

# Geographic Patterns of Disease

The study of geographic patterns of disease is part of the classic triad in **descriptive epidemiology** of “time, person, place”. Here, place is used as a surrogate for the mix of lifestyle, environmental, and possibly genetic factors that may underlie variations in rates of disease across populations. The purpose is both to *describe* such variations and to identify possible *causes* that could explain them.

Of course, apparent geographic variations in disease rates may be artifactual rather than real. Problems may occur either with the enumeration of cases (numerator) or with the population at risk (**denominator**), or both. Thus spurious geographic variations in disease could reflect differences between populations in case definition, completeness of ascertainment, diagnostic accuracy, and coding or (for mortality) survival rates. Enumeration of the population (e.g. at **census**) may be incomplete, or recent migration may distort population estimates. Great care is therefore required in interpretation. Despite these difficulties, publications such as *Cancer Incidence in Five Continents* [27], and international mortality statistics (*see Mortality, International Comparisons*) compiled by the **World Health Organization**, have provided an invaluable starting point for epidemiologic enquiry.

The analysis of geographic patterns of disease depends crucially on *scale*. Whereas broad-scale patterns may be apparent at an international level, for example, differences between developed and developing countries in the incidence of infectious diseases such as malaria, and in cardiovascular diseases [44], other patterns may only be apparent at a *local* level. These will include, for example, clusters of disease (*see Clustering*) and possible variations in disease risk near putative point sources of environmental pollution.

In this article we briefly discuss disease variations both at the broader and at the local (small-area) scale. We review issues involved in disease mapping (the usual means of presenting descriptive geographic data on disease occurrence) and discuss some of the

problems associated with geographic **correlational studies (ecologic studies)**. Here the aim is to explore geographic variation in disease in terms of underlying spatially varying “risk factors”. The emphasis is on **small-area** applications, where a number of recent advances in methodology have been made.

## International Variations in Disease

International differences in disease occurrence may give important clues as to etiology, which may then be further studied in individual-level studies (e.g. cohort–control or case–control). Thus, in the Seven Countries Study, Keys [24] described large differences in population saturated fat intakes, which were predictive of population differences in the occurrence of coronary heart disease. The INTERSALT Study found cross-population differences in average blood pressure levels, and difference in blood pressure with age, that were associated positively with average levels of salt intake (measured by urinary sodium excretion); a similar positive relationship was also found at individual level [17]. Other examples include the incidence of malignant melanoma and multiple sclerosis, both of which are strongly related to latitude. While this relationship is inverse for melanoma (i.e. a tendency for higher rates near the equator, reflecting greater exposure to sunlight [19]), it is positive for multiple sclerosis (i.e. low incidence in countries near the equator [26]).

### Migrant Studies

**Migrant studies** represent a special case of geographic study. Here, the disease experience of individuals or groups of people is examined as they move from one location or country to another. This affords a unique opportunity to examine the extent to which environmental or genetic influences might determine geographic variations in disease risk. Whereas genetic factors are important in determining which *individuals* become sick, at the population level, overwhelmingly, environmental and lifestyle factors predominate [33]. Thus, in the case of multiple sclerosis, migrants moving from a high risk to a low risk area retain their higher risk if migrating after the age of around 15 years, but attain the risk of the host country if migrating at younger ages [26]. These findings are compatible with an infectious etiology of

multiple sclerosis, with infection acquired in childhood. Another example is the low levels of blood pressure, with little or no rise with age, found among remote and isolated population groups around the world [17, 31]. Blood pressures are found to increase rapidly with migration to an urban environment [31], again indicating the overwhelming importance of environmental factors in determining the unfavorable blood pressure pattern among populations.

### Local Variations in Disease

Variations in disease incidence or mortality at national [23] or subnational [20] level have been described, usually in the form of a disease atlas (see the section on *Disease Mapping*, below). Here we briefly address the occurrence of disease at the local (**small-area**) scale. Although in this context no satisfactory definition of the term “small area” exists, Cuzick & Elliott [10] suggest a working definition as follows:

As a rough guide, any region containing fewer than about 20 cases of disease can be considered a small area . . . Many cancers have annual incidence rates of around 5 per 100 000, so for a collective period of 5 years a small area constitutes a population of around 100 000 or fewer. In some instances, such as a cluster of disease in a remote area or small village, it could be much less, but usually populations of at least 10 000 are needed to form an aggregation of minimal size.

