Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

Estimating Individual-Level Risk in Spatial Epidemiology Using Spatially Aggregated Information on the Population at Risk

Peter J. Diggle, Yongtao Guan, Anthony C. Hart, Fauzia Paize & Michelle Stanton

Estimating Individual-Level Risk in Spatial Epidemiology Using Spatially Aggregated Information on the Population at Risk

Peter J. DIGGLE, Yongtao GUAN, Anthony C. HART, Fauzia PAIZE, and Michelle STANTON

We propose a novel alternative to case-control sampling for the estimation of individual-level risk in spatial epidemiology. Our approach uses weighted estimating equations to estimate regression parameters in the intensity function of an inhomogeneous spatial point process, when information on risk-factors is available at the individual level for cases, but only at a spatially aggregated level for the population at risk. We develop data-driven methods to select the weights used in the estimating equations and show through simulation that the choice of weights can have a major impact on efficiency of estimation. We develop a formal test to detect non-Poisson behavior in the underlying point process and assess the performance of the test using simulations of Poisson and Poisson cluster point processes. We apply our methods to data on the spatial distribution of childhood meningococcal disease cases in Merseyside, U.K. between 1981 and 2007.

KEY WORDS: Estimating equations; Inhomogeneous spatial-point processes; Meningococcal disease.

1. INTRODUCTION

A fundamental problem in spatial epidemiology is to understand the relationship between the risk that an individual will experience a particular health outcome and the values of one or more risk factors associated with that individual. A widely used study design to address problems of this kind is the case-control study (Breslow and Day 1980). In its simplest form, a case-control study requires data consisting of the values of all risk factors under consideration for each individual in the study region who experiences the health outcome (cases), and for a random sample of individuals not experiencing the health outcome (controls). In contrast, an ecological study uses data in a spatially aggregated form, consisting of the numbers of cases, numbers of individuals at risk (denominators), and average values of risk factors in each of a set of spatial units that partition the study region. The attraction of the ecological study design is that the required data are often routinely available, for example, from national disease registries or censuses, whereas the individual-level data required by the case-control study design typically need to be acquired from scratch, often at considerable expense. However, it is well known that the effects estimated in ecological and case-control studies are different, a phenomenon known as ecological bias (Greenland and Morgenstern 1989; Greenland and Robins 1994). In this paper, we consider the problem of estimating individual-level effects when data on risk factors are available at the individual level for cases, but only at the spatially aggregated level for the population at risk. In many applications, the risk factors are available only for a (random) subset of the cases. Our proposed methods can still be applied, with the understanding that the interpretation of the (less interesting) intercept term in the log-linear intensity model (see Section 3) will be different.

As a specific example, we consider data relating to 864 cases of childhood meningococcal disease in the metropolitan county of Merseyside, U.K., over the period January 1, 1981 to December 31, 2007. The data available for each case include their age, sex, residential location as derived from their full postcode, and a measure of social deprivation for their residential location, the Townsend score (Townsend, Phillimore, and Beattie 1988). The corresponding data for the population at risk are available for each of the approximately 5000 census output areas that make up the Merseyside study region. Within the study period, censuses took place in the years 1981, 1991, and 2001. The scientific objectives are to investigate the effect of deprivation on the risk of meningococcal disease while accounting for the potentially confounding effects of age and sex, and to assess whether there is any residual spatial clustering of cases after accounting for these effects.

In Section 2 of the paper we review earlier work on methods for combining individual level and spatially aggregated data, and explain why these are unsuitable for our application. In Section 3 we give the details of our proposed approach, which uses a set of weighted estimating equations for the regression parameters of interest. In Section 4, we propose data-driven methods to select the weights so as to obtain reasonably efficient estimators. Our approach for inference assumes that case locations form an inhomogeneous spatial Poisson process. In Section 5, we therefore develop a method to test for residual clustering. In Section 6 we assess the performance of the proposed methods through simulations. In Section 7 we describe the application to the Merseyside meningococcal disease data. Section 8 is a concluding discussion. Some technical details are given in an Appendix.

Peter Diggle is Distinguished University Professor, School of Health and Medicine, Lancaster University, Lancaster, U.K. and Adjunct Professor, Department of Biostatistics, Johns Hopkins University School of Public Health, Baltimore (E-mail: p.diggle@lancaster.ac.uk). Yongtao Guan is Associate Professor, Division of Biostatistics, Yale University, New Haven, CT 06520-8034 (E-mail: yongtao.guan@yale.edu) and is corresponding author. Fauzia Paize is a Research Fellow, Institute of Child Health, University of Liverpool, Alder Hey Children's NHS Foundation Trust, Liverpool, U.K. (E-mail: fauziap@liv.ac.uk). C. Anthony Hart was Professor of Medical Microbiology, Division of Medical Microbiology, University of Liverpool, Liverpool, U.K. (deceased September 2007). Michelle Stanton is Research Student, School of Health and Medicine, Lancaster University, Lancaster, U.K. (E-mail: m.stanton@lancaster.ac.uk). Yongtao Guan's research was supported in part by NSF grant DMS-0845368 and by NIH/NIDCR grant UL1 DE019586. Michelle Stanton's research was supported by the award of a research studentship by the U.K. Engineering and Physical Sciences Research Council. Data on childhood meningococcal disease in Merseyside, U.K., were collected by Alistair Thomson, Omnia Marzouk, Andrew Riordan, Paul Baines, Enitan Carrol, Scott Hackett, and Niten Makwana, Alder Hey Children's Hospital, Merseyside.

2. LITERATURE REVIEW

Several authors have previously suggested methods for combining individual level and spatially aggregated epidemiological data to alleviate the well-known limitations of ecological studies that use only spatially aggregated data.

Prentice and Sheppard (1995) and Sheppard and Prentice (1995) proposed augmenting spatially aggregated data with individual-level data from a sample of the population at risk. This allowed them to develop unbiased estimating equations for the effects of individual-level risk factors, while appropriately exploiting the additional information provided by the spatially aggregated data. Wakefield (2004) proposed an approach that combines ecological data with cohort data. A practical limitation of both of these proposals is the additional expense entailed in generating the individual-level data. For the Merseyside childhood meningococcal disease data introduced in Section 1, it would not be feasible to generate retrospectively a random sample from the population at risk, because the data were observed over a 27-year period.

Best, Ickstadt, and Wolpert (2000) considered the more general problem of combining data at disparate spatial resolutions and proposed a sophisticated parametric model in which all of the data are related to a latent, spatially continuous stochastic process representing unexplained spatial variation in risk. They fit their model to data relating childhood respiratory disorders to exposure to traffic-related pollutants, using a computationally intensive Markov chain Monte Carlo (MCMC) implementation of Bayesian inference. Jackson, Best, and Richardson (2006) also used MCMC methods to fit their proposed hierarchical model for combining small area and individual-level data on exposures and health outcomes.

MCMC has had a major impact on statistical practice, chiefly through its ability to handle models whose complexity precludes fitting by analytical or non-Monte Carlo numerical methods. However, applications of MCMC algorithms often need careful tuning to generate reliably reproducible results. Although such algorithms have been commonly used for modeling areal spatial data using free software such as WinBUGS, their use in the current setting is still challenging. We believe that it is useful to develop methods that avoid the need for MCMC implementation.

Haneuse and Wakefield (2007, 2008a, 2008b) noted that the cohort-based approach of Wakefield (2004) is inherently inefficient for investigation of rare outcomes, and proposed a hybrid design in which ecological data are supplemented either with case-control data or with case data alone. We note that meningococcal disease is an extremely rare disease. Specifically, in our application the average number of cases per year is only 32, from a population at risk of the order of several hundred thousand.

An important practical distinction between methods that require individual-level data from cases only, or from both case and noncases, is that the former are typically available at little or no additional cost because the individuals concerned will have undergone detailed clinical investigation, whereas the latter require additional data to be collected as an integral component of the overall research design. As already argued earlier, this would not be feasible in applications such as ours, for which the available data consist of historical case records.

The methods proposed in Wakefield (2004) and in Haneuse and Wakefield (2007, 2008a, 2008b) are currently restricted to binary risk factors. Although binary risk factors could represent different levels of a continuous risk factor, for the Merseyside childhood meningococcal disease data we prefer to include such risk factors (e.g., age and deprivation score) directly in the model.

Our proposed method is, to the best of our knowledge, the first that can be used to fit models at the individual level with both discrete and continuous risk factors, requiring only case data at the individual level, and avoiding the need for computationally intensive MCMC methods.

