MODELLING AND ANALYSIS OF DISEASE INCIDENCE RATES BY AGE-GROUPS OVER REGIONS

Giovani L. Silva
CEAUL & DMIST - Technical University of Lisbon
gsilva@math.ist.utl.pt

Charmaine Dean
DSAS/FS - University of Western Ontario
scidean@uwo.ca

Disease mortality and incidence rates by small areas have usually been represented through maps in order to identify spatial and temporal effects and risk factors of the diseases. Because of the need for producing reliable estimates for these rates, recent developments in disease mapping have had tremendous impact in health policy and epidemiology. The aim here is to develop methods for modelling and mapping of disease mortality and incidence rates by age-groups over public health regions. These methods account for (i) over-dispersion (ii) spatial correlation (iii) nonlinear age-specific effects, using hierarchical Bayesian spatial models with age-specific smoothing. Although the associated joint posterior distribution is awkward to work with, since the marginal posterior distributions of some parameters are not easy to obtain explicitly, posterior estimates can be easily obtained using Markov chain Monte Carlo (MCMC) methods. An analysis of spatial age-specific lung cancer rates in the province of Ontario motivates and illustrates the methods developed. That cancer is the leading cause of death due to cancer in Canada, representing an estimated 30% of the cancer deaths in males and 25% of the cancer deaths in females.

Keywords: Disease mapping, Age-specific smoothing, Hierarchical spatial model, Generalized additive mixed models, Bayesian analysis.

1 INTRODUCTION

Modelling and analysis of disease mortality, morbidity or incidence rates over an area is often called **disease mapping**. Because of the need for producing reliable rate estimates for small population areas, recent developments in disease mapping have had tremendous impact in health policy and epidemiology. The idea behind the developments in spatial modelling of disease rates is essentially to model heterogeneity in the rates and better separate

systematic variability from random noise, the component that usually overshadows the SMR maps. There is a wide variety of topics on disease mapping, using both empirical Bayes and fully Bayesian methods. For a review of disease mapping or spatial analysis, see for instance Clayton and Kaldor (1987), Lawson et al. (1999), Dean and MacNab (2001), Waller and Gotway (2004), Banerjee et al. (2004), Ainsworth and Dean (2008), and Silva, Dean et al. (2008).

The rates and standard errors provided at the disease mapping are often calculated based on the assumption of a Poisson distribution assumption using traditional methods. Since counts often display over-dispersion, there is a concern that associated standard errors can be then seriously underestimated. This led to the current investigation to provide new methods which address this concern. The aim here is to present a modelling and mapping of disease mortality and incidence rates by age-groups over public health regions. We develop and validate such new methods for making inference on disease rates and for calculating reliable standard errors that account for (i) over-dispersion (ii) spatial correlation (iii) nonlinear age-specific effects, using hierarchical Bayesian spatial models with age-specific smoothing via Markov chain Monte Carlo (MCMC) methods. We will focus the illustration of the methods herein on estimating spatial age-specific lung cancer rates in the province of Ontario during the period 1995-2002, since that cancer is the leading cause of death due to cancer in Canada, representing an estimated 30% of the cancer deaths in males and 25% of the cancer deaths in females. The robust spatial age-specific models developed can be used generally for a broad range of diseases by policy makers nationally.

2 METHODOLOGY

In this section, we discuss the use of generalized additive mixed models for the spatial age-specific analysis of disease rates based on a Bayesian approach via MCMC methods. These models build upon the basic framework of the Poisson model to take account of the extra-variation of counts often observed when fitting a Poisson distribution, and introduce spatially correlated and uncorrelated random small-area effects and smoothing of the age-specific component. Wood (2006) provides a recent introduction to generalized additive models, whereas Banerjee et al. (2004) and the references therein present a comprehensive review of hierarchical Bayesian models for spatial and spatiotemporal data.

2.1 GENERALIZED ADDITIVE MIXED MODELS

Let Y_{ij} be the disease count observed out of n_{ij} individuals at risk in region i and age-group j, $i = 1, \ldots, r$, $j = 1, \ldots, a$. Denoting θ_{ij} the probability of disease (success) for any individual in region i and age-group j, a natural initial model for the outcome Y_{ij} is the binomial distribution, whose (maximum likelihood) estimate of the disease probability is y_{ij}/n_{ij} (crude disease

rate). A Poisson approximation to the binomial is an appropriate starting framework for small areas and rare diseases. Under age-standardization, the related mean is sometimes reparameterised as $\mu_i = E_i \eta_i$, where E_i is the expected number of disease counts in region i, and η_i is the associated small-area relative risk, $i = 1, \ldots, r$.

Age-standardization facilitates comparison among summary rates by accounting for different age profiles of the different regions. Under internal standardization, the expected disease count E_i becomes $\sum_j n_{ij} q_j$, where $q_j = (\sum_i y_{ij} / \sum_i n_{ij})$ is the overall rate for age-group j. The ratio of the observed and expected disease numbers for region i is a simple estimate of the small-area effect η_i, and is called the standardized mortality ratio (SMR), i.e.,

$$SMR_i = Y_i / E_i \,. \tag{1}$$

Because of the need for producing more reliable rate estimates especially for small population areas, recent developments in disease mapping have had tremendous impact in health policy and epidemiology. The idea behind the developments in spatial modelling of disease rates is essentially to model heterogeneity in the rates and better separate systematic variability from random noise, the component that usually overshadows the maps based on the quantity (1).

We propose here spatial age-specific modelling of disease rates based on a flexible spatial generalized additive mixed model developed by MacNab and Dean (2001) and adapted here for the spatial age-specific analysis from a Bayesian perspective (see also Silva, Dean et al., 2008). Conditional on the small-area effects, and assuming Y_{ij} has approximately a Poisson distribution with mean μ_{ij}, we define the spatial age-specific models as

$$\log \mu_{ij} = S_0^\star(j) + S_i^\star(j), \tag{2}$$

where $S_0^\star(j)$ represents the overall age effects, over the whole region, and $S_i^\star(j)$ represents the small-area age-specific trend effects indicating how the small-area age effects deviate from the overall age trend. Notice that, from model (2), we can obtain a simple linear model in age for the r (independent) regions, i.e., $\log \mu_{ij} = \alpha_0 + \beta j$, as well as a wide class of other Poisson models with a log link function. It is convenient in the analysis to center the age-group variable (e.g., -0.5,0,0.5 for 3 age-groups) and we use such centering in the subsequent discussion. A very flexible way to model trends in general is to partition the time axis into a few segments and fit different polynomials to each of these subintervals. These piece-wise polynomials can fit a very large class of functions. Splines are piece-wise polynomials joined at the breakpoints between segments, known as knots, with smoothness imposed at the knots. They are flexible enough to fit a wide variety of functions even those which are slowly varying. Splines can be written as a

linear combination of spline basis functions which enables straightforward representation and fitting in the Poisson regression framework.

For model (2), we adopt a convenient and flexible formulation $S_0^*(j) = \alpha_0 + S_0(j)$, where $S_0(j)$ takes the form of a spline without intercept and, in this case, α_0 represents the overall mean rate over the region at mid-period. Similarly, region-specific effects are defined as $S_i^*(j) = S_i(j) + b_i + h_i$, $S_i(j)$ modelled as a linear age trend or, more flexibly, an area-specific spline, and b_i and h_i are small-area relative risks, b_i being spatially correlated and h_i being uncorrelated random effects. In some situations epidemiologists may be interested in incorporating small-area covariates such as deprivation indices or median income in ecological studies. For such investigations, covariates may also be included in model (2) additively. Note that such studies can only be exploratory with regards the effects of covariates as individual-level data is required to move beyond the ecological framework.

If there are a moderate number of age-groups, assuming a cubic spline without intercept and with one inner knot for $S_0(t)$ and a linear trend for $S_i(t)$, provides a simple flexible model from (2):

$$\log \mu_{ij} = \alpha_0 + \sum_{k=1}^{4} \beta_{0k} p_k(j) + \delta_i j + b_i + h_i, \tag{3}$$

where $p_k(\cdot)$, $k = 1, \ldots, 4$, are the spline basis functions and β_{0k} the corresponding coefficients, δ_i is the linear age effect related to the ith region, $i = 1, \ldots, r$. B-spline smoothing is proposed for describing the nonlinear age effects and may be used for both the overall age $S_0(j)$ and the spatial age-specific $S_i(j)$ components. For the latter case, the term $\delta_i j$ in (3) is replaced by $S_i(j) = \sum_{k=1}^{4} \beta_{ik} p_k(j)$, where β_{ik}, $k = 1, \ldots, 4$, are random effects that correspond to the age effects at the regional level. Nevertheless, model (3) offers reasonable flexibility in modelling these age-specific effects, and is parsimonious especially in terms of its interpretation. It may be particularly useful for exploratory analysis of age trends, where there are fewer than 10 age-groups under observation, and if diagnostics on a small-area level do not indicate peaks or troughs over and above the overall mean trend $S_0(j)$ – which would invalidate the approximation by a simple linear trend.

