Accident Analysis and Prevention 36 (2004) 1019–1028

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Bayesian spatial and ecological models for small-area accident and injury analysis

Ying C. MacNab [a,b,*]

[a] *Department of Health Care and Epidemiology, Division of Epidemiology and Biostatistics, University of British Columbia, Vancouver, BC, Canada V6H 3V4*

[b] *Centre for Healthcare Innovation and Improvement, British Columbia Institute for Children's and Women's Health, Room E-417, University of British Columbia, 4480 Oak Street, Vancouver, BC, Canada V6H 3V4*

Received 21 January 2002; received in revised form 22 May 2002; accepted 29 May 2002

Abstract

In this article, recently developed Bayesian spatial and ecological regression models are applied to analyse small-area variation in accident and injury. This study serves to demonstrate how Bayesian modelling techniques can be implemented to assess potential risk factors measured at group (e.g. area) level. Presented here is a unified modelling framework that enables thorough investigations into associations between injury rates and regional characteristics, residual variation and spatial autocorrelation. Using hospital separation data for 83 local health areas in British Columbia (BC), Canada, in 1990–1999, we explore and examine ecological/contextual determinants of motor vehicle accident injury (MVAI) among male children and youth aged 0–24 and for those of six age groups (<1, 1–4, 5–9, 10–14, 15–19 and 20–24). Eighteen local health area characteristics are studied. They include a broad spectrum of socio-economic indicators, residential environment indicators (roads and parks), medical services availability and utilisation, population health, proportion of recent immigrants, crime rates, rates of speeding charge and rates of seatbelt violation. Our study indicates a large regional variation in MVAI in males aged 0–24 in British Columbia, Canada, in 1990–1999, and that adjusting for appropriate risk factors eliminates nearly all the variation observed. Socio-economic influence on MVAI was profoundly apparent in young males of all ages with the injury being more common in communities of lower socio-economic status. High adult male crime rates were significantly associated with high injury rates of boys aged 1–14. Seatbelt violations and excess speeding charges were found to be positively associated with the injury rates of young men aged 20–24. This and similar ecological studies shed light on reasons for regional variations in accident occurrence as well as in the resulting injuries and hospital utilisation. Thereby they are potentially useful in identifying priority areas for injury/accident prevention and in informing regional health planning and policy development.

Keywords: Bayesian spatial and ecological models; Ecological study; Hospitalisation due to motor vehicle accident injury; Group level risk factors; Accident; Injury surveillance and prevention

1. Introduction

In recent years, epidemiological investigations in which associations between disease occurrence and potential risk factors are studied over aggregated groups (e.g. areas) have gained increased popularity and recognition. The advent of recently developed Bayesian statistical methodologies enabled ecological analyses of such associations and interpretations that explore and address important issues relating to risk estimation, unmeasured confounding, spatial dependence, measurement errors, etc. (Richardson, 1992; Clayton et al., 1993; Plummer and Clayton, 1996; Best, 1999; MacNab, 2003a). Although in some situations, traditional methods such as Gaussian and Poisson regression models can be carefully implemented to examine such associations (Joshua and Garber, 1990; Roberts et al., 1992; Jolly et al., 1993; Ivan and O'Mara, 1997), these methods are limited in their scope in handling data of ecological/spatial context where unmeasured confounders and spatial autocorrelation are evident (Elliott et al., 1992; Biggeri et al., 1999; Ferrándiz et al., 1999). This study introduces a Bayesian spatial model that was proposed by MacNab and Dean (2001), extends this model to allow inclusion of ecological covariates, and presents its application in a Bayesian ecological analysis of small-area injury rates. We present a unified Bayesian hierarchical model framework that permits

* Corresponding author. Tel.: +1-604-875-3571; fax: +1-604-875-3124.
E-mail addresses: ymacnab@cw.bc.ca, ying.macnab@ubc.ca (Y.C. MacNab).

0001-4575/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2002.05.001

thorough investigation into relationships between injury occurrence and exposures, accounting for residual variation and spatial autocorrelation. The methodology is illustrated through an ecological analysis of British Columbia (BC) child and youth motor vehicle accident (MVAI) injury. Eighteen local health area characteristics are studied. These variables reflect, at the regional/community level, measures of 'at-risk' population 'exposure' to local area attributes such as neighbourhood socio-economic deprivation, residential environment (roads and parks), medical services availability and utilisation, population health, proportion of recent immigrants, crime rates, rates of speeding charges and rates of seatbelt violations.

A key feature of the Bayesian spatial and ecological regression models presented herein is that they incorporate random spatial effects into the modelling of potential associations between injury occurrence and covariate effects, both at an ecological (i.e. regional) level. These random spatial effects may reflect unmeasured confounders. The model framework makes it possible to ascertain whether residual variation remains after accounting for known and measured covariate effects and whether the residual effects suggest spatial patterns or clusters. These methods also facilitate spatial smoothing and data pooling when regions under investigation involve small-population areas (Manton et al., 1989; Dean and MacNab, 2001; MacNab and Dean, 2002). Here, the term 'small-area' is used to describe an area with small 'at-risk' population, but not necessarily small in geographical size/scale. Crude rates for areas with small 'at-risk' population are subject to high chance variation. Bayesian spatial and ecological modelling enables data sharing (i.e. risk smoothing) over space, which often results in more reliable risk prediction.

Modelling of age effects using a regression spline is also a significant feature of the Bayesian spatial and ecological regression model. Age is an important risk factor of injury. Also, age effects are often non-linear (Rosenbaum and Rubin, 1984). We will show later that in this BC study motor vehicle accident injury rate increased with age; the age effect is moderately non-linear. Failing to account for such non-linear effect may result in biased estimates of risk factor coefficients (Rosenbaum and Rubin, 1984; Biggeri et al., 1999). In this study, age effect is explicitly modelled via a regression spline. A spline represents a smooth curve of piecewise polynomial. It is often used to model non-linear patterns such as growth curves (Brumback and Rice, 1998) and non-linear effects such as time and age (Sun et al., 2000; Pickle, 2000; MacNab and Dean, 2001, 2002; MacNab, 2003b). Spline and regression spline methodologies are discussed quite extensively in Brumback and Rice (1998); Hastie and Tibshirani (1990); Green and Silverman (2000); a relatively non-technical discussion can be found in MacNab (2003b). It is worth noting that our analytic objective focuses mainly on exploring and examining potential associations between injury rates and regional characteristics. The Bayesian spatial and ecological modelling technique unifies covariate, age and spatial effects into one modelling framework so that each model component is assessed in relation to the others.

