysis of variation in time trends', *American Journal of Epidemiology* **155**(3), 257–264.

Knorr-Held, L. & Besag, J. (1998), 'Modelling risk from a disease in space and time', *Statistics in Medicine* **17**, 2045–2060.

Knox, G. (1964), 'The detection of space-time interactions', *Applied Statistics* **13**, 25–29.

Koussoulakou, A. & Kraak, M. (1992), 'Spatio-temporal maps and cartographic communication', *The Cartographic Journal* **29**, 101–108.

Krebs, J., May, R. & Stumpf, M. (2002), 'Theoretical models of sheep BSE reveal possibilities', *Nature* **415**, 115.

Kulldorff, M. (1997), 'A spatial scan statistic', *Communications in Statistics: Theory and Methods* **26**, 1481–1496.

Kulldorff, M., Feuer, E., Miller, B. & Freedman, L. (1997), 'Breast cancer clusters in the Northeast United States: A geographic analysis', *American Journal of Epidemiology* **146**(2), 161–169.

Kulldorff, M. & Nagarwalla, N. (1995), 'Spatial disease clusters: Detection and inference', *Statistics in Medicine* **14**, 799–810.

Kulldorff, M., Rand, K., Gherman, G., Williams, G. & DeFrancesco, D. (1998), 'SaTScan: Software for the spatial and space-time scan statistics Version 2.1', National Cancer Institute.

Langford, I. (1994), 'Using empirical Bayes estimates in the geographical analysis of risk', *Area* **26**, 142–190.

Lawson, A. (1989), 'Contribution to the 'cancer near nuclear installations' meeting', *Journal of the Royal Statistical Society Series A* **152**, 374–375.

Lawson, A. (1992), 'GLIM and normalising constant models in spatial and directional data analysis', *Computational Statistics and Data Analysis* **13**, 331–348.

Lawson, A. (1993), 'A deviance residual for heterogenous spatial Poisson processes', *Biometrics* **49**, 889–897.

Lawson, A. (1994), 'On using spatial Gaussian priors to model heterogeneity in environmental epidemiology', *Statistician* **43**, 69–76.

Lawson, A. (1995), 'MCMC methods for putative pollution source problems in environmental epidemiology', *Statistics in Medicine* **14**, 2473–2485.

Lawson, A. (2001a), *Statistical Methods in Spatial Epidemiology*, John Wiley and Sons Ltd, London.

Lawson, A. (2001b), 'Disease map reconstruction', *Statistics in Medicine* **20**, 2183–2204.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F. & Bertollini, R. (1999), *Disease Mapping and Risk Assessment for Public Health*, John Wiley and Sons Ltd, London.

Lawson, A., Biggeri, A. & Dreassi, E. (1999), Edge effects in disease mapping, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 85–98.

Lawson, A., Biggeri, A. & Williams, F. (1999), A review of modelling approaches in health risk assessment around putative sources, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 231–245.

Lawson, A., Böhning, D., Biggeri, A., Lesaffre, E. & Viel, J.-F. (1999), Disease mapping and its uses, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 3–13.

Lawson, A. & Clark, A. (1999), Markov chain Monte Carlo methods for putative sources of hazard and general clustering, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 119–142.

Lawson, A. & Clark, A. (2002), 'Spatial mixture relative risk models applied to disease mapping', *Statistics in Medicine* **21**(3), 359–370.

Lawson, A. & Kulldorff, M. (1999), A review of cluster detection methods, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 99–110.

Lawson, A. & Williams, F. (1993), 'Application of extraction mapping in environmental epidemiology', *Statistics in Medicine* **12**, 1249–1258.

Lawson, A. & Williams, F. (1994), 'Armadale: A case study in environmental epidemiology', *Journal of the Royal Statistical Society Series A* **157**(2), 285–298.

Lawson, A. & Williams, F. (2000), 'Spatial competing risk models in disease mapping', *Statistics in Medicine* **19**(17-18), 2451–2467.

Lawson, A. & Williams, F. (2001), *An Introductory Guide to Disease Mapping*, John Wiley and Sons Ltd, London.

Le, N., Petkau, A. & Rosychuk, R. (1996), 'Surveillance of clustering near point sources', *Statistics in Medicine* **15**, 727–740.

Leforban, Y. (2002), 'How predictable were the outbreaks of foot and mouth disease in Europe in 2001 and is vaccination the answer?', *Revue Scientifique Et Technique De L'Office International Des Epizooties* **21**(3), 549–556.

Leyland, A. (2001), Spatial analysis, *in* A. Leyland & H. Goldstein, eds, 'Multilevel Modelling of Health Statistics', John Wiley and Sons Ltd, London, pp. 143–158.

Little, R. & Rubin, D. (1987), *Statistical Analysis with Missing Data*, John Wiley and Sons Ltd, London.