3. METHODOLOGY

Let $\mathbf{s}_i$, $i = 1, \ldots, N$, be the spatial locations of a population of N individuals at risk in a spatial region D, among whom the first M members are cases, with $M \ll N$. For the ith member of the population, write $\mathbf{X}_i = \{X_{i1}, \ldots, X_{ip}\}$ for the value of an associated p-dimensional covariate vector, with $X_{i1} = 1$ for all i. We assume that $\mathbf{X}_i$ is observed for each case, $i = 1, \ldots, M$, but not for noncases, $i = M + 1, \ldots, N$. However, an aggregated quantity

$$\tilde{\mu}_{jk} = \sum_{i=1}^{N} X_{ij} I(\mathbf{s}_i \in D_k) \tag{1}$$

is available, where the subregions $D_k : k = 1, \ldots, K$, form a partition of the study region D and $I(\cdot)$ is an indicator function. Note that $\tilde{\mu}_{1k}$ is the number of individuals at risk in D_k.

We treat the (unobserved) locations of the N individuals in the population at risk as a realization of an inhomogeneous spatial Poisson process with intensity $\lambda_0(\mathbf{s})$ (Diggle 2003), representing a spatially varying population density. Let $f[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}]$ be the probability that an individual at location $\mathbf{s}$ is a case and assume, temporarily, that cases occur independently; this is often a reasonable assumption for noncontagious diseases. Then, the case locations also follow an inhomogeneous spatial Poisson process, with intensity $\lambda(\mathbf{s}; \boldsymbol{\beta}) = \lambda_0(\mathbf{s}) f[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}]$. Our first goal is to make inferences about the parameter vector $\boldsymbol{\beta}$ based on the observed covariates for the cases, that is, $\mathbf{X}(\mathbf{s}_i) : i = 1, \ldots, M$, and the spatially aggregated covariates for all members of the population at risk in each sub-region, that is, $\tilde{\mu}_{jk} : j = 1, \ldots, p$ and $k = 1, \ldots, K$.

To motivate our method, we assume that each X_{ij} can be considered as the value at $\mathbf{s}_i$ of a spatially continuous process $\{X_j(\mathbf{s}) : \mathbf{s} \in D\}$. Then, $\tilde{\mu}_{jk}$ is an unbiased estimator for

$$\mu_{jk} = \int_{D_k} X_j(\mathbf{s}) \lambda_0(\mathbf{s}) \, d\mathbf{s}. \tag{2}$$

Note that (2) can be reexpressed as

$$\mu_{jk} = \int_{D_k} \frac{X_j(\mathbf{s})}{f[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \lambda(\mathbf{s}; \boldsymbol{\beta}_0) \, d\mathbf{s},$$

where $\boldsymbol{\beta}_0$ is the true but unknown value of $\boldsymbol{\beta}$. Now define

$$\hat{\mu}_{jk}(\boldsymbol{\beta}) = \sum_{i=1}^{M} \frac{X_j(\mathbf{s})}{f[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}]} I(\mathbf{s}_i \in D_k).$$

Then, $\hat{\mu}_{jk}(\boldsymbol{\beta}_0)$ is also an unbiased estimator for μ_{jk}. Hence, we have a set of unbiased estimating equations for $\boldsymbol{\beta}$,

$$U_j(\boldsymbol{\beta}; \mathbf{W}) = \sum_{k=1}^{K} w_k[\tilde{\mu}_{jk} - \hat{\mu}_{jk}(\boldsymbol{\beta})] = 0, \qquad j = 1, \dots, p, \quad (3)$$

where $\mathbf{W} = \{w_k : k = 1, \dots, K\}$ is any set of predefined weights. The estimator $\hat{\boldsymbol{\beta}}$ obtained by solving (3) is consistent for $\boldsymbol{\beta}$, following established theoretical results for estimating equations (Crowder 1986). Typically, $\tilde{\mu}_{jk}$ converges to μ_{jk} much faster than $\hat{\mu}_{jk}$ because $N \gg M$. With a slight abuse of notation, we replace $\tilde{\mu}_{jk}$ by μ_{jk} from now on.

For ease of presentation, we suppose that the parametric part of $\lambda(\mathbf{s})$ can be described by a log-linear model, that is, $f(\cdot) = \exp(\cdot)$. This log-linear specification is widely used, but our methods apply equally to other choices for $f(\cdot)$. Now, write $\mathbf{U}(\boldsymbol{\beta}; \mathbf{W})$ for the p-element vector with jth element $U_j(\boldsymbol{\beta}; \mathbf{W})$, for $j = 1, \dots, p$. Also, let $\mathbf{U}^{(1)}(\boldsymbol{\beta}; \mathbf{W})$ be the p by p matrix of first derivatives of elements of $\mathbf{U}(\boldsymbol{\beta}; \mathbf{W})$ with respect to elements of $\boldsymbol{\beta}$. Note that

$$\mathbf{U}^{(1)}(\boldsymbol{\beta}; \mathbf{W}) = \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{M} \frac{\mathbf{X}(\mathbf{s}_i)\mathbf{X}(\mathbf{s}_i)'}{\exp[\mathbf{X}(\mathbf{s}_i)'\boldsymbol{\beta}]} I(\mathbf{s}_i \in D_k) \right],$$

hence $E[\mathbf{U}^{(1)}(\boldsymbol{\beta}_0; \mathbf{W})] = \sum_{k=1}^{K} w_k \mathbf{A}_k = \mathbf{A}(\mathbf{W})$, say, where

$$\mathbf{A}_k = \int_{D_k} \lambda_0(\mathbf{s})\mathbf{X}(\mathbf{s})\mathbf{X}(\mathbf{s})' \, d\mathbf{s}. \qquad (4)$$

Following standard arguments involving Taylor series expansions, and under suitable regularity conditions (Guyon 1995), $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically Normally distributed with mean vector zero and approximate covariance matrix

$$\boldsymbol{\Sigma}(\mathbf{W}) = [\mathbf{A}(\mathbf{W})]^{-1}\mathbf{B}(\mathbf{W})[\mathbf{A}(\mathbf{W})]^{-1}, \qquad (5)$$

where

$$\mathbf{B}(\mathbf{W}) = \sum_{k=1}^{K} w_k^2 \int_{D_k} \frac{\lambda_0(\mathbf{s})\mathbf{X}(\mathbf{s})\mathbf{X}(\mathbf{s})'}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \, d\mathbf{s}$$

$$+ \sum_{k=1}^{K} \sum_{l=1}^{K} w_k w_l \int_{D_k} \int_{D_l} \lambda_0(\mathbf{s}_1)\lambda_0(\mathbf{s}_2)\mathbf{X}(\mathbf{s}_1)\mathbf{X}(\mathbf{s}_2)'$$

$$\times [g(\mathbf{s}_1, \mathbf{s}_2) - 1] \, d\mathbf{s}_1 \, d\mathbf{s}_2. \qquad (6)$$

In (6), $g(\mathbf{s}_1, \mathbf{s}_2)$ is the pair correlation function of the spatial point process that generates the cases. For any Poisson process, $g(\mathbf{s}_1, \mathbf{s}_2) = 1$ (Møller and Waagepetersen 2004), in which case only the first term of (6) remains. Then, $\mathbf{A}(\mathbf{W})$ and $\mathbf{B}(\mathbf{W})$ can be estimated consistently by

$$\hat{\mathbf{A}}(\mathbf{W}) = \sum_{k=1}^{K} w_k \left\{ \sum_{i=1}^{M} \frac{\mathbf{X}(\mathbf{s}_i)\mathbf{X}(\mathbf{s}_i)'}{\exp[\mathbf{X}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}]} I(\mathbf{s}_i \in D_k) \right\}$$

and

$$\hat{\mathbf{B}}(\mathbf{W}) = \sum_{k=1}^{K} w_k^2 \left\{ \sum_{i=1}^{M} \frac{\mathbf{X}(\mathbf{s}_i)\mathbf{X}(\mathbf{s}_i)'}{\exp[2\mathbf{X}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}]} I(\mathbf{s}_i \in D_k) \right\},$$

respectively. For non-Poisson processes, $\mathbf{B}(\mathbf{W})$ also depends on the pair correlation function $g(\cdot)$, but in either case the covariance matrix of $\hat{\boldsymbol{\beta}}$ is affected by the choice of the weights $\mathbf{W}$. In the next section we develop data-driven methods to choose the weights, while in Section 5 we revisit the Poisson assumption.

4. CHOICE OF WEIGHTS

4.1 A General Framework

We now propose a general framework to guide the choice of weights. Let $\text{tr}[\boldsymbol{\Sigma}(\mathbf{W})]$ denote the trace of the covariance matrix $\boldsymbol{\Sigma}(\mathbf{W})$. Our aim is to find a set of weights $\mathbf{W}$ that yields the smallest possible value of $\text{tr}[\boldsymbol{\Sigma}(\mathbf{W})]$. A similar idea was used in Bevilacqua et al. (2009), who proposed minimizing the trace of a so-called Godambe information matrix in order to choose the value of a single tuning parameter in their composite likelihood estimation. Here, we consider the more complicated problem of choosing a potentially large number of weights $w_k : k = 1, \dots, K$; for example, in our Merseyside meningitis application, $K = 4586$.

