Spatial Modelling Methods

Prepared for the National Health Performance Authority

June 2016

Suggested citation

Cramb SM, Duncan EW, White NM, Baade PD, Mengersen KL, 2016. *Spatial Modelling Methods*. Brisbane: Cancer Council Queensland and Queensland University of Technology (QUT).

Author affiliations

Viertel Cancer Research Centre, Cancer Council Queensland: Susanna Cramb and Peter Baade.

ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology (QUT): Susanna Cramb, Earl Duncan, Nicole White and Kerrie Mengersen.

Acknowledgements

The authors wish to thank the following people for their helpful feedback and advice:

- Professor Joanne Aitken, Head of Research and Director of Registries, Viertel Cancer Research Centre, Cancer Council Queensland.
- Dr Paramita Dasgupta, Senior Research Officer, Viertel Cancer Research Centre, Cancer Council Queensland.

Executive Summary

Spatial information and spatial technologies can bring significant value to health agencies through improved decision support, resource management and allocation, and clinical outcomes. Disease mapping is used to explain and predict patterns of diseases outcomes across geographical areas, identify areas of increased risk, and assist in understanding the causes of diseases. As such, its use in informing policy recommendations is growing, and diseases of national importance, such as cancer, are being increasingly mapped across small regions.

Yet there are many potential methodological approaches for examining disease data over small areas, and understanding the benefits and disadvantages of any single approach when applied to a given situation is critical.

The aims of this report are threefold. First, to provide an accessible overview of the methods used in analysing spatial public health data, ranging from raw (unsmoothed) estimates through to complex Bayesian hierarchical models. Secondly, to outline the practical computational implementation of these methods. Finally, by comparing the advantages and disadvantages of these methods, to provide general guidelines and recommendations for their use.

Examples of the methods used in existing cancer atlases and other small-area analyses are also provided, as well as Bayesian approaches to incorporating multiple nested regions; considering the combined influence of related variables, such as remoteness and area-level socioeconomic disadvantage; small-area estimation from survey data; and extending the spatial analyses to also consider differences over time (spatio-temporal models).

Key issues to consider when using spatial data include data quality, including the reliability of location measures, and the degree of similarity between nearby areas (spatial correlation).

Although unsmoothed estimates such as crude or age-standardised rates may be useful for exploratory analyses, they are rarely appropriate for small-area analyses due to the small numbers involved, and should not be used when:

1. The addition of one event (disease case/death), or one more person at risk, results in a large difference (such as 25% or more) in at least one area's rates.
2. The number of events (rate numerator) is less than three for at least one area.
3. The population at risk per area is small (typically less than 500 people), and these numbers vary by an order of magnitude across the areas.

Smoothing methods may be either direct (e.g. locally-weighted, kernel smoothing) or model-based (e.g. Poisson kriging, Empirical Bayes or fully Bayesian). In general, direct smoothing methods are also more appropriate for exploratory analyses, but less useful when investigating contributing factors as they have more limited capacity for adjusting for covariates.

Model-based smoothing approaches have several advantages over the direct smoothing methods, and their use is recommended when assessing the impact of covariates is important, or the underlying pattern of risk needs to be understood.

There is no one model that represents the ultimate approach for disease mapping. The aims of the analysis, data quality, and expected results (such as disparate risks between nearby areas) can all influence the selection of the final model. Nonetheless, Bayesian hierarchical models are increasingly used in disease mapping, have been shown to perform well overall, and with the more recent application of approximation methods are able to generate results quickly.

For a cancer atlas, we generally recommend the use of Bayesian hierarchical models. The fully Bayesian approach enables the development of more complex, realistic models with reliable disease rates in low population areas, clearer summaries of spatial and temporal correlation, more precise and interpretable confidence intervals, and greater ability to account for and quantify measured sources of uncertainty than other possible approaches. The Bayesian approach also has excellent flexibility in handling changing inferential goals, such as obtaining smoothed risk maps as well as identifying motivating predictors of disease such as ethnicity or socioeconomic status.

Contents

List of Boxes

List of Tables

List of Figures

List of Abbreviations

APC	Age-Period-Cohort
ATA	Area-to-area
ATP	Area-to-point
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
BUGS	Bayesian inference Using Gibbs Sampling
BYM	Besag, York and Mollié
CAR	Conditional Autoregressive
CI	Confidence interval
DSR	Directly Standardised Rate
EB	Empirical Bayes
EBLUP	Empirical Best Linear Unbiased Predictor
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GLM	Generalised Linear Model
GLMM	Generalised Linear Mixed Model
G-NAF	Geocoded National Address File
GNU	GNU's Not Unix
GPS	Global Positioning System
GRASS	Geographic Resources Analysis Support System
HB	Hierarchical Bayes
INLA	Integrated Nested Laplace Approximation
JAGS	Just Another Gibbs Sampler
LISA	Local Indicators of Spatial Association
MCMC	Markov chain Monte Carlo
MEET	Maximised Excess Events Test
M-H	Metropolis-Hastings
MLE	Maximum Likelihood Estimation
MRF	Markov Random Field
MSE	Mean Squared Error
NHPA	National Health Performance Authority
REML	Restricted Maximum Likelihood
RER	Relative Excess Risk
SA1	Statistical Area 1
SA2	Statistical Area 2
SAS	Statistical Analysis System
SEBLUP	Spatial Empirical Best Linear Unbiased Predictor
SIR	Standardised Incidence Ratio

SLA	Statistical Local Area
SMR	Standardised Mortality Ratio
STARS	Space-Time Analysis of Regional Systems
STIS	Space-Time Information System
UK	United Kingdom
USA	United States of America

1. Introduction

"Knowing where things are, and why, is essential to rational decision making."

~ Jack Dangermond,
Environmental Systems Research Institute (ESRI)

1.1 Background

Place affects health (1). Spatial epidemiology aims to quantify and explain geographic variation in diseases and their relationship with environmental, demographic, behavioural, socioeconomic, genetic and infectious disease factors (2, 3). As such, disease mapping is an integral component of spatial epidemiology (3). Disease mapping can explain and predict patterns of diseases outcomes across geographical areas, identify areas of increased risk, and assist in understanding the causes of diseases (4).

Data used for spatial epidemiological analyses require information on the disease of interest, as well as a geographic location (Table 1.1) (5). This geographic location may be available at either the point or area level. Point level data refers to having the exact geocoded locations available, while area-level, or areal, data are only available for a region. Areal data are considered to have a constant estimate over the entire region, but commonly this is an aggregate measure such as the number of counts. Areas may consist of a regular lattice, or they may consist of irregular shapes.

Table 1.1 Examples of the types of data used in spatial epidemiological studies

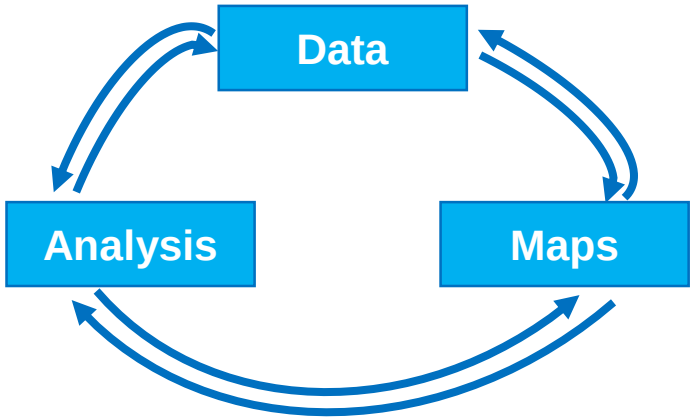
Data	Description
Health or disease	Vital statistics, notifiable diseases, patient registries, and health surveys from various international or government agencies. [Location is usually based on residential address]
Field epidemiology	Surveyed data on disease occurrences with location coordinates collected via GPS.
Spatially referenced base	Digital cartographic data available from various international or government agencies. [Often includes contours, rivers, and built environment features]
Remotely sensed	Land cover, elevation, soil type as reflected by satellite images.
Environmental and natural resources	Interpreted data on land use, water quality, air quality, climate, geology, etc.
Census or demographic	Sociodemographic and economic data.

Note: Modified version of Table 2.2 in (5), page 23. GPS=Global Positioning System.

The observed data represents one of three components involved in a spatial statistical analysis (Figure 1.1). Any one of these components may drive subsequent development of the other two, and often multiple circuits will occur before the process is complete (6). An existing map could inform data collection which then determines the appropriate statistical analysis.

Or a model could be developed, data collected and a map produced. This report concentrates on the analysis and modelling element. As areal data are commonly used by NHPA, this report focuses exclusively on methods for analysing area-level data. These data are also more readily available due to fewer data privacy restraints than point data. Another associated report "Communicating statistical outputs through maps" discusses the mapping component in detail. Cancer is a collection of diseases that are increasingly being mapped at the small-area level, and the third associated report "Grey Literature Review: Internet Published Cancer Maps" focuses on online, interactive cancer maps.

Figure 1.1 Components of a spatial statistical analysis



Note: Modified from Figure 12.4 in (6), page 179.

Many areal disease mapping methods initially arose from approaches to restoring images (that is, undoing image defects, such as motion blur, noise and/or camera misfocus), but there are important differences in public health data (7). First, areas are often irregular in shape and size (8), in contrast to the regular gridded lattice used in image restoration. There are commonly far fewer areas in public health data than pixels in an image, and these have varying rather than constant numbers of neighbours. Adjustment for important variables in the public health context is likely to influence results, whether that be population size or age, comorbidities or disease stage, while image restoration is primarily about the visual structure alone. Finally, any true boundaries in underlying risk are likely to be obscured by random noise in public health data, whereas images typically have clearly defined boundaries and multiple consecutive pixels with an identical colour (7).

1.2 Definition of small area/spatial

A small-area is defined in this report as an area that has a small population, and is not necessarily associated with their geographical size. What is a small population? This is determined by the disease of interest. A common disease, with a rate of 50%, could be quite well approximated with a population of 100. A less common disease, such as cancer, with a rate of 0.5%, would need a population of 10,000 to obtain a comparable number of cases.

Debate continues over the appropriate scale and definition of an area for a spatial analysis (9). There are no clear rules: selecting an appropriate spatial scale depends on the objective of the analysis, and (to a slightly lesser extent) data availability (9, 10). Public health estimates commonly require population data, which is often only available for administrative boundaries, and it is unusual for population estimates to be disaggregated to very fine resolution, especially by age and sex breakdowns.

Note that the terms small-area analysis and spatial analysis are used interchangeably in this report.

1.3 Aims

The aims of this report are threefold. First, to provide an accessible overview of methods used in analysing spatial public health data, ranging from raw (unsmoothed) estimates through to complex Bayesian hierarchical models. Secondly, to outline the practical computational implementation of these methods. Finally, by comparing the advantages and disadvantages of these methods, to provide general guidelines and recommendations for their use.

This report is not designed to be a comprehensive review, but instead seeks to broadly examine the key methods and models appropriate for areal data. Cancer is a complex group of diseases whose outcomes are increasingly the subject of small-area analyses. Cancer is used throughout this report to illustrate methods used in analysing spatial variation in disease outcomes.

1.4 Structure of report

This report is structured as follows. Technical details are provided in Boxes throughout the report for those interested. Additional details are also available in the Appendices, including a glossary of terms, a tutorial on Bayesian disease mapping, details on computational packages and software available, as well as further recommended reading.

In Chapter 2, methods used for analysing spatial data are presented, including unsmoothed estimates, direct smoothing methods as well as model-based smoothing approaches. No data are perfect, and this chapter outlines several statistical inference approaches to enable learning from these data. This encompasses examination of correlation (also known as cluster or hot-spot analysis), unsmoothed estimation, direct smoothing approaches such as locally-weighted averages or kernel smoothers, model-based smoothing, and computational approaches.

Chapter 3 then focuses on one group of diseases – cancer – and discusses the methods that have been applied to generate published small-area cancer estimates for screening, incidence/mortality, and survival data.

Chapter 4 presents a discussion of Bayesian approaches to spatial modelling, with four specific topics considered: (i) multiple nested geographies, (ii) combined remote and socioeconomic categories, (iii) survey data, and (iv) spatio-temporal data.

Finally, recommendations for when to use smoothing methods, and the types of smoothing or modelling methods likely to be the most appropriate, both generally and specifically for cancer atlases, are presented along with concluding remarks in Chapter 5.

2. Calculating small-area health estimates

"Everything is related to everything else, but
near things are more related than far things."

~ Waldo Tobler (First Law of Geography) (11)

Spatial data has specific characteristics that must be considered. Indeed, Fischer and Wang (12) titled their discussion around these issues as "the tyranny of spatial data".

First, data quality is critical (12). If data are geocoded, understanding the accuracy of the process is important. Although different types of geocoding exist, the most pertinent in spatial epidemiology is the geocoding of residential addresses (13). Potential errors in geocoding at the street number level include a low match rate (the completeness of geocoding to the street number level), positional error (geocoded point is not near the 'true' location), and low concordance (assignment to the correct geographic unit) (13). Geocoding is only reliable if the output is of high quality and repeatable (13). Note that repeatability can be influenced by variations in the reference data, the matching algorithms used by the geocoding software, as well as the skills and experience of geocoding personnel (13). Recently, Australia released the Geocoded National Address File (G-NAF), which is updated quarterly, although the level of uptake from health agencies is unclear.

2.1 Analysing spatial correlation

Beyond data quality, there are important inferential issues for spatial data. Perhaps the most important of these involve spatial correlation, clearly expressed in Tobler's first law of geography (11). This law posits that areas closer together are more similar than those further apart. Spatial correlation implies correlation among the same measure from different locations (14). Where spatial correlation is present, the assumptions that data are independent and identically distributed (the backbone of most traditional regression analyses) are violated (15, 16). Ignoring these spatial properties can result in false conclusions (17, 18). Any statistical techniques that assume data are independent are therefore not valid when spatial correlation is present.

A large range of options for testing for spatial correlation are available, and many GIS packages as well as standard statistical packages include the ability to conduct several tests. Popular options for area-level data include Moran's *I* (19), Geary's *C* (20), and the Localised Indicators of Spatial Association (LISA) (21) (Boxes 2.1 and 2.2). Newer options that have been shown to perform well (22) include Tango's Maximised Excess Events Test (MEET) (23) and the spatial scan statistic (24) (Boxes 2.1 and 2.2). These may assess spatial correlation throughout the entire study region (called global clustering, such as Moran's *I*), or may detect localised correlation (also called local clustering, or hot-spot analysis, such as the LISA).

Box 2.1 Global measures of spatial correlation

Moran's *I* (19) can be defined as:

$$\text{Moran's } I = \frac{I}{\sum_{i=1}^{I}(z_i - \bar{z})^2} \times \frac{\sum_{i=1}^{I}\sum_{j=1}^{I} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{I}\sum_{j=1}^{I} w_{ij}}$$

where z_i is the observed value at each area *i*=1,…I areas, $\bar{z}$ is the mean value, I is the number of areas and w_{ij} are the weights indicating which areas are adjacent/close together.

The input observed values z_i may be the original observations, or some standardisation to avoid scale dependence, such as the deviations from the mean (21). Standardised values are generally preferable.

The values of Moran's *I* generally span from -1 (dispersed) to +1 (clustering). Values around 0 indicates no spatial correlation.

Geary's *C* (20) is similar to Moran's *I*, as can be seen from the following definition using the same notation:

$$\text{Geary's } C = \frac{I-1}{\sum_{i=1}^{I}(z_i - \bar{z})^2} \times \frac{\sum_{i=1}^{I}\sum_{j=1}^{I} w_{ij}(z_i - z_j)^2}{2\sum_{i=1}^{I}\sum_{j=1}^{I} w_{ij}}$$

The values of Geary's *C* typically range between 0 and 2. A value of 1 means no spatial correlation, <1 indicates positive spatial correlation, and >1 indicates negative spatial correlation (25). Geary's *C* is considered more sensitive to local spatial correlation than Moran's *I* (5).

Tango's MEET (23) is based on the calculation of the excess events test (EET) (26), which is a weighted sum of the excess number of events (observed minus expected), with higher weights when areas are proximal, as follows:

$$EET = \sum_i \sum_j e^{-\frac{4d_{ij}^2}{\lambda^2}} \left(O_i - \frac{p_i O_{TOT}}{P_{TOT}}\right)\left(O_j - \frac{p_j O_{TOT}}{P_{TOT}}\right)$$

where o_i and p_i are the observed count and population, respectively, in each area, O_{TOT} and P_{TOT} are the overall count and population, and d_{ij} represents the distance between area *i* and area *j*. The choice of λ can influence the outcome, with large values of λ increasing sensitivity to detecting large geographical clusters, while small λ increases the sensitivity to small clusters.

Tango's MEET overcomes this by considering multiple versions of λ up to a pre-determined value. This enables clustering to be detected irrespective of geographical scale. A small p-value indicates clustering is present.

Box 2.2 Local measures of spatial correlation

One form of the LISA (Local Indicators of Spatial Association) can be considered the local equivalent of Moran's *I*, and expressed as:

$$I_i = (z_i - \bar{z}) \times \sum_{j \in J_i}^{n} w_{ij}(z_i - \bar{z})^2$$

It is also possible to have a LISA version of other common global indicators, including Geary's *C*. The LISA for each area indicates the extent of significant spatial clustering around that area, and the sum of LISAs for all areas is in proportion to the corresponding global statistic (27).

Disadvantages of the LISA include multiple testing issues as a separate statistical test is conducted for each region (14). These regions are also small, and rates are unstable, risking spurious significance. Although a Bonferroni adjustment is often used to account for multiple tests, the correlation between neighbouring LISAs (as they share some of the same observations), would cause this adjustment to be very conservative (14).

The spatial scan statistic (24) considers a large number of overlapping circles of assorted sizes and locations. Specialised software has been developed and is freely available to implement this method (SaTScan) (28). This method can be used for a range of data (count, ordinal, binomial, even multinomial and survival) and can also be adjusted for covariates (29, 30). This method of detecting clusters is based on maximising the likelihood ratio.

The spatial scan statistic is proportional to

$$\max\left(\frac{O_{IN}}{E_{IN}}\right)^{O_{IN}}\left(\frac{O_{OUT}}{E_{OUT}}\right)^{O_{OUT}}$$

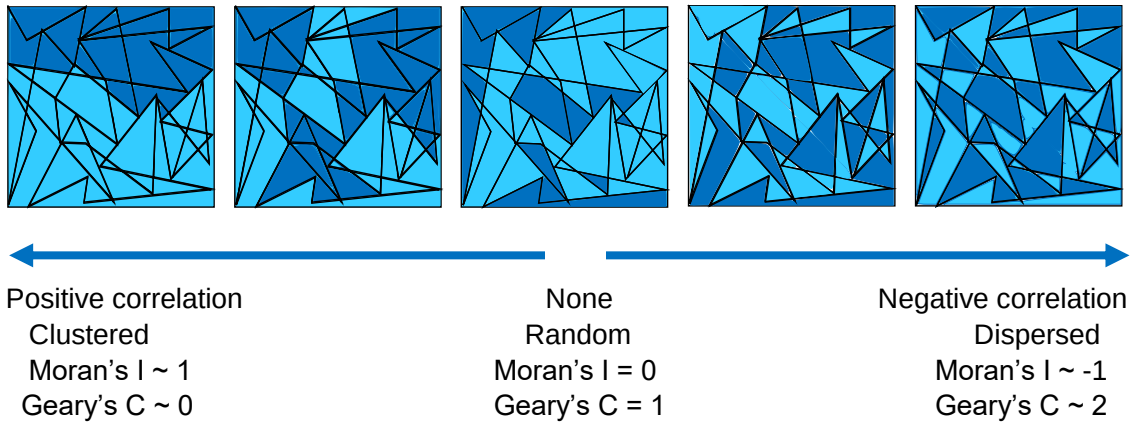
where O_{IN} and O_{OUT} are the observed counts inside and outside the circle, respectively, and E_{IN} and E_{OUT} are the respective expected counts inside and outside the circle (14).

Although the spatial scan statistic has been reported as performing well in comparison to other methods (22), others have raised concerns about the large size of the clusters detected and difficulties in detecting cluster shapes other than circles (31).

While the aim of global clustering methods is to determine if there is clustering throughout the region, the precise location of any clustering is not important (22). Instead, results may provide a general indication of overall patterns, such as whether any correlation is positive (similar values are clustered together), or negative (dissimilar values are together) (Figure 2.1).

In contrast, local clustering methods seek to detect the location of statistically significant spatial clusters and outliers in disease risk (32). This is achieved by comparing the value at one location with values at nearby locations, up to a specified threshold distance (32).

Figure 2.1 Spatial correlation patterns



Notes: Modified from Figure 3.6 in Lai *et al.* (33).

2.2 Defining the neighbourhood

One way of accounting for spatial correlation in the data is by defining a neighbourhood as part of the model. A neighbourhood is composed of surrounding areas that are considered to exert influence on the observations of an area (12).

The definition of an area-based neighbour may be based on spatial adjacency, such as those sharing a boundary, or instead may be based on the distance between the centroids (14). Here, if the distance between two area centroids is below a certain threshold distance, they are considered to be neighbours. Ways to measure the distance between the centroids include straight-line distances (the shortest distance between the two coordinates assuming they are on a flat surface), great circle distances (determining the length of the arc of the earth's surface between the two points), or using a Geographic Information System (GIS) to calculate travel distances or times (12).

Note that when there is great variation in the size of the areas, determining a suitable threshold distance value is difficult. Even just allowing for the largest areas to have at least one neighbour may result in far too many neighbours for smaller areas (34). Options for overcoming this problem include assigning a fixed number of neighbours for each area (k-nearest neighbours) (Box 2.3).

Box 2.3 Weighting mechanism examples (12, 35, 36)

Distance-based weights

$$w_{ij} = \begin{cases} 1 \text{ if } d_{ij} < \delta \\ 0 \text{ otherwise} \end{cases}$$

where d_{ij} is the distance between the centroids of regions i and j, and δ is a given critical value.

Distance-based simultaneously weighted by population

$$w_{ij} = \begin{cases} e_i e_j / d_{ij} \text{ if } d_{ij} < \delta \\ 0 \text{ otherwise} \end{cases}$$

where e_i is the standardised population for an area and e_j is the standardised population for its neighbour in area j. These can be standardised against the mean and standard deviation.

Distance-based simultaneously weighted by distance

E.g. Based on the inverse distance function

$$w_{ij} = \begin{cases} {d_{ij}}^{-\gamma} \text{ if } d_{ij} < \delta \\ 0 \text{ otherwise} \end{cases}$$

where the parameter γ specifies the declining rate of the weight, and can be set *a priori* or estimated. Common choices for γ are the values of one or two.

k-nearest neighbours (Note that w_{ij} might not be equal to w_{ji})

$$w_{ij} = \begin{cases} 1 \text{ if centroid of } j \text{ is one of the } k \text{ nearest to centroid } i \\ 0 \text{ otherwise} \end{cases}$$

Adjacency-based neighbours

$$w_{ij} = \begin{cases} 1 \text{ if regions } i \text{ and } j \text{ share a boundary} \\ 0 \text{ otherwise} \end{cases}$$

Adjacency-based weighted by the fraction of a shared border

$$w_{ij} = \begin{cases} \dfrac{l_{ij}}{l_i} \text{ if regions } i \text{ and } j \text{ share a boundary} \\ 0 \text{ otherwise} \end{cases}$$

where l_{ij} is the length of shared common boundary between regions i and j, and l_i is the perimeter of region i.

Part of assigning neighbours involves applying a measure of weighting to indicate the extent to which the information from an area's neighbours impacts on the observed estimate for that area. A weight of zero indicates no relationship, while a weight above zero indicates that areas i and j are considered to be neighbours, and some influence is expected. The weights (w_{ij}) are placed into a matrix with dimensions of the number of areas. When calculating the similarity with nearby regions, the diagonal (w_{ii}) is generally set to 0 as an area is not considered to be a neighbour of itself. (This differs from situations when the proportion is averaged over areas (see Section 2.4). Here, data in the area should be included, so $w_{ii} = 1$ (37).) The greater the weight, the more resistant they are to their neighbour's influence (35).

The derivation of this weight can be based on a range of options (Box 2.3). The weighting can also be modified according to distance (the weight decreases for more distant neighbours), or based on the population size of neighbours (larger populations receive greater weight). Often if little is known about the assumed spatial pattern, a binary weighting is assigned with 1 for neighbours and 0 otherwise (35) This is often then standardised by dividing by the number of neighbours so that the rows sum to 1. When areas are irregularly shaped, this standardised weight matrix is generally not symmetric (14). As each neighbour receives the same proportional weight, interpretation is simple as it becomes a weighted average of neighbouring values (12).

The same spatial arrangement can lead to many different neighbourhood definitions. Key considerations when selecting an approach to assigning neighbours include whether areas are regular or irregular shapes and sizes and how localised spatial dependencies are. Given the influence a neighbourhood structure can exert on a spatial analysis, checking the appropriateness of this choice and its impact on the conclusions is important (12, 38).

2.3 Unsmoothed estimates

The simplest of all techniques for generating values for small-areas are to calculate and map unsmoothed, or 'raw' estimates.

Counts may be displayed as dots on a map, randomly allocated within the area supplied. Although mapping the counts can be useful and appropriate if the aim is to inform service provision, often understanding the disease risk is of interest, and this requires some form of adjustment for population size and structure. Commonly this is achieved by using rates as a reflection of risk (14).

There are several types of rates commonly calculated. Crude rates adjust only for population size, but not structure (Box 2.4). Proportions and percentages are other commonly used measures for a crude rate. The assumption is that the risk remains constant over all age and sex categories (37), but most diseases disproportionately affect specific age groups (14). When comparing crude rates between areas, observed differences for a disease that varies with age may reflect differences in the age distribution alone.

Box 2.4 Crude rates

A crude rate for the i^{th} area (i=1,…I) can be calculated as:

$$\mathrm{CR}_i = \frac{O_i}{P_i}$$

where O_i are the observed counts in area i and P_i are the number of people residing in area i. Commonly small crude rates are multiplied by a constant and expressed as per constant (e.g. per 1,000).

An alternative approach, which also adjusts for population structure, is to calculate age-standardised rates by considering the counts of disease and the expected counts using some standard population (39). These may be either directly or indirectly standardised rates. It is also possible to further adjust for specific area-level variables, such as socioeconomic status.

Box 2.5 Directly age-standardised rates

Directly standardised rates represent the rate these areas would have if their age distribution matched that of the standard population (14).

Excluding sex for simplicity, and assuming the m age groups (m=1,…M, e.g. M=18 five-year age-groups (0-4, 5-9,..., 85+)), the directly age-standardised rate (DSR) for the i^{th} area (i=1,…I) can be calculated as (40):

$$\mathrm{DSR}_i = \sum_{m=1}^{M} \pi_m \frac{O_{im}}{P_{im}}$$

where π_m is the proportion of people in age group m from the standard population, O_{im} are the observed disease counts (number of cases for incidence, number of deaths for mortality) in area i and age group m, and P_{im} are the number of residents in area i and age group m.

Direct standardisation focuses on estimating the number of cases/deaths that would be observed in the standard population if the observed age-specific rates of disease applied (41). This is achieved by weighting the age and sex-specific rates for each small area so they correspond to the age distribution of a single standard population (Box 2.5). This method enables comparison between areas, but does require age (and sex) specific counts and populations for each area, which may not be available, or may be very unstable (5, 14). Also, the standard population is arbitrarily defined, and estimates may differ substantially between different standard population definitions (42).

In contrast, indirect standardisation focuses on estimating the number of cases/deaths that would be expected if the study population contracted/died from the disease at the same rate as the standard population (41). Indirectly standardised rates thus multiply the stratified population of each small area by the known stratified disease rates of some reference population (Box 2.6). This process produces a standardised morbidity (if using incidence data) or mortality (if using death data) ratio (SMR). This estimator is very popular, and only requires the population at risk in each age-sex group and area, as well as the total counts in each area. It also has a lower standard error (33), and is useful for small areas with unstable rates (5). However, in contrast to directly standardised rates, weights differ for each area considered, and bias can potentially result if the age-distributions differ between the areas being compared (43). Therefore indirectly standardised rates tend to not be directly comparable between different geographical regions (14).