Livestock Improvement Corporation (2003), 'MINDA Health', http://www.lic.co.nz.

Lloyd, O. (1982), Mortality in a small industrial town, *in* A. Gardner, ed., 'Current Approaches to Occupational Health — 2', Wright PSG, London, pp. 283–309.

Loader, C. (1999), *Local Regression and Likelihood*, Springer-Verlag, New York.

MacNab, Y. & Dean, C. (2000), 'Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models', *Statistics in Medicine* **19**(17-18), 2421–2435.

MAFF (1986 – 1998), *1985 – 1997 Agricultural statistics — United Kingdom*, HMSO, London.

MAFF (1999), *Bovine spongiform encephalopathy in Great Britain, a progress report*, HMSO, London.

MAFF (2001), *2000 Agricultural statistics — United Kingdom*, HMSO, London.

Maheswaran, R., Strachan, D., Dodgeon, B. & Best, N. (2002), 'A population-based case-control study for examining early life influences on geographical variation in adult mortality in England and Wales using stomach cancer and stroke as examples', *International Journal of Epidemiology* **31**(2), 375–382.

Mangen, M., Jalvingh, A., Nielen, M., Mourits, M., Klinkenberg, D. & Dijkhuizen, A. (2001), 'Spatial and stochastic simulation to compare two emergency-vaccination strategies with a marker vaccine in the 1997/1998 Dutch Classical Swine Fever epidemic', *Preventive Veterinary Medicine* **48**(3), 177–200.

Mannelli, A., Sotgia, S., Patta, C., Oggiano, A., Carboni, A., Cossu, P. & Laddomada, A. (1998), 'Temporal and spatial patterns of African swine fever in Sardinia', *Preventive Veterinary Medicine* **35**(4), 297–306.

Mantel, N. (1967), 'The detection of disease clustering and a generalized regression approach', *Cancer Research* **27**, 209–220.

Manton, K., Woodbury, M. & Stallard, E. (1981), 'A variance components approach to categorical data models with heterogenous mortality rates in North Carolina counties', *Biometrics* **37**, 259–269.

Marshall, E. & Spiegelhalter, D. (2000), 'Strategies for inference robustness in Bayesian modelling', http://www/mrc-bsu-cam.ac.uk/research/preprints.shtml.

Marshall, R. (1991*a*), 'A review of methods for the statistical analysis of spatial patterns of disease', *Journal of the Royal Statistical Society Series A* **154**, 421–441.

Marshall, R. (1991*b*), 'Mapping disease and mortality rates using empirical Bayes estimators', *Applied Statistics* **40**, 283–294.

Martuzzi, M. & Elliott, P. (1996), 'Empirical Bayes estimation of small area prevalence of non-rare conditions', *Statistics in Medicine* **15**, 1867–1873.

Matheron, G. (1965), *Les Variables Régionalisées et leur Estimation* , Masson, Paris.

Matheron, G. (1971), *The theory of regionalized variables and its applications*, Centre de Morphologie Mathématique de Fontainebleu, Paris.

McKenzie, J. (1999), The Use of Habitat Analysis in the Control of Wildlife Tuberculosis in New Zealand, PhD thesis, Massey University.

Meyers, R., Aramini, J., Lim, G., Stratton, J., Pollari, F., Sockett, P. & Majowicz, S. (2002), Using Geographic Information Systems to identify the source of a waterborne gastrointestinal outbreak in Walkerton, Ontario, Canada, May – June, 2000, *in* J. Rigby, C. Skelly & P. Wigham, eds, 'GeoHealth 2002 Proceedings of the Spatial Information Research Centre's 14th Colloquium', Victoria University, Wellington, New Zealand, pp. 71–74.

Mollié, A. (1996), Bayesian mapping of disease, *in* S. Gilks & S. Richardson, eds, 'Markov Chain Monte Carlo in Practice', Chapman and Hall, London, pp. 359–379.

Mollié, A. (2000), Bayesian mapping of disease, *in* P. Elliott, J. Wakefield, N. Best & D. Briggs, eds, 'Spatial Epidemiology Methods and Applications', Oxford University Press, London, pp. 267–285.

Monmonier, M. & De Blij, H. (1996), *How to Lie with Maps*, University of Chicago Press, Chicago.

Moore, D. (1999), 'Spatial diffusion of raccoon rabies in Pennsylvania, USA', *Preventive Veterinary Medicine* **40**(1), 19–32.

Moran, P. (1950), 'Notes on continuous stochastic phenomena', *Biometrika* **37**, 17–23.

Morris, R., Sanson, R., Stern, M., Stevenson, M. & Wilesmith, J. (2002), 'Decision-support tools for foot and mouth disease control', *Revue Scientifique Et Technique De L'Office International Des Epizooties* **21**(3), 557–567.