The statistical methods discussed address various issues surrounding the validity and potential usefulness of ecological analysis within the context of spatial disease/injury epidemiology (Richardson, 1992; Clayton et al., 1993; Plummer and Clayton, 1996; Best, 1999). Ecological studies in general, and the analysis of BC child and youth injury in particular, explore spatial patterns of disease/injury rates and the exposure–response relationships between disease/injury rates and social, economical, provision of care and environmental variables of contextual attributes. More specifically, the ecological investigations of exposure–response relationship discussed herein concern centrally with the associations at the regional or community level. With ecological effects being of primary interest, the underlying model assumption, inference and interpretation of the results are related to the 'at-risk' population and the geographical areas in which the population reside. One potential utility of our investigation is the identification of regions to which injury prevention resources may be directed. In literature, for example, Roberts et al. (1992) present a geographical analysis of child injury morbidity in Auckland, New Zealand, including vehicle occupant injury and other types of injuries. They report on strong correlations between census area socio-economic deprivation, measured by unemployment rates, and injury morbidity rates. They suggest that "injury prevention programmes should be targeted at socio-economically disadvantaged communities".

2. Methods

2.1. The data

The British Columbia motor vehicle accident injury data, provided by the BC Ministry of Health, consist of hospitalisations due to MVA injuries (including MV occupants, pedestrians, and (pedal/motor) cyclists who were in collisions with motor vehicles, ICD-9: E810-819, E822-825), excluding patients with less than 6 h hospitalisation due largely to a minor injury. In this study, we analyse injury hospitalisations (y_{ij}) that are grouped by young males aged <1, 1–4, 5–9, 10–14, 15–19, 20–24 ($i = 1$–6) and by place-of-residence in local health areas 1–83 ($j = 1$–83), for the period of 1 April 1990 to 31 March 1999 (based on the date of discharge). The analysis also considers corresponding 'at-risk' population (n_{ij}) approximated by aggregated local health area annual population estimates (BC Stats), summed over the 10-year study period. In this and previous studies (MacNab, 2002, 2003b), it was observed that, due to the rarity of the MVA injury occurrence, the BC child and youth annual injury rates are subject to high chance variation, both over time and space. The injury and 'at-risk' population are aggregated over the 10-year period to achieve adequate rate stability over the local health areas. We recognise that data aggregation over time may result in loss of important information such as time effects and trend. For example, traffic conditions and reporting practices may change substantially over the time period considered, resulting in changes in injury rate over time.

Our study began with a preliminary investigation of a much larger collection of 32 local health area 1996 census and non-census characteristic variables provided by the BC Ministry of Health and the Canadian Centre for Justice Statistics. The year 1996 is approximately halfway through the injury incidence period. These variables, measured as standardised scores, are assumed to approximate exposure values for the 10-year period (Walter et al., 1999). We observed considerable correlations among the variables. A multicollinearity study (not presented here) was therefore carried out to (a) carefully evaluate the interdependence among the explanatory variables, (b) examine the effects of such correlations on model parameter estimates and (c) reduce the number of explanatory variables so as to eliminate multicollinearity. An explanatory variable was removed from the model if correlation coefficients suggested collinearity and its regression coefficient had the opposite algebraic sign. There were 18 remaining covariates that showed no significant sign of interdependence and multicollinearity; they are the variables presented in this article and listed in Table 1. The specific variables that are used to derive the total 'neighbourhood socio-economic disadvantage score' are grades 9–13 without secondary certificate, government transfer payments, university-with degree (reversed), average household income $ (reversed), and unemployment (genders combined). In addition, the covariate 'medical health professionals' is a total score of dentists, physicians, and registered nurses. Throughout the paper, the term 'risk' refers to probability of injury. In ecological studies, we often use 'risk exposure' to express a general notion that the measurement of a regional variable reflects 'population exposure' to a contextual attribute of geographical/ecological setting. The term 'exposure', meaning population exposure from an 'ecological' perspective, is linked to the notion that the ecological regression slope represents the relationship between injury rate and population (life-style) exposure.

It should be noted that there are potential, perhaps more important, injury risk factors that are not included in this study. Exclusion of these factors may result in misclassification of exposure, although the strength of the Bayesian hierarchical model analysis is its ability to ascertain whether significant residual variation remained after accounting for measured risk factors and its potential to accommodate unmeasured covariate effects. We recognise that investigation into risk factors relating to road safety and transport, such as mode of transport, road environment, and number of hours or miles/kilometres travelled, also plays an important role in child and youth injury analysis and prevention. Although, the data are unavailable to us and not considered here, it is

Table 1
Estimates of the Bayesian ecological model parameters, British Columbia, MVAI, males age 0–24, 1990–1999