Box 2.6 The Standardised Morbidity/Mortality Ratio (SMR)

Indirect standardisation reflects whether the number of cases in an area are higher or lower than expected, given the population size and structure for that area.

The definition of an SMR for the i^{th} area (i=1,…I) is:

$$\text{SMR}_i = \frac{O_i}{E_i}$$

where O_i are the observed counts in area i and E_i are the expected counts in area i, applying the overall (reference) disease rates to the age-specific population structure P of area i and summing across all m age groups (with a maximum of M age groups), and excluding sex, as follows:

$$E_i = \sum_{m=1}^{M} \frac{O_{\text{ref}_m}}{P_{\text{ref}_m}} \times P_{im}$$

Where O_{ref_m} represents the observed disease counts (either cases or deaths) in the reference population in age group m, and P_{ref_m} is the reference population in age group m.

The corresponding standard error for the i^{th} area is: $\frac{\sqrt{O_i}}{E_i}$

These simple estimates are commonly displayed using a choropleth map. The colouring/shading of areas in a choropleth map uses a discrete scale based on the values of the estimate. Any kind of choropleth map implicitly smooths the display of results, and the fewer the number of categories, the greater this visual smoothing (44). However, the assumption of spatial independence inherent in choropleth mapping could be misleading, and caution in using with ‘raw’ (unsmoothed) estimates is advised (45).

Further, as area size diminishes, the use and interpretation of these unsmoothed estimates become increasingly difficult due to the greater variance (Box 2.7) associated with them (46). These estimates can also be prone to substantial fluctuation from year to year without there necessarily being a change in the underlying rate for a specific area. Other concerns include that when an area has no counts, the estimate is zero, regardless of denominator size (47). As such, these 'unsmoothed' methods are perhaps most useful for preliminary investigation to guide further analyses, rather than being an end in themselves.

Box 2.7 Variance and covariance

Variance can be defined as:

$$\text{var}(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

In other words, variance can be visualised as lines describing how far away each observation is from the mean on a scatter plot.

Covariance is a measure of how much two variables change together, so can be defined as:

$$\text{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})$$

Covariance can be visualised as rectangles describing how far away pairs of observations are from the mean on a scatter plot.

2.4 Direct smoothing

The objective of disease mapping is to produce an accurate estimate of the underlying rate in different areas, with noise removed (48). This 'noise' is simply additional variation in the data, and a major source is often unmeasured variables that affect the outcome (49). Smoothing methods aim to remove or minimise this noise by incorporating neighbouring information in a flexible way (40). When information from geographical neighbours are included, the information for the region is artificially inflated. This provides greater stability for the specific area as well as between areas.

2.4.1 Locally-weighted average/median

A straightforward smoothing method is averaging the values associated with neighbouring areas. First, the neighbours and appropriate weights must be selected (see Section 2.2).

Different forms of smoothing result from different weighting choices (14). If binary weights of zero and one are used, then any difference in the precision of the rates is ignored. Weighting neighbours by their population is one way to incorporate the precision (14).

Although the measures weighted are usually based on the mean of either the crude rate or the SMR, the resulting sensitivity to extreme outliers has led to extensions based on the median instead (37). To allow for differences in precision, the median crude rate could be weighted by the population, while the median SMR could be weighted by the inverse standard error (37). Here the original values are sorted then matched with both the weight (e.g. population size) and a cumulative sum of the weights (50). Whichever weight has a cumulative sum of more than half the total cumulative sum is used (50).

Despite the simplicity and range of options available for these locally-weighted methods, there are some key disadvantages. Firstly, the number and regularity of locations may influence the efficiency of the algorithm (5). There is also a risk that because neighbouring areas are "forced" to have some numerical association with each other, this method may induce spatial structure, even when the data are completely random (51).

2.4.2 Kernel smoothers

A general, non-parametric approach to smoothing rates while differentially weighting neighbours is to use a two-dimensional kernel function (37). A kernel function decreases with increasing distance (distance decay function), and the rate and range of decay is modified by the functional form of the kernel, as well as the threshold beyond which the kernel is set to zero (bandwidth) (37).

Although more commonly applied in geostatistical (point) data, kernel smoothers can be applied to areal data at a specified moving window size (25). A range of kernel smoothers exist, but one of the most commonly used is the Nadaraya-Watson kernel estimator (Box 2.8) (52, 53). Although this is a weighted average of the neighbourhood observations, the weights are controlled by the kernel function, so all observations are not treated equally (49).

Depending on the application, disadvantages of kernel smoothing can include estimates resulting in different totals across areas than in the original data (45). Boundary effects cause problems for kernel smoothers (49), while the use of cross-validation has been shown to induce over-smoothing (47). A comparison of several smoothing methods concluded that kernel smoothers performed poorly when spatial correlation was present, and suggested only using them for exploratory data analysis (47).

Box 2.8 The Nadaraya-Watson kernel estimator (47, 49, 52, 53)

This kernel smoother is simply a weighted average, so if applied to an indirectly standardised ratio such as an SMR:

$$\hat{\theta}_i = \sum_{j \neq i} w_{ij} \mathrm{SMR}_j$$

The weights are functions of neighbouring values:

$$w_{ij} = \frac{K((\mathrm{SMR}_i - \mathrm{SMR}_j)/h)}{\sum_j K((\mathrm{SMR}_i - \mathrm{SMR}_j)/h)}$$

The function $K(\cdot)$ is the kernel function: a smooth probability density function symmetric around 0 and nondecreasing on $\left[\frac{-1}{2}, 0\right]$ and h is the bandwidth and selected by minimising some goodness-of-fit measure, such as cross-validation (where data are partitioned into testing and validation sub-samples to see how it generalises to an independent dataset).

2.5 Model-based smoothing

Two standard paradigms for statistical models are:

1. The sampling model, where population characteristics are estimated from a sample or subset of the population, and
2. The measurement error model, where the focus is on estimating an underlying pattern, but the data are measured with error. This model is also applicable when complete data are observed (54).

Although these differ, in practice both approaches can be combined (54). Most models discussed in this section are measurement error models. Section 4.3 focuses on sampling models in the context of survey data.

2.5.1 Foundational approaches

In public health applications, the most widely used regression models are the generalised linear models (GLMs) and the generalised linear mixed models (GLMMs) (Box 2.9), particularly using the Poisson distribution for count data (14).

The Poisson model is appropriate when there are low disease counts and comparatively large populations in each small area (55). The counts are assumed to follow a Poisson distribution

which defines the mean and variance. The mean is dependent on two components: 1) the expected count, generally obtained through indirect standardisation, and 2) the excess risk, which is the SMR, also often referred to as the relative risk in this context (Box 2.10).

Box 2.9 GLMs and GLMMs

A generalised linear model (GLM) involves:

- A data vector $O = (O_1, O_2, \ldots, O_I)$
- Predictors X and coefficients β, to give the linear predictor $X\beta$
- A link function g, that links the linear predictor to a nonlinear transformation of the expected response (e.g. logarithm)
- An assumed data distribution (e.g. Poisson)
- Potentially other parameters involved in the predictors, link function and data distribution, such as variances, overdispersions and cutpoints (54).

A generalised linear mixed model (GLMM) involves the same components as the GLM, with the addition of random effects. The 'mixed' in the name thus refers to the model containing both fixed and random effect terms. Defining whether a term is fixed or random is seldom straightforward, but perhaps the cleanest approach is that fixed effects are constant if they are identical for all groups, while random effects are allowed to vary between groups (56).

The binomial distribution is sometimes preferred for small areas due to the Poisson distribution having some probability of obtaining more counts than persons at risk in each area (14). However, this is extremely unlikely for rare diseases, where the practical difference between Poisson and binomial distributions is negligible (14). The Poisson distribution also constrains the mean to be equal to the variance, but as small areas often have a variance greater than the mean (termed over-dispersion), some prefer to use the negative binomial distribution instead. However, the use of a fully Bayesian formulation where a prior distribution is placed on the SMR/relative risk can accommodate some over-dispersion (55).

Box 2.10 The Poisson model

The disease counts O in each of i areas (i=1,…I) is assumed to have a mean dependent on the expected count E_i and the SMR/relative risk θ_i as follows:

$$O_i \sim \text{Poisson}(E_i\theta_i)$$

The main interest is normally in modelling $\boldsymbol{\theta_i}$ in the ith area. A logarithmic link is often assumed (which ensures estimates are positive) to a linear predictor model (Box 2.9), as follows:

$$\log \theta_i = \eta_i = X_i\beta$$

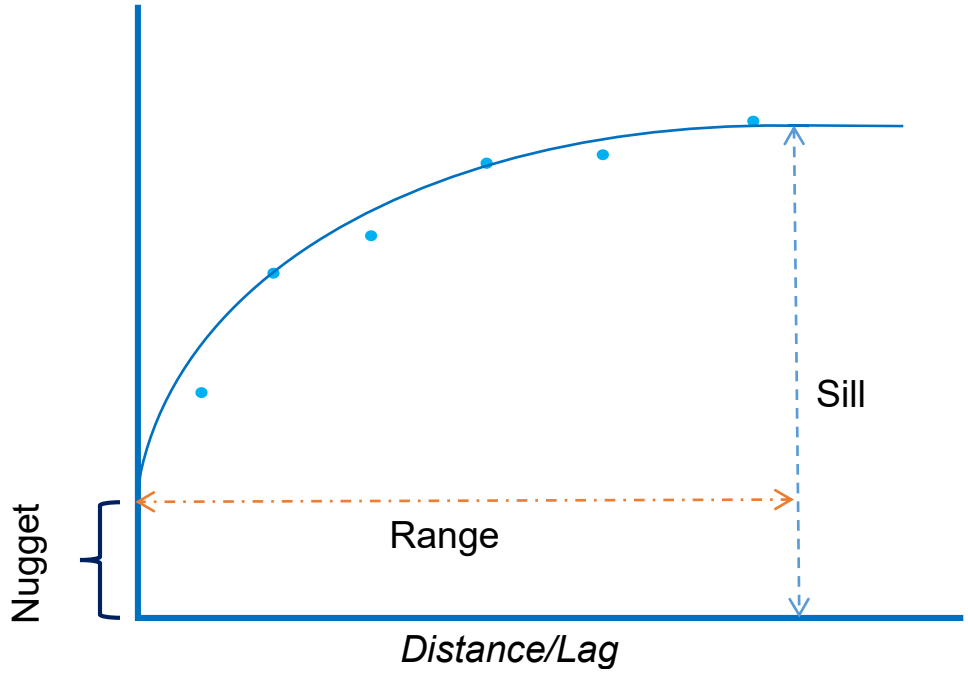
2.5.2 Poisson Kriging

Kriging was originally developed to estimate attribute values from a limited set of sampled data over a continuous spatial region (57). The weights used in kriging incorporate distance measures, as well as spatial correlation (58). Although this increases the complexity, it also increases the flexibility of the method and the reliability of predictions (59). Areal disease mapping often doesn't require the interpolative ability of kriging, but a specific variant of kriging known as Poisson kriging has been developed and is becoming increasingly used for disease maps (57, 60).

Box 2.11 The Semivariogram (5)

The semivariogram displays the semivariance, which is a measure of the level of spatial correlation, against distance or lag (Figure 2.2).

Figure 2.2 Semivariogram components



Note: Modified from the Figure in Explanation Box 6.1 in (5), page 107.

The distance at which the model first levels out is called the range. Areas separated by distances greater than this are not considered to influence each other, whereas areas separated by distances within the range are spatially correlated.

There are several variants of Poisson kriging, but as we are exploring methods appropriate for areal data, our focus is on area-to-area (ATA) and area-to-point (ATP) Poisson kriging. Goovaerts (61) extended the work of Kyriakidis (45) to introduce these. While ATA kriging can be used when both the observations and the desired predictions are over areas, ATP kriging predicts point values from areal data (45). An alternative is to simply collapse the data on the centroids, but this is not considered appropriate when the shape and/or size of areas are irregular (62).

In both ATA and ATP Poisson kriging, the risk over an area is estimated as a weighted linear combination of the rate observed for that area and neighbouring areas (61). Areas with smaller populations receive less weight. These weights are solved from a system of linear equations, but requires either the point-support covariance of the risk, or the equivalent point-support semivariogram (Box 2.11), to be modelled. This point-support model is where the spatial correlation is included.

Developing a semivariogram structure that accounts for irregularly shaped areas and varying distributions is relatively complex, and an iterative procedure is recommended (62). Approaches range from solving a set of integral equations (63), to iteratively re-weighted generalised least squares methods (64), to simulated grids within the regions of interest (62).

While kriging is a useful filter of noise, and produces uncertainty estimates, it is not designed to estimate the risk within each area (51). Nonetheless, a comparison against the popular Bayesian Besag, York & Mollié (BYM) model (see Box 2.14) found that Poisson kriging gave better discrimination between areas with high and low risks, and more precise and accurate probability intervals (65). Poisson kriging was also found to out-perform simple population-weighted averages and empirical Bayesian smoothers (see Section 2.5.3) (51).

2.5.3 Empirical Bayes

Bayesian methods differ from other statistical approaches as they consider both the data and the parameters to be random variables (37). Inference under a Bayesian approach requires specific items (Box 2.12), with the most controversial element being the selection of the prior distribution.

Some practical suggestions when selecting a prior distribution include graphing it, to ensure the shape is plausible, as well as potentially calculating the effective sample size of the prior (Box 2.13) (66). Choosing particular families of prior distributions (called ‘conjugate priors’) may assist in solving the posterior distribution without resorting to complicated integrations. For further details on the Bayesian approach, refer to Appendix B.

Empirical Bayes (EB) methods use the data to estimate the unknown information on the prior and conditional distributions (67, 68). Spatial disease patterns were initially explored using EB methods by Clayton and Kaldor (69), Cressie and Read (70) and Cressie (46).

Box 2.12 Bayesian inference

To make inference about the unknown parameter θ from the data O requires the following (71):

1. A model $f(O|\theta)$; the *likelihood*
2. A distribution for θ. This is called a *prior distribution* as it is determined before seeing the data.

The combination of the likelihood and prior distribution(s) by Bayes' rule gives the *posterior distribution*:

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

from which subsequent model-based inferences are drawn.

Empirical Bayes predictors have attractive statistical properties, provided the model is appropriate, including being the best linear unbiased predictor (BLUP) (46). The associated uncertainty for each estimate is also available. Ironically, the greatest criticism of EB models are against the uncertainty measures. Since they do not account for the additional variability in estimating the parameter values, the resulting variance tends to be too small (68, 72-74). Although methods are available to adjust the variance estimates (68, 75), fully Bayesian methods have many of the same desirable properties as an EB estimate, while adequately representing the distribution of underlying rates (72).

Box 2.13 Calculating the effective sample size of the conjugate prior distribution (66)

The conjugate distribution to the Poisson is the gamma. Say the gamma distribution is expressed as $\text{gamma}(r, v)$ where $r = \frac{m^2}{s^2}$ and $v = \frac{m}{s^2}$ where m is the prior mean and s is the prior standard deviation, and that $O_1, \ldots O_n$ is a random sample of observed counts from a Poisson distribution, $\text{Poisson}(\mu)$, so that the expected value of O has mean μ and variance $\frac{\mu}{n}$.

To check the amount of prior information entering the model, the equivalent sample size can be calculated by solving the following for n_{ESS}:

$$\frac{\mu}{n_{ESS}} = \frac{r}{v^2}$$

If the mean is set equal to the prior mean, i.e. $\mu = \frac{r}{v}$ then under the $\text{gamma}(r, v)$ prior $n_{ESS} = v$. This value represents the size of a random sample from the $\text{Poisson}(\mu)$ that is equivalent to your prior knowledge of μ. If n_{ESS} seems too large, increase the prior standard deviation s and recalculate.

2.5.4 Fully Bayesian

In contrast to Empirical Bayes where some unknown parameters are assigned a point estimate, a fully Bayesian model results when prior distributions are placed on all unknown parameters. Beyond simple models with few parameters, this results in a hierarchical model structure, where each layer defines a relationship between the observed data and/or unknown parameters. This general class of model is commonly referred to as a Bayesian hierarchical model.

There are several advantages to using Bayesian hierarchical models, including the ability to structure very complicated models from a succession of relatively simple components (76), good performance and ease of implementation (72, 77-79). They are also a natural approach to model spatially misaligned data, as occurs when the exposure and response are measured at different levels of aggregation (80).

The fully Bayesian approach enables complex, realistic models to be developed with reliable disease rates in low population areas, clear summaries of spatial and temporal correlation, precise and easily interpretable confidence intervals, and more comprehensive accounting of sources of uncertainty (77). The Bayesian approach also has excellent flexibility in handling changing inferential goals, such as obtaining smoothed risk maps as well as identifying motivating predictors of disease such as ethnicity or socioeconomic status (77).

2.5.4.1 Incidence/mortality data

The most popular Bayesian hierarchical model for disease mapping is the BYM model (Box 2.14) (81), which further developed the model of Clayton and Kaldor (69). The key feature of this model are the two random effects: one which is spatially structured, so smooths locally (towards the values of nearby areas), and one which is unstructured, so smooths globally, towards the overall average (82, 83).

Box 2.14 The BYM model

The BYM model can be expressed as follows:

$$O_i \sim \text{Poisson}(E_i \theta_i)$$

$$\log(\theta_i) = \alpha + u_i + v_i$$

where O_i is the number of disease events in the *i*th region, E_i is the expected number of cases, θ_i is the standardised incidence ratio and α is the intercept. The model incorporates extra-Poisson variability by including two spatial random effects: v_i allows for inter-area heterogeneity, while u_i is structured and represents the spatial component (84).

Commonly, the prior distributions placed on the structured spatial component is a Conditional Autoregressive (CAR) distribution. The CAR prior is characterised by an adjacency matrix, which defines the geographical neighbours of each area. The intrinsic Gaussian CAR prior (81) assumes this matrix is binary, where immediately adjacent neighbours are given the value 1, and all other pairs of areas are given the value 0. In this case, the estimated random effects for each area are smoothed towards the average of the random effects for the neighbours. The resulting simple functions of the neighbouring values and number of neighbours, n_i, equates to the following conditional distribution:

$u_i|\mathbf{u_{-i}} \sim \text{Normal}\left(\bar{\mu}_i, \frac{\omega_u^2}{n_i}\right)$ where $\bar{\mu}_i$ is the average of the neighboring regions of area *i* and the variance term ω_u^2 represents a conditional variance (so is deliberately not portrayed as σ_u^2)

$\mathbf{u_{-i}} = (u_1, \ldots u_{i-1}, u_{i+1}, \ldots, u_I)$

This can be expressed jointly as $\mathbf{u} \sim \text{Normal}_I\left(\mathbf{0}, \frac{1}{\omega_u^2}(\mathbf{D} - \mathbf{W})^{-1}\right)$

where $\mathbf{D}$ is a diagonal matrix with $D_{ii} = n_i$, $\mathbf{W}$ is a spatial weight matrix of dimension $N \times N$ with diagonal elements $w_{ii} = 0$ and off-diagonal elements $w_{ij} = 1$ if regions i and j share a boundary, and 0 otherwise.

The intrinsic CAR distribution is restricted to specifying prior distributions, as the pairwise difference joint specification results in an improper joint distribution. The computational ease is a key advantage of the intrinsic CAR formulation (85, 86).

The inter-area heterogeneity effect commonly has a vague normal prior distribution:

$v_i \sim \text{Normal}(0, \sigma_v^2)$ or expressed jointly as $\mathbf{v} \sim \text{Normal}_I(0, \sigma_v^2\mathbf{I})$ where $\mathbf{I}$ is the identity matrix (diagonals are set to 1, 0 otherwise).

Under the Bayesian hierarchical formulation, the variance components for both the CAR and the normal distributions will also receive prior distributions (termed 'hyperpriors').

Common choices include a vague gamma distribution on the inverse variance, or a uniform distribution on the standard deviation, e.g.

$\sigma_u \sim \text{Uniform}(0,10)$
$\sigma_v \sim \text{Uniform}(0,10)$

The BYM model has been shown to produce robust estimates, but results may be sensitive to the choice of priors, particularly the choice of hyperpriors (47, 65, 87). The intrinsic Gaussian CAR prior results in a spatially smooth risk surface, which has the advantage of using information from multiple areas to estimate the random effects, but is not ideal if the aim is to identify clusters of high-risk areas (87). This is because a cluster of high risk areas may have low-risk neighbours, and therefore the estimated risk for these areas becomes less distinct when geographical smoothing is used (88). Identifiability is also a concern, as the one residual component is split into two independent, additive components (89). Very sparse data,

as would be seen for smaller geographical areas such as Statistical Area 2 (SA2) or SA1, may cause difficulties when applying these models, particularly if there are also none or very few neighbours, such as on a coastline or an island (90). Finally, the spatial correlation may inflate the variance of the β components when covariates are included (91, 92).

Alternative approaches have suggested modifications to try to overcome the issues with identifiability. The Leroux CAR prior (93) (Box 2.15) outperformed the BYM model in a comparison of methods (94). MacNab's alternative convolution prior (95) (Box 2.15) also overcomes identifiability issues, at the cost of greater model complexity (89).

Box 2.15 Alternative priors to the BYM

The Leroux prior (93)

Instead of the $v_i + u_i$ components in the BYM model, the Leroux prior has the one term modelled by the multivariate normal prior, as follows:

$$\mathbf{b} \sim \text{Normal}_I(\mathbf{0}, \Sigma_b)$$

$$\frac{1}{\Sigma_b} = \frac{1}{\sigma_b^2}(\lambda(\mathbf{D}_u - \mathbf{W}) + (1-\lambda)\mathbf{I}_I)$$

Where $\boldsymbol{\lambda}$ is between 0 and 1, and referred to as the spatial correlation parameter, as it reflects the proportion of excess Poisson variation explained by spatial dependencies, D is a diagonal matrix with $\boldsymbol{D_{ii} = n_i}$, W is the spatial weight matrix and I is the identity matrix (see Box 2.14 for further details)

MacNab's alternative convolution prior (95)

This prior facilitates identifiability of the spatial and unstructured random effects. Using the same notation as above, this can be expressed as:

$$\mathbf{b} \sim \text{Normal}_I(\mathbf{0}, \Sigma_b)$$

$$\Sigma_b = \frac{\lambda\sigma_b^2}{(\mathbf{D} - \mathbf{W})} + (1-\lambda)\sigma_b^2\mathbf{I}_I$$

Allowing discontinuous risks between areas

Other approaches have focused on allowing discrete changes between areas. Some of the current approaches deal with this issue by defining the adjacency matrix such that the elements are random quantities to be estimated (96). In this way, boundaries between clusters of areas can be identified when the adjacency matrix elements for neighbouring pairs are estimated to be near zero. There are two main problems with this methodology. First, using

random quantities in an adjacency matrix of size $I \times I$ (where I represents the number of areas) means an additional I^2 parameters need to be estimated, which is usually more parameters than can be estimated reliably. And secondly, there is no constraint on boundary segments to enclose an area or cluster of areas (88).

Box 2.16 Lawson and Clark's model

Here, an intrinsic CAR prior and a difference prior act as a 'mixture of priors' (97).

Instead of the $v_i + u_i$ components in the BYM model, this model has the following terms:

$$v_i + p_i u_i + (1 - p_i) w_i$$

where $\boldsymbol{p_i}$ is interpreted as the strength of support for spatial smoothing in the *i*th area, $\boldsymbol{u_i}$ and $\boldsymbol{v_i}$ represent the spatially structured and inter-area heterogeneity terms, respectively, as in Box 2.14, and w_i represents the jump component. Note that the above equation defaults to the BYM model if $p_i = 1$, but as p_i approaches 0, the jump component is preferred.

The prior distributions on $\boldsymbol{u_i}$ and $\boldsymbol{v_i}$ are as previously defined (Box 2.14), while the prior on w_i is intended to measure spatial rates of change in risk. Although a range of options is possible, the suggested approach was:

$$\mathbf{w} \propto \frac{1}{\sqrt{\lambda}} \exp\left(-\frac{1}{\lambda} \sum_{i \sim j} |w_i - w_j| \right)$$

where λ acts as a constraining term, and j represents areas that are neighbours.

As there are only two mixing probabilities, a standard Beta distribution is used,

$$p_i \sim \text{beta}(\alpha, \alpha)$$

For higher dimensions a Dirichlet prior could be used.

New methodology to address these problems was proposed by Anderson *et al.* (98). This approach consists of two stages. In the first stage, a set of candidate cluster configurations are identified by using what is called a 'modified hierarchical agglomerative clustering algorithm'. Initially, each area is considered to be a cluster, and clusters are combined together sequentially based on how similar they are according to some metric applied to the spatial data until all areas are combined into one cluster, resulting in a total of I cluster configurations. The term 'modified' is used because the usual clustering algorithm does not necessarily produce spatially contiguous clusters, but this is enforced by only allowing clusters to be combined if the clusters share a common border. The spatial data used in this stage should not be the study data used in the second stage of the model, but should be a similar dataset, such as data on the same disease for a previous time period, data on a similar disease, or even covariate information. Each clustering configuration has a corresponding adjacency matrix, with each matrix resulting in a different degree of spatial smoothing. In

the second stage, a separate Bayesian hierarchical model is fit to the data for each of the *I* cluster configurations. These *I* models can be compared using model goodness-of-fit criterion (88, 98).

Anderson *et al.* (88) refine this methodology by fitting a single model in the second stage which is capable of estimating the cluster structure and disease risk simultaneously. This is achieved by specifying the prior for the random effects as a mixture of *I* CAR priors, where each mixture component has a different adjacency matrix with a corresponding prior weight. The most appropriate number of clusters can then be selected using the posterior mode or median. This updated methodology improves computational efficiency because little time is spent estimating the model parameters for those mixture components which correspond to an untenable cluster configuration. However, note that these methods have only been applied to diseases such as chronic obstructive pulmonary disease (COPD), which tends to be more common than diseases such as cancer.

Alternative approaches include Lawson and Clark's (99) weighted sum of spatial priors which has no single global smoothing, so the underlying risk is free to either be smoothed or to 'jump' between areas (Box 2.16). In contrast, several of the semi-parametric mixture models force the risk surface to be discontinuous, including marginal mixture models and spatial partition models (99, 100). Although some of the semi-parametric mixture models allow for a smooth underlying risk, Green and Richardson's hidden Markov model (101) smoothed the data more than the BYM model when the data had insufficient evidence to create a higher-risk group (102). Also, identified clusters might not be spatially contiguous under this model (88).

Mixture models are prone to less computational stability than the BYM, the risk of label switching and component identifiability difficulties, as well as requiring greater care in covariate selection due to their influence on risk label categorisation (103). Often the mixture models also require greater programming skills, relying on GNU and/or Fortran to run the models (101, 104).

Identifying clusters

Another research problem concerned with clustering is image segmentation, in which the goal is to classify image pixels (which represent arbitrary areas defined by a grid) into well-defined clusters (105). Hidden Potts-Markov random field (MRF) models are commonly used in Bayesian image segmentation methods (Box 2.17). However, making inferences on these types of models is difficult, hence current image segmentation methods typically rely on approximate estimators (106). A major limitation of such approaches is that they are supervised, meaning that the regularisation parameter of the Potts model must be specified *a priori*. Selecting an appropriate regularisation parameter *a priori* can be difficult since they can be highly dependent on the image. Unsupervised approaches which self-adjust the regularisation parameter are currently possible, but at an enormous computational cost (107).

The recently proposed methodology of Pereyra and McLaughlin (107) permits approximate inference on hidden Potts MRFs which is unsupervised and also computationally fast. The

crux of their approach involves dividing the problem into two simpler problems, both of which can be solved easily and relatively quickly.