Write $\mathbf{B}(\mathbf{W}) = \sum_{k=1}^{M} w_k^2 \mathbf{B}_k$, where $\mathbf{B}(\mathbf{W})$ is defined in (6). In the Poisson case, note that

$$\mathbf{B}_k = \int_{D_k} \frac{\lambda_0(\mathbf{s})\mathbf{X}(\mathbf{s})\mathbf{X}(\mathbf{s})'}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \, d\mathbf{s}. \qquad (7)$$

Now, take the derivative of $\text{tr}[\boldsymbol{\Sigma}(\mathbf{W})]$ with respect to w_k to obtain

$$T_k(\mathbf{W}) = \frac{\partial \, \text{tr}[\boldsymbol{\Sigma}(\mathbf{W})]}{\partial w_k}$$

$$= 2 \, \text{tr}\{\mathbf{A}(\mathbf{W})^{-2}[w_k \mathbf{B}_k - \mathbf{A}_k \mathbf{A}(\mathbf{W})^{-1}\mathbf{B}(\mathbf{W})]\},$$

where $\mathbf{A}_k$ is as defined in (4). To obtain the "optimal" weights, we follow Guan and Shen (2010) to solve $T_k(\mathbf{W}) = 0$ for $k = 1, \dots, K$. In general, closed-form solutions are difficult to obtain without imposing additional restrictions. We suppose that

$$\mathbf{A}_k = c_k \mathbf{B}_k \qquad (8)$$

for some constants $c_k : k = 1, \dots, K$. Note that $\mathbf{A}_k$ and $\mathbf{B}_k$ are both $p \times p$ matrices. Thus, condition (8) simply requires that all corresponding elements in $\mathbf{A}_k$ and $\mathbf{B}_k$ are being proportional. Such a condition is satisfied, if the parametric part of the intensity function is constant over each subregion D_k. Under condition (8), we have

$$T_k(\mathbf{W}) = 2 \, \text{tr}\left[\mathbf{A}(\mathbf{W})^{-2}\mathbf{B}_k \mathbf{A}(\mathbf{W})^{-1} \sum_{l=1}^{K} w_l \mathbf{B}_l(w_k c_l - c_k w_l) \right],$$

which is equal to zero if $w_k = c_k$, for $k = 1, \dots, K$. If $T_k(\mathbf{W}) = 0$ has a unique root, the resulting weights are then the optimal weights, in the sense that they minimize $\text{tr}[\boldsymbol{\Sigma}(\mathbf{W})]$ under the condition (8). Furthermore, (8) implies $\boldsymbol{\Sigma}(\mathbf{W}) = \mathbf{A}(\mathbf{W})^{-1}$ because of (4), (5), and (6). Note that when (8) holds, estimation of $\mathbf{A}(\mathbf{W})$, and hence of $\boldsymbol{\Sigma}(\mathbf{W})$, is straighforward. In the next two subsections, we try to find values of c_k such that (8) holds approximately, in the Poisson and non-Poisson cases.

4.2 The Poisson Process Case

Let $\bar{\mathbf{X}}_k$ be the average of the covariates $\{\mathbf{X}(\mathbf{s}) : \mathbf{s} \in D_k\}$. Under the working assumption that $\mathbf{X}(\mathbf{s}) \approx \bar{\mathbf{X}}_k$ for $\mathbf{s} \in D_k$, it follows from (7) that $\mathbf{A}_k \approx \exp(\bar{\mathbf{X}}_k'\boldsymbol{\beta}_0)\mathbf{B}_k$, implying the choice of weights

$$w_k = \exp(\bar{\mathbf{X}}_k'\boldsymbol{\beta}_0). \qquad (9)$$

In practice, $\bar{\mathbf{X}}_k$ is easily obtained from $\mu_{jk} : j = 1, \dots, p$, and $\boldsymbol{\beta}_0$ can be replaced by a consistent estimator based on a convenient

set of predefined weights, for example equal weights $w_k = 1$ for all k. This working assumption strictly renders our approach redundant, in the sense that if it were true then covariate information would automatically be available for every individual and a standard case-control analysis would be straightforward. Nevertheless, the resulting choice of weights (9) is intuitively appealing. To see why, imagine that $\lambda_0(\mathbf{s})$ and $\mathbf{X}(\mathbf{s})$ were both known for all $\mathbf{s} \in D$. Then, maximum likelihood estimation could be used to estimate $\boldsymbol{\beta}$. Specifically, the jth score function from the maximum likelihood would become

$$\tilde{U}_j(\boldsymbol{\beta}) = \sum_{k=1}^{K} \sum_{i=1}^{M} X_j(\mathbf{s}_i) I(\mathbf{s}_i \in D_k)$$
$$- \sum_{k=1}^{K} \int_{D_k} \lambda_0(\mathbf{s}) X_j(\mathbf{s}) \exp[\mathbf{X}(\mathbf{s})' \boldsymbol{\beta}] \, d\mathbf{s}.$$

Note that the jth estimating equation based on the proposed weights (9) is

$$U_j(\boldsymbol{\beta}; \mathbf{W}) = \sum_{k=1}^{K} \exp(\bar{\mathbf{X}}_k' \boldsymbol{\beta}_0) \sum_{i=1}^{M} \frac{X_j(\mathbf{s}_i) I(\mathbf{s}_i \in D_k)}{\exp[\mathbf{X}(\mathbf{s})' \boldsymbol{\beta}]}$$
$$- \sum_{k=1}^{K} \exp(\bar{\mathbf{X}}_k' \boldsymbol{\beta}_0) \int_{D_k} \lambda_0(\mathbf{s}) X_j(\mathbf{s}) \, d\mathbf{s}.$$

The striking similarity between $\tilde{U}_j(\boldsymbol{\beta})$ and $U_j(\boldsymbol{\beta}; \mathbf{W})$ suggests that the weights (9) essentially try to make $U_j(\boldsymbol{\beta}; \mathbf{W})$ close to $\tilde{U}_j(\boldsymbol{\beta})$. This is appealing because maximum likelihood estimators are known to be asymptotically efficient. In general, the smaller the subregions $D_k : k = 1, \ldots, K$ become, the closer $U_j(\boldsymbol{\beta}; \mathbf{W})$ will be to $\tilde{U}_j(\boldsymbol{\beta})$ and the more efficient the estimator $\hat{\boldsymbol{\beta}}$ will be. As $|D_k| \to 0$, $k = 1, \ldots, K$, $\mathbf{A}(\mathbf{W})$ and $\mathbf{B}(\mathbf{W})$ both converge to

$$\sum_{k=1}^{K} \int_{D_k} \lambda_0(\mathbf{s}) \mathbf{X}(\mathbf{s}) \mathbf{X}(\mathbf{s})' \exp[\mathbf{X}(\mathbf{s}) \boldsymbol{\beta}_0] \, d\mathbf{s}$$
$$= \int_D \lambda_0(\mathbf{s}) \mathbf{X}(\mathbf{s}) \mathbf{X}(\mathbf{s})' \exp[\mathbf{X}(\mathbf{s}) \boldsymbol{\beta}_0] \, d\mathbf{s}$$

for sufficiently smooth $\mathbf{X}(\cdot)$. Thus, the limit of the covariance matrix $\boldsymbol{\Sigma}(\mathbf{W})$ becomes

$$\tilde{\boldsymbol{\Sigma}}(\mathbf{W}) = \left\{ \int_D \lambda_0(\mathbf{s}) \mathbf{X}(\mathbf{s}) \mathbf{X}(\mathbf{s})' \exp[\mathbf{X}(\mathbf{s}) \boldsymbol{\beta}_0] \, d\mathbf{s} \right\}^{-1},$$

which is the limiting covariance matrix of the maximum likelihood estimator.