Risk factor variables	Age 0–24 (mean, S.E.)	Age < 1 (mean, S.E.)	Age 1–4 (mean, S.E.)	Age 5–9 (mean, S.E.)	Age 10–14 (mean, S.E.)	Age 15–19 (mean, S.E.)	Age 20–24 (mean, S.E.)
Legally married and not separated	−0.019 (0.049)						
Recent immigrants (1991–1996)	−0.090 (0.031)	−0.13 (0.15)	−0.113 (0.15)	0.07 (0.057)	−0.35 (0.038)	−0.18 (0.04)	−0.14 (0.03)
Female lone parent families	−0.027 (0.032)	−0.57 (0.20)	−0.17 (0.09)		−0.10 (0.05)	−0.06 (0.04)	−0.09 (0.03)
Neighbourhood socio-economic disadvantage score	0.073 (0.026)	0.39 (0.16)	0.12 (0.08)	0.19 (0.05)	0.09 (0.04)	0.06 (0.03)	0.14 (0.03)
Male life expectancy at age 0	−0.037 (0.043)		−0.06 (0.13)			−0.05 (0.06)	−0.10 (0.05)
Proportion of total admission exclude injury	0.059 (0.039)						
Low birth weight live birth rate	−0.028 (0.026)						
Public parks	−0.040 (0.030)						
Roads	−0.026 (0.033)						
Seatbelt violations, 15+	0.061 (0.030)	0.089 (0.20)	0.14 (0.09)	0.01 (0.07)	0.05 (0.05)	0.06 (0.04)	0.06 (0.03)
Excess speeding charges, 15+	−0.003 (0.028)						0.07 (0.03)
Adult day-care and residential health providers	−0.009 (0.021)	−0.23 (0.18)					
Acute and extended health care providers	0.056 (0.024)			0.04 (0.05)	0.06 (0.04)	0.10 (0.03)	0.06 (0.02)
Hospital beds	−0.005 (0.027)						
Ambulance per 1000 population	0.051 (0.043)						
Emergency services: non-threating routine calls	0.033 (0.021)				0.07 (0.04)	0.03 (0.02)	
Medical health professionals	−0.059 (0.026)						
Adult male violent crimes	0.010 (0.042)		0.24 (0.11)	0.20 (0.08)	0.14 (0.06)	0.05 (0.05)	0.02 (0.04)
Variance components							
$\hat{\sigma}^2$	0.033 (0.016)	0.26 (0.30)	0.26 (0.08)	0.17 (0.12)	0.035 (0.02)	0.06 (0.03)	0.06 (0.03)
$\hat{\lambda}$	0.031 (0.120)	0.00 (−)	0.00 (−)	0.33 (0.46)	0 (−)	0.14 (0.20)	0.11 (0.18)

probable that some of the variables included in our analysis partly account for these effects. For example, it is possible that boys of female lone parent families are less likely to have access to an automobile and travel less in number of hours or in mileage, and that socio-economically disadvantaged neighbourhoods may have poorer road environment.

2.2. The basic Bayesian spatial model with spatial autocorrelation

Our ecological analysis of injury rate and risk factors aims to assess the geographical associations between injury occurrence and covariate effects and to explore possible residual variation arising from latent or unobserved risk factors. We therefore begin the analysis with an in-depth investigation into whether there are significant regional variations in injury rates and, if the answer is yes, can such variations be explained by potential risk factors.

To investigate the injury variation alone, we first fit the following model

$$\log(\mu_{ij}^{b}) = \log(n_{ij}) + a_0 + S_0(i) + b_j \qquad (1)$$

where μ_{ij}^{b} represents the expectation of Y_{ij} conditioning on random spatial effects $\boldsymbol{b}$; $\boldsymbol{b} = (b_{1,1}, \ldots, b_{1,J})^t$, $J = 83$. The term $\log(n_{ij})$ is an offset; a_0 is a fixed effect, $\exp(a_0)$ index the mean injury rate over all local health areas; $S_0(i)$ is a fixed spline representing age effect. The $\exp(b_{\mathrm{j}})$s represent local health area age-adjusted injury ratios (also called relative risks).

There are two key assumptions underlying the proposed statistical model. Firstly, conditioning on the random effects $\boldsymbol{b}$, for the given time period, a given age group i and a given local health area j, the corresponding MVA injury hospitalisation count y_{ij} is assumed to follow a Poisson distribution, $y_{ij}|\boldsymbol{b} \sim \mathrm{Poisson}(\mu_{ij}^{b})$. Secondly, the random spatial effects are assumed to have a spatially structured (joint) prior distribution, i.e. a Markov random field (MRF) Gaussian distribution. We assume $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}j)^t \sim \mathrm{MVN}(0, \Sigma(\sigma^2, \lambda))$, where

$$\sum(\sigma^2, \lambda) = \sigma^2 D^{-1}, \qquad D = \lambda Q + (1-\lambda) I_J \qquad (2)$$

with σ^2 representing Poisson overdispersion and λ the spatial autocorrelation, $0 \leq \lambda \leq 1$; I_J is an identity matrix of dimension J; the neighbourhood matrix Q has jth diagonal element equal to the number of neighbours of the corresponding area, and the off-diagonal elements in each row equal to -1 if the corresponding areas are neighbours and 0 otherwise (Leroux et al., 1999; MacNab and Dean, 2000). In this application, neighbourhood is defined by areas which share a common border, a definition commonly used in Bayesian disease mapping. The resulting spatial model therefore depends only on the neighbourhood structure and not on, say, the distance between areas. Wakefield and Morris (1999) and MacNab and Dean (2000, 2001), among others, discuss more general formulations of neighbourhood, for example, the incorporation of distance (centroid)-based weights into the modelling of spatial dependence between areas. The rationale behind the MRF Gaussian assumptions above is to accommodate random spatial effects resulting from unmeasured/unknown risk factors that induce spatial autocorrelation between neighbouring areas. The Bayesian hierarchical model, with Poisson likelihood of injury occurrence given random spatial effects, also facilitates spatial smoothing and data sharing.

MacNab and Dean (2001) proposed hierarchical model with MRF Gaussian prior for random spatial effects and spline smoothing for time as well as age effects. The incorporation of regression splines makes it possible to explore non-linear age and time trends without having to assume certain functional form for the effects. In this particular application, we assume $S_0(i)$, the spline fitting of the age effects, to be an order-2 cubic B-spline with two inner knots. Without the intercept, four parameters are required to specify the spline. If the age effects were modelled by separate parameters for each age group, five parameters are required (without the intercept). The key feature of the spline fitting is that it requires relatively fewer (spline) parameters to give a good fit to the age effects. The advantage of spline smoothing can be much greater when an analysis involves a larger number of age groups, for example, the usual 19 age groups (<1, 1–4, 5–9,. . . , 85+) for population of all ages. In a recent investigation of injuries of all causes, we have found that the age effects follow a U-shaped trend and an order-2 cubic B-spline with two inner knots fits the effects well.

2.3. The Bayesian ecological regression model with age effect

To examine the association between injury and risk factors and to assess whether significant residual variation remains after accounting for potential risk factors, the following model is fitted to the BC MVAI data

$$\log(\mu_{ij}^{b}) = \log(n_{ij}) + a_0 + S_0(i) + \sum_{q=1}^{18} \beta_q x_{q,j} + b_j \qquad (3)$$

The β_qs are the regression coefficients while the specifications for the rest of the model components remain unchanged. Model (3) is a natural extension of model (1), i.e. the inclusion of covariate information. Within a unified model framework, we simultaneously evaluate age effects, covariate effect, residual variation, and spatial autocorrelation. We emphasise that, in model (3), the random spatial effects b_js may serve to model the effects of unmeasured risk factors (i.e. latent effects) that are independent of the variables considered.