Morris, R., Wilesmith, J., Stern, M., Sanson, R. & Stevenson, M. (2001), 'Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001', *Veterinary Record* **149**(5), 137–144.

Morrison, A., Getis, A., Santiago, M., Rigau-Perez, J. & Reiter, P. (1998), 'Exploratory space-time analysis of reported Dengue cases during an outbreak in Florida, Puerto Rico, 1991 – 1992', *American Journal of Tropical Medicine and Hygiene* **58**(3), 287–298.

Moutou, F. & Durand, B. (1994), 'Modelling the spread of foot-and-mouth disease virus', *Veterinary Research* **25**(2/3), 279–285.

Müller, I., Betuela, I. & Hide, R. (2002), 'Regional patterns of birthweights in Papua New Guinea in relation to diet, environment and socio-economic factors', *Annals of Human Biology* **29**(1), 74–88.

NAHMS (1996), *Dairy '96 Part I: Reference of 1996 Dairy Management Practices*, Center for Animal Health Monitoring, Fort Collins, Colorado.

Nielen, M., Jalvingh, A., Dijkhuizen, A. & Lattuada, R. (1996), Requirements of a geographical information system to be used during a foot-and-mouth disease outbreak, *in* 'Proceedings of a Meeting of the Society for Veterinary Epidemiology and Preventive Medicine', Society for Veterinary Epidemiology and Preventive Medicine, Roslin, UK, pp. 178–184.

Nielen, M., Jalvingh, A., Meuwissen, M., Horst, S. & Dijkhuizen, A. (1999), 'Spatial and stochastic simulation to evaluate the impact of events and control measures on the 1997-1998 classical swine fever epidemic in The Netherlands. II. Comparison of control strategies', *Preventive Veterinary Medicine* **42**(3-4), 297–317.

Noordegraaf, A., Jalvingh, A., de Jong, M., Franken, P. & Dijkhuizen, A. (2000), 'Evaluating control strategies for outbreaks in BHV1-free areas using stochastic and spatial simulation', *Preventive Veterinary Medicine* **44**(1-2), 21–42.

Nörstrom, M. & Jarp, J. (2000), Space-time clustering of acute respiratory disease in cattle herds in Norway , *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of 9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Nörstrom, M., Pfeiffer, D. & Jarp, J. (2000), 'A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds', *Preventive Veterinary Medicine* **47**(1-2), 107–119.

O'Brien, D., Kaneene, J., Getis, A., Lloyd, J., Rip, M. & Leader, R. (1999), 'Spatial and temporal distribution of selected canine cancers in Michigan, USA, 1964 – 1994', *Preventive Veterinary Medicine* **42**(1), 1–15.

O'Brien, D., Kaneene, J., Getis, A., Lloyd, J., Swanson, G. & Leader, R. (2000), 'Spatial and temporal comparison of selected cancers in dogs and humans, Michigan, USA, 1964 – 1994', *Preventive Veterinary Medicine* **47**(3), 187–204.

Oden, N. (1995), 'Adjusting Moran's I for population density', *Statistics in Medicine* **14**, 17–26.

Openshaw, S. (1984), *The Modifiable Areal Unit Problem (Concepts and Techniques in Modern Geography, No. 38)*, Geo Books, Norwich.

Openshaw, S., Charlton, M., Wymer, C. & Craft, A. (1987), 'A mark I geographical analysis machine for the automated analysis of point data sets', *International Journal of Geographical Information Systems* **1**, 335–358.

Parzen, E. (1962), 'On the estimation of probability density and mode', *Annals of Mathematical Statistics* **33**, 1065–1076.

Pascutto, C., Wakefield, J., Best, N., Richardson, S., Bernardinelli, L., Staines, A. & Elliott, P. (2000), 'Statistical issues in the analysis of disease mapping data', *Statistics in Medicine* **19**(17-18), 2493–2519.

Pasman, E., Otte, M. & Esslemont, R. (1995), 'Influences of milk yield, fertility, and health in the first lactation on the length of productive life of dairy cows in Great Britain', *Preventive Veterinary Medicine* **24**, 55–63.

Perez, A., Ward, M., Torres, P. & Ritacco, V. (2002), 'Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina', *Preventive Veterinary Medicine* **56**(1), 63–74.

Perez, E., Herrero, M., Jimenez, C., Carpenter, T. & Buening, G. (1994), 'Epidemiology of bovine anaplasmosis and babesiosis in Costa Rica', *Preventive Veterinary Medicine* **20**(1-2), 23–31.

Perry, B., Kruska, R., Lessard, P., Norval, R. & Kundert, K. (1991), 'Estimating the distribution and abundance of *Rhipicephalus appendiculatus* in Africa', *Preventive Veterinary Medicine* **11**, 261–268.