Of course, populations could be much smaller if the disease experience over many such areas is of primary interest – for example, in small-area disease mapping (see next section).

### *Disease Clusters and Clustering*

A problem commonly facing public health authorities is how to deal with reports of apparent disease excess in their locality (i.e. disease “clusters”; see **Clustering**). These reports may subsequently be linked to a putative pollution source. This complicates interpretation since, for post hoc enquires of this type, formal statistical testing is no longer valid. Although there is little potential for isolated cluster investigations to yield new information on the cause of disease, nonetheless the public health authorities often feel compelled to respond. A careful review of cases, and selection of an appropriate

**denominator** and time frame, may result in risk estimates (observed/expected ratios) that are close to 1. This is despite the potential for **bias** towards elevated risk ratios (see **Relative Risk**) – areas at apparently “low” risk do not come to the attention of the authorities! In some instances, replication of the study in other similarly polluted areas (if such can be found), or in a different time period, may be the only feasible way forward. It can also be helpful to place an alleged “cluster” in a wider context by carrying out small-area disease mapping across a larger region (see [43] for a recent example).

An alternative approach to the study of a single disease cluster is to examine more generally for evidence of clustering. Such evidence for Hodgkin’s disease has been cited in support of ideas of an infectious etiology [1], although other explanations, including artifacts related to diagnostic coding, population mobility, or variations in birth rates, are also possible [10, 18].

### *Small-area Studies Near Sources of Environmental Pollution*

Recently, high-resolution geographically referenced routine health data (see **Vital Statistics, Overview**) have become available in certain countries. Together with advances in computing and in statistical methodology, this has led to the development of largely automated systems to examine the distribution of disease near point sources of environmental pollution. In the UK, the Small Area Health Statistics Unit (SAHSU) has been established specifically to: respond rapidly to reports of disease excess (“clusters”) near sources of environmental pollution; carry out studies of health statistics more generally around sources of pollution; carry out descriptive geographic studies at small-area level; and develop the methodology. Recent studies include an investigation of cancer incidence and mortality near a pesticide factory following media reports of excess cancers in the vicinity [43], and a national study of cancer incidence near radio and television transmitters [10] following reports of a leukemia excess near one of the transmitters [13] (see **Leukemia Clusters**).

A major problem in the interpretation of such studies is the issue of socioeconomic **confounding**. Measures of social deprivation (calculated from the census statistics) have been shown to be powerful predictors of the occurrence of disease [5], including stomach and lung cancer (though not leukemia).

Deprived areas do not occur randomly throughout a region, but tend to coincide with industrial sites and correlate with higher smoking rates. Failure to account for social deprivation could thus seriously bias investigation of other lifestyle or environmental risk factors and ill-health. This is illustrated by results of a national study of cancer risk near municipal solid waste incinerators in Great Britain [16]. **Excess risk** was found for a number of cancer sites, including stomach and lung, that persisted after adjustment for deprivation at the small-area scale. However, in the areas with available data, a similar excess was found also for the period before the incinerators were operational. This indicated the presence of **residual confounding** that had not been fully accounted for in the statistical analysis [16].

### Disease Mapping

Maps have long been used to describe geographic patterns of disease (*see Mapping Disease Patterns*). For example, Stocks, in a series of atlases published in the 1930s, described the geographic variation in cancer mortality across counties in England and Wales (reproduced in [38]). A survey in 1991 [42] identified 49 international, national, and regional disease atlases; more recent examples include those by Swerdlow & dos Santos Silva [38] and Bernardinelli et al. [4]. Such maps typically show standardized mortality or incidence ratios (*see Standardization Methods*) for geographic areas such as countries, counties, or districts. The rate in area  $i$  is estimated by  $O_i/E_i$ , where  $O_i$  is the observed number of deaths or incident cases of disease in the area (assumed to follow an independent **Poisson distribution**) and  $E_i$  is the expected number of cases (calculated by applying age- and sex-specific death or disease rates to the census population counts for the area).