In this particular application, the basic unit of the ecological regression model is local health area. It is the spatial distribution of the local health area injury rates and the spatial links between local health area characteristics and injury rates that are of primary interest. The underlying assumption that we have made here, following an ecological interpretation of child and youth injury risk, is that the spatial allocation of the 'at-risk' population (males of age 0–24), based on the geographical areas in which they reside, puts them at different risk of injury and that such risk differential is associated with 'population exposures' commonly known as life-style/neighbourhood exposures such as socio-economic deprivation, female lone parent family, violent crime, etc. Past studies of child and youth injury indicate links between socio-economic indicators and injury risk. This suggests increased risk of injury among youth who live in socio-economically deprived neighbourhoods (Matheny, 1997; Valsiner and Lightfoot, 1987; Dowswell et al., 1996; Faelker et al., 2000). Various census variables and data that reflect social and economic characteristics of the areas in which children and youth live have been previously analysed to quantify associations between injury and socio-economic factors (Dougherty et al., 1990; Joly et al., 1991; Roberts et al., 1992; Faelker et al., 2000). Our analysis aims to quantify the regional experience of child and youth injury in this province and explore potential risk factors and contextual determinants with a view toward community injury prevention and control.

2.4. Bayesian ecological regression model for age-specific injury

Recognising the fact that the influence of regional/community level characteristics on child and youth injuries can vary with age, we further investigate the injury–covariates association by six age groups (<1, 1–4, 5–9, 10–14, 15–19, 20–24) respectively; the following model is fitted to the corresponding age-specific injury data

$$\log(\mu_{ij}^{b}) = \log(n_{ij}) + a_0 + \sum_{q=1}^{18} \beta_{q,i} x_{q,j} + b_j^{(i)}$$

$$\text{for } i = 1-6, \tag{4}$$

where $\{b_j^{(i)}\}_{j=1}^{83}$ represents random spatial effects for the particular population age group (the ith age group) under investigation.

3. Results

Within a generalised linear mixed model (GLMM) framework, inference for these models can be carried out using full Bayesian or empirical Bayes methods. Within the empirical Bayes inference context, various authors have discussed the implementation of penalised quasi-likelihood (PQL) estimation (Breslow and Clayton, 1993; Leroux et al., 1999; MacNab and Dean, 1999; to name a few); in this analysis, all model inferences were carried out using PQL method.

The age effect spline is fitted by replacing $S_0(i)$ with a regression (cardinal) spline term $\sum_{k=1}^{K} \alpha_{0k} B_k(i)$, where $\{B_k(i)\}$, $k = 1$–4, is a set of basis functions of an order-2 cubic B-spline that has its intercept excluded and has two inner knots at the corresponding quantiles, i is centred: $i = -2.5, -1.5, \ldots, 2.5$ (Hastie and Tibshirani, 1990; MacNab and Dean, 2001; MacNab, 2003b). We note that the spline fitting can be readily implemented as additive terms within the Bayesian spatial and ecological model framework. The resulting model inference can be implemented by statistical software (for example, SAS or Splus) with a GLMM inference capability (MacNab, 2003b).

In male children and youth of age 0–24, there were 13,660 MVA injuries that resulted in at least 6 h of hospitalisation in British Columbia during the 1990–1999 time period, about two injuries out of every 1000. The age-specific injury rate rose from two (per 10,000) in infant boys to five in boys of age 1–4, seven in boys of age 5–9, 10 in boys of age 10–14, 43 in boys of age 15–19, and 44 in young men of age 20–24; noting a sharp increase in the rate change from boys of age 10–14 to boys of age 15–19, an increase of 340%. The fitted model (1) suggests a significant regional variation in local health area injury risk in males aged 0–24. Fig. 1, a map of age-adjusted MVA injury ratios ($\hat{b}_j s$), reveals a spatial pattern of relatively high injury ratios cross the mainland from the north–west down to the south interior. A cluster of high ratios is also observed in the south–east in terior. The notable ratio variation is also reflected in the random component parameter estimates, with a relatively large overdispersion ($\hat{\sigma}^2 = 0.41$, S.E. 0.081) and a considerable spatial autocorrelation ($\hat{\lambda} = 0.93$, S.E. 0.16). Fig. 2 plots the fitted age effect spline ($\hat{S}_0(i)$) the points graph the estimated age effects ($\hat{a}_i$) using separate parameters for each age group.

The fitted model (3) depicts marked drops in the covariates-adjusted injury ratios $\hat{b}_j$ in comparison with their corresponding unadjusted ratios, i.e. the ratios of model (1). The difference in ratio variations is clearly displayed in Fig. 3, suggesting that the regional effects have been largely eliminated. From the fitted model (3), $\hat{\sigma}^2 = 0.033$ (S.E. 0.016), $\hat{\lambda} = 0.031$ (S.E. 0.12). These is also an indication of small (marginally significant) residual variation. Note that the inclusion of the area-specific covariates also removed most of the spatial autocorrelation. Table 1 lists the estimates of the regression coefficients and the variance parameters. While positive association were seen with low socio-economic status, seatbelt violations, and acute and extended health care providers, recent immigrants and medical health professionals appeared to be inversely related to the injury. The estimates of regression slopes indicate covariates of little and insignificant influence on injury variation; for simplicity (i.e. parsimony of model parameters) they may be eliminated from the model. To conserve space, however, the reduced model is not presented here.

The fitting of model (4) to the age-specific injury data enables us to examine injury–covariate associations for young males at similar developmental stages and to explore the changes of risk determinants for those of different ages.