4.3 The Non-Poisson Process Case

In the non-Poisson case, we assume that the process is second-order intensity reweighted stationary (Baddeley, Møller, and Waagepetersen 2000), hence $g(\mathbf{s}_1, \mathbf{s}_2) = g(\mathbf{s}_1 - \mathbf{s}_2)$. Define

$$\mathbf{B}_{kl} = \int_{D_k} \int_{D_l} \lambda_0(\mathbf{s}_1) \lambda_0(\mathbf{s}_2) \mathbf{X}(\mathbf{s}_1) \mathbf{X}(\mathbf{s}_2)' [g(\mathbf{s}_1 - \mathbf{s}_2) - 1] \, d\mathbf{s}_1 \, d\mathbf{s}_2.$$

Following the results in the Poisson process case, (8) holds approximately if, for some constants c_k^*,

$$\mathbf{A}_k \approx c_k^* \sum_{l=1}^{K} \frac{w_l}{w_k} \mathbf{B}_{kl}. \quad (10)$$

To see when this may be so, suppose that $g(\mathbf{t}) = 1$ if $\|\mathbf{t}\| > r$ for some small value r, where $\| \cdot \|$ is Euclidean distance. If also the spatial variation in $\mathbf{X}(\cdot)$ is sufficiently smooth, and assuming that $w_k \approx w_l$ if $d_{kl} < r$, then

$$\sum_{l=1}^{K} \frac{w_l}{w_k} \mathbf{B}_{kl} \approx \int_{D_k} \lambda_0(\mathbf{s}_1) \mathbf{X}(\mathbf{s}_1) \mathbf{X}(\mathbf{s}_1)'$$
$$\times \left[\int_D \lambda_0(\mathbf{s}_2) [g(\mathbf{s}_1 - \mathbf{s}_2) - 1] \, d\mathbf{s}_2 \right] d\mathbf{s}_1.$$

Thus, (10) holds if $\int_D \lambda_0(\mathbf{s}_2)[g(\mathbf{s}_1 - \mathbf{s}_2) - 1] \, d\mathbf{s}_2$ is approximately constant for all $\mathbf{s}_1 \in D_k$. This in turn holds if the spatial variation in $\lambda_0(\cdot)$ is sufficiently smooth, for example, $\lambda_0(\mathbf{s}) \approx \bar{\lambda}_{0k} = \mu_{1k}/|D_k|$ for all $\mathbf{s} \in D_l$ if $d_{kl} < r$. A rough approximation for c_k is then $c_k \approx \bar{\lambda}_{0k} \int_{\|\mathbf{t}\| \leq r} [g(\mathbf{t}) - 1] \, d\mathbf{t}$, which leads to the weights

$$w_k = \frac{\exp(\bar{\mathbf{X}}_k' \boldsymbol{\beta}_0)}{1 + \bar{\lambda}_{0k} \exp(\bar{\mathbf{X}}_k' \boldsymbol{\beta}_0) \int_{\|\mathbf{t}\| \leq r} [g(\mathbf{t}) - 1] \, d\mathbf{t}}. \quad (11)$$

For clustered process, the weights assigned by (11) to regions with high values of $\exp(\bar{\mathbf{X}}_k' \boldsymbol{\beta}_0)$ are smaller than those used in the Poisson process case, because typically $g(\mathbf{t}) > 1$ for such processes. This change is intuitively reasonable because the numbers of individuals in such regions are highly variable by comparison with a Poisson process with the same intensity $\lambda_0(\cdot)$, suggesting that their contributions to the estimating equation should be downweighted.

5. ASSESSING NON–POISSON BEHAVIOR

In practice, nonparametric estimation of the pair correlation function $g(\cdot)$ in conjunction with a nonparametric specification for $\lambda_0(\cdot)$ is difficult; see, for example, Baddeley, Møller, and Waagepetersen (2000) or Diggle et al. (2007). Hence, a pragmatic strategy is to conduct inference about $\boldsymbol{\beta}$ under the assumption that the point process is Poisson, but to include within the analysis a diagnostic test of this assumption. In principle, we might wish to detect different kinds of non-Poisson behavior, but in the epidemiological setting the usual concern is to detect residual spatial clustering.

To achieve this, consider the statistic

$$G_{kl}(\boldsymbol{\beta}) = \sum_{i_1, i_2 = 1}^{M} \frac{I(\mathbf{s}_{i_1} \in D_k, \mathbf{s}_{i_2} \in D_l, i_1 \neq i_2)}{\exp[\mathbf{X}(\mathbf{s}_{i_1})' \boldsymbol{\beta}] \exp[\mathbf{X}(\mathbf{s}_{i_2})' \boldsymbol{\beta}]}. \quad (12)$$

Note that if the process is Poisson, then

$$E[G_{kl}(\boldsymbol{\beta}_0)] = \int_{D_k} \int_{D_l} \lambda_0(\mathbf{s}_1) \lambda_0(\mathbf{s}_2) \, d\mathbf{s}_1 \, d\mathbf{s}_2 = \mu_{1k} \mu_{1l}, \quad (13)$$

where μ_{1k} is as defined in (1). Let d_{kl} be the distance between suitably defined centres of D_k and D_l, for example, their centroids. Motivated by (13), we define the statistic

$$\tilde{g}(u; \boldsymbol{\beta}) = \frac{\sum_{k,l=1}^{K} \kappa[(d_{kl} - u)/h] G_{kl}(\boldsymbol{\beta})}{\sum_{k,l=1}^{K} \kappa[(d_{kl} - u)/h] \mu_{1k} \mu_{1l}},$$

where $\kappa(\cdot)$ is a kernel function and h is the bandwidth. If the case process is Poisson, we expect to find $\tilde{g}(u; \hat{\boldsymbol{\beta}}) \approx 1$ because of (13) and the consistency of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$. If the process is clustered, we expect to find $\tilde{g}(u; \hat{\boldsymbol{\beta}}) > 1$ when u is small. Conversely, if the process is spatially regular, we expect to find

$\tilde{g}(u; \hat{\boldsymbol{\beta}}) < 1$ when u is small. A plot of the estimated function $\tilde{g}(u; \hat{\boldsymbol{\beta}})$ therefore gives a visual assessment of spatial clustering or regularity.

To develop a formal test, define

$$G(\boldsymbol{\beta}) = \sum_{k=1}^{K} G_{kk}(\boldsymbol{\beta}),$$

where $G_{kk}(\boldsymbol{\beta})$ is defined by (12). Note that $G_{kk}(\boldsymbol{\beta}) = \sum_{k=1} \mu_{1k}^2 \tilde{g}(0; \boldsymbol{\beta})$, if the kernel function $\kappa(x) = 0$ except at $x = 0$. We consider only $u = 0$ because small distances typically provide most of the information about non-Poisson behavior. We apply Taylor series expansion to $G(\hat{\boldsymbol{\beta}})$ and obtain

$$G(\hat{\boldsymbol{\beta}}) \approx G(\boldsymbol{\beta}_0) - 2\left[\sum_{k=1}^{K} \mu_{1k}\boldsymbol{\mu}_k'\right](\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

$$\approx G(\boldsymbol{\beta}_0) + 2\left[\sum_{k=1}^{K} \mu_{1k}\boldsymbol{\mu}_k'\right]\mathbf{A}(\mathbf{W})^{-1}\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{W}),$$

where $\boldsymbol{\mu}_k = \{\mu_{jk} : j = 1, \ldots, p\}$ is a p-element vector. See the Appendix for details.

Recall that the covariance matrix for $\hat{\boldsymbol{\beta}}$ is as given in (5). To estimate the variance of $G(\hat{\boldsymbol{\beta}})$, we therefore only need to estimate the variance of $G(\boldsymbol{\beta}_0)$ and the covariance between $\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{W})$ and $G(\boldsymbol{\beta}_0)$. Under the null hypothesis, it follows from straightforward algebra that the former is

$$2\sum_{k=1}^{K} \int_{D_k} \int_{D_k} \frac{\lambda_0(\mathbf{s}_1)\lambda_0(\mathbf{s}_2)}{\exp[X(\mathbf{s}_1)'\boldsymbol{\beta}_0]\exp[X(\mathbf{s}_2)'\boldsymbol{\beta}_0]} d\mathbf{s}_1 d\mathbf{s}_2$$

$$+ 4\sum_{k=1}^{K} (\mu_{1k})^2 \int_{D_k} \frac{\lambda_0(\mathbf{s})}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} d\mathbf{s},$$

while the latter is

$$-2\sum_{k=1}^{K} w_k\mu_{1k} \int_{D_k} \frac{\lambda_0(\mathbf{s})\mathbf{X}(\mathbf{s})}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} d\mathbf{s}.$$

See the Appendix for details. Both terms can be estimated relatively easily, for example, by

$$2\sum_{k=1}^{K} \sum_{i_1,i_2=1}^{M} \frac{I(\mathbf{s}_{i_1}, \mathbf{s}_{i_2} \in D_k, i_1 \neq i_2)}{\exp\{2[X(\mathbf{s}_{i_1})' + X(\mathbf{s}_{i_2})']\hat{\boldsymbol{\beta}}\}}$$

$$+ 4\sum_{k=1}^{K} (\mu_{1k})^2 \sum_{i=1}^{M} \frac{I(\mathbf{s}_i \in D_k)}{\exp[2X(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}]} \quad (14)$$

and

$$-2\sum_{k=1}^{K} w_k\mu_{1k} \sum_{i=1}^{M} \frac{I(\mathbf{s}_i \in D_k)\mathbf{X}(\mathbf{s}_i)}{\exp[2X(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}]}, \quad (15)$$

respectively. Combining (5), (14), and (15), the variance of $G(\hat{\boldsymbol{\beta}})$ can then be estimated straightforwardly. Assuming asymptotic normality for $G(\hat{\boldsymbol{\beta}})$, we can then apply a standard Wald-type test.