Box 2.17 Hidden Potts-Markov random field model

Let y_i is an element of $\boldsymbol{y}$ ($y_i \in \boldsymbol{y}$) be observations with latent labels $z_i \in \boldsymbol{z}$. For example, y_i might represent the intensity of the i^{th} pixel in a greyscale image, while z_i identifies a segment of the image to which the i^{th} pixel belongs, for a finite set of segments, $\{1, \dots, K\}$. Given the segment identifier z_i, the intensity of the pixels in that segment will be similar. For example, if the intensities were assumed to be Gaussian, pixels in the k^{th} segment might have the same mean and variance,

$$p(y_i|z_i = k) = \text{Normal}(\mu_k, \sigma_k^2).$$

The unobserved random variables $\boldsymbol{Z} = \{Z_i\}$ represent nodes in a hidden Markov random field, each node corresponding to a pixel y_i. For each Z_i, a neighbourhood is defined such that Z_i only depends on those nodes in the neighbourhood, and is conditionally independent of other nodes (the Markov property):

$$p\left(z_i\middle|\boldsymbol{z}_{\setminus i}\right) = p\left(z_i\middle|\boldsymbol{z}_j, j\sim i\right),$$

where $\boldsymbol{z}_{\setminus i}$ denotes all values of $\boldsymbol{z}$ except z_i, and $j\sim i$ denotes nodes i and j are in the same clique. The Hammersley-Clifford theorem states that this conditional probability distribution has the form:

$$p\left(z_i\middle|\boldsymbol{z}_j, j\sim i\right) = \frac{1}{W_i}\exp(-\beta H(z_i))$$

where W_i is the partition function, β is the regularisation parameter, and $H(z_i)$ is the energy function of z_i. In the case of the Potts MRF model, the energy function is chosen to be $\sum_{i\sim j}\mathbb{I}(z_i = z_j)$, leading to

$$p\left(z_i\middle|\boldsymbol{z}_j, j\sim i\right) = \frac{\exp\left\{-\beta\sum_{i\sim j}\mathbb{I}(z_i = z_j)\right\}}{\sum_{k=1}^{K}\exp\left\{-\beta\sum_{i\sim j}\mathbb{I}(z_j = k)\right\}}.$$

Pereyra and McLaughlin (107) compare the results obtained from this method against four state-of-the-art supervised segmentation algorithms, and one unsupervised MCMC algorithm, each applied to three different datasets of varying complexity. Visually, the resulting image segmentation of the proposed method is comparable to all five other methods. In terms of computational efficiency, the proposed method is at least one thousand times faster than the unsupervised MCMC algorithm, and only two or three times slower than the supervised algorithms. Only two or three clusters were used in these tests, so how well this proposed method works for larger numbers of clusters is yet to be quantified. However, these results indicate that the proposed method is very promising.

Change of Support

One limitation that persists in spatial modelling is the difficulty in making statistical inference on spatial support points which differ to the support points provided by the data. For example, data may be collected for each SA1, and predictions/estimates of a particular variable can be obtained from an appropriate model for these areas quite easily. But making inferences for postal areas, for example, or some arbitrarily defined area is not so straightforward. This is known as a change-of-support problem. This problem may arise when the desired support points for inference cannot be foreseen or constrained to one support type at the time of modelling, or when the support points are limited by the data available (108).

Current methods have provided solutions to this problem when the underlying data is assumed to be Gaussian. Yet disease mapping studies often have count data and are typically modelled by a Poisson or Binomial distribution. Even for this type of data, methods have been developed, such as simple areal interpolation (109), whereby inferences may be made on target support points by imputing values from surrounding data support points. However, when the data are not recorded without error, such as with survey data, the uncertainty of the estimates at target support points is unknown which limits their inferential usefulness (108).

Bradley *et al.* (108) proposed new methodology to address these problems by incorporating the estimated variance of the data in the model. The areal count data are interpreted as an aggregation of events from a latent, unobserved, spatial point process. This latent process is modelled using a Bayesian hierarchical GLMM. Specifically, the latent spatial process is modelled as a combination of additive covariate and spatial basis function effects, and the observed count data, conditional on the latent spatial process, are modelled by a Poisson distribution. Estimates of the variances of the observed data values are modelled jointly with the observed data. By estimating the model parameters at the point level, estimates can be obtained for any desired support point by aggregating the latent process. Including variance estimates in the model is not necessary, but doing so provides more precise estimates.

2.5.4.2 Survival data

Survival can be measured in several different ways. All-cause, or overall, survival captures all deaths regardless of cause. Often net survival is of more relevance, as it aims to capture only deaths resulting from the disease of interest. Net survival can be approximated by either cause-specific or relative survival. Information on the recorded cause of death is required for a cause-specific analysis, whereas for relative survival, deaths due to any cause among cancer patients are compared against background population mortality rates.

The Cox proportional hazards model is the most widely used survival model (110, 111). This model is applied to either overall or cause-specific survival, and has no assumptions regarding the nature or shape of the underlying survival distribution. Correlation can be incorporated between areas by including random effects termed ‘frailties’ (112, 113).

However, it does assume a proportional (multiplicative) relationship between the hazard and the log-linear covariate function (111), and this assumption is often violated (114).

Some Bayesian spatial cancer survival analyses have preferred to use either overall or cause-specific survival analyses, and have based their models on variants of the Cox proportional hazards model (115), or parametric models such as accelerated failure-time models (116-118). Survival for multiple cancers has also been jointly modelled with spatial frailties (119). However, these either assume accurate cause of death data (if a cause-specific analysis), which is a key disadvantage for population-based cancer data, while overall survival analyses may have confounding from unrelated differences in mortality between areas.

There have been a few different types of relative survival models incorporating spatial components within a fully Bayesian framework. Fairley *et al.* (120) expanded the additive hazard model recommended by Dickman *et al.* (121) to incorporate spatial and unstructured random effect components similar to the BYM model (Box 2.18).

Box 2.18 Bayesian spatial relative survival model

Fairley *et al.* (120) introduced the following relative survival model within a fully Bayesian context:

$$d_{ijk} \sim \text{Poisson}(\mu_{ijk})$$
$$\log(\mu_{ijk} - d^*_{ijk}) = \log(y_{ijk}) + \alpha_j + \mathrm{x}_{ijk}\beta_k + u_i + v_i$$

where d_{ijk} represents the number of deaths resulting from any cause in the *i*th area, *j*th follow up time from diagnosis interval, and *k*th age group, y_{ijk} is the person-time at risk, d^*_{ijk} is the expected number of deaths due to causes other than the cancer of interest, α_j is the intercept (which varies by follow-up year), **x** is the predictor variable vector (although proportional excess hazards are assumed, interactions can be accommodated), u_i is the spatial component assigned an intrinsic CAR prior and v_i is the unstructured component with a normal prior centred on 0.

The term $\log(\mu_{ijk} - d^*_{ijk})$ is a non-standard link function representing the log excess deaths, or the deaths considered to result from the disease of interest (121). Follow-up intervals can be of any duration, but often annual intervals are used.

Many of the advantages for this approach are similar to that of the Cox proportional hazards model (indeed, if using cause-specific or overall survival with time split at each event, the Poisson piecewise model equates to the Cox proportional hazard model (122)), including no assumption of the baseline survival shape. However, the disjointed piecewise process is biologically implausible, and covariates such as age cannot be included as continuous variables without the model becoming too cumbersome.

A similar Poisson piecewise model was combined with Bayesian geoadditive models so the baseline hazard was modelled using penalized splines (123). This flexible semiparametric model overcomes many of the limitations of the Poisson piecewise approach and incorporates a spatial effect, random effects and fixed effects, but is computationally intensive.

An alternative approach is to combine a parametric formulation with splines for flexibility in modelling the baseline hazard, producing flexible parametric survival models (122). Cramb *et al.* (124) extended Nelson's relative survival version (125) to propose the Bayesian spatial flexible parametric relative survival model. This approach combines the benefits of flexible parametric models: the smooth, well-fitting baseline hazard functions and predictive ability, with the Bayesian benefits of robust and reliable small-area estimates. Both spatially structured (with an intrinsic CAR prior) and unstructured frailty components are included. Advantages of this approach include the ease of including additional complexity, the use of individual-level input data, and the capacity to conduct overall, cause-specific and relative survival analysis within the same framework (124).

2.6 Computation

The practical aspects of producing small-area estimates is discussed in this section, both in calculating estimates and available software.

2.6.1 Numerical

Unsmoothed estimates can be calculated in any statistical software package, or in a spreadsheet.

Direct smoothing approaches such as locally-weighted averages/medians and kernel smoothing are available in many GIS packages, including commercial packages such as ArcGIS and MapInfo as well as freely-available programs such as GRASS GIS and GeoDa, among others (Appendix C).

There are also a range of options for obtaining model-based results, and most statistical software will perform these. Parameter estimates from GLMs such as simple forms of the Poisson model (see Box 2.10) can be obtained via best linear unbiased estimator (BLUE) analyses (Box 2.19). The corresponding approach for GLMMs, composed of both fixed and random effects (see Box 2.9), is via best linear unbiased predictor (BLUP) estimation (126). If the variances and covariances of random effects are estimated and used in a BLUP estimator, then it is referred to as empirical BLUP, or EBLUP (126). Including spatial structure within the random effects can improve the EBLUP estimator even further, becoming the spatial EBLUP, or SEBLUP, estimator (127).

Both empirical Bayes and EBLUPs use a similar process of estimation. First, the variance components are assumed to be known, and BLUPs or EB predictors are obtained for the

unknown parameters (128). Then, the variances and covariances are estimated by the method of fitting constants/moments, or if normality is assumed, then via maximum likelihood (Box 2.19) or restricted maximum likelihood methods (126). For further details on these and similar methods for computing empirical Bayes estimates, refer to Meza (75).

However, as complexity in the Poisson model increases, alternative methods are often required, and this can range from approximating the likelihood via 'quasi' or 'pseudo' likelihoods (14), through to sampling from the posterior distribution of a Bayesian model.

Although the BLUP and Bayes approaches theoretically produce identical point estimates for small-areas (126), in certain circumstances fully Bayesian estimates were shown to have smaller MSEs than the corresponding BLUP (129).

Specific software has been developed to enable Poisson kriging estimates to be easily calculated. Centroid-based Poisson kriging can be calculated using the freely available `poisson-kriging.exe` (51), which was written using Fortran 77. BioMedware's SpaceStat software (130) is also able to conduct Poisson kriging, including ATP Poisson kriging (in addition to many tests for spatial correlation). This software replaces the space-time information system (STIS) (131).

Box 2.19 BLUE, Maximum likelihood and Least Squares

The least squares estimate is the value that minimises the sum of squared errors (54). More formally, for the model $y_i = X_i\beta + \varepsilon_i$, the least squares estimate is the $\hat{\beta}$ that minimises $\sum_{i=1}^{n}(y_i - X_i\hat{\beta})^2$. This is also the best linear unbiased estimator (BLUE) if the variance-covariance matrix of any linear unbiased estimator $\tilde{\beta}$ is greater than or equal to the variance-covariance matrix of $\hat{\beta}$ (132). If the errors ε_i are independent with equal variance and normally distributed, then the least squares estimate is also the maximum likelihood estimate (54).

Under maximum likelihood, the probability of the likelihood (which is the joint distribution) of all observations is maximised in regards to several relevant parameters (12). Maximum likelihood estimation has several desirable attributes, including consistency and efficiency, as well as being able to handle small departures from the normality assumption (133)

Bayesian hierarchical models containing spatially structured components generally cannot be solved via numeric integration, but an alternative approach that is well suited to these models was developed in the 1950s, although it wasn't until the 1990s that this became widely applied in statistics (134). This method is Markov chain Monte Carlo (MCMC).

2.6.2 Markov chain Monte Carlo

Methods such as approximating large-sample exact solutions (asymptotic approximations), traditional numerical approaches and non-iterative Monte Carlo methods are likely to either be infeasible or produce results with low accuracy when applied to complex statistical models, many of which are Bayesian (135). MCMC methods (Box 2.20) are able to reduce complex multidimensional problems to a series of lower-dimensional problems, while not requiring conjugate structure between the likelihood and the prior distribution (135). MCMC samples from the posterior distribution of Bayesian models and has dramatically expanded the potential scope of statistical models, thanks to modern computing power (136).

Box 2.20 MCMC

A Markov chain has been described as a frog jumping on a set of lily pads (137). Assuming it must always land on a lily pad, the probability of jumping onto another (or even the same) lily pad depends only on the lily pad it is currently on. Likewise, the future behaviour of a Markov chain is dependent only its present state (137).

Provided the Markov chain has converged, the desired summary of the posterior distribution is approximated by MCMC, which are simulated random processes conditional on the previous value. A range of MCMC algorithms are available, but currently the most popular for disease mapping applications is the Gibbs sampler.

The Gibbs sampler (138, 139) is an algorithm that samples from each of the full conditional distributions $p\left(\theta_i \middle| \boldsymbol{\theta}_{j\neq i}, \boldsymbol{y}\right)$ in the model. A single new value of θ_i is generated at each iteration, conditional on all other θ's, as all proposals are accepted in Gibbs sampling (55).

The Gibbs sampler algorithm proceeds as follows for *k* parameters, given a set of starting values $\left\{\theta_1^{(0)}, \ldots, \theta_k^{(0)}\right\}$:

1. Draw $\theta_1^{(t)}$ from $p\left(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)}, \boldsymbol{y}\right)$
2. Draw $\theta_2^{(t)}$ from $p\left(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)}, \boldsymbol{y}\right)$

   ⋮

k. Draw $\theta_k^{(t)}$ from $p\left(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{k-1}^{(t)}, \boldsymbol{y}\right)$

Concerns have been raised with regards to assessing convergence, selecting starting values, and the length (and necessity) of burn-in periods for MCMC analyses (140). The computational resources required is perhaps their greatest disadvantage, with alternative methods seeking to provide good approximations in a drastically reduced timeframe (141). However, their ability to directly approximate probabilities (142), and answer a broad range of questions (143) remains unsurpassed.

Programs have been developed to assist in conducting MCMC-based analyses, including BUGS (Bayesian inference using Gibbs sampling) software (91, 144), Stan (which uses Hamiltonian Monte Carlo) (145), MLwiN (146), JAGS (Just Another Gibbs Sampler) (147) and the R package MCMCpack (148).

2.6.3 MCMC approximation methods

The computational requirements and time needed to conduct MCMC analyses can be off-putting to those considering a fully Bayesian analysis. More recently, a range of approximation methods have become available as an alternative to MCMC, with the benefit of a reduced computational burden.

The most popular of these within the disease mapping context is integrated nested Laplace approximation (INLA), and this is available in an R package (www.r-inla.org/). The approximation is broken down into smaller sub-problems, and a method of approximation known as Laplace approximation is applied when the densities are near-normal (Box 2.21) (149). A wide range of models can be approximated by INLA, including most GLMs, and it has been shown to produce good approximations to output from MCMC for cancer (simulated and real) data, provided the disease is not incredibly rare (150).

The key advantages of INLA are its speed and flexible model specification (55). Disadvantages in its current form are the somewhat restricted range of prior distributions and an inability to handle: models not expressible in log-linear form, mixture distributions, as well as certain types of missing data/measurement errors (55).

Box 2.21 INLA (149)

Critical assumptions in INLA are that:

1. The number of hyperparameters is small, and does not exceed 20. (Typically this is between two and five.)
2. The distribution of the latent field is Gaussian. When the dimension is high (10^4-10^5), this is either a Gaussian Markov random field, or close to one.
3. Each observation only depends on one component of the latent field.

2.6.4 Creating the neighbourhood matrix

Several methods are possible to create a neighbourhood matrix. Adjacency-based neighbours can be assigned using any GIS package, provided the polygon arrangement and relationships are clean. Sometimes the shapefile will have small artefacts where boundaries do not meet precisely, which would require intervention, such as ‘snapping’ vertices within a threshold distance together (151).

Some statistical programs, such as R, offer several options for creating neighbour matrices, including contiguity or distance-based options (including k-nearest neighbour and threshold distance). GeoDa also offers a wide range of options for creating and visualising neighbourhood matrices (Table 2.1). Most packages offer several export options for the resulting neighbourhood matrix, but it is worth ensuring that an appropriate format for the software used in further analyses exists.

2.6.5 Software

Tables 2.1 and 2.2 summarise the broad capabilities of common software and specific examples of software used by method, respectively. Refer to Appendix C for further details on software.

Table 2.1 Common mapping, GIS and statistical software and capabilities

		Analyses									
						Smoothing		Model-based smoothing			
Software	**Type**	**Visualise maps**	**Neighbour-hood matrix**	**Spatial correlation**	**Raw estimates**	**Locally-weighted**	**Kernel**	**Spatial regression**	**Poisson kriging**	**EB**	**HB**
Open source											
Bing Maps	Map	Y									
BUGS	Stat	Y	Y	Y	Y						Y
Epi Info	Tools	Y			Y						
GeoDa	Tools	Y	Y	Y	Y	Y	Y	Y		Y	
GRASS	GIS	Y									
Google Earth	Map	Y									
JAGS	Stat										Y
NIMBLE	Stat										Y
PySAL	Tools	Y	Y	Y	Y	Y	Y	Y		Y	
R	Stat	Y	Y	Y	Y	Y	Y	Y		Y	Y
SaTScan	Tools	Y		Y				Y			
Stan	Stat										Y*
Commercial											
ArcGIS	GIS	Y	Y		Y	Y	Y	Y			
MapInfo	GIS	Y	Y		Y	Y	Y				
MLwiN	Stat							Y			Y*
SAS	Stat	Y	Y	Y	Y	Y	Y*	Y		Y	
S-Plus	Stat	Y	Y	Y	Y						
SpaceStat	Tools	Y	Y	Y					Y		
Stata	Stat	Y	Y	Y	Y	Y	Y	Y		Y	
TerrSet	GIS	Y	Y	Y	Y	Y	Y				

Abbreviations: Stat=Statistical software, Map=Mapping software, GIS=Geographic Information Systems software, EB=Empirical Bayes, HB=Hierarchical Bayes, Y=Yes.

* Indicates limited functionality, such as lacking programmed CAR prior distributions. Note that often software can interface with other software to either provide greater functionality (e.g. between statistical packages and GIS software), or to enable programming within the language of convenience (e.g. Stan and JAGS can interface with R).

Software is considered able to perform a hierarchical Bayes analysis if a random effects term for each area can be modelled.

Table 2.2 Examples of software used by method

Analysis			Example of software used
Spatial correlation	Global	Moran's I (19)	GeoDa (152)
		Geary's C (20)	GeoDa (153)
		Tango's MEET (23)	S+ code in R (154)
	Local	LISA (21)	GeoDa (152)
		Spatial scan statistic (24)	SaTScan (155)
Unsmoothed	Crude/standardised rates		Unstated, statistical (43)
Direct smoothing	Locally-weighted		STARS (156) – now in PySAL
	Kernel smoother		R (157)
Model-based	Poisson kriging		SpaceStat, ArcGIS (60)
	Empirical Bayes		SAS (158)
	Fully Bayesian – Incidence etc.	BYM (81)	WinBUGS (159)
		Leroux CAR prior (93)	WinBUGS (160)
		MacNab alternative convolution prior (95)	Unstated, BLUE+REML (95)
		Anderson spatial pattern & cluster model (88)	R (88)
		Lawson & Clark mixture model (99)	M-H algorithm, self-coded (99)
		Spatial partition models (100)	MCMC, self-coded (100, 161)
		Green & Richardson hidden Markov model (101)	MCMC coded in Fortran (101)
		Hidden Potts-Markov random field model (107)	Self-coded, unstated (107)
		Bradley latent spatial process (108)	Self-coded, unstated (108)
	Fully Bayesian - Survival	Cox proportional hazards model	SPSS, BUGS (162)
		Bayesian spatial relative survival model (piecewise)	Stata, WinBUGS (120)
		Bayesian spatial flexible parametric relative survival model	Stata, WinBUGS, MapInfo (124)

3. Current approaches for small-area cancer estimates

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, *whatever the matter may be."*

~ William Thomson, 3 May 1883,
'Electrical Units of Measurement' lecture

The data used to generate cancer atlases may relate to cancer incidence, mortality, survival, or screening data. Methods used to generate published estimates are examined in this chapter.

3.1 Small-area cancer screening estimates

Cancer screening is the application of a test to an apparently cancer-free group to identify those people likely to have the disease (163). Cancer screening programs involve large numbers of people. As such, most studies of small-area variation used unsmoothed percentages, and occasionally additional tests were applied to determine areas that showed statistically significant evidence of higher/lower outcomes.

Cervical cancer screening was examined in small-areas of Rotterdam, although areas with <2000 residents were excluded from the analysis to prevent unstable results (164). Percentages were calculated for uptake of screening, and the association with the proportion of migrants and specific marital statuses considered.

Cervical cancer screening, breast cancer screening, and bowel cancer screening (faecal occult blood testing/colonoscopy) were examined across 205 small-areas of Peel region in Ontario, Canada (165). The average population in each area was 4,000 people (range: 2,500 to 8,000). Maps were overlaid with the proportion of South Asian people, and also used LISA (166) to objectively identify areas of extreme variation. The authors noted that they deliberately did not choose a smaller level of resolution due to several issues including the potential for unstable rates (165). Another Ontario-based analysis examined the uptake of cancer screening tests in conjunction with screening for glucose and cholesterol across 18,950 small areas (167), with funnel plots used to identify abnormal areas falling outside the 95% or 99% CI for Ontario's screening rate.

Another descriptive analysis examining a small region of Florida, USA, while not mapping screening rates, mapped the ethnicity of each small area (expressed proportionally), and showed the location of colonoscopy services on the map (168).

3.2 Small-area cancer incidence/mortality estimates

Compared to the numbers involved in a population screening program, the number of people who are diagnosed with cancer or who die from cancer in a given time period is relatively few. Small numbers mean that spatial analyses of cancer incidence and cancer mortality often require some form of smoothing being employed for small-area studies. Nonetheless, unsmoothed estimates were mapped for small-area cancer atlases in Canada (169), India (170), New York (USA) (171), Pennsylvania (USA) (172), South Australia (Australia) (173), Sweden (174), New Hampshire (USA) (175) and the USA (176). Details on some of these are available in the associated report: "Grey Literature Review: Internet Published Cancer Maps".

Poisson kriging was used to examine cervical cancer mortality rates in 118 counties across four states in Western USA (177). ATP Poisson kriging was used in another study to examine age-standardised lung and cervical cancer mortality for two different areas of the USA – one with 92 counties of reasonably similar shape and size, and another area of 118 counties with varying size and shape (61). ATA Poisson kriging was used to examine age-standardised oesophageal cancer incidence over 336 areas in Iran (60), and Poisson kriging has also been used to examine lung cancer incidence around Perth in Western Australia (57).

Empirical Bayes methods have been used in several small-area cancer incidence/mortality analyses, including:

- Endemic Burkitt's lymphoma among children in Kenya. This modelled 272 cases identified from hospital data between 1999-2004 across 324 regions (178).
- Breast cancer mortality on the island of Sardinia (covering 22 regions across 1983-1987) (179).
- Pleural cancer mortality was modelled to approximate asbestos exposure in north-western Italy, across 1,209 areas during 1980-1992 (180). Poisson regression was then used to check for an association with lung cancer mortality across the areas (180).
- Lung cancer mortality in Missouri (1972-1981), across 115 areas and 4 age groups (45-54, 55-64, 65-74, 75+) (181).
- Lung cancer mortality ratios among women in 287 central Italian regions (182).
- Gastric cancer mortality in Hungary investigating an association with nitrate exposure over 192 settlements with regularly maintained nitrate records. Proxy information was used to adjust for dietary habits, smoking prevalence and socioeconomic status (183).

Fully Bayesian models have also been employed. Geographical variation in mortality from haematological tumours (leukaemia, non-Hodgkin's lymphoma and multiple myeloma) (184), thyroid cancer (2) and pleural cancer (185) was examined over 8,077 areas in Spain using the BYM model.

Explanatory covariates are often included in models. An area-level measure of sunlight exposure was included when modelling lip cancer incidence in Scotland (83). The model

used was similar to BYM, but had only one random effect term which was spatially structured (83). Age, sex, and age-sex interactions were incorporated into a model examining lung cancer mortality in Missouri (186). Atmospheric pollutants and lung cancer mortality in Tuscany, Italy, 1995-1999, were modelled using BYM with a nested latent factor model (187). An exploration of late stage breast and colorectal cancer incidence during 1995-1997 across 87 counties in Minnesota, USA also adjusted for a range of environmental effects (188).

The inverse distance of each census tract centroid from the nearest hazardous waste site was included when modelling leukaemia incidence in upstate New York (67). Similarly, the effect of industrial pollution on lung cancer and lymphohaematopoietic cancers in Northern Italy was explored (189, 190). Cervical cancer inequalities in stage at diagnosis in the former German Democratic Republic was modelled to examine disparities in Papanicolaou testing uptake (191).

Although less common, a few analyses have also used different forms of Bayesian hierarchical models aimed at enabling disparate changes to be detected. An extension of hidden Markov models (which can be considered a generalization of a mixture model) was applied to larynx cancer mortality in France (101). Leukaemia incidence in New York was modelled using Bayesian spatial partition models (100).

3.3 Small-area cancer survival estimates

Survival measures the proportion of people expected to remain alive for a given length of time after diagnosis, and the calculations often require individual-level data. Survival is a useful measure for exploring and comparing the impact of the healthcare system over time and place (162). When analysing cancer survival estimates across small-areas, it is recognised that estimates will be unstable if the resolution is too fine (192). As survival calculations focus on the deaths within a specified time from diagnosis, numbers tend to be smaller than for either incidence or mortality analyses.

Relatively few small-area survival analyses have been performed, and these have often utilised a Bayesian approach. An exception is Huang *et al.*'s (193) analysis of lung cancer and late-stage colorectal cancer survival across small areas in California (and then a more detailed analysis of Los Angeles areas). Here the 5-year and 3-year survival estimates were mapped, but also the adjusted survival time was calculated for each region and then a spatial scan statistic applied to determine areas with higher/lower survival (193).

Empirical Bayes methods were used to model leukaemia under proportional hazards (194). Osnes and Aalen expanded a Bayesian Cox proportional hazards model using components from the BYM model, to explore regional differences in survival for breast cancer and melanoma patients in Norway (162). Acute myeloid leukaemia was modelled in northwest England across 24 districts under proportional hazards model using a range of possible correlation structures (58). Breast cancer survival in France across 377 areas was modelled

using Hennerfeind's flexible continuous time geoadditive model (195), using metastasis as a proxy for staging information (113).

The Bayesian relative survival models incorporating spatial components are growing in popularity. Fairley *et al.* (120) used their Bayesian spatial relative survival model to explore variation in prostate cancer survival across 44 regions in Northern and Yorkshire England (average population size was ~ 150,000, ranging from <70,000 to 307,000). This same model was used to examine geographical variation in breast cancer survival in Catalonia, Spain over several different area definitions, down to the level of the census tract (average population of just 604 women aged over 15 years, and a standard deviation of 302) (196). This model was also applied to cancer data to examine small-area variation across 478 areas of Queensland, Australia by Cramb *et al.* (154, 197) and modified forms were used in detailed analyses of breast cancer relative survival across Queensland by Hsieh *et al.* (198, 199).

Breast cancer relative survival was examined across north-eastern France using the Bayesian geoadditive model proposed by Hennerfeind *et al.* (123), while Cramb *et al.* (124) demonstrated their proposed Bayesian spatial flexible parametric relative survival model on breast, colorectal and lung cancer in Queensland.

3.4 Summary and conclusion

Cancer is a relatively rare disease. When cancer measures are mapped, it is important these estimates are reliable. For cancer outcomes such as incidence, mortality and survival, most analyses use some form of smoothing. Models based on GLMMs are often employed, and these also have the advantages of the ease of incorporating covariates, considering interactions and examining model fit.

Even for cancer screening data, where numbers are exponentially higher, producing unsmoothed estimates often constrains the level of resolution possible, or necessitates excluding some of the areas. Smoothing could potentially be useful for screening data as well, depending on the level of resolution of areas.