Maps convey instant visual information on the spatial distribution of disease and can identify subtle patterns which may be missed in tabular presentations. Their purpose is usually to display variations in ill-health (for example, related to the underlying sociodemography), formulate etiologic hypotheses, aid **surveillance** to detect areas of high disease incidence, and help place specific disease clusters and point source studies in proper context.

While disease maps have both visual and intuitive appeal, considerable caution is required to avoid

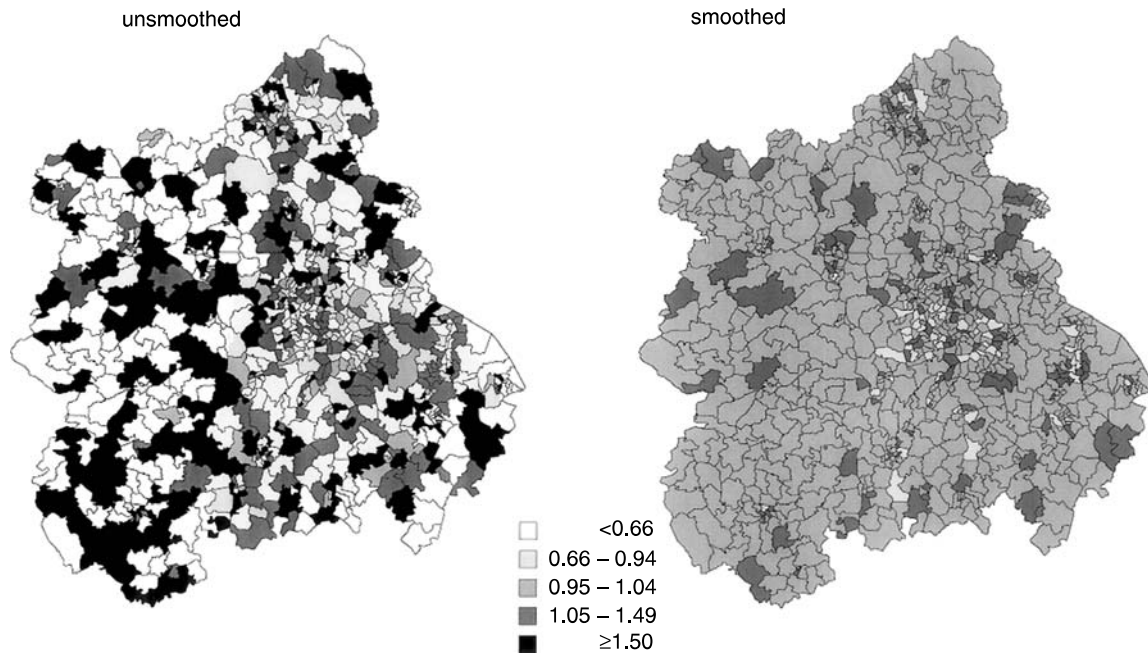
overinterpretation. Apparent geographic variation in rates may simply reflect between-area differences in the quality of reporting, diagnosis, and classification of disease, or confounding due to **ethnic** and socioeconomic factors. Furthermore, disease maps implicitly assume that risk is homogeneous within areas. This is unlikely for the large areas used in many national and international atlases, and may result in misleading inference about individual-level risk.

There is currently considerable scientific interest in exploring more *local* geographic variations in disease. For example, in the UK, small-area mapping is often carried out at the level of electoral ward (average 5000 people) and census enumeration district (400 people).

Disease mapping at the small-area level raises a number of statistical issues. For relatively rare events such as death and cancer incidence, the observed numbers of cases tend to be small in areas with low population, and typically exhibit extra-Poisson sampling variation. This may be assessed formally using the Pothoff–Whittinghill test [30] (*see Clustering*). The sparseness of population data results in unreliable estimates of the area-specific standardized rate ratios, which may create the impression of spurious geographic variation when displayed on a map. These considerations have led to the use of statistical smoothing techniques, which pool information across areas. **Empirical Bayes** [8, 15] and hierarchical Bayes [7, 39] estimates of area-specific relative risk (*see Hierarchical Models in Health Service Research*) represent a compromise between the area-specific standardized rate ratios (*see Standardization Methods*) and the overall **mean** for the whole map.