Pfeiffer, D., Duchateau, L., Kruska, R., Ushewokunze-Obatolu, U. & Perry, B. (1997), A spatially predictive logistic regression model for occurrence of theileriosis outbreaks in Zimbabwe, *in* 'Proceedings of 8th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Paris, pp. 12.12.1–3.

Piantadosi, S., Byar, D. & Green, S. (1988), 'The ecological fallacy', *American Journal of Epidemiology* **127**, 893–904.

Pickle, L. (2000), 'Exploring spatio-temporal patterns of mortality using mixed-effects models', *Statistics in Medicine* **19**(17-18), 2251–2263.

Pickle, L., Mungiole, M., Jones, G. & White, A. (1996), *Atlas of United States Mortality*, National Center for Health Statistics, Hyattsville.

Pickle, L. & White, A. (1995), 'Effects of the choice of age adjustment method on maps of death rates', *Statistics in Medicine* **14**, 615–627.

Plummer, M. & Clayton, D. (1996), 'Estimation of population exposure in ecological studies', *Journal of the Royal Statistical Society Series B* **58**, 113–126.

Pugliatti, M., Solinas, G., Sotgiu, S., Castiglia, P. & Rosati, G. (2002), 'Multiple sclerosis distribution in Northern Sardinia — spatial cluster analysis of prevalence', *Neurology* **58**(2), 277–282.

Radostits, O., Gay, C., Blood, D. & Hinchcliff, K. (2000), *Veterinary Medicine*, WB Saunders, London.

Raftery, A. & Lewis, S. (1992a), 'Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo', *Statistical Science* **7**, 493–497.

Raftery, A. & Lewis, S. (1992b), How many iterations in the Gibbs sampler?, *in* J. Benardo, J. Berger, A. Dawid & A. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, London, pp. 763–774.

Ranta, J. & Penttinen, A. (2000), 'Probabilistic small area risk assessment using GIS-based data: A case study on Finnish childhood diabetes', *Statistics in Medicine* **19**, 2345–2359.

Ranta, J., Pitkäniemi, J., Karvonen, M., Virtala, E., Rusanen, J., Colpaert, A., Naukkarinen, A. & Tuomilehto, J. (1996), 'Detection of over space-time clustering in a non-uniformly distributed population', *Statistics in Medicine* **15**, 2561–2572.

Review Committee (2001), *Review of the Origin of BSE*, HMSO, London.

Ribeiro Jr, P. & Diggle, P. (2001), 'geoR: a package for geostatistical analysis', *R News* **1/2**(15-18).

Richardson, S. & Monfort, C. (2000), Ecological correlation studies, *in* P. Elliott, J. Wakefield, N. Best & D. Briggs, eds, 'Spatial Epidemiology Methods and Applications', Oxford University Press, London, pp. 205–220.

Ripley, B. (1976), 'The second-order analysis of stationary point processes', *Journal of Applied Probability* **13**, 255–266.

Ripley, B. (1977), 'Modelling spatial patterns', *Journal of the Royal Statistical Society Series B* **39**, 172–192.

Ripley, B. (1981), *Spatial Statistics*, John Wiley and Sons Ltd, London.

Robert, C. & Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Rodríguez-Lainz, A., Hird, D., Carpenter, T., Read, D. & Lainz, A. (1996), 'Case-control study of papillomatous digital dermatitis in southern California dairy farms', *Preventive Veterinary Medicine* **28**(2), 117–131.

Rosenblatt, M. (1956), 'Remarks on some non-parametric estimates of a density function', *Annals of Mathematical Statistics* **27**, 832–837.

Rothman, K. (1981), 'Induction and latent periods', *American Journal of Epidemiology* **114**(2), 253–259.

Rothman, K. & Greenland, S. (1998), *Modern Epidemiology*, Lippincott-Raven, Philadelphia.

Rowlingson, B. & Diggle, P. (1993), 'SPLANCS: Spatial point pattern analysis code in S-PLUS', *Computers in Geosciences* **19**, 627–655.

Rudemo, M. (1982), 'Empirical choice of histograms and kernel density estimators', *Scandinavian Journal of Statistics* **9**, 65–78.

Rushton, G. & Lolonis, P. (1996), 'Exploratory spatial analysis of birth defect rates in an urban population', *Statistics in Medicine* **15**, 717–726.

Salt, J., Barnett, P., Dani, P. & Williams, L. (1998), 'Emergency vaccination of pigs against foot and mouth disease: protection against disease and reduction in contact transmission', *Vaccine* **16**(746-754).

Sankoh, O., Ye, Y., Sauerborn, R., Muller, O. & Becher, H. (2001), 'Clustering of childhood mortality in rural Burkina Faso', *International Journal of Epidemiology* **30**(3), 485–492.