Other models have been presented in the literature for describing spatial age-specific distributions of rates. Nandram et al. (2000) focus on estimation of age-specific and age-adjusted mortality rates for chronic obstructive pulmonary disease to investigate the relationship between the geographic variation in mortality rates and the corresponding variation in selected covariates. They describe the spatial patterns in the age-specific maps and relate these to patterns in potential explanatory covariates such as smoking rate, annual rainfall, population density, elevation, and measures of air quality. There have also been studies concerning the possible need for inclusion of interaction effects between regions and age-groups in disease mapping;

Dean et al. (2001) propose a simple model for including such an interaction in order to develop a test for its significance. Their motivation concerns the assumption that the spatial effect is the same over age-groups. This would be implied in investigations which consider age-standardized mortality analysis. Such an assumption offers a simpler framework for analysis but interpretations should be clear on the fact that assuming the small-area effect is the same for all age-groups is only a starting point in a careful analysis. If small-area effects are driven by pollution differences across the regions, it may be that these effects would be more prominent on the elderly or very young. Clements et al. (2005) propose generalized additive models for lung cancer rate prediction by using both one-dimensional smoothing splines for classical age-period, age-cohort and age-period-cohort models and two-dimensional smoothing splines for age and period interaction. The methods proposed here provide simple tools which offer improvement over the use of SMRs, are fairly robust and flexible, and not overly complicated. Detailed analyses of any data set would undoubtedly reveal different, likely more complex, models as more suitable in any specific instance.

2.2 Hierarchical Bayesian models

In order to make inference on disease rates based on spatial age-specific models (2), we utilize Bayesian hierarchical approaches. In a Bayesian analysis, we also consider distributions of the model parameters, so-called **prior distributions**. Empirical Bayes and Bayesian hierarchical modelling have been used to obtain reliable estimates for disease rates in small areas by using information from all areas, but especially the local neighbourhoods, to derive estimates for individual areas. Banerjee et al. (2004) provide a comprehensive review of hierarchical Bayesian models for spatial and spatiotemporal data. Notice that, in Bayesian inference, the posterior distribution reflects the conditional distribution of all unknown model parameters, given the data, and represents a compromise between the prior assumptions and data information via prior distributions and the likelihood, respectively.

In model (3), we partitioned the spatial small-area random effect as the sum of spatially structured (b_i) and unstructured (h_i) components. Besag et al. (1991) argue that the use of such a convolution prior distribution allows that the data decide how much of the residual disease risk is due to both spatially structured variation and unstructured over-dispersion. If h_i is the dominant term, then there is little spatial correlation in the rates evident; if b_i is dominant, the spatial structure is important in the analysis. In short, h_i has a normal distribution with zero mean and variance σ_h^2, denoted here by $N(0, \sigma_h^2)$, $i = 1, \ldots, r$, independently. The hyperparameter σ_h^2 controls the amount of excess heterogeneity, how much variation in the small-area effects exists. Moreover, we also assume that the small-area trend effects $\delta_i \sim N(0, \sigma_\delta^2)$, $i = 1, \ldots, r$, independently, and independent of spatial random effects b_i and h_i. When considering cubic B-splines for $S_0(j)$ and $S_i(j)$

in (2), the corresponding spline basis functions β_{0k} and β_{ik} would be assigned independent normal distributions with zero mean and variances ν_k^2 and σ_k^2, respectively, $k=1,\ldots,4$.

The joint density of the spatially correlated vector $\mathbf{b} = (b_1,\ldots,b_r)$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma_b^2\mathbf{Q}^{-1}$, where $\mathbf{Q} = \{Q_{ij}\}$ is a matrix with entries specified by the spatial correlation postulated and σ_b^2 is the variance parameter that controls the amount of spatial similarity. Although we can use this representation of a joint multivariate Gaussian distribution for the spatially structured random effects, we may alternatively express it via a set of conditional distributions of each random effect (Besag et al., 1991). This leads to a more intuitive understanding of the spatial correlation. Conditionally, we have that $b_i|\mathbf{b}_{-i},\sigma_b^2 \sim N(\bar{b}_i,\sigma_b^2/n_i)$, where $\bar{b}_i = \sum_{j\in\mathcal{N}_i} b_j/n_i$, $\mathcal{N}_i$ denotes the set of "labels" of the "neighbours" of region i, n_i is the number of neighbours of region i, and $\mathbf{b}_{-i}$ denotes $(b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_r)$. Hence conditionally, the random small-area spatially correlated effect b_i is normally distributed with mean equal to the mean effect in the local neighbourhood. In other situations, the conditional mean of b_i might be more appropriately a weighted mean of all other small-area relative risks, with weights depending on the distance between the regions - higher weight given to regions which are close to the ith. Since we define neighbours as regions which share a boundary, the simple structure lets only immediate neighbours influence the conditional mean. In this case, $\mathbf{Q}$ has diagonal element $Q_{ii} = 1/n_i$ and the off-diagonal element Q_{ij} is 1 if regions i and j are neighbours and 0 otherwise, $i,j=1,\ldots,n$. This conditional autoregressive model (the so-called "intrinsic CAR model") often serves as a useful approximation to the complicated correlation structures which may underlie the responses.

For the hyperparameters σ_b^2 and σ_h^2, the variance components, one usually assigns an inverse gamma prior, i.e., $\sigma_b^2 \sim IG(c_1,d_1)$ and $\sigma_h^2 \sim IG(c_2,d_2)$, whose kernel density is equal to $x^{-(c+1)}\exp(-d/x)$, $x>0$. If $S_i(j) = \delta_i j$ with $\delta_i \sim N(0,\sigma_\delta^2)$, one may also consider $\sigma_\delta^2 \sim IG(c_3,d_3)$. If we model small-area trends as splines, $S_i(j) = \sum_k \beta_{ik}p_k(j)$, with $\beta_{ik} \sim N(0,\sigma_k^2)$, and we assign $\sigma_k^2 \sim IG(c_{4k},d_{4k})$, $k=1,\ldots,4$, $i=1,\ldots,r$; similarly, if $S_0(j) = \sum_k \beta_{0k}p_k(j)$ with $\beta_{0k} \sim N(0,\nu_k^2)$, we designate $\nu_k^2 \sim IG(c_{0k},d_{0k})$, $k=1,\ldots,4$. These inverse gamma priors, as well as the normal priors for the B-spline coefficients β_{0k}'s, are usually assigned highly dispersed, but proper, priors, reflecting that we have little prior information.

Assuming independence amongst the prior distributions aforementioned,

the joint posterior density for model (3), is then proportional to

$$\prod_{i=1}^{r}\prod_{j=1}^{a}\left\{\exp\left(y_{ij}\log\mu_{ij}-\mu_{ij}\right)\right\}\times\left(\frac{1}{\sigma_\delta^2}\right)^{\frac{r}{2}+c_3+1}\exp\left(-\frac{1}{\sigma_\delta^2}\left(d_3+\sum_{i=1}^{r}\frac{\delta_i^2}{2}\right)\right)$$
$$\times\left(\frac{1}{\sigma_b^2}\right)^{\frac{r}{2}+c_1+1}\exp\left(-\frac{1}{\sigma_b^2}\left(d_1+\sum_{i=1}^{r}\frac{n_i(b_i-\tilde{b}_i)^2}{2}\right)\right)\times\left(\frac{1}{\sigma_h^2}\right)^{\frac{n}{2}+c_2+1}\exp\left(-\frac{1}{\sigma_h^2}\left(d_2+\sum_{i=1}^{r}\frac{h_i^2}{2}\right)\right)$$
$$\times\prod_{k=1}^{4}\left\{\left(\frac{1}{\nu_k^2}\right)^{\frac{1}{2}+c_{0k}+1}\exp\left(-\frac{1}{\nu_k^2}\left(d_{0k}+\frac{\beta_{0k}^2}{2}\right)\right)\right\}\times[\alpha_0], \tag{4}$$

where $[\cdot]$ indicates the prior density. For further details, see Silva, Dean et al. (2008).