Small-area disease data often exhibit spatial correlation due to the influence of unmeasured or unknown risk factors which themselves vary smoothly in space. Various hypothesis tests are available to assess such spatial **autocorrelation** – for example, the rank-adjusted  $D$ -statistic [23] and Smans’ test [35].

Figure 1 shows a map of “unsmoothed” (standardized incidence ratio, adjusted for age, sex, and deprivation) and smoothed (empirical Bayes) estimates of brain cancer incidence for 1974–1986 across electoral wards in the West Midlands region of England [14]. As can be seen, much of the random variability is removed by smoothing, especially the apparent high rates found in the large, sparsely populated rural areas. Overall, there is only weak evidence of



**Figure 1** Age-, sex-, and deprivation-adjusted relative risks of brain and central nervous system tumors for electoral wards in West Midlands region, England, age 15–64 years, 1974–1986. Unsmoothed risks (left) and after map smoothing (right) using empirical Bayes method. Reproduced from Eaton et al. [14] by permission of *British Journal of Cancer*

heterogeneity across the map (Potthoff–Whittinghill test;  $p = 0.04$ ), and no evidence of spatial autocorrelation [14].

Bayesian prior distributions for the area-specific relative risks which allow smoothing towards a local mean, rather than the overall map mean, are also used to model spatial interdependence in small-area studies [7, 8, 11, 21, 25, 39]. Implementation of Bayesian hierarchical–spatial models has been made feasible by recent computational [36] and *software* developments, namely BUGS [37] – involving **Markov chain Monte Carlo** simulation algorithms: this approach represents the current state of the art in small-area mapping of disease.

#### *Technical Issues Concerning Presentation of Geographic Disease Data*

Maps provide a succinct summary of geographic patterns in disease. However, visual perception may be influenced by various features of the map, such as the plotting symbols used (e.g. solid shading vs. hatching, color vs. gray scale) and the grouping of data into categories (e.g. percentiles of the distribution of

risk, and numerically equidistant cutpoints) [35]. An empirical study [41] found that the manner of data display may have at least as much effect on observer perception of spatial variation as actual differences in the data. Recently, **nonparametric** mixture distributions (*see Contagious Distributions*) have been used to model the underlying relative risk of disease in small geographic areas [34]. This approach facilitates more objective mapping of disease patterns, since areas are categorized according to statistically driven estimation of the mixture components.

The summary statistic used for presentation may also influence visual interpretation of disease maps. Common choices include standardized rate ratios, smoothed relative risks, or ***P* values**. The former tend to yield erratic maps which are visually dominated by extreme estimates of low precision in sparsely populated areas; the latter are criticized for confusing statistical significance with biological importance (*see Clinical Significance Versus Statistical Significance*) and tend to overemphasize areas of high population in which even small deviations from the expected disease rate may achieve statistical significance. Significance testing of standardized rate

ratios also suffers from the multiple decision problem (see **Multiple Comparisons**), as each ratio is considered independently of the others on the map.

In our view, maps showing Bayesian **shrinkage estimates** of relative risk represent the best compromise, although it is important to realize that these estimates are not judgment-free. For example, they depend on the functions used to describe the distribution of relative risks across the map, and to define the local neighborhood over which spatial interdependence between the small areas is assumed. However, smoothing ensures that precision of the area-specific estimates is approximately comparable across the map, and Bayesian credible intervals derived from hierarchical models are not subject to the constraints of multiple significance testing. Mapping of posterior functions of Bayesian risk estimates is also possible. For example, a map showing the posterior probability that the relative risk in each area ranks above the **median** [2, 22] conveys information about the size *and* uncertainty associated with each area-specific estimate. Further advances in the application of Bayesian methods to disease mapping, and appropriate display methods, including measures of uncertainty, are to be expected.