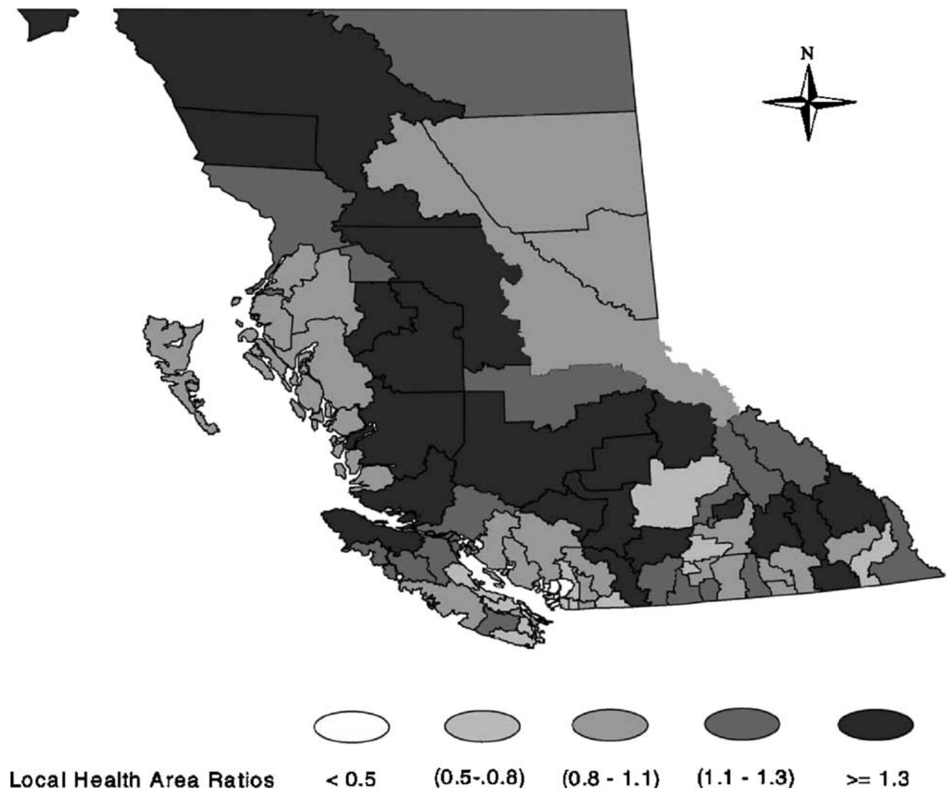

Fig. 1. Age-adjusted ratios by local health area. Hospitalisation due to motor vehicle accident injury, males aged 0–24, model (1), British Columbia, Canada, 1990–1999.

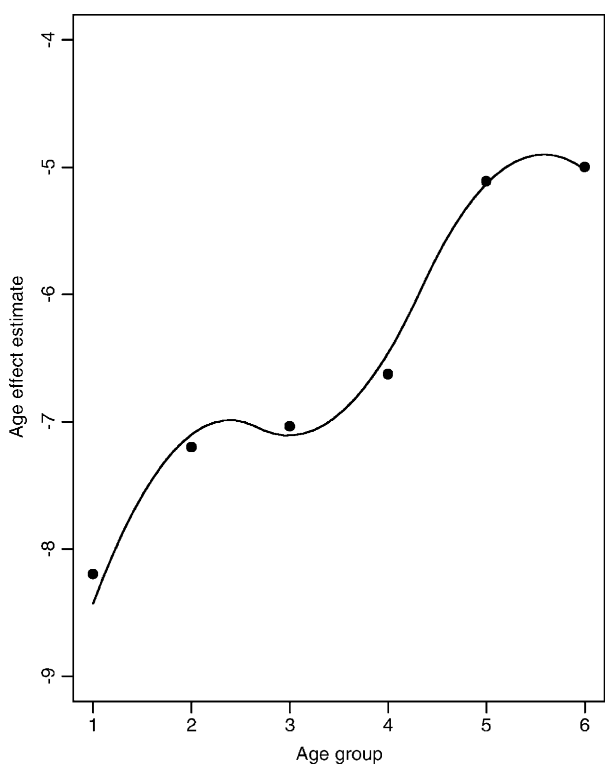

Fig. 2. Fitted age effect spline $\hat{S}_0(i)$ (line) and estimated age effects $\hat{a}_i$ (point) using separate parameters for each age group. Hospitalisation due to motor vehicle accident injury, males aged 0–24, British Columbia, Canada, 1990–1999.

Table 1 also presents the estimates of the model parameters for each of the age-specific injury data. The six fitted models each contain only those covariates of large or significant association that were derived by taking into consideration (a) parsimony of parameters, (b) reduction in residual variation, and (c) goodness of fit (based mainly on residual diagnos-

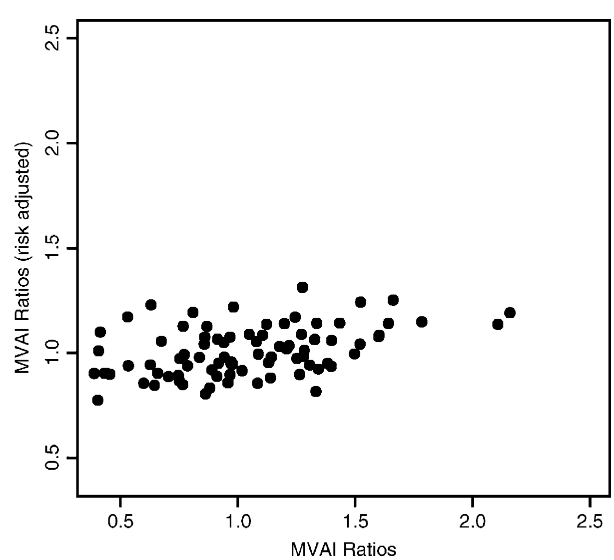

Fig. 3. Comparison of risk-adjusted model (3) and unadjusted ratios, model (1). Hospitalisation due to motor vehicle accident injury, males aged 0–24, British Columbia, Canada, 1990–1999.