The joint posterior distribution (4) is awkward to work with, since the marginal posterior distributions of some parameters are not easy to obtain explicitly. Nevertheless, these posteriors can be evaluated using MCMC methods (see, e.g., Chen et al., 2000). In particular Gibbs sampling works by iteratively drawing samples for each parameter from the corresponding full conditional distribution, which is the posterior distribution conditional upon current values of all other parameters.

Sampling values from the conditional posteriors above is straightforward, especially when using the Metropolis-Hastings algorithm (Chen et al., 2000). Implementation in many statistical programming packages is not difficult. Alternatively, Gibbs sampling via the software GeoBUGS by Thomas et al. (2004) could be used; this provides estimates of posterior quantities of interest for the current model. However, note that the practical implementation of these techniques requires care in this scenario, since the models involve many parameters which are weakly identified.

2.3 Model comparison

Model selection of postulated sub-models of model (2) is done here through three comparison measures that are easily evaluated with MCMC methods: the Deviance (D), Pearson function (Q_P) and Deviance Information Criterion (DIC). That is,

- The posterior mean of the deviance based on Poisson likelihood,

$$D(\mu) = 2\sum_{i=1}^{r}\sum_{j=1}^{a}\left(y_{ij}\log\left(\frac{y_{ij}}{\mu_{ij}}\right)-(y_{ij}-\mu_{ij})\right). \tag{5}$$

- The posterior mean of the Pearson function,

$$Q_P(\mu)=\sum_{i=1}^{r}\sum_{j=1}^{a}\left(\frac{y_{ij}-\mu_{ij}}{\sqrt{\mu_{ij}}}\right)^2. \tag{6}$$

- The standard model comparison measure, proposed by Spiegelhalter et al. (2002), the Deviance Information Criterion. This is a generalization of the Akaike information criterion (AIC) for dealing with hierarchical Bayesian models of any degree of complexity, and is computed as the sum of two components: the expected posterior deviance (5) and the effective number of parameters, measuring the goodness of fit and complexity of the model, respectively. It is often expressed as

$$DIC = 2\overline{D(\mu)} - D(\overline{\mu}), \tag{7}$$

where $\overline{D(\mu)}$ and $\overline{\mu}$ denote the posterior mean of the deviance (5) and the model parameter vector μ, respectively.

3 ILLUSTRATION

In order to illustrate the modelling and analysis of disease incidence rates by age-groups over regions, we consider lung cancer incidence data in Ontario during the period 1995-2002. Lung cancer is the leading cause of cancer death in Canada, representing an estimated 30% of the cancer deaths in males and 25% of the cancer deaths in females.

For each of the 37 public health units (PHU) in Ontario (Statistics Canada website, 2003), lung cancer incidence and population at risk were reported in 1995-2002 for 9 age-groups by gender. For males or females, let y_{ij} denote the number of lung cancer cases (incidence) and n_{ij} the corresponding population size for PHU i in Ontario and age-group j in the period under study, $i = 1, \ldots, 37$, $j = 1, \ldots, 9$. The age-groups are 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, and 85+. This data set is considered as pilot case for evaluating the methods and will be examined in-depth.

Lung cancer counts for males over the PHUs range from 3 to 1252, with a median of 56. For females they range from 3 to 756 with a lower median of 41. *Toronto* has the largest count for the 70-74 age-group for both genders, this value being substantially higher than the median value over PHUs for both genders. Note that there are cases for which the PHU data are not available and these are not considered in the analysis; they account for 1.56% of the total case counts for males and 2.09% of the total female cases.

Table 1 displays observed overall rates of lung cancer incidence ($\times 1,000$) by age-group ($\sum_{i=1}^{37} y_{ij} / \sum_{i=1}^{37} n_{ij}$), whereas Figure 1 plots the (log) overall rates of lung cancer incidence for males and females by age-group; the difference between rates for males and females are larger for the older age-groups. These rates indicate an increasing trend over earlier age-groups which slightly decreases after age-group 75-79, indicating a nonlinear effect over age-group for both genders. Because some interactions with gender were evident in preliminary studies, analyses are stratified here by gender. Notice that the overall rates for males are always higher than corresponding

rates for females, and the age-group 75-79 has the highest overall rates both for males and females.

Age-group	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
Male	0.268	0.628	1.242	2.229	3.395	4.575	5.121	4.855	4.416
Female	0.263	0.514	0.930	1.455	2.005	2.390	2.463	2.273	1.615

* Rates are multiplied by 1,000.

Table 1: Observed overall rates of lung cancer incidence by age-group and gender.

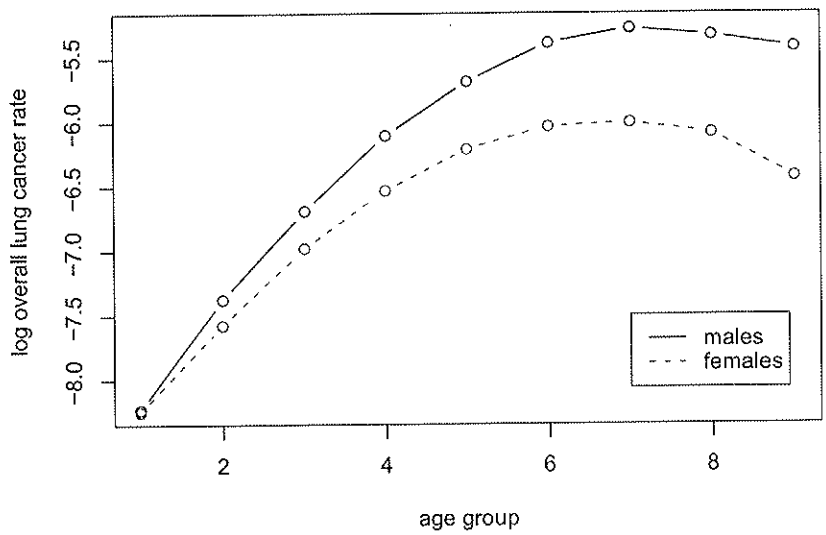

Figure 1: Log overall rates of lung cancer incidence by age-group and gender; age-group 1 corresponds to 45-49; 2 corresponds to 50-54, etc. The last age-group denoted by age-group 9 corresponds to 85 and older.

For the lung cancer incidence data, several spatial age-specific models (2) were fitted in increasing level of complexity by gender. Table 2 displays the related model comparison based on the measures: deviance D (5), Pearson function Q_P (6) and deviance information criterion DIC (7). Note that:

- $S_0(j)$ is a cubic spline with no intercept and one, two or three equally spaced inner knots, with age vector recoded as (-3.5, -2.5, -1.5, -0.5, 0, 0.5, 1.5, 2.5, 3.5). Prior distributions for age regression coefficients β (M_{1A}) and β_k (M_{1B}) are set to be vague and follow a normal distri-

bution $N(0, 10^5)$. The variable I_C denotes an indicator variable (1 if C is true and 0 otherwise).

- $\delta_i \sim N(0, \sigma_\delta^2)$ is the linear age effect related to the ith region, and spatial random effects h_i and b_i have $N(0, \sigma_h^2)$ and intrinsic CAR (σ_b^2) priors, respectively.
- α_0 is assigned a flat prior on the whole real line; highly dispersed priors are also assigned to the remaining parameters, e.g. the spatial hyperparameters σ_b^2, σ_h^2, σ_δ^2 and σ_k^2 have inverse gamma prior $IG(0.5, 0.0005)$.
- MCMC samples of size 50,000 were obtained for all models, taking every 100th iteration of the simulated sequence, after 20,000 iterations of burn-in.
- Convergence diagnostics for the simulated samples were carried out using several methods available in the Bayesian Output Analysis (BOA) program (Smith, 2007). None of them showed any distinguishing features, apart from convergence being quite slow.
- The models range in complexity. For example, Model M_{1A} is a model for which the mean rate increases linearly with age-group j; there are no small-area effects so the rates for all small-areas are the same indicating homogeneity over the province. Model M_{1B} is similar to M_{1A} in that there are no small-area effects, but the rate differs for each age-group and is not constrained to be linearly increasing or decreasing with j. The parameter α_0 is the mean rate for the first age-group and β_k measures the change in that rate for the kth age-group. Models M_{1C}, M_{1D}, and M_{1E} also have no small-area effects but model the age-group rate trends as smooth piece-wise cubic splines with increasing numbers of segments with the increasing knots. The models in the M_2 series include either spatially correlated or independent small-area random effects (small-area relative risks) or both. The term h_{ij} in M_{2G} represents independent normally distributed small-area by age-group effects – here the region-effects are not the same for each age-group and the model offers an age-group by region interaction. Model M_{4C} uses a spline to model the overall age-group trend effects, small-area splines to model the small area trend effects, and contains both spatially structured and independent spatial relative risk terms. Model M_{5J} includes a spline to model the overall age-group effects, the small-area age-group trends are approximated by a linear trend over and above the overall spline effect, and the spatially structured and independent small-area relative risks are different for each age-group.