4. Further topics in Bayesian models

"The most important questions of life are indeed,
for the most part, really only problems of probability."

~ Pierre Simon Laplace
Théorie Analytique des Probabilités, 1812

4.1 Inclusion of multiple nested geographies

Models formulated within the Bayesian framework are naturally hierarchical, given their expression of a statistical model as a series of related layers. For this reason, they are an appealing choice for the analysis of nested data structures.

The analysis of health outcomes often involves the integration of data from multiple sources, observed at different scales. In a spatial setting, it is common for these scales to be embedded in one another – for example, individuals within regions or regions within a state – resulting in a hierarchical or nested data structure.

Bayesian hierarchical models can be developed to take account of this structure, to (i) allow for spatial correlation between effects defined at the same spatial scale; and (ii) relate/compare effects defined at different levels. The latter form of inference can be achieved through careful consideration of prior distributions and, for the most part, their specification should be guided by the comparative inferences the analyst wishes to draw (200). An example of this modelling approach is provided in Box 4.1.

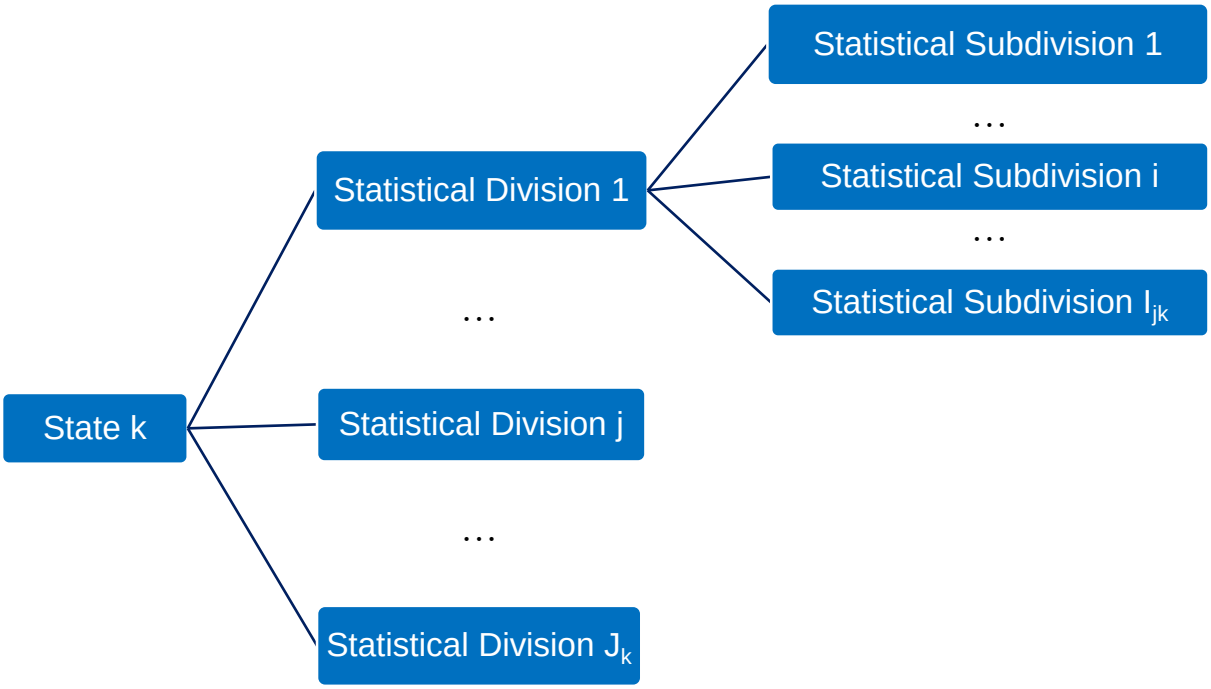
In this example, the defined model allows for three main inferences:

1. The comparison of state estimates ($\boldsymbol{\alpha}$) relative to the overall average estimate (γ).
2. The comparison of statistical division estimates ($\boldsymbol{\beta}_k$) within each state ($k = 1, \dots K$), relative to the overall estimate for state k.
3. The comparison of statistical subdivision estimates ($\boldsymbol{\theta}_{jk}$) within each statistical division ($j = 1, \dots J_k$), relative to the overall estimate for statistical division j.

Bayesian hierarchical models are often referred to as multilevel (54) or multiscale (55) models. In the non-Bayesian paradigm, multilevel models are a popular model class for analysing data of the aforementioned form. Common among these methodologies is the aim of apportioning variation in the outcome to different levels of the hierarchy (Figure 4.1). When formulated within the Bayesian setting, a multilevel model can itself be re-expressed as a hierarchical model, through the use of hierarchical centring (201); for this reason, these terms are often used interchangeably.

Box 4.1 Nested geographies as a hierarchical model

Figure 4.1 Schematic representation of a hierarchical model



Let y_{ijk} = The number of cancer cases in subdivision i, within division j, within state k.

$$O_{ijk} \sim \text{Poisson}\left(E_{ijk}\theta_{ijk}\right)$$

$$\log(\boldsymbol{\theta}_{jk}) \sim \text{Normal}_{I_{jk}}\left(\beta_{jk}, \frac{1}{\sigma_\theta^2}(\boldsymbol{D}_{\theta_{jk}} - \boldsymbol{W}_{\theta_{jk}})^{-1}\right)$$

$$\boldsymbol{\beta}_k \sim \text{Normal}_{J_k}\left(\boldsymbol{\alpha}_k, \frac{1}{\sigma_\beta^2}(\boldsymbol{D}_{\beta_k} - \boldsymbol{W}_{\beta_k})^{-1}\right)$$

$$\boldsymbol{\alpha} \sim \text{Normal}_K\left(\gamma\mathbf{1}, \frac{1}{\sigma_\alpha^2}(\boldsymbol{D}_{\boldsymbol{\alpha}} - \boldsymbol{W}_{\boldsymbol{\alpha}})^{-1}\right)$$

$$\gamma \sim p(\gamma)$$

The matrices $\boldsymbol{D}_{..}$ and $\boldsymbol{W}_{..}$ encode spatial correlation by defining the neighbourhood structure among geographic units defined as the same spatial scale. Each variance component is assigned a prior distribution, similar to the earlier Bayesian models in this report. The overall intercept (γ) is also assigned a prior distribution, denoted generically as $p(\gamma)$. Examples of prior distributions include a Uniform distribution, $\gamma \sim \text{Uniform}(-1000, 1000)$ or a Normal distribution with large variance, $\gamma \sim \text{Normal}(0, 1000)$.

In Australia, Turrell *et al.* (202) proposed a multilevel model with five levels: individuals nested in statistical local areas, statistical subdivisions, statistical divisions and States, for associating socioeconomic disadvantage with all-cause mortality (135). In the Atlas of cancer mortality in the European Union (203), Poisson regression was used to attribute variation in cancer mortality rates to age groups, countries and regions nested within countries. Models for both applications were developed in a non-Bayesian setting and correlation among spatially indexed effects was not accounted for.

The extension of these models to the Bayesian framework to allow for spatial smoothing is relatively straightforward. Lawson (55) used a Bayesian hierarchical model to analyse oral cancer incidence across the state of Georgia, USA, including both public health districts and nested counties plus the contextual effects of district on county. In this example, the joint model was slightly preferred over separate models for district and county. Another example is provided by Louie and Kolaczyk (204), who exploited the Bayesian approach to detect areas with significantly increased risk. They analysed aggregated count data across the three nested levels of region (one area), province (nine areas) and municipality (287 areas). Although their focus was not on estimation, it would be straightforward to combine this to produce a disease mapping approach with testing aspects. Bayesian multiscale analyses require careful consideration of prior selection, but have many advantages.

4.2 Inclusion of remoteness and area-level socioeconomic status

In Australia, there is substantial interaction between the geographic remoteness and socioeconomic level of an area, with more remote areas often being more socioeconomically disadvantaged as well as having higher levels of poverty (205). As such, understanding the differences between different combinations, such as comparing urban very advantaged areas to very remote very disadvantaged areas, may be desirable. Common approaches to modelling this scenario includes either including an interaction term between the levels of remoteness and socioeconomic disadvantage, or creating a composite variable and either stratifying the analysis on this term, or including the levels of the composite term as dummy variables in the model (without any remoteness/socioeconomic main effects).

Advantages of stratifying the data are the simplicity, and results can be intuitively easier for non-statisticians to grasp. Disadvantages include an inability to compare between the areas as thoroughly as when they are in the same model.

Advantages of including a composite variable is the simplicity of calculating parameter estimates, however, the disadvantages include the inability to untangle main effects, which requires careful interpretation of results (206). Previous examination of a model with main effects would be recommended before using this approach.

The key advantage of including an interaction between the levels of the variables is the flexibility. Interactions with other variables of interest, such as age groups, could also be incorporated easily, while still measuring the impact of the main effects.

The above advantages and disadvantages are true even if a Bayesian approach is not used. Using a Bayesian regression model would additionally require considering the prior choice on parameters with care, ensuring convergence and identifiability of all parameters of interest, and assessing the model via sensitivity checks (207). If spatial correlation is desired to be included in the analysis through using a structured prior, such as the CAR prior distribution, stratification would require a separate adjacency matrix to be generated for each combination, due to the varying number of included areas. The Bayesian approach additionally facilitates comparison of non-nested models, so can assist in choosing between model options.

4.3 Use with survey data

Sample surveys are commonly used to obtain a variety of information over time for both the total population, as well as a variety of subpopulations (126). Although these subpopulations can be any domain, such as sociodemographic groups, our focus in this section will be on geographic subpopulations.

Due to cost, as well as unanticipated uses of survey data, often a sample size is not sufficiently large to enable reliable estimates for all domains. Spatial subpopulations thus require the use of small-area estimation methods, which may involve statistical models (126). Note that in contrast to our earlier definition of small-areas (Section 1.2), which was based on population size, a small-area for survey data is based on having a small (and insufficient) sample size, regardless of population (208).

The focus of small-area estimation is on producing reliable estimates of means, counts, quantiles, as well as the associated error, for areas with limited/no sample data (209). When outcome data are lacking, auxiliary covariate information (such as obtained from censuses or disease registries) with good predictive power, becomes critical (209). Auxiliary variables are thus used to 'borrow strength' (208).

A recent comparison of a range of procedures, spanning from weighted 'raw' estimates through to models with random effect components, found that model-based estimates were generally the 'more effective' approach (208). In practice, and especially in the Australian context, direct estimates are often suppressed for at least some areas due to small numbers and high uncertainty.

Providing the model is appropriate and the sampling is robust, there are several advantages to modelling small-area estimates from survey data, such as (126, 210):

1. The assumed model allows 'optimal' estimators to be obtained
2. Each estimator can have area-specific measures of variability
3. Sample data can be used to validate models
4. Complicated data structures (such as spatial correlation) can be examined by a variety of models.

When modelling survey data, the predominant paradigm employed is that of the sampling model (see Section 2.5) (54). Most models for survey data are mixed effects models built on the model developed by Fay and Herriot (211). It is now standard practice to include not just the variation in auxiliary variables across small areas, but to also add random area effects to further account for between area variability (210). Linear estimators in GLMMs can be estimated by EBLUP, empirical Bayes, or hierarchical Bayesian models. Both empirical and hierarchical Bayesian methods are also appropriate for a broader range of modelled outcomes, whether binary or count data, and alternate model structures (126).

Hierarchical Bayes approaches are now extensively utilised for small-area estimation (126). In addition to the advantages mentioned in Section 2.5.4, benefits within the sampling context include obtaining smaller coefficient of variations for direct estimates, especially for areas with smaller populations (126). They also avoid the problems that EBLUP or EB can have if the restricted maximum likelihood (REML) model estimate variance is estimated to be around zero, which results in all the estimates of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ being given a weight of zero (212). Any small-area estimation model can be expanded to the hierarchical Bayesian context (Box 4.2). This extends to unmatched sampling and linking models, or incorporating spatial correlation.

Box 4.2 The basic area-level model (126)

This model can be expressed as:

$$\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i$$

where $\hat{\theta}_i$ is an estimate of the *i*th area parameter $\theta_i = g(Y_i)$, z_i is a vector of area-level covariates, b_i is a known positive constant, v_i are area effects that are considered to be independent and identically distributed with a mean and variance of $(0, \sigma^2)$ and are independent of the sampling errors e_i which are independently distributed with a mean of 0 and known variance ψ_i.

The hierarchical Bayes version of this model has the addition of priors on the following levels, for instance:

$$\hat{\theta}_i \sim \text{Normal}(\theta_i, \psi_i)$$
$$\theta_i \sim \text{Normal}(\mathbf{z}_i^T \boldsymbol{\beta}, b_i^2 \sigma_v^2)$$
$$\sigma_v^2 \sim \text{Uniform}(-\infty, \infty)$$

Note that using the flat prior shown on σ_v^2 may not be ideal when the sampling variances differ substantively over areas.

The main disadvantage of using a Bayesian analysis is that it should be conditional on all variables that affect the probability of inclusion and non-response, and this can rapidly result in extremely complicated models, especially when aiming to produce population estimates from sample survey data that are not representative of the population (213). Although weighting is often used in this situation, producing appropriate weights can be difficult, and empty cells can cause additional difficulties for weighting in the small-area context (213).

Suggestions have included using multiple Bayesian hierarchical models and then averaging over the posterior distribution, although this remains an area of active research (214).

4.4 Spatio-temporal data

Spatio-temporal data consists of data points which are stratified both by space and time. The rate at which spatio-temporal data is generated and collected is ever-increasing, and new methods are persistently being developed to deal with this type of data. Spatio-temporal models can be seen as a natural extension of spatial models, but this extension increases the complexity, both in terms of notation and computation, and introduces new complications to be addressed, such as how to account for interactions between space and time. Moreover, difficulties in spatial modelling, such as the handling of missing data, are exacerbated in spatio-temporal modelling (44, 215).

Nonetheless, spatio-temporal models have many benefits in interpretation of overall patterns of risk and dynamics, as well as improved accuracy compared with purely spatial models (216-218).

Naturally, much of the earliest work on Bayesian spatio-temporal models focused on extending the BYM model. The CAR prior used in the BYM model can define neighbourhood structures across space and time, so that an area's neighbours includes spatial neighbours as well as its own value in the previous and following time periods (77).

One of the earliest Bayesian approaches was the Bernardinelli space-time model (Box 4.3) (219). This has been applied to diseases such as insulin-dependent diabetes mellitus (220), and leishmaniasis (221). Covariates have been included (82, 222), and in some cases, errors in the estimates of indirectly observed covariates (such as, for example, estimating cigarette smoking prevalence from survey data) have also been incorporated (222, 223).

Box 4.3 The Bernardinelli spatio-temporal model

Let O_{it} denote observations from area $i = 1, \dots, I$ at time $t = 1, \dots, T$. The Bernardinelli model can then be expressed as follows:

$$O_{it} \sim \text{Poisson}(E_{it}\theta_{it})$$

$$\log(\theta_{it}) = \alpha + u_i + \gamma_t + \delta_{it}$$

where Y_{it} are the observed cases for the i^{th} area and t^{th} time interval, E_{it} are the expected number of cases, θ_{it} are the underlying relative risks, α is the mean log-rate over all areas, u_i represents the area effect and follows an intrinsic CAR distribution, γ_t is the mean linear time trend over all areas and δ_{it} represents the difference between the area-specific trend and the mean trend β_t (219). In this model, the intercept is the sum of $\alpha + u_i$, while the trend is the sum of $\gamma_t + \delta_{it}$ (219). The prior for $\delta_{it}$was a modified CAR distribution that allowed for correlation between the intercept and trend.

However, the restriction to linear trends over time in the Bernardinelli model was an important limitation (219). Further extensions have been proposed to overcome this, including using quadratic instead of linear time trends (221, 224). In contrast, Waller *et al.* (225) applied the BYM model to each time point separately. Although this allowed the spatial structure to evolve over time, it essentially treated time as exchangeable (225, 226). This may not be ideal for modelling a disease such as cancer since it would be unlikely to have a separate spatial distribution within each time period (227).

Spatio-temporal interactions have also been incorporated. Sun *et al.* (227) and Kim *et al.* (228) included random spatial and spatio-temporal interaction effects when modelling cancer mortality in Missouri, but the temporal component was still restricted to a linear form (229). Abellan (216) included a space-time interaction term in a BYM-type model to capture any departure from predictable patterns based on the overall time trend and the overall spatial risk surface. Further extensions allowed for random spatial, temporal and spatio-temporal interaction terms, and was used to examine prostate cancer incidence in Iowa over six time periods of 5-year groupings (229).

Mixture models have also been extended to a spatio-temporal formulation, which were applied to lung cancer incidence and mortality in Germany for 30 years (divided into three time periods) across 215 counties (230).

The BYM model has also been combined with dynamic models (231). Dynamic models allow estimates to 'borrow' strength from adjacent timepoints, so do not assume linearity or stationarity, but instead enable non-parametric estimation of temporal trends (226, 231). This means time-changing effects of covariates can be included (231). In principle this model allows for estimation of any age-period interaction, including cohort effects (224). This model was demonstrated on Ohio lung cancer mortality data, stratified by age, gender, race for each year (of 21 years) and each county (of 88 counties) (231).

Specific age-period-cohort (APC) Bayesian hierarchical spatio-temporal models have also been proposed as a method to jointly study the spatial pattern of disease risk and evolution in time (232). Generally the BYM model again forms the basis, with additional time main effects defining age, period and cohort specific parameters; space-time interactions as specified in Knorr-Held (226); or cohort effects (232). Time effects are assumed to vary smoothly over space (232). These models have been applied to lung cancer in Tuscany (232), and stomach cancer in Germany (224). A broader version of this model was proposed which incorporated age-area and age-time effects (233). However, the inclusion of cohort effects increases model complexity, and cohort effects in small areas may be tenuous, particularly if there are high rates of migration between areas which would dilute cohort by birthplace effects (233).

While methods for modelling spatio-temporal data have only transpired in the last few decades, a plethora of spatio-temporal models now exist and continue to grow in number. The complex nature of spatio-temporal data and the underlying processes that give rise to such data necessitates complex models. Bayesian hierarchical models are particularly well-suited for this task, as they provide a flexible way to describe and relate model parameters. The use of prior

distributions also makes it easy to account for spatial and/or temporal heterogeneity (i.e. autocorrelation and/or clustering), as well as uncertainty and expert knowledge (87, 215).

5. Recommendations

"Any approach to scientific inference which seeks to legitimize an *answer in response to complex uncertainty is, for me, a totalitarian parody of a would-be rational learning process."*

~ Adrian F. M. Smith, in (234)

This section summarises the issues and outlines some approaches for determining an appropriate method of analysis of spatial data in a given situation.

5.1 When should smoothing/modelling replace direct estimation?

If a raw, unsmoothed estimate possesses a sufficient level of reliability for the desired purpose then more detailed methods may not be necessary. Nonetheless, the definition of 'statistical reliability' varies between different agencies and countries, even when used for similar purposes (208). Often the suppression of estimates is dependent on both the underlying counts as well as the uncertainty in the estimate (47), and attempts to increase counts to sufficient levels may involve aggregating over the regions of interest.

The key advantages of smoothing/modelling are that rates can be stabilised at the resolution of interest, and noise in the rates resulting from differences in population size is reduced (47).

Waller and Gotway (47) suggested that smoothing should be considered when:

1. The addition of one event (disease case/death), or one more person at risk, results in a large difference (such as 25% or more) in at least one area's rates.
2. The number of events (rate numerator) is less than three for at least one area.
3. The population at risk per area is small (for instance, less than 500 people), and these numbers vary by an order of magnitude across the areas.

Even if the raw estimate meets confidentiality/reliability/precision guidelines, modelling is recommended when it is desirable to:

- Include covariates
- Understand the underlying pattern of risks.

Validation of results (either external or internal) is important regardless of the method chosen (126).

5.2 What type of smoothing/modelling should be used?

The accuracy of the method of smoothing – whether model-based or not – is critical. Areas of high and low risk should be correctly identified, while artificially elevated, unstable rates

should be reduced (235). No trends or patterns should be induced by the method. Uncertainty should also be quantified.

Building on suggested practical guidelines from Griffith (236) as well as Waller and Gotway (14) regarding the choice of spatial proximity:

1. Using any reasonable method for modelling spatial correlation is preferable to assuming the data are independent.
2. Exploratory spatial data analysis can ensure the choice of spatial dependence is supported by the data.
3. Comparing the results from several different types of spatial models is also useful.
4. Spatial correlation reduces the amount of information – the effective sample size. A very rough rule of thumb is to assume it will halve the information contained in the data. So, if 30 data values are needed assuming independent and identically distributed data, 60 correlated values should be used.
5. It is vital that the method used accounts for population heterogeneity.
6. Parsimony is still important. Choose the simplest model that adequately describes the data without compromising interpretation.

When the aim is to explore the data, simplicity, speed and ease of use is preferable (47). When the aim is to perform more detailed inferential analyses involving adjustment for confounders, hypothesis tests, and/or ranking of areas, Bayesian methods offer several advantages (47). Although no method perfectly compensates for small counts (237), some approaches perform better than others. Table 5.1 provides an overview of the main approaches discussed in this report.

Note that it is impossible to select the ideal model prior to examining the data. The amount of smoothing that occurs is dependent on both the model and the data (86). Generally, smaller counts will result in greater smoothing, and vice versa.

5.3 What methods should be used for a cancer atlas?

A cancer atlas may be purely descriptive or it may have a purpose or goal specific to a particular group of users. Therefore, the first step is to obtain input from potential end users to identify their requirements in using the maps (238). Specific methods may be better suited to different purposes, whether for providing an accurate overview, guiding further epidemiological studies, uncovering cancer hot-spots, or comparing regions.

However, in most situations we recommend the use of Bayesian hierarchical models for the reason that their output is useful in decision-making (239). A Bayesian model is able to rank estimates, compare between regions, and provide robust, reliable estimates with associated uncertainty (239). These models also have more flexibility in adjusting to changing purposes and aims.

Table 5.1 Summary of key methods

Method			Data privacy	Robust risk estimates	Can include covariates	Identifies high-risk areas	Quantifies uncertainty	Recommended for
Spatial correlation	Global	Moran's I						Exploratory/Significance of results
		Geary's C						Exploratory/Significance of results
		Tango's MEET						Exploratory/Significance of results
	Local	LISA						Exploratory
		SaTScan						Exploratory
Unsmoothed estimates		Count						Exploratory
		Crude rate						Exploratory
		ASR						Exploratory
		SMR						Exploratory
Direct smoothing		Locally-weighted average						Exploratory
		Locally-weighted median						Exploratory
		Kernel smoothers						Exploratory
Model-based smoothing		Poisson kriging						Final results
		EBLUP						Final results
		Empirical Bayes						Final results
	Fully Bayesian	BYM						Final results
		Anderson's spatial pattern & cluster model						Final results
		Mixture models						Final results
		Spatial partition models						Final results

Legend		Yes		Somewhat		No

5.4 Conclusion

There is no universal approach to analysing spatial data. The characteristics of the data, the presence of spatial correlation, and the purpose of the analysis are all important considerations.

In public health, spatial analyses are increasing in importance and popularity. Increasingly, decisions regarding resource and service allocation are influenced by mapped estimates. It is therefore vital that the small-area estimates used are robust, accurate and reliable.

Our recommendations are for unsmoothed estimates as well as directly smoothed estimates to be calculated and mapped as part of the exploratory data analysis. For producing final estimates, modelling that incorporates smoothing has many advantages.

References

1. Poland B, Lehoux P, Holmes D, Andrews G. How place matters: unpacking technology and power in health and social care. Health & Social Care in the Community. 2005;13(2):170-180.
2. Lope V, Pollan M, Perez-Gomez B, Aragones N, Ramis R, Gomez-Barroso D, *et al.* Municipal mortality due to thyroid cancer in Spain. BMC Public Health. 2006;6.
3. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environmental Health Perspectives. 2004;112(9):998-1006.
4. Shen W, Louis TA. Triple-goal estimates for disease mapping. Statistics in Medicine. 2000;19(17-18):2295-2308.
5. Lai P-C, So F-M, Chan K-W. Spatial Epidemiological Approaches in Disease Mapping and Analysis. Baton Rouge: CRC Press; 2008.
6. Ord JK. Spatial Autocorrelation: A Statistician's Reflections. In: Anselin L, Rey JS, editors. Perspectives on Spatial Data Analysis. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 165-180.
7. Ma H, Carlin BP, Banerjee S. Hierarchical and Joint Site-Edge Methods for Medicare Hospice Service Region Boundary Analysis. Biometrics. 2010;66(2):355-364.
8. Zhang J, Atkinson P, Goodchild MF. Lattice data and scale models. In: Zhang J, Atkinson P, Goodchild MF, editors. Scale in Spatial Information and Analysis. Boca Raton, FL: CRC Press; 2014.
9. Vázquez EF, Morollón FR. Preface. In: Vázquez EF, Morollón FR, editors. Defining the Spatial Scale in Modern Regional Analysis: New Challenges from Data at Local Level. Berlin: Springer; 2012.
10. Kang SY, McGree J, Baade P, Mengersen K. An investigation of the impact of various geographical scales for the specification of spatial dependence. Journal of Applied Statistics. 2014; 41(11):2515-2538.
11. Tobler WR. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography. 1970;46:234-240.
12. Fischer MM, Wang J. Spatial Data Analysis: Models, Methods and Techniques. Berlin: Springer; 2011.
13. Zandbergen PA. Geocoding Quality and Implications for Spatial Analysis. Geography Compass. 2009;3(2):647-680.
14. Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Chichester: John Wiley & Sons, Inc; 2004.
15. Demsar U, Harris P, Brunsdon C, Fotheringham AS, McLoone S. Principal component analysis on spatial data: An overview. Annals of the Association of American Geographers. 2013; 103(1):106-128.
16. Anselin L, Griffith DA. Do spatial effects really matter in regression analysis? Papers in Regional Science. 1988;65(1):11-34.
17. Choo L, Walker SG. A new approach to investigating spatial variations of disease. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2008;171(2):395-405.
18. Wakefield J, Elliott P. Issues in the statistical analysis of small area health data. Statistics in Medicine. 1999;18(17-18):2377-2399.
19. Moran PAP. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1-2):17-23.
20. Geary RC. The Contiguity Ratio and Statistical Mapping. The Incorporated Statistician. 1954; 5(3):115-146.
21. Anselin L. Local Indicators of Spatial Association—LISA. Geographical Analysis. 1995; 27(2):93-115.
22. Kulldorff M, Song C, Gregorio D, Samociuk H, DeChello L. Cancer map patterns: are they random or not? American Journal of Preventive Medicine. 2006;30(2 Suppl):S37-S49.
23. Tango T. A test for spatial disease clustering adjusted for multiple testing. Statistics in Medicine. 2000;19(2):191-204.
24. Kulldorff M. A spatial scan statistic. Communications in Statistics - Theory and Methods. 1997;26(6):1481-1496.