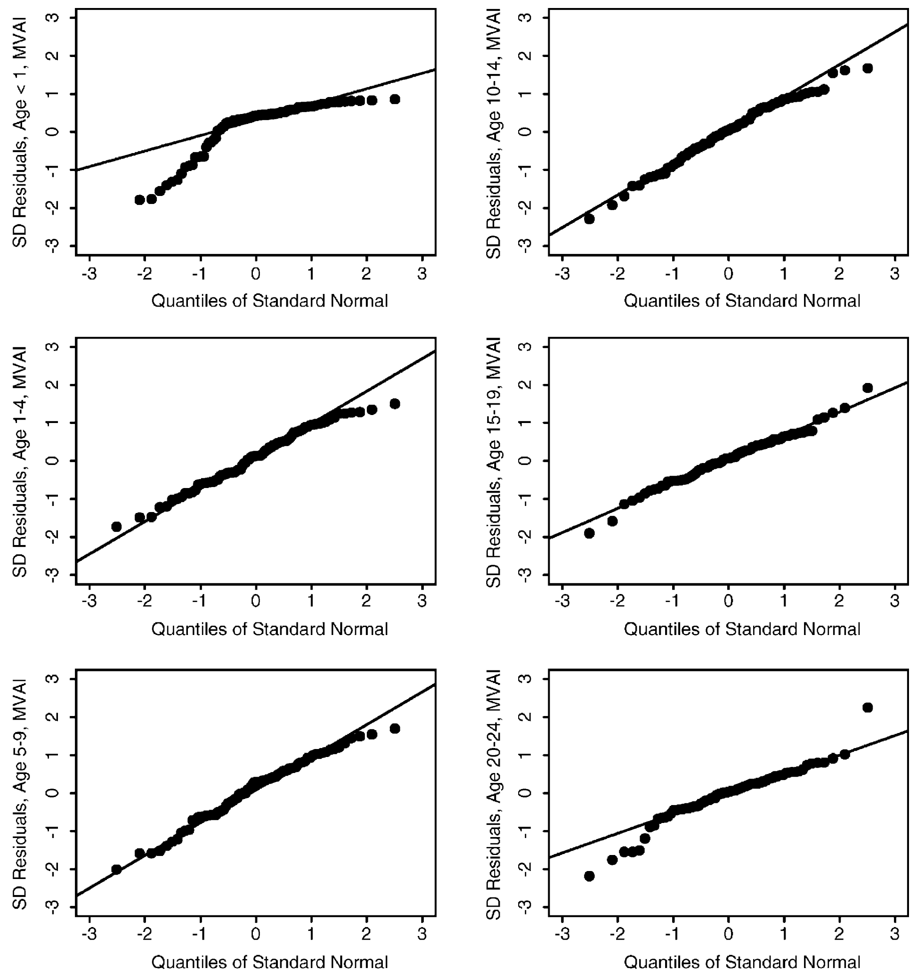

Fig. 4. Probability plots of the standardised residuals, Bayesian spatial and ecological models (4). Hospitalisation due to motor vehicle accident injury, young males of six age groups, British Columbia, Canada, 1990–1999.

tic plots). Fig. 4 presents residual diagnostic plots, i.e. QQ normal with line plots, for the six fitted models. The QQ plots graph the quantiles of the standardised residuals (on the y-axes) against the quantiles of standard normal (on the x-axes). This figure displays six probability plots of standardised residuals $(\hat{y}_i - y_i)/\sqrt{\mathrm{Var}(\hat{y}_i)}$, with each plot corresponding to the fitted model of a specified age group i. Table 1 lists the estimates of dispersion and spatial autocorrelation parameters. They correspond to the residual variations modelled by the random spatial effects b_js, for each specific age group. Overall, the final models fit adequately for boys of age groups 2–6 as adjusting for the risk factors eliminated nearly all injury variation and spatial autocorrelation. After accounting for risk factors considered here, moderate regional effects remained only for boys of age 1–4. Fig. 5 maps the risk-adjusted ratios. These estimated ratios appear to show no particular spatial pattern. This is also indicated in the estimate of the spatial autocorrelation parameter, which is 0 (Table 1). MVA injury occurs very rarely in infants. During the 1990–1999 time period, 60 out of 83 local health areas had no injury to male infants. As a result, considerable departure from normality was observed from its residual plot (see Fig. 4).

In summary, MVA injury was generally more common in male children and youth living in a lower rated socio-economic area and it appeared to be so for all age groups. This finding is consistent with those reported in the literature (Roberts et al., 1992; Jolly et al., 1993; Faelker et al., 2000). Among all age groups except those age 5–9, a negative association was seen with the rate of female lone parent families. A possible explanation is that children of female lone parent families are less likely to have access to and travel less in an automobile. They would therefore be less likely to be involved in a MV accident. The study suggested a positive association between injury rate and adult male crime rate, but only for boys of age 1–14. More studies are needed to provide adequate explanation. Seatbelt violaions and excess speeding charges were shown to be positively associated with MVA injury. The statistically significant regression slope for young men of age 20–24, who are at the legal age to drive, may have a behavioural explanation. It is commonly known that seatbelt violation

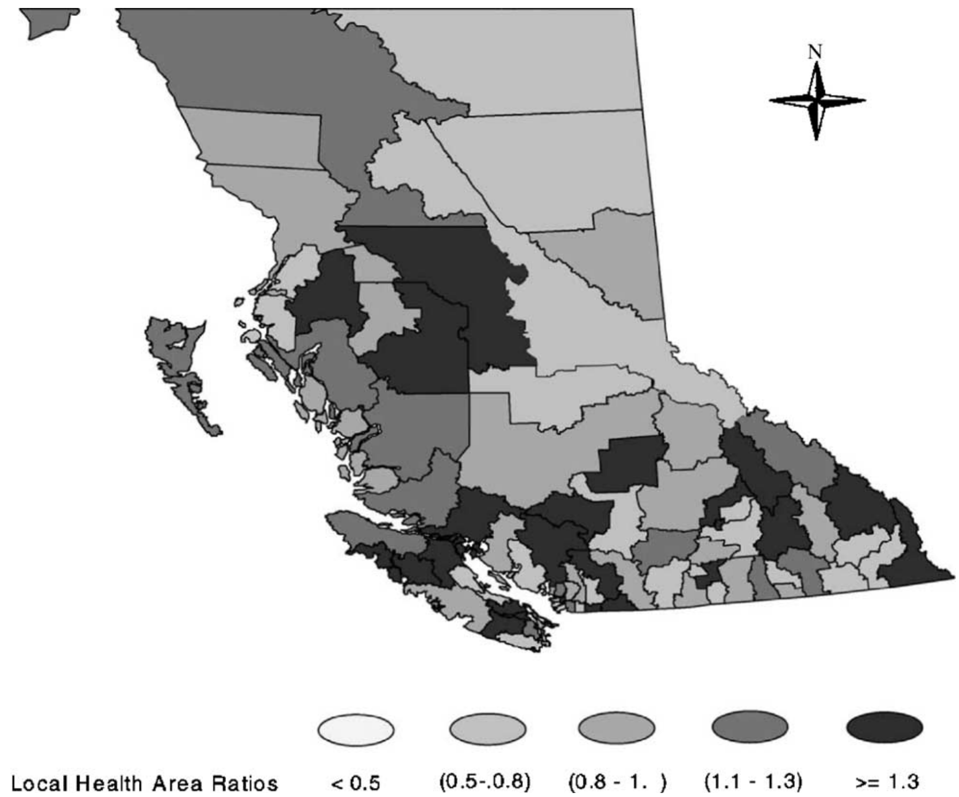

Fig. 5. Risk-adjusted ratios by local health area. Hospitalisation due to motor vehicle accident injury, boys aged 1–4, British Columbia, Canada, 1990–1999.