Based on the comparison measures shown in Table 2, there is evidence that:

- When exploring overall trends in the models in the M_1 series, 2-knot cubic spline models fit reasonably well and better than 1-knot and 3-knot models for both genders. Hence the remaining spline models in the table were fitted using two knots.

- Although saturated models (identified by the term $(\sum_{k=2}^{9} \beta_k I_{a=k})$ have perfect fit in comparison with other models, cubic B-spline models are more parsimonious in term of the number of parameters.

- Models which seem to fit very well are M_{3C} and M_{3D}. For simplicity, we suggest the use of the former because of the simpler interpretation and implementation of this model. For other comparisons, we also consider M_{2C}, M_{4C} and M_{5H} as the gold standard models.

Figure 2 displays the overall age relative risk ($\exp(\alpha_0 + S_0(j))$) for males and females, respectively, based on the spatially uncorrelated models M_{1A} (simple linear age effect), M_{1B} (dummy-variable based age effect), M_{1C} (1-knot cubic spline age effect), M_{1D} (2-knots cubic spline age effect), and M_{1E} (3-knots cubic spline age effect), as well as the corresponding observed overall age rates. Spline models M_{1C}, M_{1D} and M_{1E} essentially have the same estimated overall age rates as the saturated model M_{1B}, whereas the linear model M_{1A} gives a poor fit and departs considerably from the observed age rates after age-group 4 (60-64) for both genders.

Another quantity of interest is the small-area age trend effect, defined by $\exp(\delta_i)$, for the selected models M_{3C} and M_{3D}. Recall this represents the small-area age-group trends over and above the spline overall trends. Regions for which $\exp(\delta_i)$ is large correspond to those for whom the age effects are more striking: there are larger increases in mortality for older age-groups. Conversely, regions for which $\exp(\delta_i)$ is small correspond to those regions for which the increases in rates are smaller than the overall provincial average for the older age-groups. The terms $\exp(\delta_i)j$ operate multiplicatively on the overall means plotted in Figure 2 for the jth age-group. Figure 3 plots 95% highest posterior density (HPD) credible intervals for the spatial age trend effect based on model M_{3C}, while Figure 4 maps the related posterior means (these estimates are also provided in the first and third columns of Table 3). *Halton*, *York Regional* and *Peel Regional* have significant spatial age trend effects (in decreasing order) for males, whereas for females increasing age trend effects over and above the overall mean are significant for *York Regional*, *Toronto* and *Halton*. *Eastern Ontario* is the only region with significantly low spatial age trend effects for both genders.

In order to quantify the unobserved spatially correlated and uncorrelated heterogeneity, we present in Table 4 the estimates of the variance components for selected spatial age-specific models, including the posterior mean, standard deviation and 95% HPD credible interval. Based on the posterior

means of variance components for model M_{3C}, note that the spatially correlated component, σ_b^2, is almost 13 and 8 times bigger than uncorrelated one, σ_h^2, for males and females, respectively, indicating an important unobserved spatially correlated effect. Figure 5 plots the associated credible intervals over age-groups based on model $M5F$, which permits these two variance components to differ over age-groups. Again we see that the spatial effects dominate the map (over the random uncorrelated small-area risks). There is generally more spatially correlated variability for earlier age-groups for both genders. The precision of the spatially correlated variance component corresponding to the older age-groups tends to be much higher than that from younger age-groups.

Spatial or small-area (residual) relative risk, which is here defined as $\exp(h_i + b_i)$, is another quantity of interest. Figure 6 maps the spatial relative risk of lung cancer incidence for males and females based on model M_{3C} (the posterior means are provided in the fifth and seventh columns of Table 3). First note that these effects are quite large and there seems to be urban/rural differences in the risks, perhaps particularly for males. In order to identify regional/spatial effects, in particular regions associated with extreme rates or relative risks, so-called "hotspots" (highest rates) or "lowspots" (lowest rates), we ranked the posterior means of both the spatial age trend and the spatial relative risk by gender based on model M_{3C} (see Table 3 in increasing order of male spatial age trend). There are some noteworthy gender differences in some regions. The highest spatial relative risks are basically located in Northern Ontario (*Timiskaming, Porcupine, Algoma*) and *Eastern Ontario* for males and females. The latter also has the lowest spatial age trend, indicating that these high-risk regions do not increase for older age-groups over the above the overall provincial averages. In contrast, *York Regional, Peel Regional* and *Halton* have low spatial relative risks but the highest spatial age trends so though the risks are low for these three regions, risks for older adults tend to be somewhat higher than expected based on the low small-area risk and the overall provincial age trends. We choose to discuss these specific areas since their risk effects are significant. However, policymakers interested in effects for any specific small area can similarly use these exploratory analyses to obtain relative standing of that area's effects.

To understand the interaction between the small-area and age-group effects observed in the selected model M_{3C}, we compute the spatial age-specific relative risk as

$$\exp(S_i(j) + b_i + h_i),$$

where $S_i(j) = \delta_i j$ for model M_{3C}. These are the small-area relative risks for each age-group over the above the provincial mean age trends. Figure 7 maps the spatial age-specific relative risk of lung cancer incidence for three age-groups 50-54, 65-69 and 80-84 by gender based on model M_{3C}. *Timiskaming, Eastern Ontario, Porcupine* and *Algoma* in essence have

the highest spatial risks over the nine age-groups for both genders. Note changes in the spatial effect over age: *Thunder Bay District* and *North Bay and District* have increasing estimates of spatial age-specific relative risks for males, starting at the 15th and 23rd ranks respectively for the 45-49 age-group and ending at the 36th and 35th ranks for the 85+ age-group. *Eastern Ontario* and *North Bay and District* occupy the 37th and 36th corresponding ranks, respectively, for the first four age-groups for females but the 18th and 23rd ranks for the last age-group; the ranks of the estimates of the spatial age-specific relative risks for *Algoma* are increasing over the age-groups from the 25th to the 35th.

Figures 8, 9 and 10 plot spatial age-specific relative risks based on the models M_{3C} and M_{4C} for several PHUs, as well as the standardized crude rate that is defined by

$$y_{ij}/(n_{ij} \times q_j),$$

where $q_j = \sum_i y_{ij} / \sum_i n_{ij}$, $i = 1, \ldots, r$, $j = 1, \ldots, a$. Note that the scales for these plots differ for males and females. Recall that these age-region relative risks are over and above the overall provincial trend provided in Figures 2 and represent age-region discrepancies with respect to the overall trend. PHUs displayed in Figure 8 have the highest age trend effects (model M_{3C}) for males: *Halton*, *York Regional* and *Thunder Bay District* (in decreasing order), though region effects are low for these areas. The plots in Figure 9 display regions with the lowest age trend effects (model M_{3C}) for males and females: *Eastern Ontario*, *Elgin-St.Thomas* and *Bruce-Grey-Owen Sound*. In fact, *Eastern Ontario* also has the second highest spatial relative risk for both genders. Figure 10 displays regions with higher spatial risks, for example *Timiskaming*; these high risks seem to be the same for all the age-groups.

4 Discussion

In general, the use of spatial age-specific models (2) which incorporate smoothing splines to model age-group effects and include spatially correlated and uncorrelated random effects seems to be a promising option for modelling and mapping disease rates by age-groups over public health regions. They yield simple and informative interpretation of the disease mortality and incidence data, and provide information relevant to the national, provincial and local health area administrators.

The spatial model (3) provided a fair fit to the lung cancer incidence data. It seems flexible enough at the small-area level, and works well for modelling the overall mean, if there are sufficient age-groups to employ a spline trend for modelling the overall age-group rates. That model has some advantages over both the dummy-variable model ($S_0(j) = \sum_{k=2}^{9} \beta_k I_{j=k}$) and the model with splines at both the overall mean and small-area levels (splines for $S_0(j)$ and $S_i(j)$) because of parsimony in the use of parameters. In addition, the small-area effects in model (3) are expressed as linear trends ($\delta_i j + b_i +$

h_i), i.e., there is a straightforward interpretation of the model in terms of the contribution of spatial age trend and spatial relative risk for each public health regions. Though the linear small-area trend may not be adequate for handling long time series, it seems to give a fair enough fit when handling the estimation of trends across 10 or so age-groups.

We recommend that when there are fewer than 8 age-groups, adopting cubic spline models with two knots as used here is not recommended. Dummy-variable models with separate age-group effects for each of the age-groups are preferable; alternatively, a simple cubic spline with one knot may provide similar fits.