25. Sankey TT. Statistical Descriptions of Spatial Patterns. In: Shekhar S, Xiong H, editors. Encyclopedia of GIS. Boston, MA: Springer US; 2008. p. 1135-1141.
26. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. Statistics in Medicine. 1995;14(21-22):2323-2334.
27. Oyana TJ, Margai F. Spatial Analysis : Statistics, Visualization, and Computational Methods. Boca Raton: CRC Press; 2015.
28. Kulldorff M, Information Management Services Inc. SaTScan v9.4.2: Software for the spatial, temporal and space-time scan statistics. www.satscan.org/; 2015.
29. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. Statistics in Medicine. 2010;29(18):1910-1918.
30. Huang L, Kulldorff M, Gregorio D. A Spatial Scan Statistic for Survival Data. Biometrics. 2007;63(1):109-118.
31. Tango T. A Spatial Scan Statistic with a Restricted Likelihood Ratio. Japanese Journal of Biometrics. 2008;29(2):75-95.
32. Samarasundera E, Walsh T, Cheng T, Koenig A, Jattansingh K, Dawe A, *et al.* Methods and tools for geographical mapping and analysis in primary health care. Primary Health Care Research & Development. 2012;13(1):10-21.
33. Inskip H, Beral V, Fraser P, Haskey J. Methods for age-adjustment of rates. Statistics in Medicine. 1983;2(4):455-466.
34. Anselin L. Under the hood. Issues in the specification and interpretation of spatial regression models. Agricultural Economics. 2002;27(3):247-267.
35. Bavaud F. Models for Spatial Weights: A Systematic Look. Geographical Analysis. 1998; 30(2):153-171.
36. Cliff AD, Ord JK. Spatial Processes: Models and Applications. London: Pion; 1981.
37. Anselin L, Lozano N, Koschinsky J. Rate transformations and smoothing. Urbana, IL: Department of Geography, University of Illinois, 2006.
38. Earnest A, Morgan G, Mengersen K, Ryan L, Summerhayes R, Beard J. Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. International Journal of Health Geographics. 2007;6:54.
39. Lawson AB, Williams FLR. An Introductory Guide to Disease Mapping. New York: John Wiley & Sons, Ltd; 2001.
40. Kafadar K. Smoothing geographical data, particularly rates of disease. Statistics in Medicine. 1996;15(23):2539-2560.
41. Mausner JS, Kramer S. Epidemiology: An Introductory Text. Philadelphia: W. B. Saunders; 1985.
42. Krieger N, Williams DR. Changing to the 2000 Standard Million: Are Declining Racial/Ethnic and Socioeconomic Inequalities in Health Real Progress or Statistical Illusion? American Journal of Public Health. 2001;91(8):1209-1213.
43. Semenciw RM, Le ND, Marrett LD, Robson DL, Turner D, Walter SD. Methodological issues in the development of the Canadian Cancer Incidence Atlas. Statistics in Medicine. 2000; 19(17-18):2437-2449.
44. Banerjee S, Carlin BP, Gelfand AE. Hierarchical Modeling and Analysis for Spatial Data, Second Edition. Boca Raton, FL, USA: Chapman and Hall/CRC; 2014.
45. Kyriakidis PC. A Geostatistical Framework for Area-to-Point Spatial Interpolation. Geographical Analysis. 2004;36(3):259-289.
46. Cressie N. Smoothing Regional Maps Using Empirical Bayes Predictors. Geographical Analysis. 1992;24(1):75-95.
47. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, *et al.* Disease mapping models: an empirical evaluation. Statistics in Medicine. 2000;19(17-18):2217-2241.
48. Lawson AB. Statistical Methods in Spatial Epidemiology, Second Edition. West Sussex: Wiley; 2006.
49. Qiu P. Basic Statistical Concepts and Conventional Smoothing Techniques. In: Qiu P, editor. Image Processing and Jump Regression Analysis. Chichester, UK: John Wiley & Sons, Inc.; 2005. p. 13-54.

50. Mungiole M, Pickle LW, Simonson KH. Application of a weighted head-banging algorithm to mortality data maps. Statistics in Medicine. 1999;18(23):3201-3209.
51. Goovaerts P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. International Journal of Health Geographics. 2005;4:31.
52. Nadaraya EA. On Estimating Regression. Theory of Probability & Its Applications. 1964; 9(1):141-142.
53. Watson GS. Smooth regression analysis. Sankhya (Series A). 1964;26:359-372.
54. Gelman A, Hill J, Alvarez RM, Beck NL, Wu LL. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2006.
55. Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Second edition. Boca Raton: CRC Press; 2013.
56. Gelman A. Analysis of variance - why it is more important than ever. The Annals of Statistics. 2005;33(1):1-53.
57. Shao C. Approaches to the spatial modelling of cancer incidence and mortality in metropolitan Perth, Western Australia, 1990-2005. [PhD Thesis]. Perth, WA: Edith Cowan University; 2011.
58. Henderson R, Shimakura S, Gorst D. Modeling spatial variation in leukaemia survival data. Journal of the American Statistical Association. 2002;97:965-972.
59. Montero J-M, Fernández-Avilés G, Mateu J. Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. Chichester, UK: John Wiley & Sons, Ltd; 2015.
60. Asmarian NS, Ruzitalab A, Amir K, Masoud S, Mahaki B. Area-to-Area Poisson Kriging analysis of mapping of county- level esophageal cancer incidence rates in Iran. Asian Pacific Journal of Cancer Prevention. 2013;14(1):11-13.
61. Goovaerts P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. International Journal of Health Geographics. 2006;5(1):1-31.
62. Goovaerts P. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. Mathematical Geology. 2008;40(1):101-128.
63. Mockus A. Estimating Dependencies from Spatial Averages. Journal of Computational and Graphical Statistics. 1998;7(4):501-513.
64. Gotway CA, Young LJ. A geostatistical approach to linking geographically aggregated data from different sources. Technical report # 2004-012. Gainesville, FL: Department of Statistics, University of Florida, 2004.
65. Goovaerts P, Gebreab S. How does Poisson kriging compare to the popular BYM model for mapping disease risks? International Journal of Health Geographics. 2008;7:6.
66. Bolstad WM. Bayesian Inference for Poisson. Introduction to Bayesian Statistics. Hoboken, NJ: John Wiley & Sons, Inc.; 2007. p. 183-198.
67. Ghosh M, Natarajan K, Waller LA, Kim D. Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping. Journal of Statistical Planning and Inference. 1999; 75(2):305-318.
68. Devine OJ, Louis TA, Halloran ME. Empirical Bayes methods for stabilizing incidence rates before mapping. Epidemiology. 1994;5(6):622-630.
69. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. Biometrics. 1987;43(3):671-681.
70. Cressie N, Read TRC. Spatial data analysis of regional counts. Biometrics Journal. 1989;6: 699-719.
71. Ghosh JK, Delampady M, Samanta T. An Introduction to Bayesian Analysis: Theory and Methods. New York, NY: Springer; 2006.
72. Louis TA, Shen W. Innovations in Bayes and empirical Bayes methods: Estimating parameters, populations and ranks. Statistics in Medicine. 1999;18(17-18):2493-2505.
73. Maiti T. Hierarchical Bayes estimation of mortality rates for disease mapping. Journal of Statistical Planning and Inference. 1998;69(2):339-348.
74. Lahiri P, Rao JNK. Robust estimation of mean squared error of small area estimators. Journal of the American Statistical Association. 1995;90:758-766.

75. Meza JL. Empirical Bayes estimation smoothing of relative risks in disease mapping. Journal of Statistical Planning and Inference. 2003;112(1-2):43-62.
76. Mugglin AS, Cressie N, Gemmell I. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. Statistics in Medicine. 2002;21(18):2703-2721.
77. Carlin BP, Xia H. Assessing environmental justice using Bayesian hierarchical models: two case studies. Journal of Exposure Analysis and Environmental Epidemiology. 1999;9(1):66-78.
78. Dunson DB. Commentary: Practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology. 2001;153(12):1222-1226.
79. Thompson JA, Carozza SE, Zhu L. An evaluation of spatial and multivariate covariance among childhood cancer histotypes in Texas (United States). Cancer Causes & Control. 2007;18(1): 105-113.
80. Greco FP, Lawson AB, Cocchi D, Temples T. Some interpolation estimators in environmental risk assessment for spatially misaligned health data. Environmental and Ecological Statistics. 2005;12(4):379-395.
81. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. 1991;43:1-59.
82. Kim H, Sun DC, Tsutakawa RK. Lognormal vs. gamma: Extra variations. Biometrical Journal. 2002;44(3):305-323.
83. Bell BS, Broemeling LD. A Bayesian analysis for spatial processes with application to disease mapping. Statistics in Medicine. 2000;19(7):957-974.
84. Ocana-Riola R. The misuse of count data aggregated over time for disease mapping. Statistics in Medicine. 2007;26(24):4489-4504.
85. Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics. 2003;4(1):11-25.
86. Waller LA, Carlin BP. Disease mapping. In: Gelfand AE, Diggle PJ, Guttorp P, Fuentes M, editors. Handbook of spatial statistics. Boca Raton: CRC Press; 2010.
87. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Statistical Methods in Medical Research. 2005;14(1):35-59.
88. Anderson C, Lee D, Dean N. Bayesian cluster detection via adjacency modelling. Spatial and Spatiotemporal Epidemiology. 2016;16:11-20.
89. MacNab YC. On Gaussian Markov random fields and Bayesian disease mapping. Statistical Methods in Medical Research. 2011;20(1):49-68.
90. Kokki E, Ranta J, Penttinen A, Pukkala E, Pekkanen J. Small area estimation of incidence of cancer around a known source of exposure with fine resolution data. Occupational and Environmental Medicine. 2001;58(5):315-320.
91. Hughes J, Haran M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013;75(1):139-159.
92. Clayton DG, Bernardinelli L, Montomoli C. Spatial correlation in ecological analysis. International Journal of Epidemiology. 1993;22(6):1193-1202.
93. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran ME, Berry D, editors. Statistical models in epidemiology, the environment and clinical trials. New York: Springer; 2000. p. 135-178.
94. Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. Spatial and Spatiotemporal Epidemiology. 2011;2(2):79-89.
95. MacNab YC, Dean CB. Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. Statistics in Medicine. 2000;19(17-18):2421-2435.
96. Lu H, Reilly CS, Banerjee S, Carlin BP. Bayesian areal wombling via adjacency modeling. Environmental and Ecological Statistics. 2007;14(4):433-452.
97. White NM. Review of statistical methods for disease mapping. Available online: http://eprints.qut.edu.au/56859/. 2012.
98. Anderson C, Lee D, Dean N. Identifying clusters in Bayesian disease mapping. Biostatistics. 2014;15(3):457-469.
99. Lawson AB, Clark A. Spatial mixture relative risk models applied to disease mapping. Statistics in Medicine. 2002;21(3):359-370.

100. Denison DGT, Holmes CC. Bayesian partitioning for estimating disease risk. Biometrics. 2001;57(1):143-149.
101. Green PJ, Richardson S. Hidden Markov models and disease mapping. Journal of the American Statistical Association. 2002;97(460):1055-1070.
102. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives. 2004;112(9):1016-1025.
103. Hossain MM, Lawson A, B. Mixtures and Latent Structure in Spatial Epidemiology. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD, editors. Handbook of Spatial Epidemiology. Chapman & Hall/CRC Handbooks of Modern Statistical Methods: Chapman and Hall/CRC; 2016. p. 349-361.
104. Fernández C, Green PJ. Modelling spatially correlated data via mixtures: a Bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64(4): 805-826.
105. Cremers D. Image Segmentation with Shape Priors: Explicit Versus Implicit Representations. In: Scherzer O, editor. Handbook of Mathematical Methods in Imaging. New York, NY: Springer New York; 2015. p. 1909-1944.
106. Zach C, Häne C, Pollefeys M. What Is Optimized in Convex Relaxations for Multilabel Problems: Connecting Discrete and Continuously Inspired MAP Inference. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014;36(1):157-170.
107. Pereyra M, McLaughlin S. Fast unsupervised Bayesian image segmentation with adaptive spatial regularisation. E-print on Arxiv (arxiv.org/abs/1502.01400v3). 2016.
108. Bradley JR, Wikle CK, Holan SH. Bayesian Spatial Change of Support for Count-Valued Survey Data with Application to the American Community Survey. Journal of the American Statistical Association. 2015:1-43.
109. Flowerdew R, Green M. Areal interpolation and types of data. In: Fotheringham S, Rogerson P, editors. Spatial Analysis and GIS. London: Taylor and Francis; 1994. p. 121-145.
110. Dunson DB, Herring AH. Bayesian model selection and averaging in additive and proportional hazards models. Lifetime Data Analysis. 2005;11(2):213-232.
111. Omurlu IK, Ozdamar K, Ture M. Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets. Expert Systems with Applications. 2009; 36(8):11341-11346.
112. Banerjee S, Carlin BP. Semiparametric spatio-temporal frailty modeling. Environmetrics. 2003;14(5):523-535.
113. Sauleau EA, Hennerfeind A, Buemi A, Held L. Age, period and cohort effects in Bayesian smoothing of spatial cancer survival with geoadditive models. Statistics in Medicine. 2007;26(1):212-229.
114. Banerjee T, Chen MH, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. Lifetime Data Analysis. 2007;13(2):241-260.
115. Henderson R, Shimakura S, Gorst D. Modeling spatial variation in leukemia survival data. Journal of the American Statistical Association. 2002;97:965-972.
116. Zhang J, Lawson AB. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. Journal of Applied Statistics. 2011;38(2):591-603.
117. Li L, Hanson T, Zhang J. Spatial extended hazard model with application to prostate cancer survival. Biometrics. 2015;71(2):313-322.
118. Wang S, Zhang J, Lawson AB. A Bayesian normal mixture accelerated failure time spatial model and its application to prostate cancer. Statistical Methods in Medical Research. 2016;25(2): 793-806.
119. Diva U, Dey DK, Banerjee S. Parametric models for spatially correlated survival data for individuals with multiple cancers. Statistics in Medicine. 2008;27(12):2127-2144.
120. Fairley L, Forman D, West R, Manda S. Spatial variation in prostate cancer survival in the Northern and Yorkshire region of England using Bayesian relative survival smoothing. British Journal of Cancer. 2008;99(11):1786-1793.
121. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. Statistics in Medicine. 2004;23(1):51-64.

122. Royston P, Lambert PC. Flexible parametric survival analysis using Stata: beyond the Cox model. College Station, Texas: StataCorp LP; 2011.
123. Hennerfeind A, Held L, Sauleau EA. A Bayesian analysis of relative cancer survival with geoadditive models. Statistical Modeling. 2008;8(2):117-139.
124. Cramb SM, Mengersen KL, Lambert P, Ryan L, Baade PD. A flexible parametric approach to examining spatial variation in relative survival In: Cramb S, editor. Spatio-temporal modelling of cancer data in Queensland using Bayesian methods PhD by Publication. Brisbane: Queensland University of Technology; 2015.
125. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. Statistics in Medicine. 2007;26(30):5486-5498.
126. Rao JNK, Molina I. Small Area Estimation, Second Edition. New Jersey: John Wiley & Sons, Inc; 2015.
127. Pratesi M, Salvati N. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Statistical Methods and Applications. 2008;17(1):113-141.
128. Datta GS, Ghosh M. Bayesian prediction in linear models: applications to small area estimation. The Annals of Statistics. 1991;19(4):1748-1770.
129. Arora V, Lahiri P. On the superiority of the Bayesian method over the BLUP in small area estimation problems. Statistica Sinica. 1997;7(4):1053-1063.
130. BioMedware. Spacestat v4.0. Ann Arbor, MI: BioMedware, 2014.
131. AvRuskin GA, Jacquez GM, Meliker JR, Slotnick MJ, Kaufmann AM, Nriagu JO. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. International Journal of Health Geographics. 2004;3(1):26.
132. Konishi S. Introduction to Multivariate Analysis : Linear and Nonlinear Modeling. Bosa Roca: CRC Press; 2014.
133. LeSage JP, Pace RK. Introduction. In: LeSage JP, Pace RK, editors. Spatial and Spatiotemporal Econometrics. Amsterdam: Elsevier; 2004. p. 1-32.
134. Brooks S, Gelman A, Jones G, Meng X. Handbook of Markov Chain Monte Carlo. Boca Raton, FL: Chapman & Hall/CRC; 2011.
135. Cowles MK, Carlin BP. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. Journal of the American Statistical Association. 1996;91(434):883-904.
136. Gelfand AE. Gibbs sampling. Journal of the American Statistical Association. 2000;95(452): 1300-1304.
137. Howard RA. Dynamic Probabilistic Systems. Volume I: Markov models. Mineola, NY: Dover Publications, Inc; 2007.
138. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984;PAMI-6(6):721-741.
139. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association. 1990;85(410):398-409.
140. Geyer CJ. Introduction to Markov Chain Monte Carlo. In: Brooks S, Gelman A, Jones G, Meng X, editors. Handbook of Markov Chain Monte Carlo. Boca Raton, FL: Chapman & Hall/CRC; 2011.
141. Taylor BM, Diggle PJ. INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. Journal of Statistical Computation and Simulation. 2013;84(10):2266-2284.
142. Besag J, Green P, Higdon D, Mengersen K. Bayesian Computation and Stochastic Systems. Statistical Science. 1995;10(1):3-41.
143. Haran M. Gaussian Random field models for spatial data. In: Brooks S, Gelman A, Jones G, Meng X, editors. Handbook of Markov Chain Monte Carlo. Boca Raton, FL: Chapman & Hall/CRC; 2011.
144. Lunn DJ, Thomas A, Best N, Spiegelhalter DJ. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000;10:325-337.
145. Stan Development Team. Stan Modeling Language User's Guide and Reference Manual. Version 2.6.1. 2015.

146. Browne WJ. MCMC Estimation in MLwiN v2.1. Bristol: Centre for Multilevel Modelling, University of Bristol, 2009.
147. Plummer M. rjags: Bayesian Graphical Models using MCMC. R package version 4-6, 2016.
148. Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo Package. R package version 1.3-4: 2014.
149. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian Computing with INLA: A Review. arXiv:1604.00860v1, 2016.
150. De Smedt T, Simons K, Van Nieuwenhuyse A, Molenberghs G. Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. Archives of Public Health. 2015; 73(Suppl 1):O2-O2.
151. Bivand RS, Pebesma E, Gómez-Rubio V. Areal Data and Spatial Autocorrelation. In: Bivand RS, Pebesma E, Gómez-Rubio V, editors. Applied Spatial Data Analysis with R. New York: Springer; 2008. p. 237-272.
152. Sadeq M. Spatial patterns and secular trends in human leishmaniasis incidence in Morocco between 2003 and 2013. Infectious Diseases of Poverty. 2016;5:48.
153. Chaikaew N, Tripathi NK, Souris M. Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. International Journal of Health Geographics. 2009;8:36.
154. Cramb SM, Mengersen KL, Baade PD. Atlas of Cancer in Queensland: geographical variation in incidence and survival, 1998 to 2007. Brisbane: Viertel Centre for Research in Cancer Control, Cancer Council Queensland, 2011.
155. Ozdenerol E, Williams BL, Kang SY, Magsumbol MS. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. International Journal of Health Geographics. 2005;4(1):1-10.
156. Rey SJ, Janikas MV. STARS: Space–Time Analysis of Regional Systems. Geographical Analysis. 2006;38(1):67-86.
157. Nguyen P, Brown PE, Stafford J. Mapping Cancer Risk in Southwestern Ontario with Changing Census Boundaries. Biometrics. 2012;68(4):1228-1237.
158. Baade PD, Fritschi L, Aitken JF. Geographical differentials in cancer incidence and survival in Queensland: 1996 to 2002. Brisbane: Viertel Centre for Research in Cancer Control, Queensland Cancer Fund, 2005.
159. Cramb SM, Mengersen KL, Baade PD. Developing the atlas of cancer in Queensland: methodological issues. International Journal of Health Geographics. 2011;10:9.
160. Congdon P. Bayesian models for area health and mortality variations: an overview [ppt presentation]. Lodon, UK: Centre for Statistics and Department of Geography, Queen Mary University of London, 2007.
161. Costain DA. Bayesian partitioning for mapping disease risk using a matched case-control approach to confounding. Biostatistics. 2013;14(1):99-112.
162. Osnes K, Aalen OO. Spatial smoothing of cancer survival: A Bayesian approach. Statistics in Medicine. 1999;18(16):2087-2099.
163. Moss S. General Principles of Cancer Screening. In: Chamberlain J, Moss S, editors. Evaluation of Cancer Screening. London: Springer; 1996. p. 1-13.
164. Kreuger FA, van Oers HA, Nijs HG. Cervical cancer screening: spatial associations of outcome and risk factors in Rotterdam. Public Health. 1999;113(3):111-115.
165. Lofters AK, Gozdyra P, Lobb R. Using geographic methods to inform cancer screening interventions for South Asians in Ontario, Canada. BMC Public Health. 2013;13(1):1-8.
166. Anselin L. Local indicators of spatial association - LISA. Geographical Analysis. 1995;27.
167. Fernandes KA, Sutradhar R, Borkhoff CM, Baxter N, Lofters A, Rabeneck L, *et al.* Small-area variation in screening for cancer, glucose and cholesterol in Ontario: a cross-sectional study. CMAJ Open. 2015;3(4):E373-381.
168. Gwede CK, Ward BG, Luque JS, Vadaparampil ST, Rivers D, Martinez-Tyson D. Application of geographic information systems and asset mapping to facilitate identification of colorectal cancer screening resources. Online Journal of Public Health Informatics. 2010; 2(1):ojphi.v2i1.2893.
169. Le N, Marrett LD, Robson DL, Semenciw RM, Turner D, Walter SD. Canadian Cancer Incidence Atlas. Ottawa: Ministry of Supply and Services, 1995.

170. Nandakumar A, Gupta PC, Gangadharan P, Visweswara RN. Development of an Atlas of Cancer in India: First all India report 2001-2002. Bangalore: National Cancer Registry Programme (ICMR), 2004.
171. New York State Cancer Registry. Maps of Cancer Incidence by County. New York: New York State Department of Health; 2016. Available from: http://www.health.state.ny.us/statistics/cancer/registry/cntymaps/index.htm.
172. Penn State Health. Pennsylvania Cancer Atlas: a CDC/GeoVISTA prototype. Pennsylvania: Penn State GeoVISTA. Available from: http://www.geovista.psu.edu/grants/CDC.
173. SA Department of Health. The Geography of Cancer in South Australia,1991-2000. Adelaide: SA Department of Health, 2005.
174. Swedish Oncological Centres. Atlas of Cancer Incidence in Sweden. Stockholm: Swedish Oncological Centres, 1995.
175. University of New Hampshire's Institute for Health Policy and Practice. MapNH Health: Insights for our future. Available at: mapnhhealth.org/. Concord, NH: University of New Hampshire, 2014.
176. Devesa SS, Grauman DJ, Blot WJ, Pennello G, Hoover RN, Fraumeni JFJ. Atlas of cancer mortality in the United States, 1950-94. Washington: US Government Printing Office, 1999.
177. Goovaerts P. Geostatistical Analysis of Health Data: State-of-the-Art and Perspectives. In: Soares A, Pereira MJ, Dimitrakopoulos R, editors. geoENV VI – Geostatistics for Environmental Applications: Proceedings of the Sixth European Conference on Geostatistics for Environmental Applications. Dordrecht: Springer Netherlands; 2008. p. 3-22.
178. Rainey JJ, Omenah D, Sumba PO, Moormann AM, Rochford R, Wilson ML. Spatial clustering of endemic Burkitt's lymphoma in high-risk regions of Kenya. International Journal of Cancer. 2007;120(1):121-127.
179. Militino AF, Ugarte MD, Dean CB. The use of mixture models for identifying high risks in disease mapping. Statistics in Medicine. 2001;20(13):2035-2049.
180. Martuzzi M, Comba P, De Santis M, Iavarone I, Di Paola M, Mastrantonio M, *et al.* Asbestos-related lung cancer mortality in Piedmont, Italy. American Journal of Industrial Medicine. 1998;33(6):565-570.
181. Lu WS, Tsutakawa RK. Analysis of mortality rates via marginal extended quasi-likelihood. Statistics in Medicine. 1996;15(13):1397-1407.
182. Chellini E, Gorini G, Martini A, Giovannetti L, Costantini AS. Lung cancer mortality patterns in women resident in different urbanization areas in central Italy from 1987-2002. Tumori. 2006; 92(4):271-275.
183. Sandor J, Kiss I, Farkas O, Ember I. Association between gastric cancer mortality and nitrate content of drinking water: Ecological study on small area inequalities. European Journal of Epidemiology. 2001;17(5):443-447.
184. Prieto RR, Garcia-Perez J, Pollan M, Aragones N, Perez-Gomez B, Lopez-Abente G. Modelling of municipal mortality due to haematological neoplasias in Spain. Journal of Epidemiology and Community Health. 2007;61(2):165-171.
185. Lopez-Abente G, Hernandez-Barrera V, Pollan M, Aragones N, Perez-Gomez B. Municipal pleural cancer mortality in Spain. Occupational and Environmental Medicine. 2005;62(3):195-199.
186. Ghosh P, Huang L, Yu BB, Tiwari RC. Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data. Computational Statistics & Data Analysis. 2009;53(12):4073-4082.
187. Biggeri A, Bonannini M, Catelan D, Divino F, Dreassi E, Lagazio C. Bayesian ecological regression with latent factors: Atmospheric pollutants emissions and mortality for lung cancer. Environmental and Ecological Statistics. 2005;12(4):397-409.
188. Thomas AJ, Carlin BP. Late detection of breast and colorectal cancer in Minnesota counties: an application of spatial smoothing and clustering. Statistics in Medicine. 2003;22(1):113-127.
189. Parodi S, Stagnaro E, Casella C, Puppo A, Daminelli E, Fontana V, *et al.* Lung cancer in an urban area in Northern Italy near a coke oven plant. Lung Cancer. 2005;47(2):155-164.
190. Parodi S, Vercelli M, Stella A, Stagnaro E, Valerio F. Lymphohaematopoietic system cancer incidence in an urban area near a coke oven plant: an ecological investigation. Occupational and Environmental Medicine. 2003;60(3):187-194.

191. Knorr-Held L, Raβer G, Becker N. Disease mapping of stage-specific cancer incidence data. Biometrics. 2002;58(3):492-501.
192. Rachet B, Coleman MP. Cancer survival indicators for the National Health Service in England: exploration of alternative geographic units of analysis – the Primary Care Organisation and Strategic Health Authority. London: London School of Hygiene and Tropical Medicine, 2004.
193. Huang L, Pickle LW, Stinchcomb D, Feuer EJ. Detection of spatial clusters: application to cancer survival as a continuous outcome. Epidemiology. 2007;18(1):73-87.
194. Chen DG, Lio YL. Comparative studies on frailties in survival analysis. Communications in Statistics - Simulation and Computation. 2008;37(8):1631-1646.
195. Hennerfeind A, Brezger A, Fahrmeir L. Geoadditive survival models. Journal of the American Statistical Association. 2006;101:1065-1075.
196. Li Y, Brown P, Rue H, al-Maini M, Fortin P. Spatial modelling of lupus incidence over 40 years with changes in census areas. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2012;61(1):99-115.
197. Cramb SM, Mengersen KL, Turrell G, Baade PD. Spatial inequalities in colorectal and breast cancer survival: premature deaths and associated factors. Health Place. 2012;18(6):1412-1421.
198. Hsieh JCF, Cramb SM, McGree JM, Dunn NAM, Baade PD, Mengersen KL. Does geographic location impact the survival differential between screen- and interval-detected breast cancers? Stochastic Environmental Research and Risk Assessment. 2016;30(1):155-165.
199. Hsieh JC-F, Cramb SM, McGree JM, Baade PD, Dunn NAM, Mengersen KL. Bayesian Spatial Analysis for the Evaluation of Breast Cancer Detection Methods. Australian & New Zealand Journal of Statistics. 2013;55(4):351-367.
200. Walsh CD, Mengersen KL. Ordering of Hierarchies in Hierarchical Models: Bone Mineral Density Estimation. In: Alston CL, Mengersen KL, Pettitt AN, editors. Case Studies in Bayesian Statistical Modelling and Analysis. Chichester, UK: John Wiley & Sons, Ltd; 2012. p. 159-170.
201. Papaspiliopoulos O, Roberts GO, Sköld M. A general framework for the parametrization of hierarchical models. Statistical Science 2007;22(1):59-73.
202. Turrell G, Kavanagh A, Draper G, Subramanian SV. Do places affect the probability of death in Australia? A multilevel study of area-level disadvantage, individual-level socioeconomic position and all-cause mortality, 1998-2000. Journal of Epidemiology and Community Health. 2007;61(1):13-19.
203. Robertson C, Mazzetta C, D'Onofrio A. Regional variation and spatial correlation. In: Boyle P, Smans M, editors. Atlas of cancer mortality in the European Union and the European Economic Area 1993-1997 IARC Scientific Publication No 159. Lyon, France: International Agency for Research on Cancer; 2008.
204. Louie MM, Kolaczyk ED. Multiscale detection of localized anomalous structure in aggregate disease incidence data. Statistics in Medicine. 2006;25(5):787-810.
205. National Rural Health Alliance, Australian Council of Social Service. A snapshot of poverty in rural and regional Australia. Canberra: NRHA and ACOSS, 2013.
206. Ntzoufras I. Bayesian Modeling Using WinBUGS. Hoboken, NJ: John Wiley & Sons, Inc.; 2008.
207. Congdon P. Bayesian methods and Bayesian estimation. Applied Bayesian Modelling, Second edition. Chichester: John Wiley & Sons, Ltd; 2014. p. 1-33.
208. Hindmarsh DM. Small area estimation for health surveys [PhD thesis]. Wollongong: School of Mathematics and Applied Statistics, University of Wollongong; 2013.
209. Pfeffermann D. New important developments in small area estimation. Statistical Science. 2013;28(1):40-68.
210. Chambers R, Clark R. An Introduction to Model-Based Survey Sampling with Applications. Oxford: Oxford University Press; 2012.
211. Fay RE, Herriot RA. Estimates of income for small places: An application of James-Stein procedures to census data. Journal of the American Statistical Association. 1979;74:269–277.
212. Bell WR. Accounting for uncertainty about variances in small area estimation. Bulletin of the International Statistical Institute. 1999;52nd session, Helsinki.
213. Gelman A. Struggles with survey weighting and regression modeling. Statistical Science. 2007;22(2):153-164.