### Geographic Correlation Studies

Geographic correlation studies are a valuable means of formulating and testing etiologic hypotheses: disease patterns are compared with the geographic distribution of environmental and lifestyle exposures. They are particularly useful when individual-level measurements of exposure are either difficult or impossible to obtain for use in epidemiologic study (for example, air pollution) or are measured imprecisely (for example, diet, and sunlight exposure). (See [19] for further discussion and [32] for a review of the statistical methods.)

Examples of broad-scale **ecologic studies** are given in the section on *International Variations in Disease*. In some cases – for example, sunlight and melanoma, salt and blood pressure – the ecologic relationships have also been demonstrated at individual level. However, the potential for *bias* in such ecologic studies [19, 32] should be recognized. Exposure *within* areas is often heterogeneous; thus the ecologic (average group-level) association between exposure and disease may not equate to the relationship in individuals. To assume otherwise is to commit

the **ecologic fallacy** [28]. Small-area studies may be less prone than broad-scale geographic studies to ecologic bias since the group data are closer to the level of the individual. Nonetheless, positive findings arising from ecologic analyses usually require replication in other data sets and, where possible, at individual level.

As already noted, a major problem in small-area disease studies is the potential for confounding by socioeconomic variables. Adjustment may be made by including, say, a deprivation score such as the Carstairs [6] index (based on small-area census statistics) as a **covariate** in the ecologic regression analysis. Alternatively, indirect standardization of the expected small-area disease counts can be done by stratifying on the socioeconomic status of the areas as well as on age and sex (see **Stratification**). Modeling of spatial autocorrelation between small areas in an ecologic regression study also provides some control for the effect of confounding due to location [9], but further development of these methods is required, and in particular their application to “real” data sets.

Interest has focused on ecologic designs which combine data on the general population with individual-level survey data to improve estimation of group exposure [29, 40]. Methods to adjust for random measurement error in exposure are also receiving attention [3]. Such techniques should enhance the ability of ecologic analyses to estimate the *size* of exposure–disease relationships, not merely to identify the possible presence of such associations.

### Summary and Conclusions

The study of geographic patterns of disease plays a central role in descriptive epidemiology, and has led to some notable etiologic insights. However, geographic studies are associated with major problems of data quality, bias, confounding, and presentation which can seriously complicate their interpretation. The methodologic challenge is clear: to produce objective, statistically valid analyses of geographic variations in ill-health and its determinants, with particular emphasis on developments to combine the best features of individual-level and ecologic studies. Recent advances, particularly in methods for small-area studies, have begun to address these issues. As such techniques become routinely available, they should enhance our ability to quantify the effects

of environmental pollution (see **Environmental Epidemiology**) and lifestyle characteristics on human health.