6. SIMULATIONS

We now show the performance of our method using simulations of inhomogeneous Poisson and Poisson cluster processes. We generated simulated realizations on square regions $D = [0, n] \times [0, n]$ with $n = 1$ or $n = 2$. The intensity function for both processes was $\lambda(\mathbf{s}; \alpha, \boldsymbol{\beta}) = \lambda_0(\mathbf{s})\exp[\alpha + \boldsymbol{\beta}X(\mathbf{s})]$, where $\lambda_0(\mathbf{s}) = \exp[\boldsymbol{\beta}^*X^*(\mathbf{s})]$, $\boldsymbol{\beta} = \boldsymbol{\beta}^* = 0.5$, and $X(\mathbf{s})$ and $X^*(\mathbf{s})$ are independent realizations of a stationary, isotropic Gaussian process with covariance $\exp(-\gamma u)$ for $\gamma = 5$ or 10. Note that a smaller value of γ yields a smoother covariate surface. The average number of events per realization was 200 and 800 for $n = 1$ and $n = 2$, respectively. For the Poisson cluster process, we used a stationary Poisson process with intensity 100 to generate the parent locations and a radially symmetric Gaussian distribution with standard deviation σ in each coordinate direction to generate the offspring locations relative to their parents (Diggle 2003; Waagepetersen 2007). We used values of $\sigma = 0.02$ and 0.04 to correspond to relatively strong and weak clustering, respectively.

We assume that $X(\mathbf{s})$, but not $X^*(\mathbf{s})$, is observed at every simulated event location. Aggregated covariate informaton is observed in the form of $\mu_{1k} = \int_{D_k} \lambda_0(\mathbf{s}) d\mathbf{s}$ and $\mu_{2k} = \int_{D_k} \lambda_0(\mathbf{s})X(\mathbf{s}) d\mathbf{s}$ where the $D_k : k = 1, \ldots, K$, are equal subsquares that partition D. For the 1×1 region D, we used $K = 25, 100, 400$, and for the 2×2 region, $K = 100, 400, 1600$. To estimate the parameters α and $\boldsymbol{\beta}$ we used both equal weights and the data-driven weights determined by the methods described in Section 3. Note that the estimator based on equal weights do not change with K. For the cluster process, we used $\max[\hat{K}(r) - \pi r^2, 0]$ as an approximation to $\int_{\|\mathbf{t}\| \leq r}[g(\mathbf{t}) - 1] d\mathbf{t}$, where $\hat{K}(r)$ is an estimate of the inhomogeneous K-function (Baddeley, Møller, and Waagepetersen 2000) and $r = 3\sigma$. For comparison, we also implemented two other methods on each simulated data-set: Poisson maximum likelihood estimation under the unrealistic assumption that both $\lambda_0(\mathbf{s})$ and $X(\mathbf{s})$ are completely observed for all $\mathbf{s} \in D$; and an ecological analysis using only spatially aggregated data. For the latter, we assume that $\lambda_0(\mathbf{s})$ and $X(\mathbf{s})$ are both constants within each subregion; the estimates $\hat{\lambda}_0(\mathbf{s}) = \mu_{1k}/|D_k|$ and $\hat{X}(\mathbf{s}) = \mu_{2k}/\mu_{1k}$ are then used to form the Poisson maximum likelihood instead of $\lambda_0(\mathbf{s})$ and $X(\mathbf{s})$, where $\mathbf{s} \in D_k$.

Table 1 gives the empirical mean squared errors (MSEs) of the resulting estimators for $\boldsymbol{\beta}$ from 1000 simulations. Even in the best scenario with $K = 400n^2$ subregions, the estimator based on ecological-data only analysis has substantially much larger MSE than the rest estimators in all cases. The large MSE is mostly due to a large bias of the estimator, which becomes more severe with smaller K. In contrast, all other estimators are approximately unbiased, and the empirical biases (not shown) are consistent with this. As would be expected, for all of the estimators the MSE decreases as the study region becomes larger. Also, the proposed estimator based on the data-driven weights is more efficient than that based on equal weights, and the MSE of the weighted estimator decreases as K increases. In the Poisson process case, the MSE approaches that of the maximum likelihood estimator with known $\lambda_0(\mathbf{s})$, as is consistent with the theoretical comparison given in Section 4.2.

Table 1. Mean squared errors, multiplied by 10,000, from 1000 simulations of various estimators for the regression coefficient β

Process	γ	n	EST_1	EST_2			EST_3	EST_4
				$K = 25n^2$	$100n^2$	$400n^2$		
Poisson	5	1	98	81	74	72	71	192
		2	23	16	15	15	14	81
	10	1	105	78	68	61	57	325
		2	20	16	15	14	13	221
Cluster 1	5	1	226	207	202	197	203	353
		2	45	37	36	35	36	107
	10	1	157	139	130	127	125	417
		2	37	33	32	31	30	240
Cluster 2	5	1	260	232	218	215	263	457
		2	50	42	41	40	48	116
	10	1	177	159	147	144	159	461
		2	42	38	36	35	39	265

NOTE: EST_1 and EST_2 are the proposed estimator based on equal weights and the "optimal" weights, EST_3 is the maximum likelihood estimator with $\lambda_0(\mathbf{s})$ and $X(\mathbf{s})$ known for all $\mathbf{s} \in D$, and EST_4 is the estimator from ecological analysis when $K = 400n^2$, where K is the number of subregions D_k and n defines the region size. Clusters 1 and 2 are the inhomogeneous Poisson cluster process with $\sigma = 0.04, 0.02$, respectively. γ determines the smoothness of the covariates (see Section 6).

In the cluster process case with $\sigma = 0.02$, the MSE of the weighted estimator is consistently smaller than that of the Poisson maximum likelihood estimator. This indicates the benefit of accounting for non-Poisson behavior. When $\sigma = 0.04$, the weighted estimator can still out-perform the maximum likelihood estimator, but to a lesser extent because the spatial clustering is now weaker.

Table 2 presents the empirical size and power of the proposed test for non-Poisson behavior from 1000 simulations, using a nominal significance level of 10%. The empirical size is close to the nominal size in all cases. The empirical power generally increases when the study region becomes larger, when the cluster-

Table 2. Empirical size and power from 1000 simulations for the proposed test for residual spatial clustering, at the 10% nominal significance level

Process	γ	n	$K = 25n^2$	$100n^2$	$400n^2$
Poisson	5	1	0.089	0.094	0.109
		2	0.101	0.093	0.120
	10	1	0.083	0.100	0.134
		2	0.111	0.089	0.096
Cluster 1	5	1	0.490	0.686	0.571
		2	0.890	0.993	0.971
	10	1	0.562	0.760	0.571
		2	0.943	0.997	0.968
Cluster 2	5	1	0.629	0.981	0.998
		2	0.966	1.000	1.000
	10	1	0.746	0.983	0.997
		2	0.992	1.000	1.000

NOTE: Regression parameters are estimated by the proposed estimation method using equal weights. Clusters 1 and 2 are the inhomogeneous Poisson cluster process with $\sigma = 0.04, 0.02$, respectively, n denotes the region size, γ determines the smoothness of the covariates (see Section 6) and K is the number of subregions D_k.

ing strength becomes stronger, and when the covariate process becomes less smooth. When K increases, so that the D_k become smaller, the power generally increases for the case $\sigma = 0.02$, but first increases and then decreases for the case $\sigma = 0.04$. Taken together, these results suggest that the test is most powerful when the size of each subregion D_k is well matched to the scale of the spatial clustering.

7. APPLICATION: CHILDHOOD MENINGOCOCCAL DISEASE IN MERSEYSIDE, U.K.

7.1 Individual-Level Data

Individual-level data were obtained for all meningococcal disease patients admitted to Alder Hey Children's Hospital, Liverpool, U.K. during the period January 1, 1981 to December 31, 2007. Alder Hey is the primary hospital in Merseyside for childhood illnesses. All patients admitted to Alder Hey during this time period were aged 16 or under. Data for each case included date of admission, full unit post-code of residence, age, and sex. Each post-code was converted to a grid reference using the online tool *GeoConvert*, developed by the Census Dissemination Unit at the University of Manchester, U.K. (*http://cdu.mimas.ac.uk*).