214. Gelman A. Rejoinder: Struggles with survey weighting and regression modeling. 2007:184-188.
215. Cressie N, Wikle CK. Statistics for spatio-temporal data. Hoboken, NJ: Wiley; 2011.
216. Abellan JJ, Richardson S, Best N. Use of space-time models to investigate the stability of patterns of disease. Environmental Health Perspectives. 2008;116(8):1111-1119.
217. Christakos G, Lai JJ. A study of the breast cancer dynamics in North Carolina. Social Science & Medicine. 1997;45(10):1503-1517.
218. Marshall RJ. A review of methods for the statistical analysis of spatial patterns of disease. Journal of the Royal Statistical Society Series A (Statistics in Society). 1991;154(3):421-441.
219. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. Statistics in Medicine. 1995;14(21-22):2433-2443.
220. Songini M, Bernardinelli L, Clayton D, Montomoli C, Pascutto C, Ghislandi M, *et al.* The Sardinian IDDM study: 1. Epidemiology and geographical distribution of IDDM in Sardinia during 1989 to 1994. Diabetologia. 1998;41(2):221-227.
221. Assuncao RM, Reis IA, Oliveira CD. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. Statistics in Medicine. 2001;20(15): 2319-2335.
222. Bernardinelli L, Pascutto C, Best NG, Gilks WR. Disease mapping with errors in covariates. Statistics in Medicine. 1997;16(7):741-752.
223. Xia H, Carlin BP. Spatio-temporal models with errors in covariates: Mapping Ohio lung cancer mortality. Statistics in Medicine. 1998;17(18):2025-2043.
224. Schmid V, Held L. Bayesian extrapolation of space-time trends in cancer registry data. Biometrics. 2004;60(4):1034-1042.
225. Waller L, Carlin BP, Xia H, Gelfand AE. Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association. 1997;92(438):607-617.
226. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine. 2000;19(17-18):2555-2567.
227. Sun DC, Tsutakawa RK, Kim H, He ZQ. Spatio-temporal interaction with disease mapping. Statistics in Medicine. 2000;19(15):2015-2035.
228. Kim H, Sun DC, Tsutakawa RK. A bivariate bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. Journal of the American Statistical Association. 2001;96(456):1506-1521.
229. Kim H, Oleson JJ. A Bayesian dynamic spatio-temporal interaction model: An application to prostate cancer incidence. Geographical Analysis. 2008;40(1):77-96.
230. Bohning D, Dietz E, Schlattmann P. Space-time mixture modelling of public health data. Statistics in Medicine. 2000;19(17-18):2333-2344.
231. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. Statistics in Medicine. 1998;17(18):2045-2060.
232. Lagazio C, Biggeri A, Dreassi E. Age-period-cohort models and disease mapping. Environmetrics. 2003;14(5):475-490.
233. Congdon P. A model framework for mortality and health data classified by age, area, and time. Biometrics. 2006;62(1):269-278.
234. Smith AFM. Present Position and Potential Developments: Some Personal Views: Bayesian Statistics. Journal of the Royal Statistical Society Series A (General). 1984;147(2):245-259.
235. Kafadar K. Choosing among two-dimensional smoothers in practice. Computational Statistics and Data Analysis. 1994;18(4):419-439.
236. Griffith DA. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In: Arlinghaus SL, editor. Practical Handbook of Spatial Statistics. Boca Raton, FL: CRC Press; 1996. p. 65–82.
237. Jerrett M, Gale S, Kontgis C. Spatial Modeling in Environmental and Public Health Research. International Journal of Environmental Research and Public Health. 2010;7(4):1302-1329.
238. Pickle LW. A history and critique of U.S. mortality atlases. Spatial and Spatiotemporal Epidemiology. 2009;1(1):3-17.

239. Kang SY, Cramb SM, White NM, Ball SJ, Mengersen KL. Making the most of spatial information in health: a tutorial in Bayesian disease mapping for areal data. Geospatial Health. 2016;11(2):428.
240. Rytkönen MJP. Not all maps are equal: GIS and spatial analysis in epidemiology. International Journal of Circumpolar Health. 2004;63(1):9-24.
241. Burrough PA, McDonnell R. Principles of Geographical Information Systems. Oxford: Oxford University Press; 1998.
242. Elliot P, Wakefield JC, Best NG, Briggs DJ. Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press; 2000.
243. English PB. An Introductory Guide to Disease Mapping. American Journal of Epidemiology. 2001;154(9):881-882.
244. López-Abente G, Aragonés N, García-Pérez J, Fernández-Navarro P. Disease mapping and spatio-temporal analysis: importance of expected-case computation criteria. Geospatial Health. 2014;9:27-35.
245. Catelan D, Biggeri A. Multiple testing in disease mapping and descriptive epidemiology. Geospatial Health. 2010;4:219-229.
246. Goovaerts P. Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. International Journal of Health Geographics. 2006;5:7.
247. International Agency for Research on Cancer. World Cancer Report 2014. Geneva: IARC, World Health Organization, 2014.
248. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2015;65(2):87-108.
249. Australian Institute of Health and Welfare. Cancer in Australia: an overview. Cat. no. CAN 88. Canberra: AIHW, 2014.
250. Wilkinson D, Cameron K. Cancer and Cancer Risk in South Australia: What Evidence for a Rural–Urban Health Differential? Australian Journal of Rural Health. 2004;12(2):61-66.
251. Woods LM, Rachet B, Coleman MP. Origins of socio-economic inequalities in cancer survival: a review. Annals of Oncology. 2006;17(1):5-19.
252. Ernst J, Zenger M, Schmidt R, Schwarz† R, Brähler E. [Medical and psychosocial care needs of cancer patients: a systematic review comparing urban and rural provisions]. Dtsch med Wochenschr. 2010;135(31/32):1531-1537.
253. Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF. Atlas of Cancer Mortality for US Counties: 1950-1969. Washington: US Govt. Printing Office, 1975.
254. Koch T. Disease Maps: Epidemics on the Ground. Chicago: University of Chicago Press; 2011.
255. Borrell C, Marí-Dell'Olmo M, Serral G, Martínez-Beneito M, Gotsens M. Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). Health & Place. 2010;16(4):703-711.
256. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. Biometrics. 1987;43.
257. Cressie N, Chan NH. Spatial Modeling of Regional Variables. Journal of the American Statistical Association. 1989;84(406):393-401.
258. Ancelet S, Abellan JJ, Del Rio Vilas VJ, Birch C, Richardson S. Bayesian shared spatial-component models to combine and borrow strength across sparse disease surveillance sources. Biometrical Journal. 2012;54(3):385-404.
259. Bernardo J, Smith A. Bayesian Theory. Chichester, UK: John Wiley & Sons, Ltd, 2000.
260. Johnson GD. Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. International Journal of Health Geographics. 2004; 3(1):29.
261. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. Third edition. Boca Raton, FL: Chapman & Hall/CRC; 2014.
262. Gurrin LC, Kurinczuk JJ, Burton PR. Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. Journal of Evaluation in Clinical Practice. 2000;6(2):193-204.

263. Wakefield J. Disease mapping and spatial regression with count data. Biostatistics. 2007;8(2): 158-183.
264. ApSimon HM, Warren RF, Kayin S. Addressing uncertainty in environmental modelling: a case study of integrated assessment of strategies to combat long-range transboundary air pollution. Atmospheric Environment. 2002;36(35):5417-5426.
265. Li Y, Brown P, Gesink DC, Rue H. Log Gaussian Cox processes and spatially aggregated disease incidence data. Statistical Methods in Medical Research. 2012;21(5):479-507.
266. Gotway CA, Young LJ. Combining Incompatible Spatial Data. Journal of the American Statistical Association. 2002;97(458):632-648.
267. Yu B. A Class of Transformation Covariate Regression Models for Estimating the Excess Hazard in Relative Survival Analysis. American Journal of Epidemiology. 2013;177(7):708-717.
268. Congdon P. Estimating diabetes prevalence by small area in England. Journal of Public Health. 2006;28(1):71-81.
269. Thomas DC. Statistical Methods in Environmental Epidemiology. Oxford: Oxford University Press; 2014.
270. Wakefield J. Sensitivity analyses for ecological regression. Biometrics. 2003;59(1):9-17.
271. Wakefield J. Ecological inference for 2 × 2 tables (with discussion). Journal of the Royal Statistical Society: Series A (Statistics in Society). 2004;167(3):385-445.
272. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin. 1995;118(3):392-404.
273. Browning CR, Cagney KA, Wen M. Explaining variation in health status across space and time: implications for racial and ethnic disparities in self-rated health. Social Science & Medicine. 2003;57(7):1221-1235.
274. Fahrmeir L, Kneib T. Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data. Oxford: Oxford University Press; 2011.
275. Wang F. Quantitative Methods and Applications in GIS. Boca Raton, FL: CRC Press; 2006.
276. Paciorek CJ. Spatial models for point and areal data using Markov random fields on a fine grid. Electronic Journal of Statistics. 2013;7:946-972.
277. Lawson AB, Browne WJ, Vidal Rodeiro CL. Disease Mapping with WinBUGS and MLwiN. Chichester: John Wiley & Sons, Ltd; 2004.
278. Clayton D, Bernardinelli L. Bayesian methods for mapping disease risk. In: Elliott P, Cuzick J, English D, Stern R, editors. Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford: Oxford University Press; 1996. p. 205-220.
279. Mollié A. Bayesian mapping of disease. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov Chain Monte Carlo in Practice. London: Chapman & Hall; 1996. p. 359–379.
280. Wakefield JC, Best NG, Waller LA. Bayesian approaches to disease mapping. In: Elliot P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press; 2001. p. 104–127.
281. Assunção R, Krainski E. Neighborhood Dependence in Bayesian Spatial Models. Biometrical Journal. 2009;51(5):851-869.
282. Junaidi, Stojanovski E, Nur D, editors. Prior sensitivity analysis for a hierarchical model. Proceedings of the Fourth Annual ASEARC Conference, 17–18 February 2011; 2011; University of Western Sydney, Paramatta, Australia.
283. Crainiceanu CM, Ruppert D, Wand MP. Bayesian Analysis for Penalized Spline Regression Using WinBUGS. Journal of Statistical Software. 2005;14:1-24.
284. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book: A Practical Introduction to Bayesian Analysis. Boca Raton, FL: Chapman and Hall/CRC Press; 2012.
285. Lykou A, Ntzoufras I. WinBUGS: a tutorial. Wiley Interdisciplinary Reviews: Computational Statistics. 2011;3(5):385-396.
286. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual, version 1.4. London: MRC Biostatistics Unit, Cambridge and Imperial College School of Medicine, 2003.
287. The R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012.

288. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009;71(2):319-392.
289. Schrödle B, Held L. A primer on disease mapping and ecological regression using INLA. Computational Statistics. 2011;26(2):241-258.
290. Schrödle B, Held L. Spatio-temporal disease mapping using INLA. Environmetrics. 2011;22(6):725-734.
291. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. Spatial and Spatiotemporal Epidemiology. 2013;4:33-49.
292. Rue H MS, Lindgren F. The R-INLA Project. www.r-inla.org. 2012.
293. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to Spatial Data Analysis. In: Fischer MM, Getis A, editors. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 73-89.
294. Bivand R, Altman M, Anselin L, Assunção R, Berke O, Bernat A, *et al.* spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.6-5. 2011.
295. Athens JK, Catlin BB, Remington PL, Gangnon RE. Using Empirical Bayes Methods to Rank Counties on Population Health Measures. Preventing Chronic Disease. 2013;10:E129.
296. Jaccard J, Becker MA, Wood G. Pairwise multiple comparison procedures: A review. Psychological Bulletin. 1984;96(3):589-596.
297. Brewer CA, Pickle L. Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. Annals of the Association of American Geographers. 2002;92(4):662-681.
298. Lawson AB, Biggeri AB, Böhning D, Lesaffre E, Viel J-F, Bertollini R, editors. Disease Mapping and Risk Assessment for Public Health. West Sussex: John Wiley & Sons; 1999.
299. McNamee R. Regression modelling and other methods to control confounding. Occupational and Environmental Medicine. 2005;62(7):500-506.
300. Statewide Health Service Strategy and Planning Unit. Cancer care services statewide health service strategy 2014. Brisbane: Queensland Health, 2014.

Appendix A Glossary

Asymptotic	Here, referring to large sample.
Bandwidth	In kernel smoothing, it refers to the maximum distance from an area that its influence is expected to extend. Beyond this, the kernel is set to zero (see Box 2.8).
Boundary effects	Areas on the 'edge' of the analysis region have fewer neighbours as neighbours beyond the boundary are excluded.
Choropleth map	Displays the value of interest on a set of regions within the study area.
Convergence	In MCMC analysis, the point at which it is reasonable to believe that samples are truly representative of the underlying stationary distribution of the Markov chain.
Covariance	A measure of the dependence between two random variables, and how they change together (see Box 2.7).
Covariate	In statistics, a covariate is a variable that is possibly predictive of the outcome under study. A covariate may be of direct interest or it may be a confounding or interacting variable.
Cross-validation	A model validation technique assessing how the results from a statistical analysis will generalise to an independent data set.
Direct method of standardisation	Apply stratum-specific rates observed in the populations of interest to a standard population. The ratio of two directly standardised rates is called the comparative incidence ratio (see Box 2.5).
Gaussian	An alternative term for the Normal distribution, which is a symmetrical bell-shaped curve.
Geostatistical data	Point-referenced data.
Gibbs sampling	See Box 2.20. A common form of MCMC sampling.
Hamiltonian Monte Carlo	Original name is Hybrid Monte Carlo, as this is a hybrid between traditional dynamical simulation and the Metropolis algorithm. Used by Stan software.

Hidden Markov model	Represents probability distributions over sequences of observations. Assumes that the state of the process generating the data is hidden, and that the current state is independent of all others except for the one immediately prior to it.
Hierarchical model	A model written in a hierarchical form or in terms of sub-models.
Hyperparameter	A parameter in a prior distribution.
Hyperprior distribution	A prior distribution on a hyperparameter, i.e., on a parameter of a prior distribution.
Incidence	A measure of the risk of developing a disease within a specified period of time.
Indirect method of standardisation	Apply stratum-specific reference rates to the populations of interest. The ratio of two indirectly standardised rates is called the SMR (see Box 2.6).
Kernel function	A kernel is a weighting function used in non-parametric estimation techniques. Common types of kernel functions include uniform, triangular, Gaussian, quadratic and cosine (see Box 2.8).
Likelihood	The probability of the evidence given the parameters. It is the probability of a given sample being randomly drawn regarded as a function of the parameters of the population.
Marginal mixture models	A model based on the assumption that the total space can be split into local regions where the responses come from the same distribution. Similar to spatial partition models.
Markov chain	A mechanism for generating plausible parameter value, whereby the value to be drawn depends on the previously drawn value.
Markov chain Monte Carlo (MCMC)	A class of algorithms for sampling from probability distributions by constructing a Markov chain that has the desired distribution as its equilibrium distribution (see Box 2.20).
Monte Carlo methods	A broad class of computational algorithms that use repeated random sampling to obtain numerical results.

Non-parametric model	The structure of the model is not fixed, but determined by the data. The number and nature of parameters are flexible. (Compare against parametric.)
Over-dispersion	In the statistical context, the presence of greater variability in a data set than expected using a given statistical model.
Parameter	A value used to represent a certain population characteristic which is usually unknown and therefore has to be estimated.
Parametric model	Assumes there is an underlying probability distribution based on a fixed set of parameters.
Posterior distribution	A probability distribution on the values of an unknown parameter that combines prior information about the parameter contained in the observed data to give a composite picture of the final judgements about the values of the parameter.
Predictor	A predictor variable is also known as an independent variable.
Prevalence	The number or proportion of cases or events or conditions in a given population.
Prior distribution	A probability distribution that represents the uncertainty about the parameter before the current data are examined.
Random effects	Effects that account for differences among the individual observational units in the sample, which are randomly sampled from the population. These effects usually conform to a specified distribution (typically a Normal distribution) and have a mean of zero.
Regression	A statistical technique for estimating the relationships among variables.
Relative survival	A standard estimate of net survival (measuring survival from the disease of interest) in population based disease survival studies (see Box 2.18).
Risk factors	An aspect of personal behaviour or lifestyle, an environmental exposure, or an inborn or inherited characteristic that is associated with an increased occurrence of disease or other health-related event or condition.

Semi-parametric model	A model containing parametric and nonparametric components. Often the nonparametric components are not of interest, such as the baseline hazard in the Cox proportional hazards model.
Sensitivity checks	These check the influence of model inputs (such as prior distributions) on the variation in model output.
Spatial partition models	A model based on the assumption that the total space can be split into local regions where the responses come from the same distribution. Similar to marginal mixture models.
Variance	A measure of how far values are spread out from their mean. The variance is the square of the standard deviation and the covariance of a random variable with itself (see Box 2.7).

Note: Several of these definitions were obtained from or based on the glossary in (239).

Appendix B Bayesian disease mapping tutorial

The following is from (239):

Making the most of spatial information in health: A tutorial in Bayesian disease mapping for areal data

Abstract

Disease maps are effective tools for explaining and predicting patterns of disease outcomes across geographical space, identifying areas of potentially elevated risk, and formulating and validating aetiological hypotheses for a disease. Bayesian models have become a standard approach to disease mapping in recent decades. This article aims to provide a basic understanding of the key concepts involved in Bayesian disease mapping methods for areal data. It is anticipated that this will help in interpretation of published maps, and provide a useful starting point for anyone interested in running disease mapping methods for areal data. The article provides detailed motivation and descriptions on disease mapping methods by explaining the concepts, defining the technical terms, and illustrating the utility of disease mapping for epidemiological research by demonstrating various ways of visualising model outputs using a case study. The target audience includes spatial scientists in health and other fields, policy or decision makers, health geographers, spatial analysts, public health professionals, and epidemiologists.

Introduction

Disease mapping is a flourishing field due to the growing amount of routinely collected health information worldwide (240). Advances in geographic information systems have greatly aided the analytical manipulation and visual representation of spatial data (241). Spatial information in health is especially useful for informing the locations of disease occurrences and the onus is on making the best possible use of this information.

Some excellent introductory guides for disease mapping are available in the literature. Nonetheless, many of these are either not intended for non-statistical audiences, or lack specific details. For instance, Elliot *et al.* (242) present a comprehensive review of the recent developments in spatial epidemiology but the statistical methods require a level of background knowledge which may not be suitable for beginners. Marshall (218) covers a broad range of methods for the analysis of the geographical distribution of disease, rather than upskill the reader in using particular methods. Lawson and Williams (39) provide a broad overview of the issues concerning disease mapping but is short on specifics (243). Banerjee *et al.* (44) presents a fully model-based approach to all types of spatial data, including point level, areal, and point pattern data. Cramb *et al.* (159) offer insight into the decisions made in generating a health atlas, but is not intended as an entry-level article for a

non-statistical audience. This article fills the niche by providing motivation, definition and description at a general level, and illustrating these ideas via a substantive case study. Although disease mapping has been undertaken in various forms for over 100 years, the opportunity now exists to use model-based maps that acknowledge uncertainty in inputs and outputs (244, 245), take account of the spatial nature of the data to 'borrow strength' from neighbouring areas in order to improve small area estimates, and can provide probability statements (246). In this article, we describe Bayesian disease mapping for areal data (39, 55) as an approach that addresses these issues. We focus on a running example of mapping cancer, although the methods are applicable to other diseases.

The primary purpose of this article is to provide a basic understanding of the key concepts involved in Bayesian statistical models for disease mapping of areal data. We commence with a discussion of why disease model-based mapping methods are required. Background on Bayesian methods typically used for disease mapping is then provided, and then some of the cartographic outputs commonly used are discussed, including methods for indicating statistical uncertainty in relative risk of disease.

Case Study: Cancer in Australia

Cancer is now the world's and Australia's biggest killer (247). The number of cases diagnosed continues to increase worldwide due to population growth and aging, with the increasing prevalence of physical inactivity, poor diet and reproductive changes (such as later parity) also contributing (248). In Australia, cancer accounts for almost one-fifth (19%) of the total disease burden (249).

Disparities in cancer outcomes across broad socioeconomic status and urban/rural categories have been reported internationally (250-252). Within Australia, there are disparities in cancer outcomes with respect to geographic remoteness and socioeconomic status (249). Cancers such as cervical and lung had higher incidence and mortality as remoteness or area-level disadvantage increased. Furthermore, the five-year relative survival from all cancers combined decreased with greater remoteness and greater socioeconomic disadvantage.

Understanding disparities in these broad areas, while useful, is unlikely to accurately reflect the heterogeneity in outcomes at the local level. Efforts to monitor and reduce cancer disparities can benefit greatly from quantifying variation across population groups and pertinent, small geographical areas. An understanding of the geographic patterns of cancer enables health decision-making by health service planners, clinicians, epidemiologists and industry groups to be more accurate and effective, for example by targeting policy development and resource allocation at areas of greater need (22, 253).

Cramb *et al.* (154) produced the first Atlas of Cancer in Queensland to describe geographical variation in cancer incidence and survival across small areas in Queensland, using routinely-collected health information from the Queensland Cancer Registry. For the first time,

Bayesian model-based cancer incidence and survival maps for Queensland were systematically presented at a comprehensive level. The Atlas significantly contributed to the understanding of geographical variation of cancer incidence and survival across Queensland, and subsequently influenced government policy decisions.

Methods

Disease maps are a visual representation of disease outcomes. The use of disease maps to aid decision making in epidemiological and medical research is well recognized (254). Disease maps are effective tools for explaining and predicting patterns of disease outcomes across geographical space, identifying areas of potentially elevated risk, and formulating and validating aetiological hypotheses for a disease (4). They are able to uncover local-level inequalities frequently masked by health estimates from large areas such as states, regions or cities (255), enabling the development of disease reduction and prevention programs targeting high-risk populations, see for instance, Mason *et al.* (253) and Kulldorff *et al.* (22) who have used cancer maps to depict the geographic patterns of cancer outcomes.

Disease mapping encompasses small area studies that use data aggregated over small areas and take into account local spatial correlation, see for example, Clayton and Kaldor (256); Cressie and Chan (257); Besag *et al.* (81) and Bernardinelli *et al.* (222). Data sparseness is common in small area analyses, especially when working with less common diseases. A small number of observed and expected disease occurrences leads to unstable risk estimates (258).

The problem of potentially unstable risk estimates for sparse spatial data needs to be mitigated to obtain reliable estimates. In practice, this is achieved by implementing spatial smoothing techniques. Spatial smoothing effectively "borrows strength" across small areas, so that the disease rate estimated for an area with a small population denominator would be weighted towards the estimated disease rate of neighbouring areas that have larger denominators. The estimates obtained by smoothing information from neighbouring areas are more reliable and robust due to the increased precision in the risk estimates in areas with few observations (258). In the context of disease mapping for small areas, the implementation of spatial smoothing is commonly achieved via the incorporation of a conditional autoregressive prior distribution for the spatial effects (see Lee (94) and the "Bayesian Spatial Statistical Models" section for details).

A disease mapping model is essentially a regression model that links a disease outcome to a set of risk factors. An important concept in disease mapping models (which is common to many other regression models) is the use of random effects. In this context, random effects provide a way of estimating variation in disease risk between areas that is not otherwise captured by known risk factors (e.g. age, sex, socioeconomic status, etc.).

Why Bayesian?

Bayesian statistics takes its name from the English clergyman Thomas Bayes (1702-1761), although the key concepts were also contemporaneously established by Laplace and embedded in the general view of 'inverse probability' at that time (259). It is an approach to data analysis that focuses on relating observed and unknown quantities using conditional probabilities, which are measures of the probability of an event given that another event has occurred.

In a Bayesian model (Box B.1), an unknown parameter is represented using a distribution rather than a single point estimate (260). The model parameters have distributions and are probabilistic (e.g. parameters representing coefficients associated with covariates in a regression model might be given a Normal distribution (Box B.2)). These distributions are known as prior distributions. These prior distributions can be considered as representing the uncertainty about the parameter before the data are seen. The parameters in the prior distributions (e.g. the mean and variance of the prior on a regression coefficient) can also have distributions which are known as hyperprior distributions. Again, these distributions also represent uncertainty about our knowledge of these values.

The combination of the prior information and the data results in a posterior distribution. The posterior distribution can be thought of as a probability distribution on the values of an unknown parameter that combines prior knowledge about the parameter and the observed data. The Bayesian model thus consists of parameters related to one another in the form of a hierarchy. The complex nature of spatial data can be captured using this hierarchical structure (4, 87).

Box B.1 Bayesian model

Given Bayes' theorem (261),

$$\boldsymbol{P(A|B) \propto P(A)P(B|A)}$$

The posterior distribution ($\boldsymbol{P(A|B)}$) is proportional to the prior distribution for parameters ($\boldsymbol{P(A)}$) multiplied by the data-based distribution given parameters (also known as the likelihood, $\boldsymbol{P(B|A)}$).

- Posterior estimates (model output) are a combination of the prior information and the data.
- Parameters in the model are assigned prior distributions.
- A prior distribution is the probability distribution that represents the uncertainty about the parameter before the current data are examined.
- Parameters in the prior distribution can also be assigned distributions.
- Parameters in the prior distribution (called 'hyperparameters') can also be assigned distributions.

Random effects are generally included in these models. Typically, a random effect is specified as being normally distributed, whereby a few areas are allowed to have a disease incidence much lower than expected based on these risk factors, a few areas much higher, but most are close to expected (following a bell curve). For spatial data, we assume that sites closer to each other are more similar, so we can use information from neighbouring sites to obtain better estimates of disease risk. Hence, when we fit a spatially-correlated random effect, the variation at a particular site is normally distributed relative to the mean of its neighbours. These random effects thus relate disease risk estimates to neighbouring estimates, producing a 'smoothing' effect across the area of interest.

Box B.2 Normal distribution

A distribution contains information on every possible observation and its associated probability. For instance, a Normal distribution is a continuous distribution that is "bell-shaped", at which data are most likely to be distributed around the mean and are less likely to be farther away from the mean.

A Normal distribution is often specified in terms of its mean (μ) and variance (σ^2) and can be written in the form of $\text{Normal}(\mu, \sigma^2)$. A parameter can be assigned a Normal distribution with mean 0 and variance 100 which can be denoted as $\text{Parameter} \sim \text{Normal}(0,100)$.