Sanson, R. (1993), The development of a decision support system for an animal disease emergency, PhD thesis, Massey University.

Sanson, R., Liberona, H. & Morris, R. (1991), 'The use of a geographical information system in the management of a foot-and-mouth disease epidemic', *Preventive Veterinary Medicine* **11**(3-4), 309–313.

Sanson, R., Morris, R. & Stern, M. (1999), 'EpiMAN-FMD: a decision support system for managing epidemics of vesicular disease', *Revue Scientifique Et Technique Office International Des Epizooties* **18**(3), 593–605.

Sanson, R. & Ryan, J. (1997), *User manual for the BSE release database*, Epidemiology Department, Central Veterinary Laboratory, New Haw, Addlestone Surrey.

Sayers, A. (1999), 'Dose response of cattle infected orally with BSE as measured by the incubation period of the disease', Epidemiology Unit, Veterinary Laboratories Agency, Addlestone, Surrey.

Sayers, A. (2000), 'Dose response of cattle artificially infected with BSE as measured by the incubation period of the disease', Epidemiology Unit, Veterinary Laboratories Agency, Addlestone, Surrey.

Schafer, J. (1997), *Analysis of Incomplete Multivariate Data — Monographs on Statistics and Applied Probability 72*, Chapman and Hall, London.

Schimert, J., Schafer, J., Hesterberg, T., Fraley, C. & Clarkson, D. (2001), *Analyzing data with missing values in S-PLUS*, Insightful Corporation, Seattle, Washington.

Schlattmann, P. (1996a), 'The computer package DismapWin', *Statistics in Medicine* **15**, 931.

Schlattmann, P. (1996b), 'Covariate adjusted mixture models and disease mapping with the program DismapWin', *Statistics in Medicine* **15**, 919–929.

Schober, E., Rami, B., Waldhor, T. & Karimian-Teherani, D. (2001), 'Regional distribution of childhood onset diabetes mellitus in Austria — analysis of national registration from 1989 to 1999', *Wiener Klinische Wochenschrift* **113**(13-14), 491–495.

Schulman, J., Selvin, S. & Merrill, D. (1988), 'Density equalized map projections: A method for analysing clustering around a fixed point', *Statistics in Medicine* **7**, 491–505.

Scudamore, J. (2002), *Partnership, Priorities and Professionalism. A proposed strategy for enhancing veterinary surveillance in the UK*, Veterinary Surveillance Division, Department for Environment, Food and Rural Affairs, London.

Sellers, R. (1971), 'Quantitative aspects of the spread of foot-and-mouth disease', *Veterinary Bulletin* **41**, 431–439.

Selvin, S. & Merrill, D. (2002), 'Adult leukemia: A spatial analysis', *Epidemiology* **13**(2), 151–156.

Selvin, S., Merrill, D., Erdmann, C., White, M. & Ragland, K. (1998), 'Breast cancer detection: Maps of two San Francisco Bay area counties', *American Journal of Public Health* **88**(8), 1186–1192.

Selvin, S., Merrill, D., Schulman, J., Sacks, S., Bedell, L. & Wong, L. (1988), 'Transformations of maps to investigate clusters of disease', *Social Science and Medicine* **26**(2), 215–221.

Selvin, S., Schulman, J. & Merrill, D. (1992), 'Distance and risk measures for the analysis of spatial data — a study of childhood cancers', *Social Science and Medicine* **34**(7), 769–777.

SERAD (2001), *2000 Agricultural statistics — Scotland*, Department of Agriculture, Environment and Fisheries, Edinburgh.

Silman, A., Bankhead, C., Rowlingson, B., Brennan, P., Symmons, D. & Gatrell, A. (1997), 'Do new cases of rheumatoid arthritis cluster in time or in space?', *International Journal of Epidemiology* **26**(3), 628–634.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simes, R. (1986), 'An improved Bonferroni procedure for multiple tests of significance', *Biometrika* **73**, 751–754.

Singer, R., Case, J., Carpenter, T., Walker, R. & Hirsh, D. (1998), 'Assessment of spatial and temporal clustering of ampicillin- and tetracycline-resistant strains of *Pasteurella multocida* and *P haemolytica* isolated from cattle in California', *Journal of the American Veterinary Medical Association* **212**(7), 1001–1005.

Snow, J. (1855), *On the Mode of Communication of Cholera*, John Churchill, London.

SOAEFD (1987 – 1997), *1986 – 1996 Agricultural statistics — Scotland*, Department of Agriculture, Environment and Fisheries, Edinburgh.

Spence, S., Murison, R. & Harden, S. (1998), 'Rate of decline of chlorfluazuron concentration in the fat of cattle', *Australian Veterinary Journal* **76**(1), 54–56.

Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society Series B* **64**(583-616).

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (2000), 'BUGS: Bayesian Inference Using Gibbs Sampling Version 1.3', MRC Biostatistics Unit.

Staines, A. & Järup, L. (2000), Health event data, *in* P. Elliott, J. Wakefield, N. Best & D. Briggs, eds, 'Spatial Epidemiology Methods and Applications', Oxford University Press, London, pp. 15–29.

Stevenson, M. (2000), 'Disease incidence in dairy herds in the Southern highlands district of New South Wales, Australia', *Preventive Veterinary Medicine* **43**(1), 1–11.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lawson, A., Pfeiffer, D. & Lin, D. (2000b), 'A descriptive spatial analysis of the bovine spongiform encephalopathy epidemic in Great Britain', *Veterinary Record* **147**(14), 379–384.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lockhart, J., Lin, D. & Jackson, R. (2000a), 'Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: Individual animal-associated risk factors for disease', *Veterinary Record* **147**(13), 349–354.

Stone, C. (1974), 'Cross-validatory choice and assessment of statistical predictions', *Journal of the Royal Statistical Society Series B* **36**, 111–147.

Stoyan, D. & Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields*, John Wiley and Sons Ltd, London.

Sun, D., Tsutakawa, R., Kim, H. & He, Z. (2000), 'Spatio-temporal interaction with disease mapping', *Statistics in Medicine* **19**, 2015–2035.

Sun, D., Tsutakawa, R. & Speckman, P. (1999), 'Posterior distribution of hierarchical models using CAR(1) distributions ', *Biometrika* **86**(2), 341–350.

Swayne, D., Cook, D. & Buja, A. (1998), 'XGobi: Interactive Dynamic Data Visualization in the X Window System', *Journal of Computational and Graphical Statistics* **7**(1), –.

Symanzik, J., Kötter, T., Schmelzer, S., Klinke, S., Cook, D. & Swayne, D. (1998), 'Spatial Data Analysis in the Dynamically Linked ArcView/XGobi/XploRe Environment', *Computing Science and Statistics* **29**(1), 561–569.

Talbot, T., Kulldorff, M., Forand, S. & Haley, V. (2000), 'Evaluation of spatial filters to create smoothed maps of health data', *Statistics in Medicine* **19**(17-18), 2399–2408.

Taylor, D., Fernie, K., McConnell, I., Ferguson, C. & Steele, P. (1998), 'Solvent extraction as an adjunct to rendering: The effect on BSE and scrapie agents of hot solvents followed by dry heat and steam', *Veterinary Record* **143**, 6–9.

Taylor, D., Woodgate, S. & Atkinson, M. (1995), 'Inactivation of bovine encephalopathy agent by rendering procedures', *Veterinary Record* **137**, 605–610.

Teekayuwat, T., Pfeiffer, D. & Hayes, D. (2000), Spatial clustering of enzootic bovine leukosis in New Zealand, *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of 9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Thompson, D., Muriel, P., Russell, D., Osborne, P., Bromley, A., Rowland, M., Creigh-Tyte, S. & Brown, C. (2002), 'Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001', *Revue Scientifique Et Technique De L'Office International Des Epizooties* **21**(3), 675–687.

Thulke, H., Tischendorf, L., Staubach, C., Selhorst, T., Jeltsch, F., Muller, T., Schluter, H. & Wissel, C. (2000), 'The spatio-temporal dynamics of a post-vaccination resurgence of rabies in foxes and emergency vaccination planning', *Preventive Veterinary Medicine* **47**(1-2), 1–21.

Timmander, L. & McLafferty, S. (1998), 'Breast cancer in West Islip, NY', *Social Science and Medicine* **46**(12), 1623–1635.

Tinline, R., Rosatte, R. & MacInnes, C. (2002), 'Estimating the incubation period of raccoon rabies: a time-space clustering approach', *Preventive Veterinary Medicine* **56**(1), 89–103.

Toledano, M., Jarup, L., Best, N., Wakefield, J. & Elliott, P. (2001), 'Spatial variation and temporal trends of testicular cancer in Great Britain', *British Journal of Cancer* **84**(11), 1482–1487.

Torok, T., Kilgore, P. & Clarke, M. (1997), 'Visualizing geographic and temporal trends in rotavirus activity in the United States, 1991 to 1996', *Pediatric Infectious Disease Journal* **16**(10), 941–946.

Tsutakawa, R. (1988), 'A mixed model for analyzing geographic variability in mortality rates', *Journal of the American Statistical Association* **83**(401), 37–42.

Van Larebeke, N., Covaci, A., Schepens, P. & Hens, L. (2002), 'Food contamination with polychlorinated biphenyls and dioxins in Belgium. Effects on the body burden', *Journal of Epidemiology and Community Health* **56**(11), 828–830.