and speeding are high-risk behaviours that are linked to serious MV accident and injury. Studies have shown that male gender and youth are significant correlates of such behaviours (Knight et al., 1991). It is worth noting that rate of recent immigrants was negatively associated with rate of MVA injury; the regression slope was statistically significant for males of age 15–24. Although, the association is not widely reported and well understood, cultural and life-style characteristics of the immigrant population may be conducive to children less engaging in risk-taking or aggressive behaviours. Provision of non-hospital care appeared to have a considerable influence on MVA injury hospitalisation among infants; communities with higher rates of adult day-care and residential health providers per capita had lower rates of injury-related infant hospitalisations.

Overall, this study suggests that, among young males of all age groups, there were considerable local health area variations in MVA injury and these variations were largely explained by the mid-period census and non-census regional variables considered. There was a general consistency in the estimates of the regression coefficients across the age groups. It is to be expected that for boys of age 0–19 (age groups 1–5), those involved in MV accident are most likely the passengers and victims, not the drivers. As the at-risk population enters young adulthood and the legal age to drive, the proportion of injured while behind the wheel should rise. Also, it is anticipated that the number of hours or miles children travel in automobiles increases with age. These and similar factors may have been reflected in the gradual changes in the exposure risk relations observed; as the age increased, the injury rate rose, and more covariates were identified as important variables and potential sources of the injury variation.

4. Discussion

This article illustrates a Bayesian statistical model frame-work within which ecological analysis of accident and injury variations, covariate effects, random spatial effects and age effects may be considered simultaneously. We demonstrate that the modelling techniques enable us to assess spatial variation of injury rates and the association between injury rate and risk exposures in a cohesive way. Its advantage over conventional methods lies in its ability to account for extra variation over space. Also, its extended scope and model flexibility accommodate spatial effects, as well as non-linear age effects without imposing rigid parametric and functional assumptions. Although, the statistical framework and similar models are presented and discussed in recent literature, their application to accident and injury analysis remains relatively unexplored. The significance of Bayesian ecological studies in general is the prediction of spatial patterns of 'outcome' (say, injury or mortality) variation and the potential utility of such prediction in identifying 'areas in need'. In this study, we exemplify the methods and their application through an ecological analysis of British Columbia child and youth motor vehicle accident injury rates, focusing mainly on the evaluation of ecological determinates of socio-economical and health services context.

Our study reveals geographical/spatial patterns in injury ratios but does not indicate any areas with exceptionally elevated injury rate, once appropriate area-level attributes and non-linear age effect are taken into account. In boys and young men of different age groups, moderate regional effects remained only for boys of age 1–4. It is therefore important to note that this study suggests area-level characteristics that are important contextual determinates of the children and youth MVA injury and that these variables explained majority of the regional MVA injury variations. As the analysis is based on data of MVA injury hospitalisation, our findings are relevant to the severity of injuries and the resultant hospital utilisation. Ecological regression studies of this and similar kind shed light on reasons for regional variations in accident occurrence and thereby are potentially useful in targeting priority areas for injury/accident prevention and in informing regional health planning and policy development.

The Bayesian spatial and ecological model framework should be applicable to accident and injury data of various forms and sources. The modelling technique may be readily applied for accident and injury prevention and evaluation cycles. Regional occurrence could be routinely monitored and potential area-level (or ecological) risk factors examined and assessed. The modelling framework also applies to surveillances where accident and injury occurrence are monitored for collection of event locations. For instance, one may incorporate ecological risk factors, such as external physical environment (road condition, traffic light system at major road intersections) into the analysis of accident and injury incidence where group-level exposure–response association would be examined for place-of-occurrence. In addition, the model framework can be readily extended to include time component and time-varying/seasonal covariates so that a systematic evaluation of accident/injury prevention (e.g. intervention study) may be implemented.

We note that an important issue, commonly known as 'error-in-variable', may arise in ecological regression studies of disease epidemiology when within-area measurement errors of the risk factor of interest are large in comparison with between-area differences in mean exposure, resulting in what we call covariate measurement error. Measurement error may also occur in other situations when, for example, covariates are measured imprecisely or surrogates are used for the true covariates that are unmeasurable or difficult to measure. Failing to account for such errors may result in biased estimates of the true regression slope. Best (1999); Elliott et al. (1992); Lawson et al. (1999); Wakefield and Morris (1999) discuss Bayesian ecological models that account for covariate measurement errors (i.e. Bayesian 'error-in-variable' models). In this study, the covariate measurements are assumed to be reasonably accurate; a preliminary study appears to support this. The model framework presented here can be extended to make appropriate adjustment for 'error-in-variable' and this extension is currently under investigation.

Acknowledgements

This work was partially supported by a research grant from the Natural Sciences and Engineering Research Council of Canada and by an establishment grant from the British Columbia Research Institute for Children's and Women's Health. The author thanks two anonymous referees for their thorough reviews, comments and suggestions which have led to substantial improvements of the article. The author thanks the BC Ministry of Health, BC stats, the Canadian Centre for Justice statistics, and the BC Injury Research and Prevention Unit for provision of the data.

References

Best, N., 1999. Bayesian ecological modelling. In: Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R., (Eds.), Disease Mapping and Risk Assessment for Public Health. Wiley, New York, pp. 194–201.