Alternatively, instead of specifying the values $(0,100)$, uncertainty about these parameters can also be described probabilistically. For example, instead of specifying '100' for the variance, the prior distribution could be written as $\text{Normal}(\mu, \sigma_0^2)$, and μ set to 0 while σ_0^2 is described by another probability distribution. Here σ_0^2 is termed a hyperparameter and the distribution on σ_0^2 a hyperprior distribution.

There are many reasons why the Bayesian approach is a useful framework for disease mapping. Firstly, Bayesian smoothing methods produce robust and reliable estimation of health outcomes of interest in a small area, even when based on small sample sizes (258). Within these small areas, the sample sizes are sometimes too small to yield estimates with adequate precision and reliability. Bayesian smoothing techniques improve the estimation by using information from neighbouring areas.

Secondly, the use of prior distributions (usually based on existing knowledge or expert opinion) in disease mapping models helps strengthen inferences about the true value of the parameter and ensures that all relevant information is included (262). These can be 'uninformative' (e.g. set to be Normally distributed with a mean of zero and a very large variance) or 'informative' if there is other information about the effect of this risk factor (given the other risk factors in the model). Thirdly, the Bayesian approach allows for quantification of the uncertainty related to the health estimates from the posterior distributions (67, 263). Spatial uncertainties added to the resulting risk maps depict local

details of the spatial variation of the risk and provide valuable information for policy makers to make decisions about thresholds and public health (246, 260, 264).

Lastly, direct probabilistic statements can be made about the underlying and unobserved parameters of interest using their posterior probability distributions. In disease mapping, it might be of interest to make probability statements about areas of high risk for a disease. For instance, computing and mapping probabilities that the risk in an area exceeds certain thresholds can be done using the posterior probability distributions (101). This probability of exceedance can then be used to decide whether an area should be classified as having excess risk of a disease (102). It is straightforward to make these kind of statements in a Bayesian context, since they are directly obtained from the corresponding posterior distribution.

Box B.3 Selecting regional scale

Important questions to consider when deciding on an appropriate area scale to conduct the analysis include:

1. Is there a risk of patient confidentiality being compromised?
2. Are population data available at the same scale as disease occurrences?
3. Will boundaries change over time? If so, what options are possible for keeping your data consistent?
4. Is there a digital boundary file available?
5. Will areas have a practical and relevant interpretation?
6. How does the size of the areas compare relative to the spatial pattern of the variation? If there is a lot of variation in an environmental effect within areas, this will limit the scope to measure the effect.
7. How many areas will there be? This affects computational time.
8. Are some areas likely to have zero population? This is likely to cause difficulties in modelling and estimation, e.g., zero denominator causes difficulties when using a Poisson distribution.
9. What scale have other similar studies used?
10. What spatial scale is available for covariate data? If spatial variation that takes fixed effects into account is of interest, it is not necessary to have a spatial scale finer than the available covariate data.

Data

Often health data are only available with location data supplied as a small area (known as areal data), rather than a street address geocoded to a latitude/longitude point. Determining the most appropriate region size to use involves several considerations (Box B.3). This article

focuses on the application of disease mapping methods for areal data aggregated over small areas and omits the discussion of other forms of spatial data such as geostatistical and point patterns data. As an alternative, health outcome data may also be analysed at the individual level, while incorporating spatial information at any geographical scale such as a point or an area.

The data described in the Atlas (154) focused on Queensland cancer data aggregated to the SLA level, which was the smallest area with annual population data available. However, consistent with most administrative regions, the areas are of varying sizes, and larger areas tend to dominate the map. An alternative approach is to aggregate disease data with continuous coordinate information to regular grid cells; see Li *et al.* (196, 265) and Kang *et al.* (10). Such an approach allows modelling of disease data at a fine spatial scale, independent of administrative boundaries while preserving patient confidentiality. Using this approach, the spatial scale can be manipulated to a practically, geographically and computationally sensible scale. It does, however, require individual level geocoded data, which may not be accessible due to confidentiality concerns. Spatial data may also be available at various geographical scales and hence there is a need to combine information from multiple sources (see Gotway and Young (266) for further details).

Box B.4 Data required to produce incidence estimates

Given a disease of interest, the information required to produce incidence estimates includes:

- Number of disease cases among people within a certain time period for each small area
- Estimated population counts by age group, sex, year and small area of residence − this is used as the denominator for calculating rates and for age-standardisation
- Geographical boundaries − this is used to compute the adjacency matrix required for spatial smoothing
- *Optional*: any desired small area level covariates (if available) such as rurality and socioeconomic status

Cramb *et al.* (154) mapped two health outcome measures in the Atlas, namely the incidence estimates and the relative survival estimates (discussed in the following Section). Incidence is a measure of the risk of developing a disease within a specified period of time. Relative survival is the standard measure of survival from a disease in population-based disease survival studies (267). Each of these outcomes require specific input data (refer to Boxes B.4 and B.5).

Although other estimates of disease, such as prevalence, are beyond the scope of this article, Bayesian mapping approaches are described in Congdon (268).

Box B.5 Data required to produce survival estimates

To produce relative survival estimates of a disease of interest, the input data required include:

- From the patients with the disease of interest (if not available for each individual then aggregated over each small area, any covariates and follow-up time intervals):
  - The observed number of deaths (from any cause) within a certain time period
  - Person-time at risk (the length of time between diagnosis and either death or censoring)
- General population mortality data used to calculate the expected number of deaths, which represents deaths due to causes other than the disease of interest for each small area, sex and broad age group
- Geographical boundaries − this is used to compute the adjacency matrix required for spatial smoothing
- *Optional*: individual or area-level covariates, including age, tumour stage, or area rurality and socioeconomic status

Bayesian Spatial Statistical Models

A response variable is the event studied and expected to vary whenever the independent variable is altered. It is also known as a dependent variable. Here we consider two response variables, namely the number of cancers diagnosed (incidence model) and the number of deaths within *x* years of a cancer diagnosis (relative survival model). Because both response distributions are counts, and the disease is less common, a Poisson distribution is used to model them (Box B.6).

Box B.6 Probability distributions used in epidemiology

For common diseases, the Binomial distribution models the number of disease occurrences in a sample size *n* from a population size *N*. The Binomial distribution is also commonly used in the analysis of disease prevalence data and case-control studies (269).

- When the disease is rare or less common (i.e., the probability of a disease is small), the Poisson distribution is used as an approximation to a Binomial distribution (270, 271). A Poisson distribution expresses the probability of a given number of events occurring in a fixed interval of time and/or space.
- For over-dispersed count distributions (where the data admit more variability than expected under the assumed distribution), a Negative Binomial distribution may be appropriate (272).
- For empirical data that show more zeroes than would be expected, zero-inflated models may be employed (272).

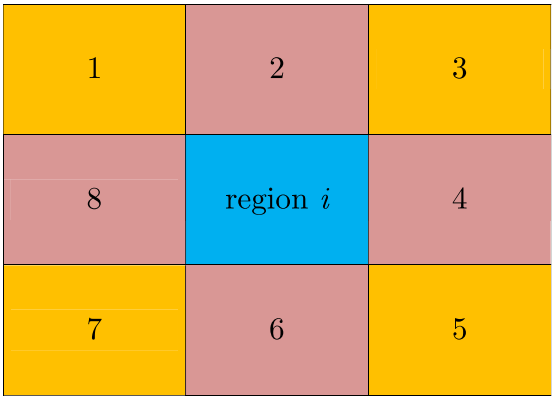
The resulting estimate for the incidence of a disease is known as the standardised incidence ratio or SIR, which is an estimate of relative risk within each area based on the population size, that compares the observed incidence against the expected incidence. The SIR explains if the observed incidence in a particular area is higher or lower than the average across all areas included, given the age and sex distribution and population size of the area.

The relative survival of a disease is modelled using an excess mortality model that contrasts the mortality in the background population with disease mortality. The survival model results in an excess hazard, which is called the relative excess risk (RER). The RER informs the relative survival of a disease within each area, by reporting the risk of death within a certain number of years of diagnosis after adjusting for broad age groups, compared to the average.

Small-area disease data typically exhibit spatial correlation due to spatial structure in the unknown risk factors. The presence of spatial correlation can be caused by a combination of socio-demographic clustering and environmental effects (273). Traditional regression models assume independence of random effects and so ignore the potential presence of spatial correlation. This may lead to false conclusions regarding covariate effects and unstable risk estimates (274).

The spatial correlation can be accounted for using spatial smoothing techniques, by estimating the effect of interest at a location using the effect values at nearby locations (275). Spatial smoothing approaches based on neighbourhood dependence are widely employed in disease mapping where areas with a common boundary are treated as neighbours (276). By accounting for the spatial correlation, model inference, prediction and estimation can be improved (143). The effect of the arbitrary geographical boundaries can also be reduced via spatial smoothing. Other smoothing techniques include interpolation methods, kernel regression, kriging and partition methods (61, 277).

Figure B.1 The representation of neighbourhood structure of area *i*.



Note: Based on the Rook method, neighbours for area *i* include areas 2, 4, 6 and 8, while the Queen method defines regions $1-8$ as neighbours of area *i*.

Two popular ways of defining a neighbourhood structure for the modelling of spatial correlation are the Queen definition and the Rook definition. The Rook method defines that two areas are considered neighbours if they share a common boundary whereas the Queen method specifies that two areas are termed neighbours if they share a common boundary or vertex. Following Earnest *et al.* (38), the illustration of these two methods for defining a neighbourhood structure is given in Figure B.1. Such information can be used to calculate the average of spatially correlated random effects of neighbours for area *i*.

The following Bayesian spatial models take the spatial correlation into account by incorporating spatially correlated random effects. Both the incidence and relative survival models assume a Poisson distribution for the observed data and contain spatial and unstructured (non-spatial) random effects. The well-known Bayesian BYM model (81) is widely used to model disease incidence (Box B.7) as it has desirable properties for disease mapping, particularly in modelling the geographical dependence between neighbouring areas (87). The incidence model can also be used to model mortality.

Box B.7 The incidence model

Given a set of *n* areas, the model for area *i* (*i* = 1,…, *n*) can be written as follows:

Observed counts in area *i* ~ Poisson(expected counts of area *i* × SIR of area *i*)

log(SIR of area *i*) = intercept term + coefficient × predictor variable vector for area *i* + spatial random effect of area *i* + unstructured random effect of area *i*.

Apply stratum-specific reference rates to the populations of interest.

The ratio of two indirectly standardised rates is called the SIR.

With regard to relative survival, the excess mortality can be modelled via a GLM, using exact survival times (121). The excess mortality is the mortality that is attributable to a particular disease. It is a measure of the deaths which occur over and above those that would be expected for a given population. Such a Bayesian relative survival model (Box B.8) has been used by Fairley *et al.* (120) and Cramb *et al.* (154). See Boxes 2.14 and 2.18 for the statistical models for incidence and relative survival, respectively.

In both models, the spatial random effect is the component that accounts for spatial correlation between neighbouring areas. The unstructured or non-spatial random effect accounts for the unexplained variation in the model.

In a Bayesian analysis, it is assumed that all parameters arise from a probability distribution. As such, distributions representing the likely spread of values are placed on each parameter. Commonly, a vague Normal distribution such as one with mean 0 and variance 1.0×10^6or Normal(0, 1.0×10^6) is used for the intercept or coefficients of predictor terms. Vague priors

refer to distributions with high spread, such as a Normal distribution with extremely large variance. Such a distribution gives similar prior value over a large range of parameter values.

Box B.8 The relative survival model

The model can be written as below, where for area *i*, follow-up interval *j*, and age group *k*,

$$\text{Number of deaths}_{ijk} \sim \text{Poisson}(\text{expected number of deaths}_{ijk})$$

$$\log(\text{expected number of deaths}_{ijk} - \text{expected number of deaths due to other causes}_{ijk}) = \log(\text{person time at risk}_{ijk}) + \text{intercept}_j + \text{coefficient}_k \times \text{predictor variable vector} + \text{spatial random effect of area}_i + \text{unstructured random effect of area}_i$$

Generally, the unstructured (non-spatial) random effects and the spatial random effects are both assigned a prior distribution with additional hyperparameters (Box B.9). To allow for spatial correlation, commonly an intrinsic conditional autoregressive (CAR) distribution is used. The CAR prior models the spatial dependence in a study region by effectively borrowing information from neighbouring areas than from distant areas and smoothing local rates toward local, neighbouring values. The method provides some shrinkage and spatial smoothing of the raw relative risk estimates (69). This results in a more stable estimate of the pattern of the underlying disease risk than that provided by the raw estimates. Consequently, the variance in the associated estimates is reduced and the spatial effect of geographical differences can be identified. This prior has been widely employed in disease mapping to study the geographical variation of disease risk (278-280), and works particularly well to smooth out variability not relevant to the underlying risk (281).

Box B.9 Prior distributions for the random effects

Unstructured

The unstructured random effects are assumed to follow a Normal distribution with mean zero and a hyperparameter for variance.

Unstructured random effect of area *i* ~ Normal(0, variance hyperparameter).

Spatial

The spatial random effects are assumed to follow a CAR prior (81) with some hyperparameters, as follows:

Spatial random effect of area *i* ~ Normal (average of spatial effects of neighbours of area *i*, variance hyperparameter / number of neighbours of area *i*).

Commonly, both of the precision (inverse of the variance) hyperparameters are assigned a Gamma distribution. Alternative hyperprior distributions may include placing either a Uniform or half-Normal distribution on the standard deviation (square root of the variance) (54).

The prior distributions used for the parameters may influence the results and therefore should be carefully considered and compared. There are two issues to consider when deciding on a prior distribution (54): (a) what information is going into the prior distribution; and (b) the impact on the resulting posterior distribution. A sensitivity analysis (282) can be used to investigate the dependence of the posterior distribution on prior distributions by comparing posterior inferences under different reasonable choices of prior distribution. A literature review is usually helpful to determine the prior distributions being used in similar Bayesian models.

Computation

The complexity of these models mean they cannot be solved analytically. Instead, some method of approximation is required. One approach is to use Markov chain Monte Carlo (MCMC) methods, which sample from the posterior distribution. A variety of software is available to conduct MCMC, including BUGS (Bayesian inference Using Gibbs Sampling), JAGS (Just Another Gibbs Sampler), Stan and BACC (Bayesian Analysis, Computation & Communication). WinBUGS is one of the most popular options (134) that provides great flexibility in Bayesian modelling, has a simple programming language (283) and interfaces with multiple statistical software, including R, Matlab, Stata and SAS. See Additional Information B.1 for the WinBUGS code for the discussed models. Some useful resources to help learn WinBUGS include Lawson *et al.* (277), Lunn *et al.* (284), Ntzoufras (206), Lykou and Ntzoufras (285), and Spiegelhalter (286).

Bayesian computation for the above models can also be conducted in R (287), by calling the `inla` program and adopting the integrated nested Laplace approximation (INLA) approach proposed by Rue *et al.* (288). The INLA approach performs Bayesian inference for spatial models and is able to return accurate parameter estimates in a much shorter time than MCMC. The use of R-INLA for statistical analysis in various disciplines is increasingly common in recent years, including disease mapping. Additional Information B.3 provides R-INLA code to perform computation for the discussed models. Some useful resources for getting started with R-INLA include Schrödle and Held (289, 290), Blangiardo *et al.* (291), and Rue *et al.* (292).

To incorporate neighbourhood dependence into the Bayesian models, a neighbourhood matrix is required. The neighbourhood matrix contains a list of neighbours for an area. Freely available software programs that will calculate a neighbourhood matrix include GeoDa (293), the spdep R package (294), or within WinBUGS.

Making Decisions

Perhaps the greatest advantage of Bayesian methods is the diversity of options available to assist in the decision making process. Communicating results in a way that is easily interpretable and accurate enables informed decisions to be made. Here we outline some of the ways modelled estimates can be used and visualized.

The SIR and RER estimates produced using the methods described in the previous sections are two commonly seen measures of disease risk. The estimates produced by Bayesian models give great flexibility in reporting results, including comparison of the risk estimates against the average, ranking estimates, and/or examining the uncertainty around the estimates.

Ranking of disease estimates ensures that public health investigations or interventions are prioritized correctly (4). In the Bayesian context, the posterior distributions of health outcome measures (such as SIR and RER) allow for the calculation of rank estimates of each area (47, 256). For instance, Athens *et al.* (295) use five health outcome measures to obtain county rank estimates for a composite health outcome measure. The five health outcome measures are converted to a score, and then ranked by weighted means. The ranking of health outcomes is useful for representing health performance of each area which can then be used to inform health decision making.

Moreover, comparison between two areas can be made easily in the Bayesian framework. Outside of Bayesian methods, it may be difficult and problematic to conduct a large number of pairwise comparisons for all areas using post-hoc tests (296). The problem is that by conducting so many comparisons, the probability of finding some of the differences statistically significant by chance alone increases. The Bayesian context eliminates this issue with pairwise comparisons of the posterior distributions.

Bayesian methods produce measures of uncertainty for each modelled estimate. The uncertainty attached to the spatial distribution of risk values across the study region can be known as spatial uncertainty (246). It is valuable to visualize spatial uncertainty as it provides local details of the spatial variation of the risk, as well as an input to resource allocation, management and policy strategies. Several methods have been proposed to describe the uncertainty attached to the smoothed rates, including mapping the 95% credible interval of the posterior distribution of smoothed rates (260) and the probability that the risk in each small area exceeds a certain threshold (102).

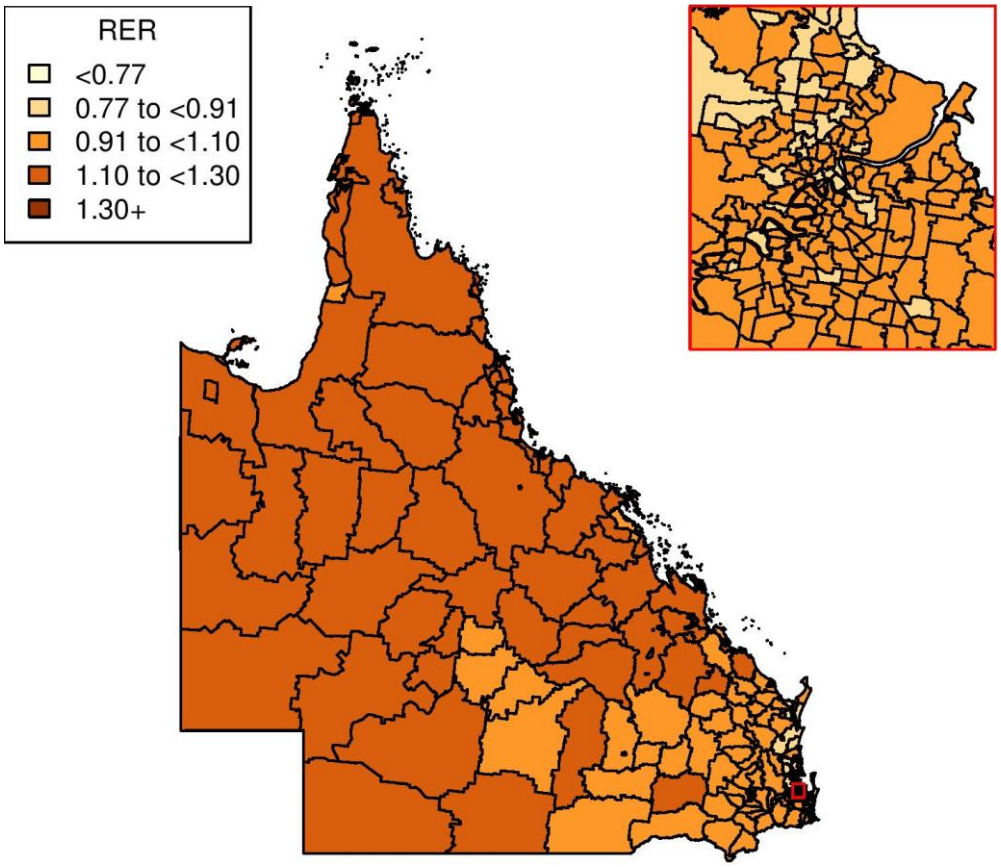
Under the Bayesian paradigm, there is great flexibility in communicating and visualising results. Options include maps or graphs of the smoothed estimates, their associated uncertainty, or the probabilities of being above/below certain values. Mapping of disease rates or outcomes facilitates comparison of spatial patterns in disease rates between males and females, between age groups, between races, over time, and motivates comparison with patterns of potential causes (297). By comparing disease rates of different areas, clues to

possible causation may be found and this serves as a starting point for further investigation.

The purpose of this Section is to showcase various visualisations that can be produced using the outputs obtained from Bayesian modelling techniques and the associated interpretation. This is demonstrated on a common cancer with poor survival: male lung cancer in Queensland. Figures B.2 to B.7 present an array of maps or plots based on the results from modelled survival (RER of death within 5 years of diagnosis) for each SLA that are useful for communicating the results of statistical analysis via the Bayesian paradigm. The RER expresses the risk of cancer patients dying from their cancer within five years of diagnosis in an SLA compared to the Queensland average (RER = 1), and therefore should not be directly compared between two SLAs. The figures were produced using `R` software, package `maptools`.

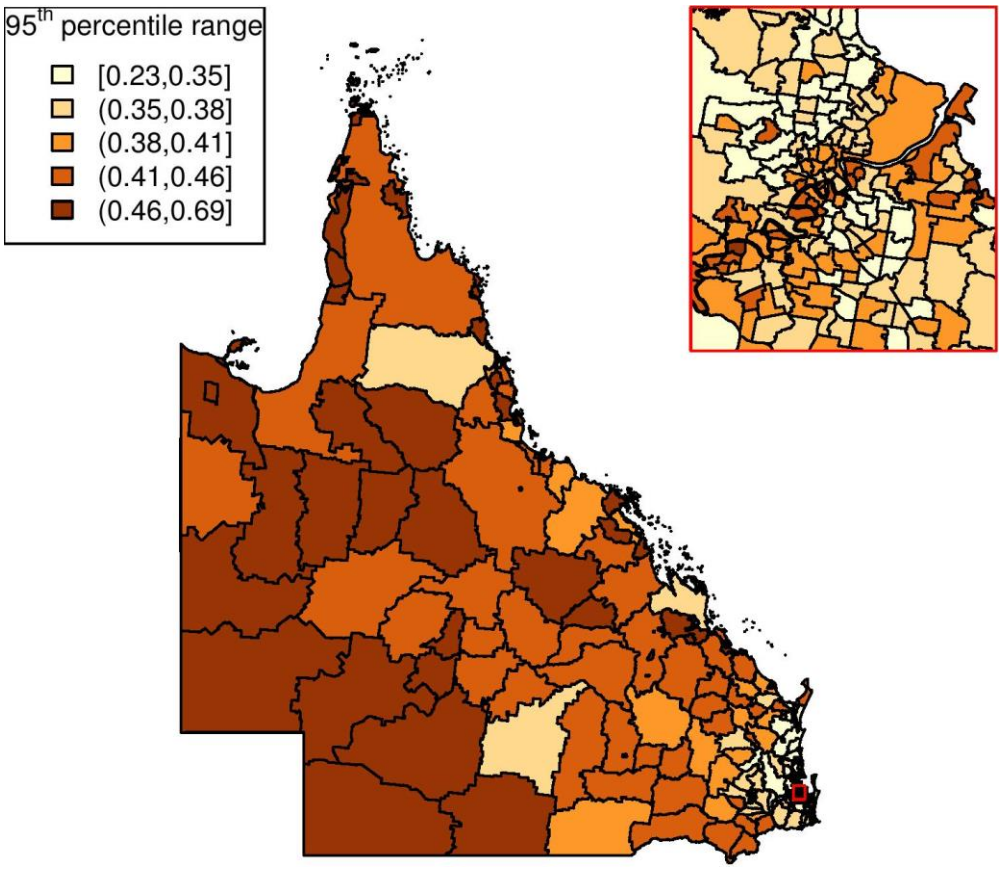
Figure B.2 maps the posterior distribution of SLA-level RER and provides a picture of the spatial pattern of the underlying risk. Figure B.3 depicts the uncertainty associated with the Bayesian estimates of RER by mapping the 95th percentile range of the 10,000 values sampled from the posterior distribution of RER for each SLA. A graph showing the ranked RER with the associated 95% credible interval for each SLA is provided in Figure B.4. Horizontal box plots of the RER estimates by socioeconomic status and rurality are provided in Figure B.5 to provide additional information about where the extent of variability across the Queensland state. Figure B.6 maps the SLAs having a 90% probability of RER being higher than the Queensland average (RER = 1) (highlighted in red) and the SLAs having at least a 90% probability of RER being lower than the Queensland average (RER = 1) (highlighted in blue). Figure B.7(a) depicts the probability of the SLAs having RER exceeding 1 and Figure B.7(b) depicts the probability of the SLAs having RER exceeding 1.2.

Figure B.2 Bayesian smoothed estimate of RER



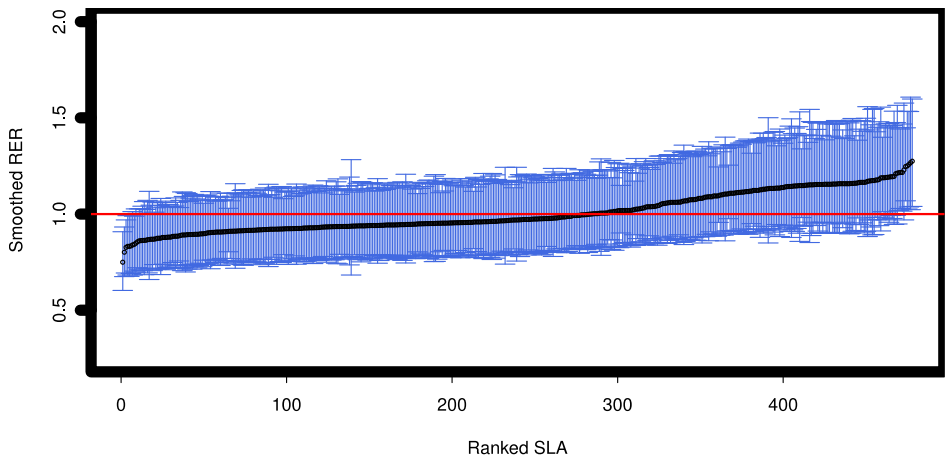
Notes: To show the spatial pattern of the underlying risk, the median of the posterior distribution of SLA-level RER is mapped. An inset of South-East Queensland is provided for greater detail as this region has a large number of SLAs. Thematic categories are based on fixed breaks method.

Figure B.3 Uncertainty of Bayesian smoothed estimate of RER



Notes: This map depicts the uncertainty associated with the estimates of relative risk. The 95^{th} percentile range (97.5 minus the 2.5 percentile) of the 10,000 values sampled from the posterior distribution of RER for each SLA is mapped here. An inset of South-East Queensland is provided for greater detail as this region has a large number of SLAs. Thematic categories are based on quintiles.