Although data were available for all meningococcal disease cases aged 16 or under, for reasons explained below, we analyze only cases aged 0 to 14 years old, inclusive. Patients who were admitted to Alder Hey, yet resided outside the Merseyside boundary were also excluded from the subsequent analysis, leaving 864 cases for the analysis.

7.2 Area-Level Data

Spatially aggregated control data from the 1981, 1991, and 2001 censuses for Merseyside were obtained from the Census Dissemination Unit via their website (*http://casweb.mimas.ac.uk*). The finest spatial resolution at which data were available was enumeration district (1981, 1991) or output area (2001). The average size of an enumeration district in England and Wales is approximately 450 residents, whereas the recommended size of each output area is approximately 300 residents. Population counts for each small area are available for each year of age and each sex. For the purposes of this analysis, the population of interest consists of the age range 0–15 years.

The Townsend score, a measure of social deprivation (Townsend, Phillimore, and Beattie 1988), was derived for each small area. This uses census variables relating to unemployment, car ownership, owner occupation, and overcrowding.

Owing to boundary changes across the three censuses of interest (1981, 1991, 2001), spatially aggregated census data are not strictly comparable across censuses. To resolve this, we needed to harmonize the boundaries such that all population counts correspond to the 2001 census output area level; the Merseyside study region consists of 4586 such output areas. We therefore redistributed counts from the 1981 and 1991 censuses at enumeration district level among the 2001 output areas in proportion to their respective areas of overlap.

Midyear population estimates are available from the Office of National Statistics (ONS) at local authority level by sex and quinary age group (0 years, 1–4 years, 5–9 years, 10–14 years, 15–19 years, etc.) from 1981 to 2007 (Office of National Statistics 2004). To obtain population estimates at output area level

for each year, we assumed that the annual percentage changes applied equally to all output areas within each local authority.

As midyear estimates of the variables required to calculate the Townsend score were not available, we estimated Townsend scores at output area level for each year by linearly interpolating the values calculated in each of the three census years and extrapolating at a constant level from 2001 onwards.

7.3 Results

We denote the years 1981 to 2007 by $t = 1, 2, \ldots, 27$ and write $\mathbf{X}(\mathbf{s}, t)$ for the vector of covariates associated with a child living at $\mathbf{s}$ in year t. Each $\mathbf{X}(\mathbf{s}, t)$ includes age, sex, Townsend score, and indicator variables for the five local authorities that make up the county of Merseyside, namely Knowsley, Liverpool, St. Helens, Sefton, and Wirral. We model the case intensity as

$$\lambda(\mathbf{s}, t) = \lambda_0(\mathbf{s}, t) \exp[\mathbf{X}(\mathbf{s}, t)' \boldsymbol{\beta} + h(t)],$$

where $\lambda_0(\mathbf{s}, t)$ is the spatially varying population density at year t and $h(t)$ describes the temporal trend in risk. We specify $h(t)$ as a cubic spline with knots at $t = 4, 8, 12, 16$, and 20 with $t = 1$ for the year 1981. These knots are selected based on visual examination of the histogram of the yearly counts. We have also tried other choices for the knots and obtained very similar results.

We first fit the above model to our data using equal weights, to obtain estimates $\hat{\boldsymbol{\beta}}_e$ and $\hat{h}_e(t)$. We then calculate weights $w(\mathbf{s}, t) = \exp[\bar{\mathbf{X}}_k(t)' \hat{\boldsymbol{\beta}}_e + \hat{h}_e(t)]$, where $\bar{\mathbf{X}}_k(t)$ is the vector of average covariates in the output area D_k that includes $\mathbf{s}$ and at year t. Table 3 gives the estimates for the regression coefficients and their associated standard error estimates on the assumption that the case process is an inhomogeneous Poisson process. Figure 1 plots the fitted temporal trend in risk, $\exp[\hat{h}(t)]$.

The formal test of the Poisson assumption described in Section 4 yields a test statistic -0.8009, corresponding to a two-sided p-value 0.4232. Figure 2 shows the estimate $\tilde{g}(u; \hat{\boldsymbol{\beta}})$ for separation distances $u = 1, 2, \ldots, 10$ km, obtained using bandwidth of 1 km. The estimate is approximately constant, taking a value reasonably close to one throughout the plotted range, again consistent with the Poisson assumption.

The results support the hypothesis of a positive association between area-level social deprivation and individual-level meningococcal disease risk. After adjusting for the effects of age, sex, and local authority, a unit change in the Townsend
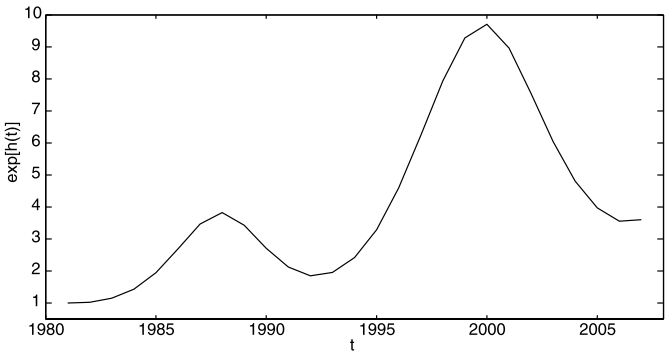

Figure 1. Estimated temporal trend component in the model for meningococcal disease risk in Merseyside, U.K.

score at output-area level increases the risk by a factor of 1.1 (95% confidence interval 1.0493 to 1.1466).

The results also revealed a significant local authority effect, with highest risk in Liverpool and lowest in St. Helens after controlling for other risk factors. The local authorities that consitute Merseyside vary greatly in terms of their social composition and degree of urbanization.

Age, as anticipated, is a significant risk factor for the disease, with risk decreasing by a factor of 0.8146 (95% confidence interval 0.7928 to 0.8370) for each year of increase in age after controlling for other risk factors. We find no significant association between meningococcal disease risk and sex.

The estimated temporal trend in risk over the study period indicates a large increase over the period 1995 to 2000 and a decrease thereafter. The decrease is likely to be due in part to the introduction of the MenC vaccine in November 1999, after which the incidence of serogroup C meningococcal disease in the target age groups began to decline (Trotter et al. 2004).

PCR testing was introduced as a meningococcal disease confirmation test in October 1996. Following this, there were improvements in the ascertainment of meningococcal disease cases, which may explain this increase (Carrol et al. 2000; Gray et al. 2006).

We are unable to explain the secondary peak in estimated risk centred on 1988. There is a known relationship between
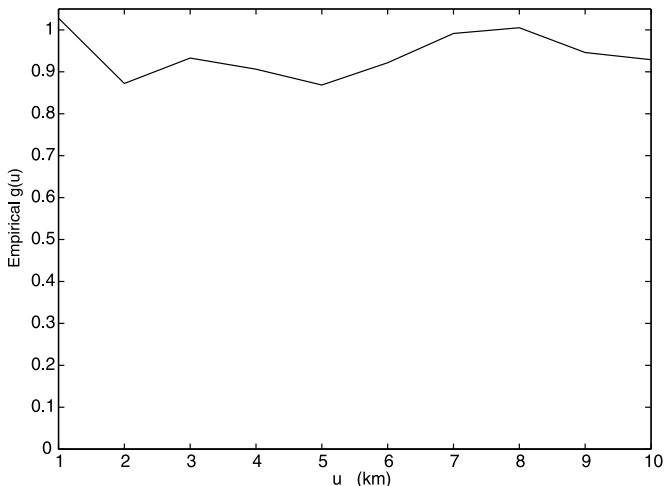
Table 3. Estimates of covariate effects in the model for meningococcal disease risk in Merseyside, U.K.

Covariates	EST	STD	95% CI
Knowsley	−9.74	0.35	(−10.43, −9.06)
Liverpool	−9.39	0.33	(−10.03, −8.76)
St. Helens	−11.38	0.51	(−12.38, −10.38)
Sefton	−9.74	0.37	(−10.46, −9.02)
Wirral	−11.26	0.43	(−12.10, −10.42)
Age	−0.21	0.01	(−0.24, −0.18)
Gender	0.25	0.13	(−0.00, 0.50)
Townsend	0.09	0.02	(0.05, 0.14)

NOTE: Figures in parentheses are 95% confidence intervals.

Figure 2. Estimated pair correlation function for the point process of case locations of meningococcal disease cases in Merseyside, U.K.

previous viral upper respiratory tract infections, including influenza, and meningococcal disease risk (Stuart et al. 1996; Jensen et al. 2004). During the 1989/90 winter season, an influenza epidemic in England and Wales was followed by an increase in the national incidence of meningococcal disease cases. Although no direct evidence of causality is available, Cartwright and Jones (1991) demonstrated a possible association between the two events. However, we do not have data relating to the Merseyside area that would allow us to investigate the relationship between local incidence of flu or flu-like illnesses and meningococcal disease risk.

