Spatial and Spatio-temporal Epidemiology 2 (2011) 173–183

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

The detection of spatially localised outbreaks in campylobacteriosis notification data

Simon E.F. Spencer [a,*], Jonathan Marshall [a], Ruth Pirie [b], Donald Campbell [c], Nigel P. French [a]

[a] Molecular Epidemiology and Veterinary Public Health Laboratory, Hopkirk Research Institute, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand
[b] Institute of Environmental Science and Research, Kenepuru Science Centre, 34 Kenepuru Drive, Porirua, New Zealand
[c] New Zealand Food Safety Authority, Wellington, New Zealand

ARTICLE INFO

Article history:
Available online 20 July 2011

Keywords:
Bayesian analysis
Campylobacter
Disease outbreak
Outbreak detection
Spatial
Epidemiology

ABSTRACT

This paper applies a Bayesian hierarchical model designed to identify potential outbreaks of campylobacteriosis from a background of sporadic cases. We assume that such outbreaks are characterized by spatially-localised periods of increased incidence. As well as calculating an outbreak probability for each potential disease cluster, the model simultaneously estimates the underlying spatial and temporal distribution of sporadic cases. The model is applied to notification data from a region of New Zealand for the period 2001–2007 and correctly identifies known outbreaks, whilst highlighting an appropriate number of potential outbreaks for further investigation. Using simulated data, we show that if additional epidemiological information is included in the construction of the model then it can outperform an established method.

1. Introduction

Campylobacteriosis is New Zealand's most common notifiable disease, with 7176 cases reported in 2009 (ESR, 2010). The epidemiology is complex with several possible transmission routes (Müllner et al., 2009; Wilson et al., 2008). Examples include eating undercooked meat, especially chicken; contaminated drinking water; and environmental and occupational exposures. The proportion of cases attributable to each source is not fully characterized, but consumption of poultry is widely regarded as being a prominent cause (Müllner et al., 2010). Campylobacteriosis case rates are highly seasonal (Baker et al., 2007), but with a large variation between years in the timing, duration and peak incidence of the summer epidemic. There have also been sustained, wide-spread, winter surges of unconfirmed aetiology, for example in 2006 (McTavish et al., 2008). In addition, there have been brief localised outbreaks in which the cases shared a common identified exposure (Graham et al., 2005). Identifying these short periods of increased incidence through the close proximity of the cases in space and time is the main subject of this paper. Once these outbreaks have been identified, further investigations can be initiated to determine the exposure source, providing information that may be used to prevent future cases from occurring.

It may be that a much higher proportion of cases share a common exposure than has traditionally been recognised, due to the fact that a high proportion of campylobacteriosis cases fail to reach notification (Wheeler et al., 1999). Since multiple *Campylobacter* strains – such as multilocus sequence types (Müllner et al., 2009) – coexist in many sources of infection, even detailed subtyping of notified cases may not accurately determine which shared a common exposure. Consequently automatic outbreak detection

* Corresponding author. Present address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK. Tel.: +44 24 7615 0862; fax: +44 24 7652 4532.
E-mail addresses: s.e.f.spencer@warwick.ac.uk (S.E.F. Spencer), j.c.marshall@massey.ac.nz (J. Marshall), ruth.pirie@esr.cri.nz (R. Pirie), donald.campbell@nzfsa.govt.nz (D. Campbell), n.p.french@massey.ac.nz (N.P. French).

1877-5845/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sste.2011.07.008

methods that use only the time and spatial location of the cases – such as the one suggested in this paper – can increase our understanding of the epidemiology of infection.

The high degree of variability in case rates presents unique statistical challenges for surveillance. The background temporal and spatial trends must be well-understood before anomalous outbreaks can be identified as such. For a disease such as campylobacteriosis, outbreak-free periods do not exist due to the seasonal and non-seasonal epidemics and the occurrence of transient localised outbreaks. We define a spatially-localised outbreak as an excess of cases in a small area after accounting for the estimated spatial and temporal pattern of notifications. In this paper a model based approach is adopted, allowing the outbreak risk to be separated from the highly variable temporal and spatial risk caused by sporadic cases. Two versions of a Bayesian hierarchical model are proposed, both of which include 'outbreak indicators', first proposed in disease modelling by Knorr-Held and Richardson (2003). These spatio-temporal parameters change from zero to one during a period of increased incidence. The posterior distribution of these indicator variables consists of a probability that an outbreak is occurring at each point in space and time. This intuitive model output makes the results easy to interpret by non-statisticians – an essential feature of any model that is to be implemented in a public health setting.