### References

- [1] Alexander, F.E. (1990). Clustering and Hodgkin's disease, *British Journal of Cancer* **62**, 708–711.
- [2] Bernardinelli, L., Clayton, D.G. & Montomoli, C. (1995). Mapping disease risk: how important are priors?, *Statistics in Medicine* **14**, 2411–2431.
- [3] Bernardinelli, L., Pascutto, C., Best, N.G. & Gilks W.R. (1997). Disease mapping with errors in covariates, *Statistics in Medicine* **16**, 741–752.
- [4] Bernardinelli, L., Maida, A., Marinoni, A., Clayton, D.G., Romano, G., Montomoli, C., Fadda, D., Solinas, M.G., Castiglia, P., Cocco, P.L., Ghislandi, M., Berzuini, C., Pascutto, C., Nerini, M., Styles, B., Capocaccia, R., Lispi L. & Mallardo, E. (1994). *Atlas of Cancer Mortality in Sardinia, 1983–87*. FATMA-CNR.
- [5] Carstairs, V. (2000). Socio-economic factors at area level and their relationship with health, Ch. 4 in *Spatial Epidemiology: Methods and Application*, P. Elliott, J.C. Wakefield, N.G. Best, & D.J. Briggs, eds. Oxford University Press, Oxford, pp. 51–67.
- [6] Carstairs, V. & Morris, R. (1991). *Deprivation and Health in Scotland*. Aberdeen University Press, Aberdeen.
- [7] Clayton, D.G. & Bernardinelli, L. (1992). Bayesian methods for mapping disease risk, in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, P. Elliott, J. Cuzick, D. English & R. Stern, eds. Oxford University Press, Oxford, pp. 205–220.
- [8] Clayton, D.G. & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics* **43**, 671–682.
- [9] Clayton, D.G., Bernardinelli, L. & Montomoli, C. (1993). Spatial correlation in ecological analysis, *International Journal of Epidemiology* **22**, 1193–1202.
- [10] Cuzick J. & Elliott, P. (1992). Small area studies: purpose and methods, in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, P. Elliott, J. Cuzick, D. English & R. Stern, eds. Oxford University Press, Oxford, pp. 14–21.
- [11] Denison, D.G.T. & Holmes, C.C. (2001). Bayesian partitioning for estimating disease risk, *Biometrics* **57**, 143–149.
- [12] Dolk, H., Elliott, P., Shaddick, G., Walls, P. & Thakrar, B. (1997). Cancer incidence near radio and television transmitters in Great Britain. II. All high power transmitters, *American Journal of Epidemiology* **145**, 10–17.
- [13] Dolk, H., Shaddick, G., Walls, P., Grundy, C., Thakrar, B., Kleinschmidt, I. & Elliott, P. (1997). Cancer incidence near radio and television transmitters in Great Britain. I. Sutton Coldfield transmitter, *American Journal of Epidemiology* **145**, 1–9.
- [14] Eaton, N., Shaddick, G., Dolk, H. & Elliott, P. (1997). Small-area study of the incidence of neoplasms of the brain and central nervous system among adults in the West Midlands Region, 1974–86, *British Journal of Cancer* **75**, 1080–1083.
- [15] Efron, B. & Morris, C. (1975). Data analysis using Stein's estimation and its generalisation, *Journal of the American Statistical Association* **70**, 311–319.
- [16] Elliott, P., Shaddick, G., Kleinschmidt, I., Jolley, D., Walls, P., Beresford, J. & Grundy, C. (1996). Cancer incidence near municipal solid waste incinerators in Great Britain, *British Journal of Cancer* **73**, 702–710.
- [17] Elliott, P., Stamler, J., Nichols, R., Dyer, A.R., Stamler, R., Kesteloot, H. & Marmot, M. (1996). INTERSALT revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations, *British Medical Journal* **312**, 1249–1253.
- [18] Elliott, P. & Wakefield, J.C. (2000). Bias and confounding in spatial epidemiology, Ch. 5 in *Spatial Epidemiology: Methods and Applications*, P. Elliott, J.C. Wakefield, N.G. Best & D.J. Briggs, eds. Oxford University Press, Oxford, pp. 68–84.
- [19] English, D. (1992). Geographical epidemiology and ecological studies, in *Geographical and Environmental Epidemiology: Methods for Small-Area studies*, P. Elliott, J. Cuzick, D. English & R. Stern, eds. Oxford University Press, Oxford, pp. 3–13.
- [20] Gardner, M.J., Winter, P.D. & Barker, D.J.P. (1984). *Atlas of Mortality from Selected Diseases in England and Wales 1968–1978*. Wiley, Chichester.
- [21] Green, P.J. & Richardson, S. (2002). Hidden Markov models and disease mapping, *JASA*, to appear.
- [22] Jarup, L., Best, N.G., Toledano, M.B., Wakefield, J.C. & Elliott, P. (2002). Geographical epidemiology of prostate cancer in Great Britain, *International Journal of Cancer* **97**, 695–699.
- [23] Kemp, I., Boyle, P., Smans, M. & Muir C. (1985). *Atlas of Cancer in Scotland, 1975–1980, Incidence and Epidemiologic Perspective*. IARC Scientific Publication No. 72, International Agency for Research on Cancer, Lyon.
- [24] Keys, A., ed. (1970). *Coronary Heart Disease in Seven Countries*. American Heart Association Monograph no. 29. American Heart Association, New York.
- [25] Knorr-Held, L. & Rasser, G. (2000). Bayesian detection of clusters and discontinuities in disease maps, *Biometrics* **56**, 13–21.
- [26] Kurtzke, J.F. (1985). Neurological system, in *Oxford Textbook of Public Health*, Vol. 4. Oxford University Press, Oxford, Chapter 12, pp. 203–249.
- [27] Muir, C., Waterhouse, J., Mack, T., Powell, J. & Whelan, S., eds. (1987). *Cancer Incidence in Five Continents*, Vol. V. IARC Scientific Publication No. 88, International Agency for Research on Cancer, Lyon.