With regard to our main finding of an association with deprivation, earlier ecological studies in Gwent, S. Wales at enumeration district level (Fone et al. 2003), in the eastern region of England at local authority ward level (Williams et al. 2004) and in North East Thames at local authority ward level (Jones et al. 1997) have also reported a significant positive association between the Townsend score and meningococcal disease risk. However, in each of these studies the target for inference was the ecological, rather than individual, association (Greenland and Morgenstern 1989). Also, none of the earlier studies allowed for temporal trends in risk.

Haynesa and Gale (1999) has recognized that the relationship between health and deprivation is not uniform across the U.K., but varies according to the geographical type of area. In particular, the relationship between deprivation measures such as the Townsend score and health tends to be weaker in rural areas than in urban areas (Townsend, Phillimore, and Beattie 1988). For example, car ownership contributes to the Townsend score but is more weakly related to deprivation in rural than in urban areas. We therefore tested for interaction between Townsend score and the indicator variables for the local authorities, but the result was not significant at the 5% level.

8. DISCUSSION

We have proposed a method for analyzing spatially referenced, individual-level health outcomes that does not require individual-level control data. Instead, the method uses small-area level information on the population at risk, consisting of numbers of individuals and average values of relevant covariates for each small-area unit. The method delivers estimates of individual-level risk-factor effects, an inferential procedure based on the assumption that case locations are a realization of an inhomogeneous Poisson point process model, and a diagnostic test for departure from the Poisson assumption. The ability to handle this kind of data structure is essential for our application to the Merseyside childhood meningococcal disease data.

Simulation studies show that our method works well in practice. Firstly, it yields estimators that can be competitive in terms of efficiency by comparison with maximum likelihood estimation under the unrealistic assumption that the continuous spatial variation in the density of the population at risk is known exactly. Secondly, the diagnostic test has the correct size and detects residual spatial clustering of cases, with power which depends in a sensible way on the strength of the residual clustering.

Our analysis of the Merseyside childhood meningococcal disease data adds to the evidence of a positive association between small-area level social deprivation and disease risk

and provides, for the first time, an estimate of the size of the individual-level effect.

In some situations there is more information than simply the mean of each predictors in each subregion. For example, statistics related to higher-order moments of each predictor, such as its standard deviation, may be available. Each higher-order moment will lead to a new set of estimating equations. The availability of such information would allow us to consider higher-order models, for example models with linear and quadratic terms in each of the predictors. Our proposed methods can be directly applied in such cases by treating each higher-order term as a new predictor. However, when only a linear model is fitted, it is not obvious how best to incorporate into the analysis the additional estimating equations that are generated by the higher-order moments; in particular, it is difficult to develop intuitively appealing weights.

One potential use of the higher-order information would be to refine the choice of the weights. Our current proposal amounts to expanding $\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]$ at $\mathbf{X}(\mathbf{s}) = \bar{\mathbf{X}}_k$ for $\mathbf{s} \in D_k$. Alternatively, we might consider

$$\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0] \approx \exp(\bar{\mathbf{X}}_k'\boldsymbol{\beta}_0)\{1 + \boldsymbol{\beta}_0'[\mathbf{X}(\mathbf{s}) - \bar{\mathbf{X}}_k] + \boldsymbol{\beta}_0'[\mathbf{X}(\mathbf{s}) - \bar{\mathbf{X}}_k][\mathbf{X}(\mathbf{s}) - \bar{\mathbf{X}}_k]'\boldsymbol{\beta}_0\}.$$

If a covariance matrix $\mathbf{S}_k$ is available for each of the predictors in D_k, we can instead define $w_k = \exp(\bar{\mathbf{X}}_k'\hat{\boldsymbol{\beta}})(1 + \hat{\boldsymbol{\beta}}'\mathbf{S}_k\hat{\boldsymbol{\beta}})$. Theese new weights are likely to be useful when the covariates show a high level of spatial variation within subregions. However, in our simulations we found that they made almost no difference to the resulting estimates of $\boldsymbol{\beta}$.

APPENDIX: TECHNICAL DETAILS

Taylor series expansion for $G(\hat{\boldsymbol{\beta}})$: Recall that

$$G(\boldsymbol{\beta}) = \sum_{k=1}^{K} G_{kk}(\boldsymbol{\beta}),$$

where

$$G_{kk}(\boldsymbol{\beta}) = \sum_{i_1, i_2 = 1}^{M} \frac{I(\mathbf{s}_{i_1} \in D_k, \mathbf{s}_{i_2} \in D_k, i_1 \neq i_2)}{\exp[\mathbf{X}(\mathbf{s}_{i_1})'\boldsymbol{\beta}]\exp[\mathbf{X}(\mathbf{s}_{i_2})'\boldsymbol{\beta}]}.$$

A Taylor series expansion gives

$$G(\hat{\boldsymbol{\beta}}) \approx G(\boldsymbol{\beta}_0)$$
$$- 2\sum_{k=1}^{K}\sum_{i_1,i_2=1}^{M} \frac{I(\mathbf{s}_{i_1}, \mathbf{s}_{i_2} \in D_k, i_1 \neq i_2)\mathbf{X}(\mathbf{s}_{i_1})'}{\exp[\mathbf{X}(\mathbf{s}_{i_1})'\boldsymbol{\beta}_0]\exp[\mathbf{X}(\mathbf{s}_{i_2})'\boldsymbol{\beta}_0]}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$\approx G(\boldsymbol{\beta}_0)$$
$$+ 2\sum_{k=1}^{K}\sum_{i_1,i_2=1}^{M} \frac{I(\mathbf{s}_{i_1}, \mathbf{s}_{i_2} \in D_k, i_1 \neq i_2)\mathbf{X}(\mathbf{s}_{i_1})'}{\exp[\mathbf{X}(\mathbf{s}_{i_1})'\boldsymbol{\beta}_0]\exp[\mathbf{X}(\mathbf{s}_{i_2})'\boldsymbol{\beta}_0]}$$
$$\times \mathbf{A}(\mathbf{W})^{-1}\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{W})$$
$$\approx G(\boldsymbol{\beta}_0) + 2\left[\sum_{k=1}^{K} \mu_{1k}\boldsymbol{\mu}_k'\right]\mathbf{A}(\mathbf{W})^{-1}\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{W}),$$

where $\boldsymbol{\mu}_k = \{\mu_{jk} : j = 1, \ldots, p\}$ is a p-element vector, and the last approximation is due to the fact that under the null hypothesis of a Poisson process,

$$E\left\{\sum_{k=1}^{K}\sum_{i_1,i_2=1}^{M} \frac{I(\mathbf{s}_{i_1}, \mathbf{s}_{i_2} \in D_k, i_1 \neq i_2)\mathbf{X}(\mathbf{s}_{i_1})}{\exp[\mathbf{X}(\mathbf{s}_{i_1})'\boldsymbol{\beta}_0]\exp[\mathbf{X}(\mathbf{s}_{i_2})'\boldsymbol{\beta}_0]}\right\} = \sum_{k=1}^{K} \mu_{1k}\boldsymbol{\mu}_k'.$$

Expression for the variance of $G(\boldsymbol{\beta}_0)$: Under the null hypothesis, it follows from Campbell's theorem (Daley and Vere-Jones 1988) that

$$\mathrm{var}[G_{kk}(\boldsymbol{\beta}_0)] = 2 \int_{D_k} \int_{D_k} \frac{\lambda_0(\mathbf{s}_1)\lambda_0(\mathbf{s}_2)}{\exp[X(\mathbf{s}_1)'\boldsymbol{\beta}_0]\exp[X(\mathbf{s}_2)'\boldsymbol{\beta}_0]} \, d\mathbf{s}_1 \, d\mathbf{s}_2$$

$$+ 4 \int_{D_k} \int_{D_k} \int_{D_k} \frac{\lambda_0(\mathbf{s}_1)\lambda_0(\mathbf{s}_2)\lambda_0(\mathbf{s}_3)}{\exp[\mathbf{X}(\mathbf{s}_1)'\boldsymbol{\beta}_0]} \, d\mathbf{s}_1 \, d\mathbf{s}_2 \, d\mathbf{s}_3.$$

Because $\int_{D_k} \lambda_0(\mathbf{s}) \, d\mathbf{s} = \mu_{1k}$, the last term is equal to

$$4(\mu_{1k})^2 \int_{D_k} \frac{\lambda_0(\mathbf{s})}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \, d\mathbf{s}.$$

This immediately yields the variance of $G(\boldsymbol{\beta}_0)$ since $\mathrm{var}[G(\boldsymbol{\beta}_0)] = \sum_{k=1}^{K} \mathrm{var}[G_{kk}(\boldsymbol{\beta}_0)]$.