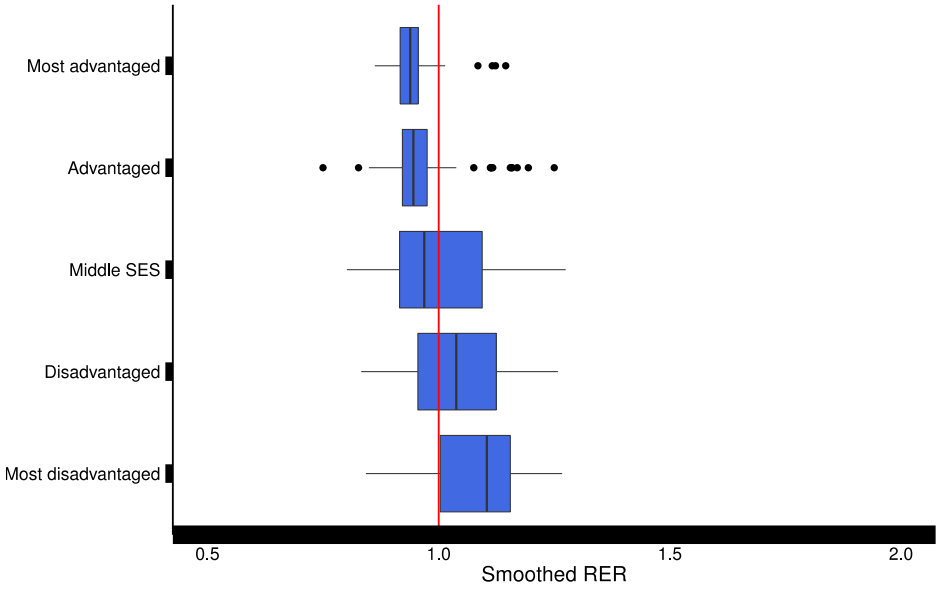
Figure B.4 Uncertainty of Bayesian smoothed estimate of RER



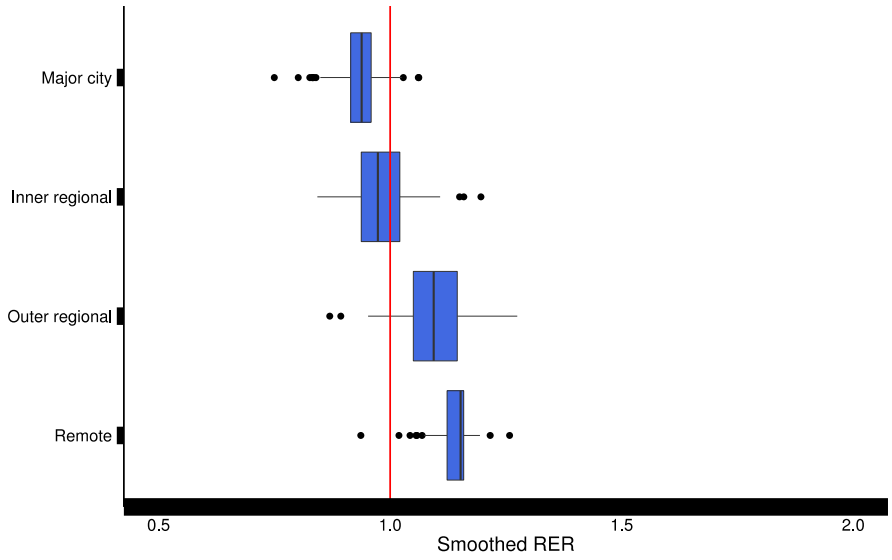
Notes: The 95% credible interval (97.5 − 2.5 percentile) of the 10,000 values sampled from the posterior distribution of RER for each SLA is plotted here. This plot shows how much reliance can be placed on the estimates. The black line is the median RER for each SLA. The blue vertical lines are the 95% credible intervals, and indicate the amount of uncertainty associated with each estimate. The red line shows the Queensland average (set to 1).

Figure B.5 Distribution of smoothed RER estimates according to (a) Socioeconomic status (b) Rurality

(a)

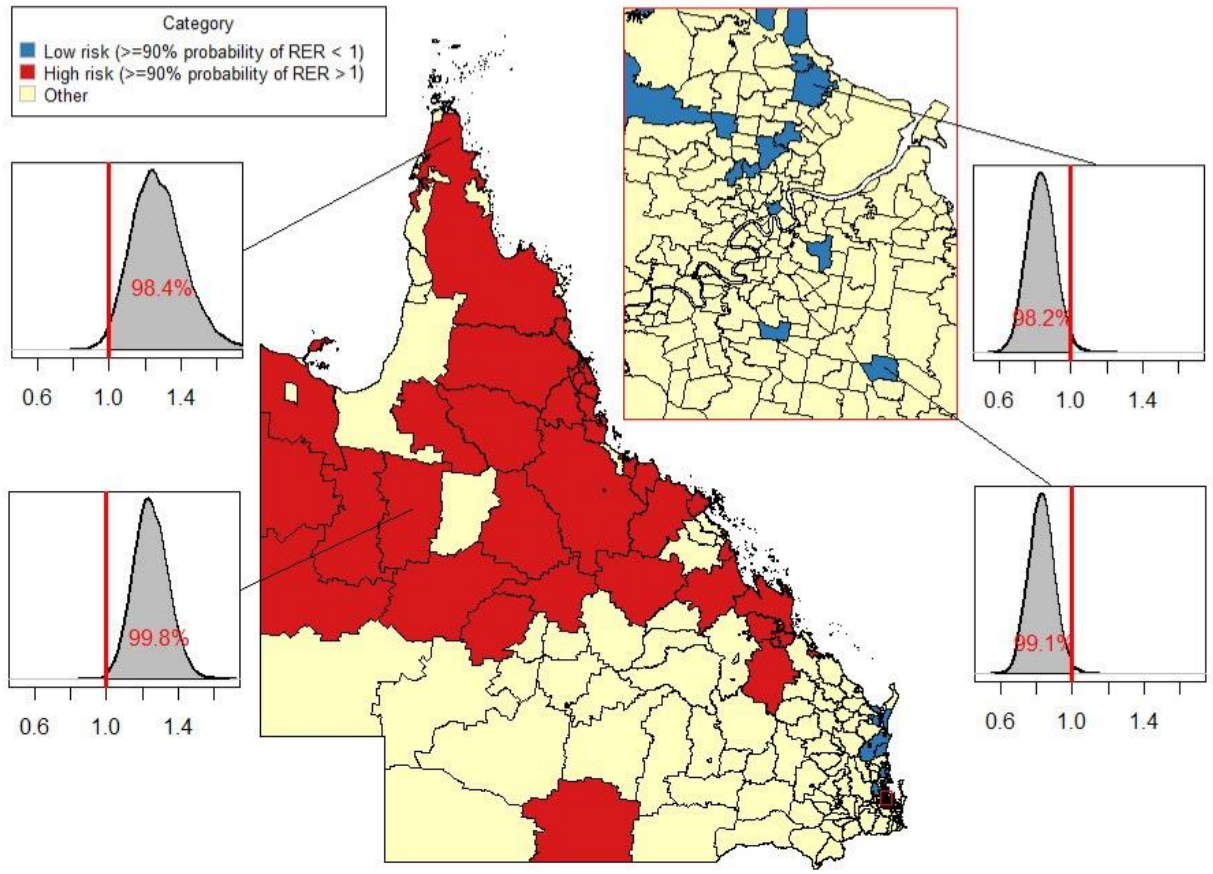


(b)



Notes: The distributional plots reflect the general patterns in the smoothed RER estimates across the area-based categories of socioeconomic status and rurality. These plots show the proportion of RER estimates that are above or below the Queensland average (vertical red line) within each of the area-based categories. The plots only present the range of point estimates, and so do not take the amount of uncertainty associated with each SLA-specific estimate into account.

Figure B.6 Using posterior probabilities to classify risk



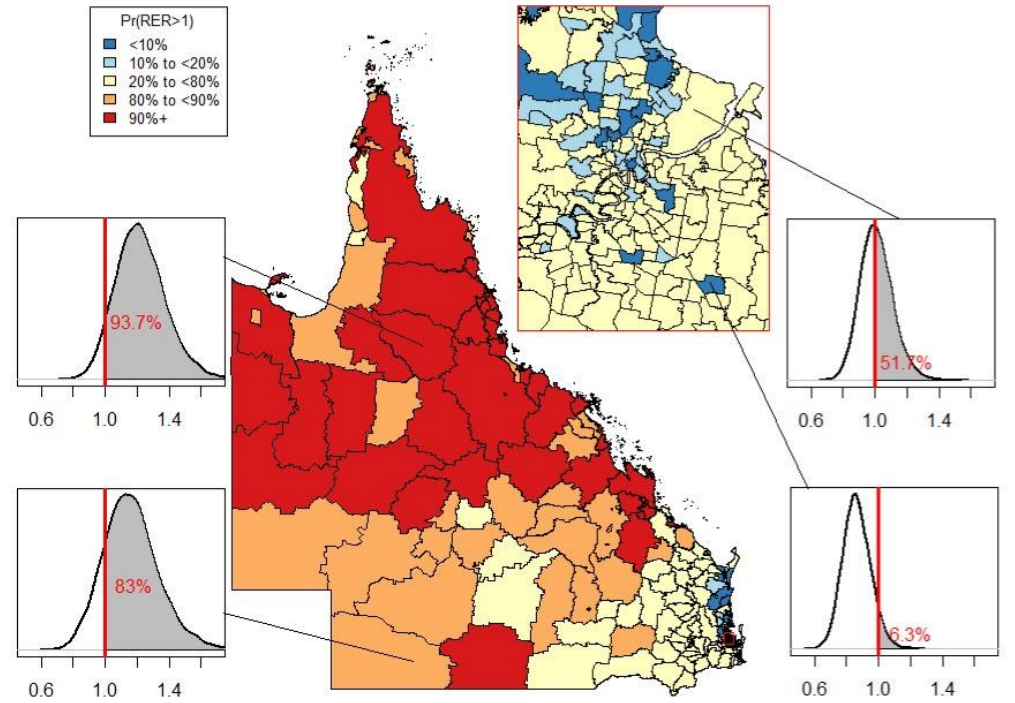
Notes: In the Bayesian paradigm, the SLAs highlighted in red have a 90% probability of RER being higher than the Queensland average (RER = 1). This means that the lower 10th percentile of the posterior distribution of RER exceeds 1. The SLAs highlighted in blue express at least a 90% probability of RER being lower than the Queensland average (RER = 1). This means that the upper 90th percentile of the posterior distribution of RER is less than 1. The density plots show the posterior distribution of RER for four randomly chosen SLAs where the x-axis is the RER values. The two density plots on the left show that there is more than 90% chance for the RER to be higher than 1. The two density plots on the right show that there is more than 90% chance for the RER to be lower than 1. The percentage of low risk or high risk for each SLA is also given in each density plot. An inset of South-East Queensland is provided for greater detail as this region has a large number of SLAs.

Discussion

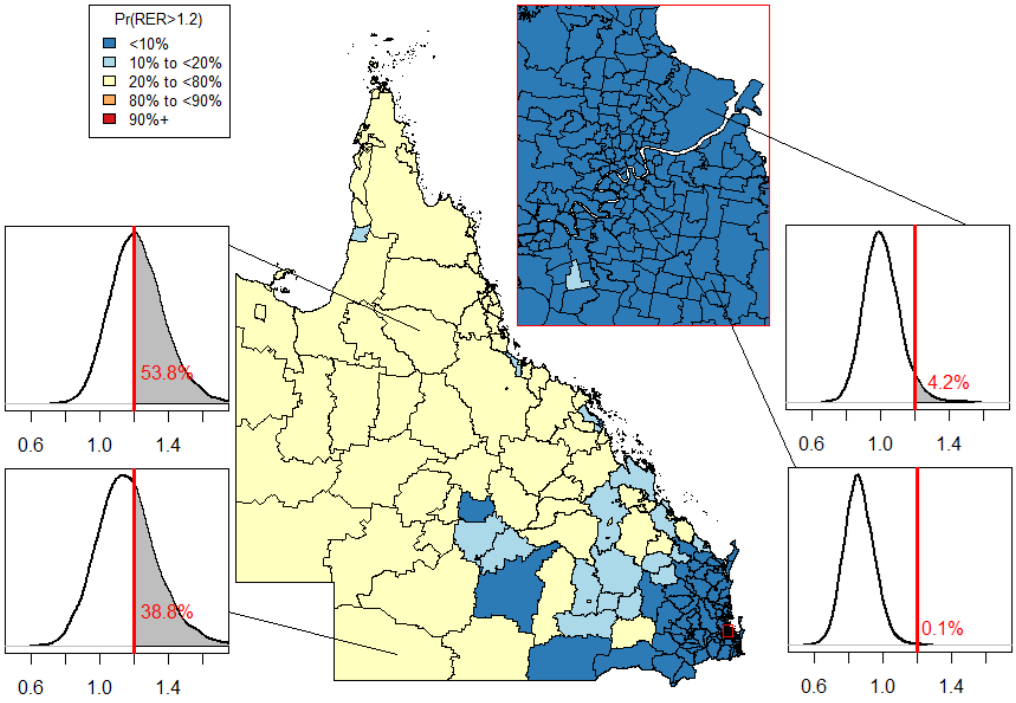
In this article we have outlined the benefits of Bayesian models for both analysis and visualization. The public health arena regularly makes practical decisions affecting people's health. To facilitate decisions, it is vital that the analysis is conducted appropriately, and results are communicated effectively.

Bayesian methods are increasingly being used to analyse routinely collected data. The Bayesian framework is now the tool of choice in many applied statistical areas, including disease mapping (298). In small area studies, Bayesian methods often have better model fit than non-Bayesian smoothing methods (47). Greater flexibility in distributional assumptions is possible under Bayesian methods than in traditional regression models (14).

Figure B.7 (a) Thematic map depicting the probability of RER exceeding 1, (b) Thematic map depicting the probability of RER exceeding 1.2



(a)



(b)

Notes: The threshold 1.2 was chosen to reflect high risk as it lies in the fifth quintile. Four SLAs are chosen to demonstrate how the probabilities change when the thresholds change. An inset of South-East Queensland is provided for greater detail as this region has a large number of SLAs.

Whether to standardise response rates depends on the study objectives. For the cancer atlas, it was desirable to remove the influence of age, so that differences were not due to different age structures between areas. For incidence, we used the standardised incidence ratio (SIR), which adjusts for the area-specific age and sex structure. An alternative method to standardisation for dealing with confounders is via the use of regression models (299). These can be particularly useful when multiple confounders need to be controlled for simultaneously. For relative survival, we included age in the regression equation to remove its influence on the results. However, if the purpose of a study is to identify where the highest rates of disease are, such as for service provision, then there is no need to standardise (or otherwise adjust) the incidence rates. This is because the cause of the variation (whether sex, age or other factors), is inconsequential.

Visualising disease patterns through maps remains an effective method to convey a large amount of information in an engaging way. Few modern day visualisations include uncertainty measures, yet this greatly assists in decision making. Online, interactive visualisations can dynamically link maps (e.g. Figure B.2 showing the smoothed Bayesian RER), with plots of the uncertainty (e.g. Figure B.3 showing the 95% credible interval for each area). Selecting an area would then highlight the corresponding region in both plots, providing much greater information to the user.

There are limitations associated with using routinely collected data. Determining the direction of causation may not be possible. Often there is a lag time between exposure and disease detection, and patients may move during this time. Bayesian methods also have certain limitations, including greater computational time if using Markov chain Monte Carlo approaches, and requiring sensitivity analyses to ensure priors are not exerting undue effect. With regard to computation using R-INLA, models must be expressible in the linear model format and there are restrictions on the types of prior distributions that can be assumed.

However, we believe the advantages outlined in this article outweigh any limitations. Routinely collected data exist to enable disease monitoring and control. Appropriate analyses convert this data into information, which once communicated, enables action. Bayesian methods not only enable appropriate analyses to be performed, they also provide greater flexibility in visual communications.

Can descriptive studies really influence government policy? The disparities identified in the cancer atlas resulted in the Queensland government including a specific objective aimed at reducing the geographic disparities in cancer outcomes in their Strategic Directions (300). Results were also used in lobbying to increase the amount of financial assistance the government provided to remote patients to offset travel and accommodation costs while obtaining treatment away from home, and the amount provided was subsequently increased. Our experience is that routinely collected data, when appropriately analysed and communicated, facilitate appropriate government action.

We hope this article will enable greater understanding, and potentially uptake, of Bayesian

methods in disease mapping, along with available options for communicating estimates and their uncertainty.

Additional Information

B.1 WinBUGS code

WinBUGS code for the incidence model

```
Model
{
for (i in 1 : N) {
# Likelihood
O[i] ~ dpois(mu[i])
Opred[i] ~ dpois(mu[i])
log(mu[i]) <- log(E[i]) + alpha + u[i] + v[i]
# Area-specific relative risk (for maps)
RR[i] <- exp(alpha + u[i] + v[i])
# Prior distribution for the uncorrelated heterogeneity
v[i] ~ dnorm(0, tauv)
}
# CAR prior distribution for spatial random effects
u[1 : N] ~ car.normal(adj[], weights[], num[], tauu)
for(k in 1:sumNumNeigh) {
weights[k] <- 1
}
# Other priors:
alpha ~ dflat()
# Hyperpriors on precisions
tauu ~ dgamma(0.1, 0.1)
tauv ~ dgamma(0.001, 0.001)
sigmau <- sqrt(1 / tauu)
sigmav <- sqrt(1 / tauv)
#Standard deviations
sdv <- sd(v[]) #marginal SD of heterogeneity
sdu <- sd(u[]) #marginal SD of clustering
}
```

WinBUGS code for the relative survival model

```
Model
{
# Likelihood
for (i in 1 : datarows) {
d[i]  ~ dpois(mu[i])
mu[i]<-d_star[i] + excessd[i]
log(excessd[i]) <-  log(y[i])+ alpha[RiskYear[i]] + beta[1]*agegp2[i]
+ beta[2]*agegp3[i]+ u[slaNo[i]] + v[slaNo[i]]
```

```
for (j in 1:N_RiskYear){
alpha[j] ~ dnorm (0, 0.001)
}
}
# CAR prior for spatial effects
u[1:Nsla] ~ car.normal(adj[], weights[], num[], tauu)
for (k in 1:sumNumNeigh) {weights[k] <- 1 }
for (i in 1:Nsla) {
# Prior distribution for the uncorrelated heterogeneity
v[i] ~ dnorm(0, tauv)
logRER[i]<-u[i]+v[i]
RER[i]<-exp(logRER[i])
}
# Other priors
tauu ~ dgamma(0.5, 0.001)
tauv ~ dgamma(0.5, 0.001)
varv <- 1/tauv
varu_con <-1/tauu
varu_marg<-sd(u[])*sd(u[])
}
```

B.2 R-INLA code

R-INLA code for the incidence model

Assume that data are available for a set of areas as $\{y_i,e_i,x_{1i},x_{2i}\}$ for $i = 1,...,n$, where y_i is a count, e_i is an expected count, and x_{1i} and x_{2i} are two predictors/covariates. These data should be read into `R` as vectors and can be held in a list. In the code below, `n` represents the number of areas, `obs` represents disease count, `expe` represents expected count, `cov1` and `cov2` represent the covariates, `u` represents the spatial random effects, and `v` represents the unstructured (non-spatial) random effects.

```
u=seq(1:n)
v=seq(1:n)
data.incid = list(obs=obs, expe=expe, cov1=cov1, cov2=cov2, u=u, v=v)
formula1 = obs ~ cov1 + cov2
 + f(u, model="besag", graph="queensland.graph", param=c(0.1, 0.1))
 + f(v, model="iid", param=c(0.001, 0.001))
result1 = inla(formula1, family="poisson", data=data.incid,
control.compute=list(dic=TRUE, cpo=TRUE, mlik=TRUE), E=expe)
summary(result1)
```

R-INLA code for the relative survival model

In the code below, `n` represents the number of areas, `d` represents the number of deaths (d_{ijk}), `d_star` represents the expected number of deaths due to causes other than the disease of interest (d^*_{ijk}), `y` represents the person-time at risk (y_{ijk}), `cov1` and `cov2` represent the covariates, `u` represents the spatial random effects, and `v` represents the unstructured (non-spatial) random effects.

```
u=seq(1:n)
v=seq(1:n)
data.surv = list(d=d, d_star=d_star, y=y, cov1=cov1, cov2=cov2, u=u, v=v)
formula2 = d ~ offset(d_star) + cov1 + cov2
 + f(u, model="besag", graph="queensland.graph", param=c(0.5, 0.001))
 + f(v, model="iid", param=c(0.5, 0.001))
result2 = inla(formula2, family="poisson", data=data.surv,
   control.compute=list(dic=TRUE, cpo=TRUE, mlik=TRUE), E=y)
summary(result2)
```

Appendix C Computational software

Free resources

BUGS (includes WinBUGS and OpenBUGS), available from: http://www.mrc-bsu.cam.ac.uk/software/bugs/ Enables running of Bayesian models using (predominately) Gibbs sampling. The built-in GeoBUGS can be used to generate neighbourhood matrices.

GeoDa Easy-to-use software featuring various smoothing and regression models, as well as generation of various types of neighbourhood matrices. Available from: https://geodacenter.asu.edu/software/downloads.

JAGS Has a cross-platform engine for the BUGS language, but also allows users to write their own distributions, functions etc. Available from: http://mcmc-jags.sourceforge.net/.

The National Cancer Institute has developed several resources, all of which are freely available at gis.cancer.gov/tools/nci_tools.html including:

- Plug-ins for using with ESRI ArcGIS ArcMap include, among others:
  - ColorTool (Assists in using ColorBrewer colours for chloropleth maps)
  - Head-Bang (Smooths data within ArcMap using the Head-Bang smoothing algorithms. These are semi-related to the locally-weighted median discussed in Section 2.4.1.)

- Linked MicroMaps (a graphing program written in Java, allowing easy comparison of statistics across regions and time. Multiple variables can be examined interactively)

- HD*Calc (statistical software for evaluating health disparities. Originally developed for cancer data, so can be used as an extension of SEER*Stat software, but also with any dataset. Generates tables and/or graphs containing calculated summary measures of disparities.)

- SaTScan (aims to detect clusters in spatial, temporal., or spatio-temporal data using scan statistics and evaluate their significance.) www.satscan.org/

NIMBLE http://r-nimble.org/ Can be used as an extension of the BUGS language to write flexible statistical models, or can also be used without BUGS models as a way to compile simple code similar in form to R into C++, which is then compiled and loaded into R.

PySAL www.pysal.org An open source library of spatial analysis functions written in Python.

R R is statistical software, available from: https://www.r-project.org/ A myriad of packages enable spatial analysis within R, including methods appropriate for point- and area-level data. Useful packages for areal data could include:

- bdsmatrix: routines for block diagonal symmetric matrices
- CARBayes: spatial GLMMs for areal data
- CARBayesST: spatio-temporal GLMMs for areal data
- coda: output analysis and diagnostics for MCMC
- colorspace: maps between a variety of colour spaces (e.g. RGB, HSV, CIELAB)
- DCluster: detection of spatial clusters of diseases
- epitools: for epidemiology data and graphics
- fields: curve, surface and function fitting with an emphasis on splines, spatial data and spatial statistics
- gdistance: calculates distances and routes on geographic grids
- glmmBUGS: pass spatial models to WinBUGS
- geoR: geostatistical analysis
- geosphere: computes distances and related measures for geocoordinates
- geospacom: generates distance matrices from shape files and plots data on maps
- gwrr: fits geographically weighted regression models with diagnostic tools
- INLA (available from www.r-inla.org/, not CRAN): Integrated Nested Laplace Approximation
- INLABMA: Bayesian model averaging with INLA
- lmtest: testing linear regression models
- locfit: local regression, likelihood and density estimation
- maps: draw geographical maps
- maptools: tools for reading and handling spatial objects
- Matrix: sparse and dense matrix classes and methods
- MCMCpack: functions to perform Bayesian inference using posterior simulation for a number of statistical models
- McSpatial: nonparametric spatial data analysis
- mgcv: mixed generalised additive model with multiple smoothing parameter estimation
- nlme: linear and nonlinear mixed effects models
- pixmap: import, export and other functions of bitmapped images
- plotGoogleMaps: plot spatial or spatio-temporal data over Google maps
- PReMiuM: for profile regression (a Dirichlet process Bayesian clustering model)
- raster: enables many GIS methods
- R2BayesX: interfaces R with BayesX (performs Bayesian inference in structured additive regression models
- R2WinBUGS: interfaces R with WinBUGS
- RandomFields: simulation and analysis of Gaussian fields, as well as extreme value random fields
- RColorBrewer: provides colour schemes for maps as described at colorbrewer2.org

- RPyGeo: ArcGIS processing in R via Python
- sandwich: robust covariance matrix estimators
- shapefiles: read and write ESRI shapefiles
- sp: classes and methods for spatial data
- spacetime: classes and methods for spatio-temporal data
- spaMM: spatial GLMMs
- sparr: estimates kernel-smoothed relative risk and subsequent inference
- SparseM: Basic linear algebra for sparse matrices
- spatcounts: Spatial count regression via customised MCMC
- SpatialEpi: cluster detection and disease mapping functions, including Bayesian cluster detection
- spatsurv: Bayesian inference for parametric proportional hazards spatial survival models
- spBayes: Univariate and multivariate spatio-temporal models with MCMC
- spBayesSurv: Bayesian modelling and analysis of spatially correlated survival data
- spdep: useful functions to create spatial weights matrix objects from polygon contiguities, and various tests for global and spatial correlation
- spgrass6: interfaces R with GRASS 6+ GIS
- sphet: Estimation of spatial autoregressive models with and without heteroskedastic innovations
- tmap: thematic maps

Stan Can be used for Bayesian modelling with either MCMC or approximate Bayesian inference, or penalised MLE. Available from: http://mc-stan.org/.

Commercial software

ArcGIS Comprehensive GIS software from ESRI. Further details at: www.arcgis.com/.

BoundarySeer Statistical analysis software from BioMedware that enables detection and analysis of geographic boundaries. Further details at: www.biomedware.com/?module=Page&sID=boundaryseer-overview

ClusterSeer Statistical analysis software from BioMedware that enables detection and analysis of event clusters. Further details at: www.biomedware.com/?module=Page&sID=clusterseer-overview

MapInfo Comprehensive GIS software from Pitney Bowes. Further details at: www.mapinfo.com.

MLwiN Statistical software for fitting multilevel (hierarchical) models via either maximum likelihood estimation or MCMC methods. Further details at: www.bristol.ac.uk/cmm/software/mlwin/

SAS (Statistical Analysis System). This is a software suite developed by SAS Institute for advanced data analysis and management. Further details at: www.sas.com. The SAS-ESRI bridge enables ArcGIS functionality. Other useful commands include:

- Proc mapimport – converts a shapefile to a dataset
- Proc gmap – for creating maps (includes choropleth maps)
- WinBUGSio – A user-written macro for interfacing SAS with WinBUGS

Stata Comprehensive software for data analysis and statistical analyses developed by StataCorp. Further details at: www.stata.com/. User-written programs for spatial analyses include:

- geocode3 – Using Google geocoding can either geocode addresses into coordinates or reverse geocode coordinates into addresses to examine the quality of geocoding
- shp2dta – imports .shp data to stata formats
- spatgsa – calculates global spatial autocorrelation measures
- spatlsa – calculates local spatial autocorrelation measures
- spatwmat – generates a matrix of weights
- spmap – generates a large variety of thematic maps
- spgrid – generates two-dimensional grids
- spkde – uses datasets generated by spgrid to perform a variety of kernel estimators
- traveltime3 – uses Google Distancematrix to retrieve distance and travel time between two locations (either geocoded coordinates or addresses)
- winbugs – A suite of commands starting with “wb” that allow Stata to interface with WinBUGS.

SpaceStat Statistical analysis software from BioMedware that enables visualisation, analysis, modelling and exploration of spatiotemporal data. Further details at: www.biomedware.com/?module=Page&sID=spacestat-features

Appendix D Recommended further reading

Lai P-C, So F-M, Chan K-W. Spatial Epidemiological Approaches in Disease Mapping and Analysis. Baton Rouge: CRC Press; 2008.

Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Second edition. Boca Raton: CRC Press; 2013.

Lawson AB. Statistical Methods in Spatial Epidemiology, Second Edition. West Sussex: Wiley; 2006.

Lawson AB, Browne WJ, Vidal Rodeiro CL. Disease Mapping with WinBUGS and MLwiN. Chichester: John Wiley & Sons, Ltd; 2004.

Lawson AB, Williams FLR. An Introductory Guide to Disease Mapping. New York: John Wiley & Sons, Ltd; 2002.

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book: A Practical Introduction to Bayesian Analysis. Boca Raton, FL: Chapman and Hall/CRC Press; 2012.

Mengersen KL. Bayes for Beginners. Brisbane: Queensland University of Technology; 2011.

Ntzoufras I. Bayesian Modeling Using WinBUGS. Hoboken, NJ: John Wiley & Sons, Inc.; 2008.

Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Chichester: John Wiley & Sons, Inc; 2004.