There are several existing approaches to cluster and outbreak detection in disease notification data, with a diverse range of applications. The Farrington Algorithm (Farrington et al., 1996) is a method of determining exceedance thresholds for weekly time series notification data which is appropriate for a wide range of diseases and levels of disease incidence. Multiple significance test approaches for spatial data were developed by Openshaw et al. (1987) and Besag and Newell (1991). Scan statistics, originally introduced by Naus (1965) and extended to include spatial dimensions by Kulldorff (1997), have been used to detect clusters in point process data. Spatial and spatio-temporal scan statistic approaches have since been extended in many directions with a large number of applications, see for example Kulldorff et al. (2007), Takahashi et al. (2008), Jung et al. (2010) and many others. Kelsall and Diggle (1995) use a non-parametric kernel smoothing method to estimate the intensity of a spatial point process, and provide exceedance thresholds to identify significant clusters. This approach was extended to adaptive kernel smoothing by Davies and Hazelton (2010). Bayesian approaches have also been developed for spatial point process models (see for example Lawson, 2006) as well as spatially discrete Markov Random Field models (Knorr-Held and Richardson, 2003). Since the data used for this current study were spatially discrete, we have chosen to adopt the latter for the basis of our model.

2. Methods

First the campylobacteriosis notification data are described and then a Bayesian hierarchical model for outbreak detection is proposed. Finally two schemes for generating simulated data containing outbreaks on the borderline of detectability are discussed.

2.1. Data

The data used in this study were extracted from the national notifiable diseases database, EpiSurv (New Zealand Ministry of Health, 2007). The dataset comprised all of the reported laboratory confirmed cases of campylobacteriosis during a 7 year period (2001–2007) which nominated an address in MidCentral District Health Board (DHB) as their home address. The spatial location of each case was taken to be the 2006 census meshblock (available from Statistics New Zealand) that was computer matched to the home address of the case. Meshblocks are the smallest areas defined in the 5-yearly New Zealand census, and generally have a population of between 0 and about 200 individuals. Their spatial resolution is therefore much finer in urban areas than in rural areas. At this stage of analysis 317 of the 2696 cases were excluded as the address matching was unable to obtain a unique meshblock from the home address. Denominator information on the number of people usually resident in each meshblock in the study region was obtained from the 2006 New Zealand census (available from Statistics New Zealand), and was assumed to be constant throughout the study period.

2.2. Bayesian model formulation

Denote the number of cases in meshblock i in week t by $Y_{i,t}$ and assume that $Y_{i,t} \sim \text{Poisson}(n_i\lambda_{i,t})$, where the offset n_i is the population in meshblock i and $\lambda_{i,t}$ is the risk associated with meshblock i in week t. Next, we decompose the log of the risk into three components: a purely temporal component, R_t; a purely spatial component, U_i; and the spatio-temporal component $W_{i,t}$. Thus,

$$\log(\lambda_{i,t}) = R_t + U_i + W_{i,t}. \tag{1}$$

The following prior distributions were assumed. For the temporal trend we assumed a Gaussian second order random walk, with a conjugate gamma hyperprior for the precision. For the spatial component we assumed a Gaussian Markov Random Field (also called a Gaussian intrinsic autoregression, see Besag et al. (1991)), again with a conjugate gamma hyperprior for the precision. For $t \geqslant 2$,

$$\begin{aligned} R_{t+1} - R_t &\sim N(R_t - R_{t-1}, \kappa_R^{-1}), & \kappa_R &\sim \Gamma(a_R, b_R), \\ U_i &\sim N\left(\sum_{j\in n(i)} \frac{U_j}{|n(i)|}, \kappa_U^{-1}|n(i)|^{-1}\right), & \kappa_U &\sim \Gamma(a_U, b_U), \end{aligned} \tag{2}$$

where $n(i)$ is the set of indices of meshblocks neighbouring meshblock i, and any (non-equal) pair of distributions are independent. Although both the spatial and temporal components have Gaussian priors, the posterior distributions need not have the same form. We assumed flat priors for R_1 and R_2, which allowed the baseline level of risk to be absorbed by the temporal component.

The spatio-temporal component $W_{i,t}$, is designed to capture short term localised periods of increased risk that characterize the outbreaks we aim to detect. Firstly we collected together groups of neighbouring meshblocks that share similar risk factors into regions, for example small towns and suburbs become a single region. This massively