Although spatially correlated random effects seem to dominate these analyses, models which include an independent random small-area effect only are preferable to those with no means of incorporating overdispersion. There seems to be quite a bit of heterogeneity in these sorts of data and simple Poisson models tend to provide considerably poorer fits.

Bayesian methods seem quite useful for these types of analyses. There is a long history of work in that field, e.g., see references and reviews in Ainsworth and Dean (2008) and Silva, Dean et al (2008). These papers and references cited in them discuss the advantages of estimating small-area disease rates based on the types of spatial models considered here. Comparisons with the use of the standardized mortality ratio (SMR) are provided including simulations showing gains in precision and in correctly identifying hotspots.

We have presented rank estimates somewhat informally in this document to rate regional/spatial effects. In the next phase, we will consider more carefully inference concerning ranks including the estimation of standard errors of ranks in the spatial analysis of risks. We will also consider rules for identifying hotspots.

ACKNOWLEDGEMENTS

Special Thanks to Elizabeth Juarez Colunga, Robert Semenciw, and Nathaniel Bell for providing some SAS codes related to standard error estimates of the overdispersion parameter, for supplying the data and interesting comments, and for producing the maps, respectively. This paper was partially supported by Pest-OE/MAT/UI0006/2011.

REFERENCES

[1] Ainsworth LM and Dean CB, Detection of local and global outliers in mapping studies, *Environmetrics*, **19**, 21-37, 2008.

[2] Banerjee S, Carlin BP and Gelfand AE, *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, 2004.

[3] Besag J, York J and Mollié, Bayesian image restoration, with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, **43**, 1-59, 1991.

[4] Chen MH, Shao QH and Ibrahim JG, *Monte Carlo Methods in Bayesian Computation*, Sprnger-Verlag, 2000.

[5] Clayton D and Kaldor J, Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics*, **43**, 671-681, 1987.

[6] Clements M, Armstrong BK and Moolgavkar SH, Lung cancer rate predictions using generalized additive models, *Biostatistics*, **6**, 576-589, 2005.

[7] Dean CB, Ugarte MD and Militino AF, Detecting interaction between random region and fixed age effects in disease mapping, *Biometrics*, **57**, 197-202, 2001.

[8] Dean CB and MacNab YC, Modelling of rates over a hierarchical health administrative structure, *The Canadian Journal of Statistics*, **29**, 405-419, 2001.

[9] Lawson AB, Biggeri A, Böhning D, Lesaffre E, Viel J and Bertollini R (eds), *Disease Mapping and Risk Assessment for Public Health*. Wiley: New York, 1999.

[10] MacNab YC and Dean CB, Autoregressive spatial smoothing and temporal spline smoothing for mapping rates, *Biometrics*, **57**, 949-956, 2001.

[11] McCullagh P and Nelder JA, *Generalized Linear Models*, Chapman&Hall/CRC, 1989.

[12] Nandram B, Sedransk J and Pickle LW, Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease, *Journal of the American Statistical Association*, **95**, 1110-1118, 2000.

[13] Silva G, Dean CB, Niyonsenga T and Vanasse A, Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines, *Statistics in Medicine*, **27**, 2381-2401, 2008.

[14] Smith BJ, *BOA User Manual* (version 1.1.6), Department of Biostatistics, College of Public Health, University of Iowa, 2007. (available online in http://www.public-health.uiowa.edu/boa/)

[15] Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society B*, **64**, 583-639, 2002.

[16] Thomas A, Best N, Lunn D, Arnold R and Spiegelhalter D, *GeoBUGS User Manual* (version 1.2), Department of Epidemiology and Public Health, Imperial College, St Mary's Hospital London, 2004. (available on-line in http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs.shtml)

[17] Waller LA and Gotway CA, *Applied Spatial Statistics for Public Health Data*, Wiley, 2004.

[18] Wood S, *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC, 2006.

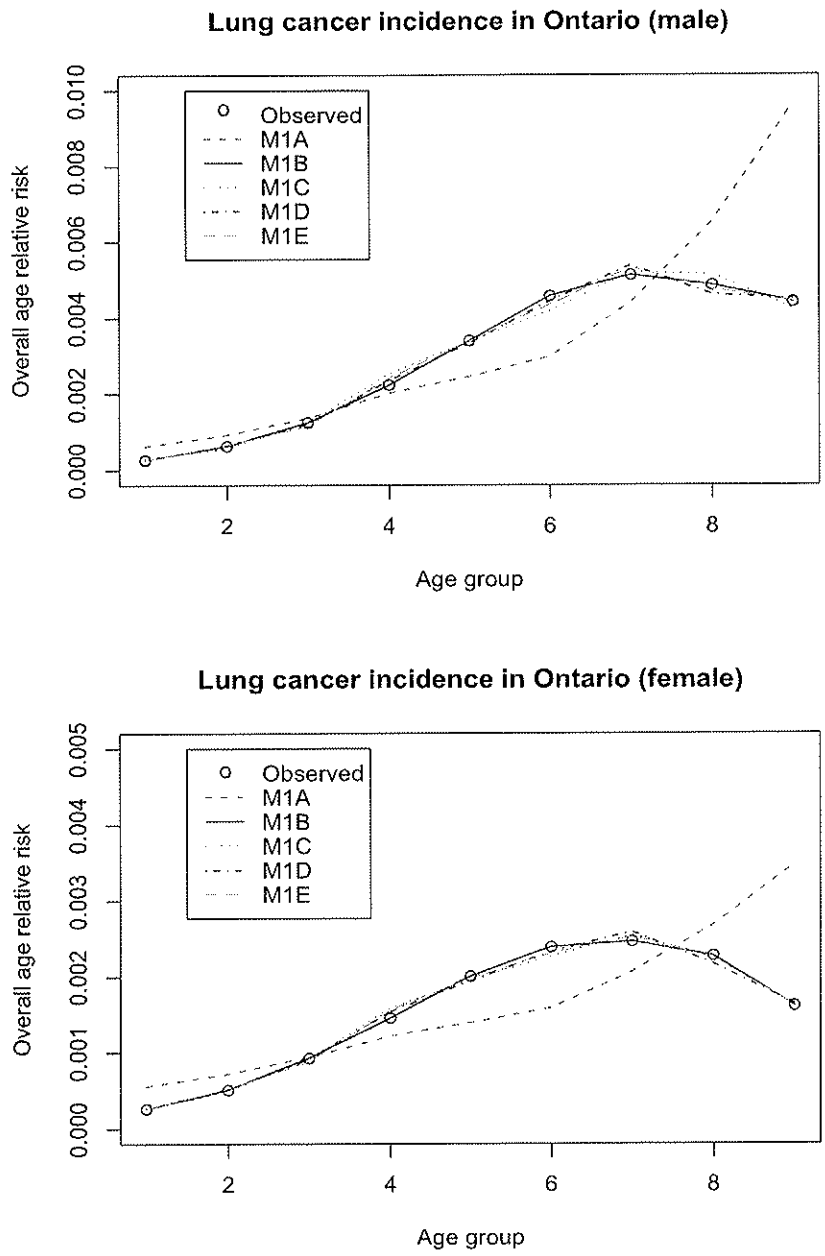

Figure 2: Overall age effect for lung cancer incidence for males (top) and females (bottom) based on several spatial age-specific models.