Biggeri, A., Divino, F., Frigessi, A., Lawson, A., Böhning, D., Lesaffre, E., Viel, J.-F., 1999. Introduction to spatial models in ecological analysis. In: Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R. (Eds.), Disease Mapping and Risk Assessment for Public Health. Wiley, New York, pp. 181–192.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalised linear mixed models. J. Am. Stat. Assoc. 88, 9–25.

Brumback, B.A., Rice, J.A., 1998. Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). J. Am. Stat. Assoc. 93 (443), 961–994.

Clayton, D.G., Bernardinelli, L., Montomoli, C., 1993. Spatial correlation and ecological analysis. Int. J. Epidemiol. 22, 1193–1201.

Dean, C., MacNab, Y.C., 2001. Modelling of rates over a hierarchical health administrative structure. Can. J. Stat. 29, 405–419.

Dougherty, G., Pless, I.B., Wilkins, R., 1990. Social class and the occurrence of traffic injuries and deaths in ubran children. Can. J. Public Health 81, 204–209.

Dowswell, T., Towner, E.M.L., Simpson, G., Jarvis, S.N., 1996. Preventing childhood unintentional injuries: what works? A literature review. Inj. Prev. 2, 140–149.

Elliott, P., Cuzick, J., English, D., Stern, R. (Eds.), 1992. Geographical and Environmental Epidemiology: Methods for Small-Area Studies. Oxford University Press, London.

Faelker, T., Pickett, W., Brison, R.J., 2000. Socioeconomic differences in childhood injury: a population based epidemiologic study in Ontario Canada. Inj. Prev. 6, 203–208.

Ferrándiz, J., López, A., Sanmartin, P., 1999. Spatial regression models in epidemiological studies. In: Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R. (Eds.), Disease Mapping and Risk Assessment for Public Health. Wiley, New York, pp. 203–215.

Green, P.P., Silverman, B.W., 2000. Nonparametric Regression and Generalised Linear Models: A Roughness Penalty Approach. Chapman & Hall, New York.

Hastie, T.J., Tibshirani, R.J., 1990. Generalised Additive Models. Chapman & Hall, New York.

Ivan, J., O'Mara, P., 1997. Prediction of traffic accident rates using Poisson regression. In: Proceedings of the 76th Annual Meeting of the Transportation Research Board.

Jolly, D.L., Moller, J.N., Volkmer, R.E., 1993. The socio-economic context of child injury in Australia. J. Paediatr. Child Health 29 (6), 438–444.

Joly, M.F., Foggin, P.M., Pless, I.B., 1991. Geographical and socio-ecological variations of traffic accidents among children. Soc. Sci. Med. 33, 765–769.

Joshua, S., Garber, N., 1990. Estimating truck accident rate and involvement using linear and Poisson regression models. Transportation Plann. Technol. 15, 41–58.

Knight, K.K., Fielding, J.E., Goetzel, R.Z., 1991. Correlates of motor-vehicle safety behaviours in working populations. J. Occup. Med. 33 (6), 705–710.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R., (Eds.), 1999. Disease Mapping and Risk Assessment for Public Health. Wiley, New York.

Leroux, B.G., Lei, X., Breslow, N., 1999. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology, the Environment and Clinical Trials. Springer, New York, pp. 135–178.

MacNab, Y.C., Dean, B.C., 2000. Parametric bootstrap and penalised quasi-likelihood inference in conditional autoregressive models. Stat. Med. 19 (15–16), 15–30.

MacNab, Y.C., Dean, B.C., 2001. Autoregressive spatial smoothing and temporal smoothing B-splines for mapping rates. Biometrics 57 (3), 949–956.

MacNab, Y.C., Dean, B.C., 2002. Spatio-temporal modelling of rates for the construction of disease maps. Stat. Med. 21 (3), 347–358.

MacNab, Y.C., 2003a. Hierarchical Bayesian Spatial Modelling of Small-Area Rates of Non-Rare Diseases. Stat. Med. 22 (10), 1761–1773.

MacNab, Y.C., 2003b. A Bayesian hierarchical model for accident and injury surveillance. Accid. Anal. Prevent 35 (1), 91–102.

MacNab, Y.C., 2002. Generalised additive mixed models with splines: Bayesian disease mapping and spatio-temporal surveillance. In: Proceedings of the 2002 Hawaii International Conference on Statistics, pp. 1–28.

Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.B., Pellom, A.C., 1989. Empirical Bayes procedures for stabilising maps of US cancer mortality rates. J. Am. Stat. Assoc. 84, 637–650.

Matheny, A.P., 1997. Psychological characteristics of childhood accidents. J. Soc. Issues 43 (2), 45–60.

Pickle, L.W., 2000. Exploring spatio-temporal patterns of mortality using mixed effects models. Stat. Med. 19 (17–18), 2251–2264.

Plummer, M., Clayton, D., 1996. Estimation of population exposure in ecological studies. J. R. Stat. Soc. Ser. B 58, 113–126.

Richardson, S., 1992. Statistical methods for geographical correlation studies. In: Elliott, P., Cuzick, J., English, D., Stern, R. (Eds.), Geographical and Environmental Epidemiology: Methods for Small-Area Studies. Oxford University Press, London, pp. 181–204.

Roberts, I., Marshall, R., Norton, R., Borman, B., 1992. An area analysis of child injury morbidity in Auckland. J. Paediatr. Child Health 28 (6), 438–441.

Rosenbaum, P.R., Rubin, D.B., 1984. Difficulties with regression of age-adjusted rates. Biometrics 40, 437–443.

Sun, D., Tsutakawa, R.K., Kim, H., He, Z., 2000. Spatio-temporal interaction with disease mapping. Stat. Med. 19 (15), 2015–2036.

Valsiner, J., Lightfoot, C., 1987. Process structure of parent–child-environment relations and the prevention of children's injuries. J. Soc. Issues 43, 61–72.

Wakefield, J., Morris, S., 1999. Spatial dependence and error-in-variables in environmental epidemiology. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F. (Eds.), Bayesian Statistics, vol. 6. Clarendon Press, Oxford, pp. 657–684.

Walter, S.D., Taylor, S.M., Marrett, L.D., 1999. An analysis of determinants of regional variation in cancer incidence: Ontario, Canada. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F. (Eds.), Bayesian Statistics, vol. 6. Clarendon Press, Oxford, pp. 657–684.