Viel, J., Arveux, P., Baverel, J. & Cahn, J. (2000), 'Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels', *American Journal of Epidemiology* **152**(1), 13–19.

Viel, J., Pobel, D. & Carré, A. (1995), 'Incidence of leukaemia in young people around the La Hague nuclear waste reprocessing plant: A sensitivity analysis', *Statistics in Medicine* **14**, 2459–2472.

Vinten-Johansen, P., Brody, H., Paneth, N., Rachman, S. & Rip, M. (2003), *Cholera, Chloroform, and the Science of Medicine – A Life of John Snow*, Oxford University Press, London.

Vlachonikolis, I., Aletra, T. & Georgoulias, V. (2002), 'Incidence of breast cancer on Crete, 1994 – 1995', *European Journal of Cancer* **38**, 574–577.

Wakefield, J. & Elliott, P. (1999), 'Issues in the statistical analysis of small area health data', *Statistics in Medicine* **18**, 2377–2399.

Wakefield, J., Kelsall, J. & Morris, S. (2000), Clustering, cluster detection, and spatial variation in risk, *in* P. Elliott, J. Wakefield, N. Best & D. Briggs, eds, 'Spatial Epidemiology Methods and Applications', Oxford University Press, London, pp. 128–152.

Wakefield, J. & Morris, S. (1999), Spatial dependence and errors-in-variables in environmental epidemiology, *in* J. Bernardo, J. Berger, A. Dawid & A. Smith, eds, 'Bayesian Statistics 6: Proceedings of the Sixth Valencia Meeting on Bayesian Statistics', Oxford University Press, Oxford, pp. 657–84.

Waller, L., Carlin, B., Xia, H. & Gelfand, A. (1997), 'Hierarchical spatio-temporal mapping of disease rates', *Journal of the American Statistical Association* **92**(438), 607–617.

Walter, S., Martin Taylor, S. & Marrett, L. (1999), An analysis of determinants of regional variation in cancer incidence: Ontario, Canada, *in* A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel & R. Bertollini, eds, 'Disease Mapping and Risk Assessment for Public Health', John Wiley and Sons Ltd, London, pp. 365–381.

Ward, M. (2000), Application of the Scan Statistic to detect time-space clustering of animal disease in the presence of confounding, *in* M. Salman, P. Morley & R. Ruch-Gallie, eds, 'Proceedings of 9th Symposium of the International Society for Veterinary Epidemiology and Economics', International Society for Veterinary Epidemiology and Economics, Fort Collins, Colorado, pp. –.

Ward, M. (2002), 'Clustering of reported cases of leptospirosis among dogs in the United States and Canada', *Preventive Veterinary Medicine* **56**(3), 215–226.

Ward, M. & Carpenter, T. (1995), 'Infection of cattle in Queensland with bluetongue viruses. II. Distribution of antibodies', *Australian Veterinary Journal* **72**(217-220).

Ward, M. & Carpenter, T. (2000a), 'Techniques for analysis of disease clustering in space and in time in veterinary epidemiology', *Preventive Veterinary Medicine* **45**(3-4), 257–284.

Ward, M. & Carpenter, T. (2000b), 'Analysis of time-space clustering in veterinary epidemiology', *Preventive Veterinary Medicine* **43**(4), 225–237.

Ward, M., Carpenter, T. & Johnson, S. (1996), 'Spatial analysis of seroconversion of sentinel cattle to bluetongue viruses in Queensland', *Australian Veterinary Journal* **74**(2), 128–131.

Wells, G., Hancock, R., Cooley, W., Richards, M., Higgins, R. & David, G. (1989), 'Bovine spongiform encephalopathy: Diagnostic significance of vacuolar changes in selected nuclei of the medulla oblongata', *Veterinary Record* **125**, 521–524.

Wells, G., Scott, A., Johnson, C., Gunning, R., Hancock, R., Jeffrey, M., Dawson, M. & Bradley, R. (1987), 'A novel progressive spongiform encephalopathy in cattle', *Veterinary Record* **121**(419-420).

White, M., Schukken, Y. & Tanksley, B. (1989), 'Space-time clustering of, and risk factors for, farmer diagnosed winter dysentery in dairy cattle', *Canadian Veterinary Journal* **30**, 948–951.

Whittle, P. (1958), 'On the smoothing of probability density functions', *Journal of the Royal Statistical Society Series B* **55**, 549–557.

WHO (1977), *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, WHO, Geneva.

WHO (1992), *ICD10 International Statistical Classification of Diseases and Related Health Problems Volume 1*, WHO, Geneva.

Wilesmith, J. (1991a), 'Epidemiology of bovine spongiform encephalopathy', *Seminars in Virology* **2**, 239–245.

Wilesmith, J. (1991b), 'Origins of BSE', *Veterinary Record* **128**, 310.