Gender	Models defined from $log\ \mu_{ij}$	Q_P	D	DIC
Male	M_{1A}: $\alpha_0+\beta\, j$	5290	7367	7368.73
	M_{1B}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}$	1168	3071	3080.21
	M_{1C}: $\alpha_0+S_0(j)$ (1 knot)	1319	3225	3230.34
	M_{1D}: $\alpha_0+S_0(j)$ (2 knots)	1203	3108	3113.53
	M_{1E}: $\alpha_0+S_0(j)$ (3 knots)	1234	3135	3141.58
	M_{2A}: $\alpha_0+S_0(j)+h_i$ (2 knots)	448.8	2383	2422.56
	M_{2B}: $\alpha_0+S_0(j)+b_i$ (2 knots)	448.1	2384	2421.28
	M_{2C}: $\alpha_0+S_0(j)+h_i+b_i$ (2 knots)	447.1	2382	2419.33
	M_{2D}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+h_i$	413.9	2348	2390.84
	M_{2E}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+b_i$	413.4	2349	2389.45
	M_{2F}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+h_i+b_i$	412.2	2348	2387.41
	M_{2G}: $\alpha_0+S_0(j)+h_{ij}$ (2 knots)	345.5	2279	2478.42
	M_{3C}: $\alpha_0+S_0(j)+\delta_i\, j+h_i+b_i$ (2 knots)	*358.2*	*2298*	*2358.64*
	M_{3D}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\delta_i\, j+h_i+b_i$	*323.6*	*2264*	*2327.06*
	M_{4C}: $\alpha_0+S_0(j)+S_i(j)+h_i+b_i$ (2 knots)	361.4	2300	2371.28
	M_{5C}: $\alpha_0+S_0(j)+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$ (2 knots)	342.2	2280	2448.12
	M_{5F}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$	339.7	2277	2441.51
	M_{5G}: $\alpha_0+S_0(j)+h_i+\sum_{k=1}^{9} b_k I_{j=k}$ (2 knots)	384.4	2320	2403.55
	M_{5H}: $\alpha_0+S_0(j)+b_i+\sum_{k=1}^{9} h_k I_{j=k}$ (2 knots)	344.7	2283	2383.53
	M_{5I}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\delta_i\, j+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$	336.1	2274	2444.45
	M_{5J}: $\alpha_0+S_0(j)+\delta_i\, j+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$ (2 knots)	334.4	2272	2438.73
Female	M_{1A}: $\alpha_0+\beta\, j$	4393	6276	6277.86
	M_{1B}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}$	1352	3174	3183.01
	M_{1C}: $\alpha_0+S_0(j)$ (1 knot)	1388	3213	3218.06
	M_{1D}: $\alpha_0+S_0(j)$ (2 knots)	1375	3200	3205.97
	M_{1E}: $\alpha_0+S_0(j)$ (3 knots)	1376	3201	3207.77
	M_{2A}: $\alpha_0+S_0(j)+h_i$ (2 knots)	475.5	2332	2371.52
	M_{2B}: $\alpha_0+S_0(j)+b_i$ (2 knots)	475.9	2334	2371.10
	M_{2C}: $\alpha_0+S_0(j)+h_i+b_i$ (2 knots)	474.3	2332	2368.55
	M_{2D}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+h_i$	451.6	2307	2349.48
	M_{2E}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+b_i$	452.3	2309	2349.16
	M_{2F}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+h_i+b_i$	450.4	2307	2346.31
	M_{2G}: $\alpha_0+S_0(j)+h_{ij}$ (2 knots)	325.6	2174	2378.50
	M_{3C}: $\alpha_0+S_0(j)+\delta_i\, j+h_i+b_i$ (2 knots)	*366.8*	*2224*	*2283.28*
	M_{3D}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\delta_i\, j+h_i+b_i$	*343.2*	*2199*	*2261.39*
	M_{4C}: $\alpha_0+S_0(j)+S_i(j)+h_i+b_i$ (2 knots)	358.5	2214	2285.34
	M_{5C}: $\alpha_0+S_0(j)+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$ (2 knots)	327.5	2178	2346.54
	M_{5F}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$	329.0	2179	2346.01
	M_{5G}: $\alpha_0+S_0(j)+h_i+\sum_{k=1}^{9} b_k I_{j=k}$ (2 knots)	363.2	2218	2305.29
	M_{5H}: $\alpha_0+S_0(j)+b_i+\sum_{k=1}^{9} h_k I_{j=k}$ (2 knots)	339.8	2195	2298.96
	M_{5I}: $\alpha_0+\sum_{k=2}^{9}\beta_k I_{j=k}+\delta_i\, j+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$	324.6	2174	2346.74
	M_{5J}: $\alpha_0+S_0(j)+\delta_i\, j+\sum_{k=1}^{9}(h_k+b_k)\, I_{j=k}$ (2 knots)	326.0	2176	2346.37

Table 2: Comparison of the fits of several gender-specific models for trends in rates over age-groups and spatial effects.

Spatial age trend effect				Spatial relative risk				
male		female		male		female		
mean	rank	mean	rank	mean	rank	mean	rank	PHU Name
0.947	1	0.928	1	1.308	36	1.261	36	Eastern Ontario
0.961	2	0.967	4	0.940	13	0.867	9	Bruce-Grey-Owen Sound
0.964	3	1.039	33	0.783	4	0.891	10	Perth District
0.966	4	0.964	3	1.097	28	0.909	11	Kent-Chatham
0.975	5	0.987	12	1.004	19	0.955	15	Haldimand-Norfolk
0.976	6	0.987	13	0.866	7	0.791	5	Waterloo Regional
0.977	7	0.985	11	1.041	23	1.135	26	Peterborough County City
0.980	8	0.961	2	0.991	17	0.977	17	Elgin-St. Thomas
0.982	9	0.979	8	1.160	31	1.244	35	Hasting-Prince Edward County
0.983	10	0.994	18	0.960	14	0.940	14	Muskoka-Parry Sound
0.985	11	0.990	14	1.040	22	1.216	32	Kingston-Frontenac-Lennox-Addington
0.986	12	0.994	17	1.144	30	1.171	29	Leeds-Grenville-Lanark
0.987	13	0.984	9	0.977	16	0.913	12	Middlesex-London
0.989	14	1.010	27	1.039	21	1.132	25	Simcoe County District
0.990	15	0.972	6	0.927	10	0.866	8	Oxford County
0.992	16	1.007	24	0.935	12	0.939	13	Niagara Regional
0.994	17	0.985	10	0.995	18	1.008	18	Hamilton-Wentworth
0.995	18	0.992	15	1.090	27	1.201	30	Renfrew County District
0.995	19	1.074	36	0.826	6	0.707	2	Toronto
0.996	20	1.011	28	1.238	34	1.170	28	Algoma
0.996	21	0.978	7	0.876	8	0.861	7	Huron
0.996	22	0.997	20	1.251	35	1.229	33	Porcupine
0.998	23	1.017	29	1.045	24	1.082	23	Haliburton-Kawartha-Pine Ridge
1.000	24	1.007	25	1.208	33	1.213	31	Sudbury and District
1.008	25	1.008	26	0.803	5	0.824	6	Wellington-Dufferin-Guelph
1.014	26	0.997	21	1.415	37	1.274	37	Timiskaming
1.015	27	1.036	32	1.032	20	1.047	21	Northwestern
1.017	28	1.028	31	0.916	9	1.025	19	Durham Regional
1.019	29	0.996	19	0.966	15	0.968	16	Brant County
1.019	30	1.000	23	1.071	26	1.097	24	Lambton
1.021	31	0.999	22	1.068	25	1.028	20	Windsor-Essex
1.028	32	0.968	5	1.175	32	1.239	34	North Bay and District
1.030	33	1.024	30	0.933	11	1.049	22	Ottawa-Carleton Regional
1.042	34	1.041	34	0.768	3	0.673	1	Peel Regional
1.042	35	0.992	16	1.134	29	1.164	27	Thunder Bay District
1.081	36	1.080	37	0.698	1	0.730	3	York Regional
1.084	37	1.050	35	0.767	2	0.790	4	Halton

Table 3: Estimates and associated ranks for both the small-area age trend effect and the small-area relative risk by gender based on model M_{3C}.