- [28] Piantadosi, S., Byar, D.P. & Green, S.B. (1988). The ecological fallacy, *American Journal of Epidemiology* **127**, 893–904.
- [29] Plummer, M. & Clayton, D. (1996). Estimation of population exposure in ecological studies, *Journal of the Royal Statistical Society, Series B* **58**, 113–126.
- [30] Pothoff R.F. & Whittinghill, M. (1966). Testing for homogeneity. II. The Poisson distribution, *Biometrika* **53**, 183–190.
- [31] Poulter, N.R., Khaw, K.T., Hopwood, B.E.C., Mugambi, M., Peart, W.S., Rose, G. & Sever, P.S. (1990). The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure, *British Medical Journal* **300**, 967–972.
- [32] Richardson, S. & Monfort, C. Ecological correlation studies, Ch. 11 in *Spatial Epidemiology: Methods and Applications*, P. Elliott, J.C. Wakefield, N.G. Best & D.J. Briggs, eds. Oxford University Press, Oxford, pp. 205–219.
- [33] Rose, G. (1985). Sick individuals, sick populations, *International Journal of Epidemiology* **14**, 32–38.
- [34] Schlattmann, P., Dietz, E. & Bohning, D. (1996). Covariate adjusted mixture models and disease mapping with the program DismapWin, *Statistics in Medicine* **15**, 919–929.
- [35] Smans, M. & Esteve, J. (1992). Practical approaches to disease mapping, in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, P. Elliott, J. Cuzick, D. English & R. Stern, eds. Oxford University Press, Oxford, pp. 141–157.
- [36] Smith, A.F.M. & Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society, Series B* **55**, 3–23.
- [37] Spiegelhalter, D.J., Thomas, A., Best, N.G. & Lunn, D. (2002). WinBUGS Bayesian inference Using Gibbs Sampling Manual Version 1.4, Imperial College, London and MRC Biostatistics Unit, Cambridge, available from <http://www.mrc-bsu.cam.ac.uk/bugs>.
- [38] Swerdlow, A. & dos Santos Silva, I. (1993). *Atlas of Cancer Incidence in England and Wales 1968–85*. Oxford University Press, Oxford.
- [39] Wakefield, J.C., Best, N.G. & Waller, L.A. (2000). Bayesian approaches to disease mapping, Ch. 7 in *Spatial Epidemiology: Methods and Applications*, P. Elliott, J.C. Wakefield, N.G. Best & D.J. Briggs, eds. Oxford University Press, Oxford, pp. 104–127.
- [40] Wakefield, J.C. & Salway, R. (2001). A statistical framework for ecological and aggregate studies, *Journal of the Royal Statistical Society, Series A* **164**, 119–138.
- [41] Walter, S.D. (1993). Visual and statistical assessment of spatial clustering in mapping data, *Statistics in Medicine* **12**, 1275–1291.
- [42] Walter S.D. & Birnie, S.E. (1991). Mapping mortality and morbidity patterns: an international comparison, *International Journal of Epidemiology* **20**, 678–689.
- [43] Wilkinson, P., Thakrar, B., Shaddick, G., Stevenson, S., Pattenden, S, Landon, M., Grundy, C. & Elliott, P. (1997). Cancer incidence and mortality around the Pan Britannica Industries pesticide factory, Waltham Abbey, *Occupational and Environmental Medicine* **54**, 101–107.
- [44] World Bank (1993). *World Development Report 1993, Investing in Health*. Oxford University Press, Oxford.

PAUL ELLIOTT &amp; NICOLA BEST