Expression for the covariance of $G(\boldsymbol{\beta}_0)$ and $\mathbf{U}(\boldsymbol{\beta}_0; \mathbf{W})$: Recall that $\mathbf{U}(\boldsymbol{\beta}; \mathbf{W})$ is a p-element vector with jth element $U_j(\boldsymbol{\beta}; \mathbf{W}) : j = 1, \ldots, p$, where

$$U_j(\boldsymbol{\beta}; \mathbf{W}) = \sum_{k=1}^{K} w_k[\mu_{jk} - \hat{\mu}_{jk}(\boldsymbol{\beta})]$$

$$= \sum_{k=1}^{K} w_k \mu_{jk} - \sum_{k=1}^{K} w_k \sum_{i=1}^{M} \frac{X_j(\mathbf{s}_i)}{\exp[\mathbf{X}(\mathbf{s}_i)'\boldsymbol{\beta}_0]} I(\mathbf{s}_i \in D_k).$$

Under the null hypothesis, it follows from Campbell's theorem that

$$\mathrm{cov}[G_{kk}(\boldsymbol{\beta}_0), U_j(\boldsymbol{\beta}; \mathbf{W})] = -2w_k \int_{D_k} \int_{D_k} \frac{\lambda_0(\mathbf{s}_1)\lambda(\mathbf{s}_2)X_j(\mathbf{s}_1)}{\exp[\mathbf{X}(\mathbf{s}_1)'\boldsymbol{\beta}_0]} \, d\mathbf{s}_1 \, d\mathbf{s}_2$$

$$= -2w_k \mu_{1k} \int_{D_k} \frac{\lambda_0(\mathbf{s})X_j(\mathbf{s})}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \, d\mathbf{s}.$$

This immediately implies that

$$\mathrm{cov}[G(\boldsymbol{\beta}_0), \mathbf{U}(\boldsymbol{\beta}; \mathbf{W})] = -2 \sum_{k=1}^{K} w_k \mu_{1k} \int_{D_k} \frac{\lambda_0(\mathbf{s})\mathbf{X}(\mathbf{s})}{\exp[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}_0]} \, d\mathbf{s}.$$

[Received May 2009. Revised June 2010.]

REFERENCES

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000), "Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns," *Statistica Neerlandica*, 54 (3), 329–350. [1397,1398]

Best, N., Ickstadt, K., and Wolpert, R. (2000), "Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Resolutions," *Journal of the American Statistical Association*, 95, 1076–1088. [1395]

Bevilacqua, M., Mateu, J., Porcu, E., Zhang, H., and Zini, A. (2009), "Weighted Composite Likelihood-Based Tests for Space-Time Separability of Covariance Functions," *Statistics and Computing*, 20, 283–293. [1396]

Breslow, N., and Day, N. (1980), *Statistical Methods in Cancer Research: The Analysis of Case-Control Studies*, Lyon: International Agency for Research on Cancer. [1394]

Carrol, E., Thomson, A., Riordan, F., Fellick, J., Shears, P., Sills, J., and Hart, C. (2000), "Increasing Microbiological Confirmation and Changing Epidemiology of Meningococcal Disease on Merseyside, England," *Clinical Microbiology and Infection*, 6, 259–262. [1400]

Cartwright, K. A. V., and Jones, D. M. (1991), "Influenza A and Meningococcal Disease," *The Lancet*, 338, 554–557. [1401]

Crowder, M. (1986), "On Consistency and Inconsistency of Estimating Equations," *Econometric Theory*, 2, 305–330. [1396]

Daley, D. J., and Vere-Jones, D. (1988), *An Introduction to the Theory of Point Processes*, New York: Springer-Verlag. [1402]

Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*, New York: Academic Press. [1395,1398]

Diggle, P. J., Gómez-Rubio, V., Brown, P. E., Chetwynd, A. G., and Gooding, S. (2007), "Second-Order Analysis of Inhomogeneous Spatial Point Processes Using Case-Control Data," *Biometrics*, 63 (2), 550–557. [1397]

Fone, D. L., Harries, J., Lester, N., and Nehaul, L. (2003), "Meningococcal Disease and Social Deprivation: A Small Area Geographical Study in Gwent, UK," *Epidemiology and Infection*, 130, 53–58. [1401]

Gray, S. J., Trotter, C. L., Ramsay, M. E., Guiver, M., Fox, A. J., Borrow, R., Mallard, R. H., and Kaczmarski, E. B. (2006), "Epidemiology of Meningococcal Disease in England and Wales 1993/94 to 2003/04: Contribution and Experiences of the Meningococcal Reference Unit," *Journal of Medical Microbiology*, 55, 887–896. [1400]

Greenland, S., and Morgenstern, H. (1989), "Ecological Bias, Confounding and Effect Modification," *International Journal of Epidemiology*, 18, 269–274. [1394,1401]

Greenland, S., and Robins, J. (1994), "Ecological Studies: Biases, Misconceptions and Counterexamples," *American Journal of Epidemiology*, 139, 747–760. [1394]

Guan, Y., and Shen, Y. (2010), "A Weighted Estimating Equation Approach for Inhomogeneous Spatial Point Processes," *Biometrika*, to appear. [1396]

Guyon, X. (1995), *Random Fields on a Network: Modeling, Statistics, and Applications*, New York: Springer-Verlag. [1396]

Haneuse, S., and Wakefield, J. (2007), "Hierarchical Models for Combining Ecological and Case-Control Data," *Biometrics*, 63, 128–136. [1395]

——— (2008a), "The Combination of Ecological and Case-Control Data," *Journal of the Royal Statistical Society, Ser. B*, 70, 73–93. [1395]

——— (2008b), "Geographic-Based Ecological Correlation Studies Using Supplemental Case-Control Data," *Statistics in Medicine*, 27, 864–887. [1395]

Haynesa, R., and Gale, S. (1999), "Mortality, Long-Term Illness and Deprivation in Rural and Metropolitan Wards of England and Wales," *Health and Place*, 5, 301–312. [1401]

Jackson, C., Best, N., and Richardson, S. (2006), "Improving Ecological Inference Using Individual-Level Data," *Statistics in Medicine*, 25, 2136–2159. [1395]

Jensen, E. S., Lundbye-Christensen, S., Samuelsson, S., Sørensen, H. T., and Schønheyder, H. C. (2004), "A 20-Year Ecological Study of the Temporal Association Between Influenza and Meningococcal Disease," *European Journal of Epidemiology*, 19, 181–187. [1401]

Jones, I. R., Urwin, G., Feldman, R. A., and Banatvala, N. (1997), "Social Deprivation and Bacterial Meningitis in North East Thames Region: Three Year Study Using Small Area Statistics," *British Medical Journal*, 314, 794. [1401]

Møller, J., and Waagepetersen, R. (2004), *Statistical Inference and Simulation for Spatial Point Process*, New York: Chapman & Hall. [1396]

Office of National Statistics (2004), "A Short Guide to Population Estimates," available at *http://www.statistics.gov.uk/downloads/theme_population/Short_Guide_revision_Nov_04_final.pdf*. [1399]

Prentice, R., and Sheppard, L. (1995), "Aggregate Data Studies of Disease Risk Factors," *Biometrika*, 82, 113–125. [1395]

Sheppard, L., and Prentice, R. (1995), "On the Reliability and Precision of Within- and Between-Population Estimates of Relative Rate Parameters," *Biometrics*, 51, 853–863. [1395]

Stuart, J. M., Robinson, P. M., Cartwright, K., and Noah, N. D. (1996), "Antibiotic Prescribing During an Outbreak of Meningococcal Disease," *Epidemiology and Infection*, 117, 103–105. [1401]

Townsend, P., Phillimore, P., and Beattie, A. (1988), *Health and Deprivation: Inequality and the North*, London: Routledge. [1394,1399,1401]

Trotter, C. L., Andrews, N. J., Kaczmarski, E. B., Miller, E., and Ramsay, M. E. (2004), "Effectiveness of Meningococcal Serogroup C Conjugate Vaccine 4 Years After Introduction," *The Lancet*, 364, 365–367. [1400]

Waagepetersen, R. P. (2007), "An Estimating Function Approach to Inference for Inhomogeneous Neyman–Scott Processes," *Biometrics*, 63 (1), 252–258. [1396]

Wakefield, J. (2004), "Ecological Inference for 2×2 Tables" (with discussion), *Journal of the Royal Statistical Society, Ser. A*, 167, 385–445. [1395]

Williams, C. J., Willocks, L. J., Lake, I. R., and Hunter, P. R. (2004), "Geographic Correlation Between Deprivation and Risk of Meningococcal Disease: An Ecological Study," *BMC Public Health*, 4, 30. [1401]