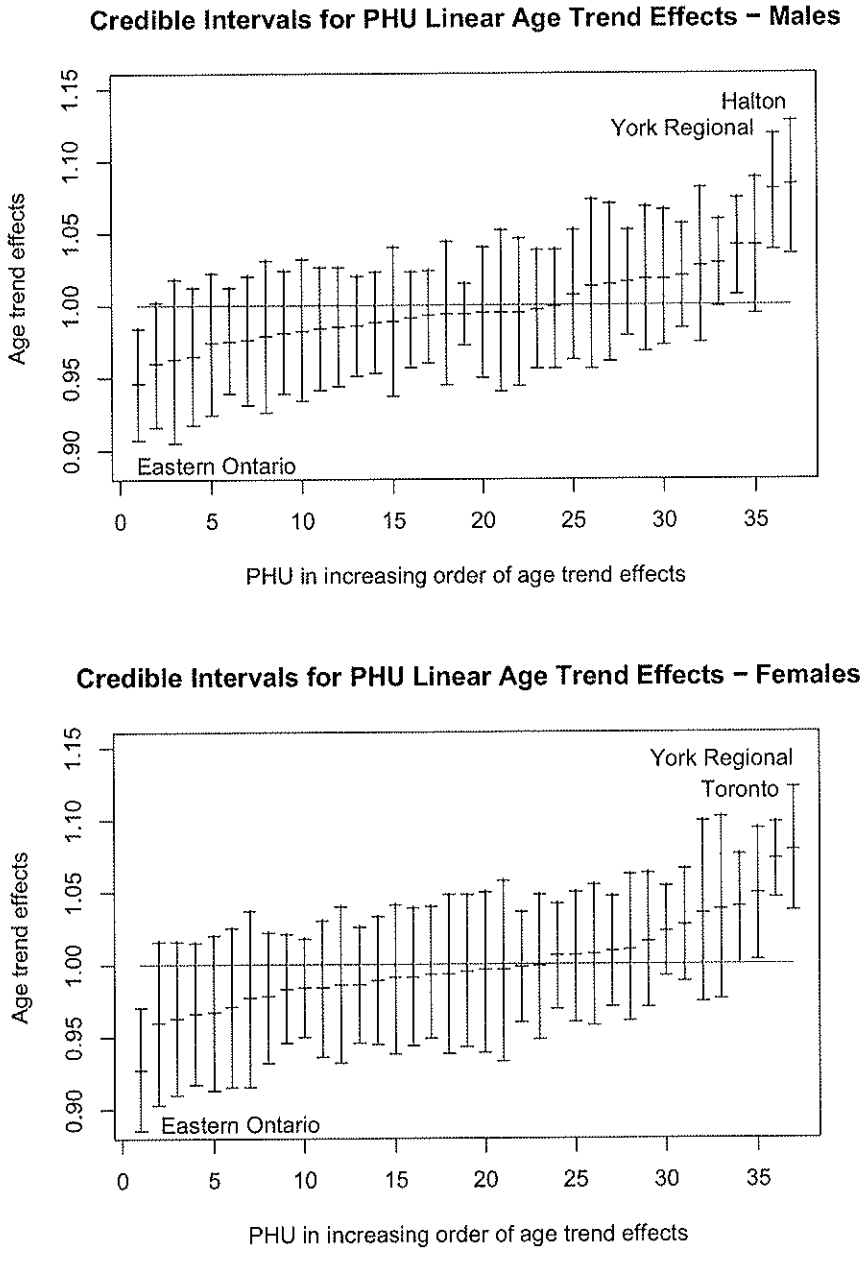

Figure 3: Credible intervals of the small-area age trend effects for the analysis of lung cancer incidence for males (top) and females (bottom) based on model M_{3C}.

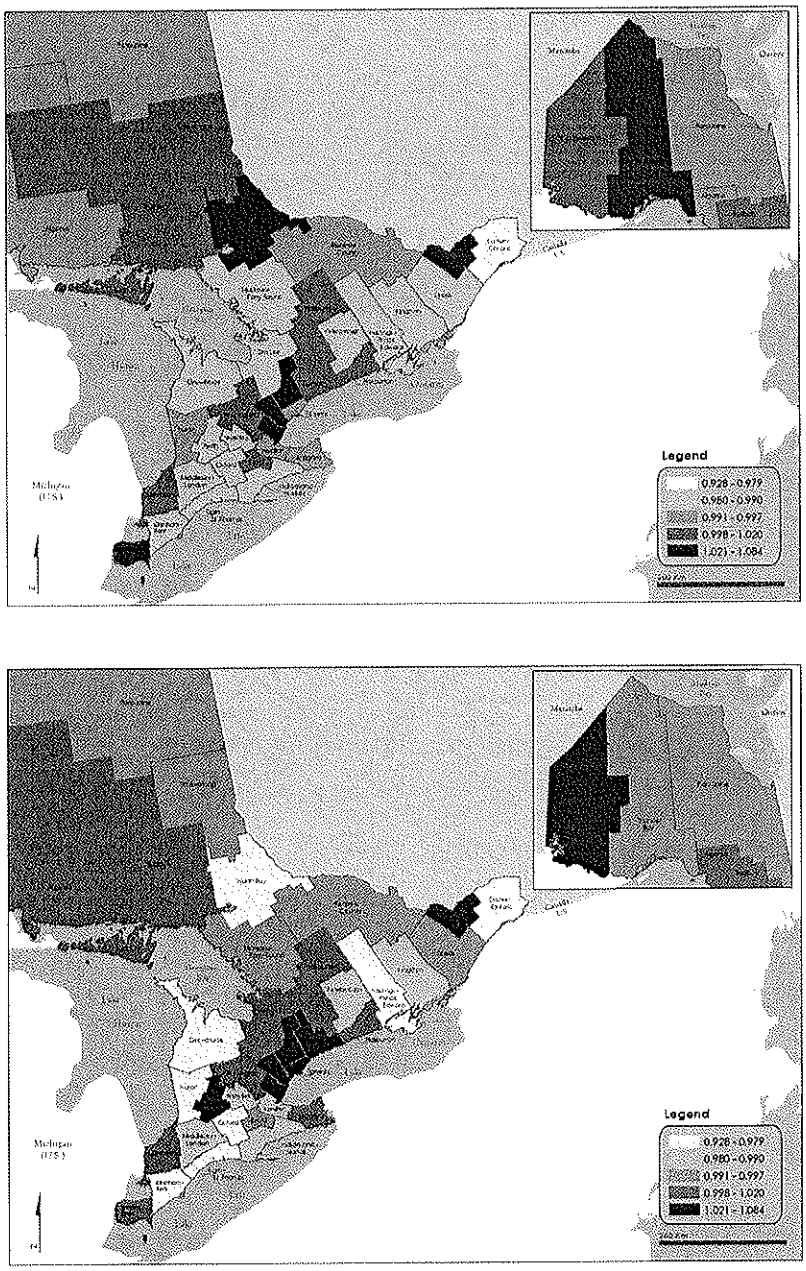

Figure 4: Maps of the small-area age trend estimates for the analysis of lung cancer incidence for males (top) and females (bottom) based on model M_{3C}.

Model	parameter	male			female		
		mean	st.dev.	95% HPD CI	mean	st.dev.	95% HPD CI
M_{2C}	$\sigma^2_{b_1}$	0.00265	0.00279	(0.00021,0.01029)	0.00538	0.00473	(0.00030,0.01733)
	σ^2_{b}	0.03592	0.01368	(0.01329,0.06686)	0.03852	0.01835	(0.00988,0.08015)
M_{3C}	$\sigma^2_{b_1}$	0.00276	0.00291	(0.00021,0.01068)	0.00516	0.00475	(0.00028,0.01723)
	σ^2_{b}	0.03726	0.01423	(0.01366,0.06972)	0.04083	0.01882	(0.01110,0.08362)
M_{3D}	$\sigma^2_{b_1}$	0.00275	0.00292	(0.00007,0.00856)	0.00515	0.00475	(0.00008,0.01449)
	σ^2_{b}	0.03721	0.01429	(0.01055,0.06530)	0.04065	0.01876	(0.00709,0.07676)
M_{4C}	$\sigma^2_{b_1}$	0.00143	0.00158	(0.00017,0.00581)	0.00274	0.00289	(0.00020,0.01066)
	σ^2_{b}	0.02203	0.00976	(0.00817,0.04588)	0.02333	0.01277	(0.00610,0.05493)
M_{5F}	$\sigma^2_{b_1}$	0.06596	0.05684	(0.00069,0.20810)	0.07500	0.05619	(0.00058,0.20840)
	$\sigma^2_{b_2}$	0.04549	0.03440	(0.00064,0.12760)	0.12130	0.06251	(0.00635,0.26420)
	$\sigma^2_{b_3}$	0.08486	0.05769	(0.00057,0.20750)	0.08398	0.03770	(0.02254,0.17220)
	$\sigma^2_{b_4}$	0.04488	0.03329	(0.00066,0.12100)	0.06990	0.04308	(0.00109,0.16550)
	$\sigma^2_{b_5}$	0.03171	0.01661	(0.00598,0.07132)	0.05391	0.02620	(0.01433,0.11690)
	$\sigma^2_{b_6}$	0.02930	0.01349	(0.00957,0.06172)	0.03054	0.02010	(0.00355,0.07992)
	$\sigma^2_{b_7}$	0.02293	0.01257	(0.00625,0.05415)	0.02153	0.02662	(0.00039,0.09644)
	$\sigma^2_{b_8}$	0.01630	0.01075	(0.00338,0.04409)	0.00883	0.00994	(0.00034,0.03603)
	$\sigma^2_{b_9}$	0.00701	0.01198	(0.00021,0.04102)	0.01048	0.01199	(0.00033,0.04305)
	$\sigma^2_{b_1}$	0.01120	0.01726	(0.00022,0.06105)	0.01350	0.02111	(0.00021,0.07455)
	$\sigma^2_{b_2}$	0.00913	0.01210	(0.00023,0.04283)	0.00925	0.01688	(0.00021,0.06286)
	$\sigma^2_{b_3}$	0.01470	0.01984	(0.00023,0.06676)	0.00469	0.00827	(0.00019,0.02874)
	$\sigma^2_{b_4}$	0.01183	0.01208	(0.00028,0.04224)	0.01146	0.01488	(0.00024,0.05260)
	$\sigma^2_{b_5}$	0.00350	0.00441	(0.00020,0.01625)	0.00458	0.00614	(0.00021,0.02180)
	$\sigma^2_{b_6}$	0.00199	0.00258	(0.00018,0.00910)	0.00722	0.00714	(0.00027,0.02559)
	$\sigma^2_{b_7}$	0.00273	0.00314	(0.00020,0.01143)	0.02206	0.01477	(0.00065,0.05615)
	$\sigma^2_{b_8}$	0.00189	0.00243	(0.00017,0.00873)	0.00305	0.00397	(0.00019,0.01432)
	$\sigma^2_{b_9}$	0.00483	0.00643	(0.00021,0.02318)	0.00280	0.00410	(0.00018,0.01428)
M_{5H}	$\sigma^2_{b_1}$	0.00581	0.00798	(0.00021,0.02839)	0.00653	0.00875	(0.00022,0.03158)
	$\sigma^2_{b_2}$	0.00210	0.00262	(0.00018,0.00946)	0.00748	0.00798	(0.00027,0.02900)
	$\sigma^2_{b_3}$	0.01102	0.00830	(0.00053,0.03138)	0.00703	0.00607	(0.00037,0.02274)
	$\sigma^2_{b_4}$	0.00714	0.00505	(0.00058,0.01952)	0.01038	0.00583	(0.00235,0.02473)
	$\sigma^2_{b_5}$	0.00122	0.00125	(0.00016,0.00468)	0.00116	0.00126	(0.00016,0.00465)
	$\sigma^2_{b_6}$	0.00105	0.00107	(0.00015,0.00398)	0.00141	0.00150	(0.00017,0.00554)
	$\sigma^2_{b_7}$	0.00344	0.00296	(0.00032,0.01117)	0.00924	0.00848	(0.00037,0.03086)
	$\sigma^2_{b_8}$	0.00539	0.00555	(0.00027,0.02035)	0.01605	0.00958	(0.00307,0.03997)
	$\sigma^2_{b_9}$	0.01117	0.01095	(0.00031,0.03899)	0.00590	0.00683	(0.00025,0.02466)
	σ^2_{b}	0.04086	0.01248	(0.02229,0.07059)	0.05512	0.01662	(0.03010,0.09451)