Wilesmith, J. (1994), 'An epidemiologist's view of bovine spongiform encephalopathy', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **343**(1306), 357–361.

Wilesmith, J. (1994b), 'Bovine spongiform encephalopathy: Epidemiological factors associated with the emergence of an important new animal pathogen in Great Britain', *Seminars in Virology* **5**(2), 179–187.

Wilesmith, J. (1996), Recent observations on the epidemiology of bovine spongiform encephalopathy, *in* C. Gibbs, ed., 'Bovine spongiform encephalopathy: The BSE dilemma', Springer-Verlag, New York, pp. 45–55.

Wilesmith, J. (1997), What Can We Learn From BSE?, *in* , ed., 'Proceedings of the 30th Annual Convention of The American Association of Bovine Practitioners', American Association of Bovine Practitioners, Quebec, Canada, pp. 71–75.

Wilesmith, J. (1998), *Manual on bovine spongiform encephalopathy*, Food and Agriculture Organization of the United Nations, Rome.

Wilesmith, J. (2002), 'Preliminary epidemiological analyses of the first 16 cases born after July 31, 1996, in Great Britain', *Veterinary Record* **151**, 451–452.

Wilesmith, J. & Ryan, J. (1992), 'Bovine spongiform encephalopathy: Recent observations on the age-specific incidences', *Veterinary Record* **130**, 491–492.

Wilesmith, J. & Ryan, J. (1993), 'Bovine spongiform encephalopathy: Observations on the incidence during 1992', *Veterinary Record* **132**(12), 300–301.

Wilesmith, J. & Ryan, J. (1997), 'Absence of BSE in the offspring of pedigree suckler cows affected by BSE in Great Britain', *Veterinary Record* **141**, 250–251.

Wilesmith, J., Ryan, J. & Atkinson, M. (1991), 'Bovine spongiform encephalopathy: Epidemiological studies on the origin', *Veterinary Record* **128**, 199–203.

Wilesmith, J., Ryan, J. & Hueston, W. (1992), 'Bovine spongiform encephalopathy: Case control studies of calf feeding practices and meat and bone meal inclusion in proprietary concentrates', *Research in Veterinary Science* **52**, 325–331.

Wilesmith, J., Ryan, J., Hueston, W. & Hoinville, L. (1992), 'BSE: Epidemiological features 1985 to 1990', *Veterinary Record* **130**, 90–94.

Wilesmith, J., Ryan, J., Stevenson, M., Morris, R., Pfeiffer, D., Lin, D., Jackson, R. & Sanson, R. (2000), 'Temporal aspects of the bovine spongiform encephalopathy

epidemic in Great Britain: Holding-associated risk factors for disease', *Veterinary Record* **147**(12), 319–325.

Wilesmith, J. & Wells, G. (1991), Bovine spongiform encephalopathy, *in* B. Chesebro, ed., 'Transmissible Spongiform Encephalopathies. Current Topics in Microbiology and Immunology', Springer-Verlag, New York, pp. 21–38.

Wilesmith, J., Wells, G., Cranwell, M. & Ryan, J. (1988), 'Bovine spongiform encephalopathy: Epidemiological studies', *Veterinary Record* **123**, 638–644.

Williams, E., Smith, P., Day, N., Geser, A., Ellice, J. & Tukei, P. (1978), 'Space-time clustering of Burkitt's lymphoma in the West Nile district of Uganda 1961 – 1975', *British Journal of Cancer* **27**(109-122).

Xia, H., Carlin, B. & Waller, L. (1997), 'Hierarchical models for mapping Ohio lung cancer rates', *Environmetrics* **8**, 107–120.

Yu, B. (1994), *Monitoring the convergence of Markov samplers based on L^1 error, Technical Report 9409*, Department of Statistics, University of California, Berkeley, California.

Yu, B. & Mykland, P. (1998), 'Looking at Markov samplers through CUSUM path plots: A simple diagnostic idea', *Statistics and Computing* **8**, 275–286.

Yu, P., Habtemariam, T., Oryang, D., Obasa, M., Nganwa, D. & Robnett, V. (1995), 'Integration of temporal and spatial models for examining the epidemiology of African trypanosomiasis', *Preventive Veterinary Medicine* **24**(2), 83–95.

Zeman, P. (1997), 'Objective assessment of risk maps of tick-borne encephalitis and Lyme borreliosis based on spatial patterns of located cases', *International Journal of Epidemiology* **26**(5), 1121–1130.

Zheng, X., Pierce, G., Reid, D. & Jolliffe, I. (2002), 'Does the North Atlantic current affect spatial distribution of whiting? Testing environmental hypotheses using statistical and GIS techniques', *Ices Journal of Marine Science* **59**(2), 239–253.