Table 4: Estimates of the variance components for selected spatial age models.

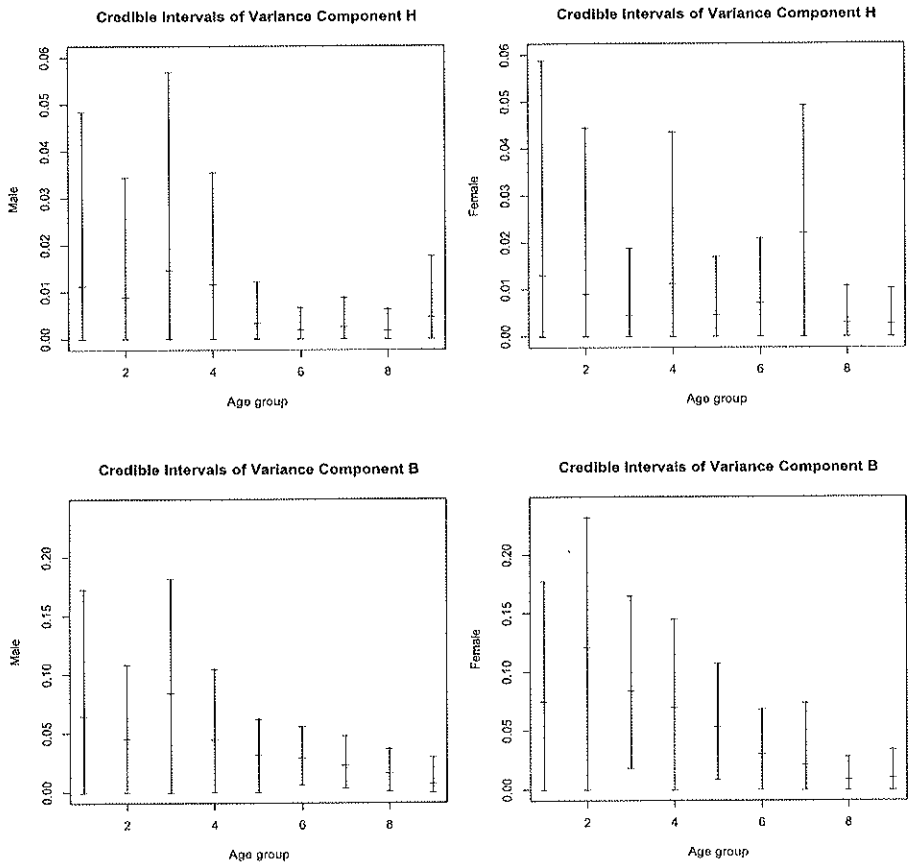

Figure 5: Credible intervals of the uncorrelated (top) and spatially correlated (bottom) variance components from the analysis of lung cancer incidence for males (left) and females (right) over age-groups based on model $M5F$.

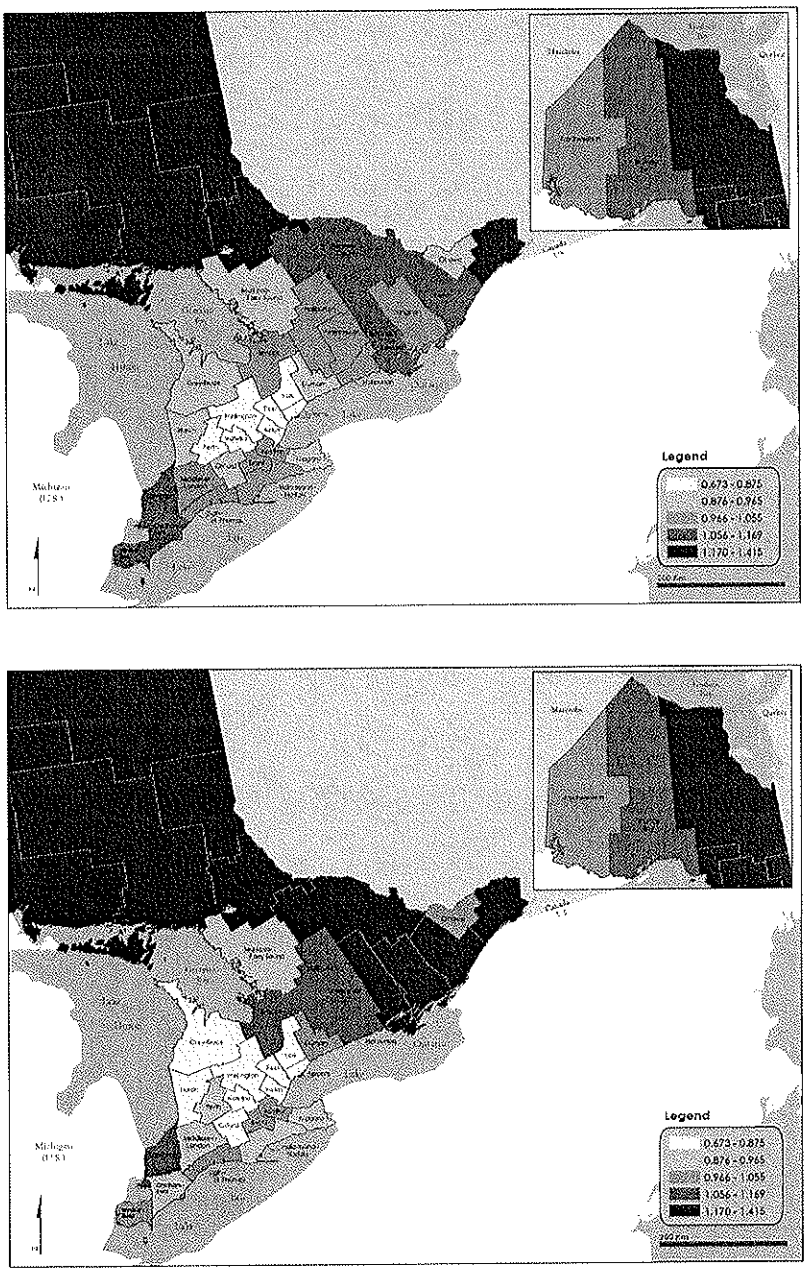

Figure 6: Maps of the spatial relative risk of lung cancer incidence for males (top) and females (bottom) based on model M_{3C}.

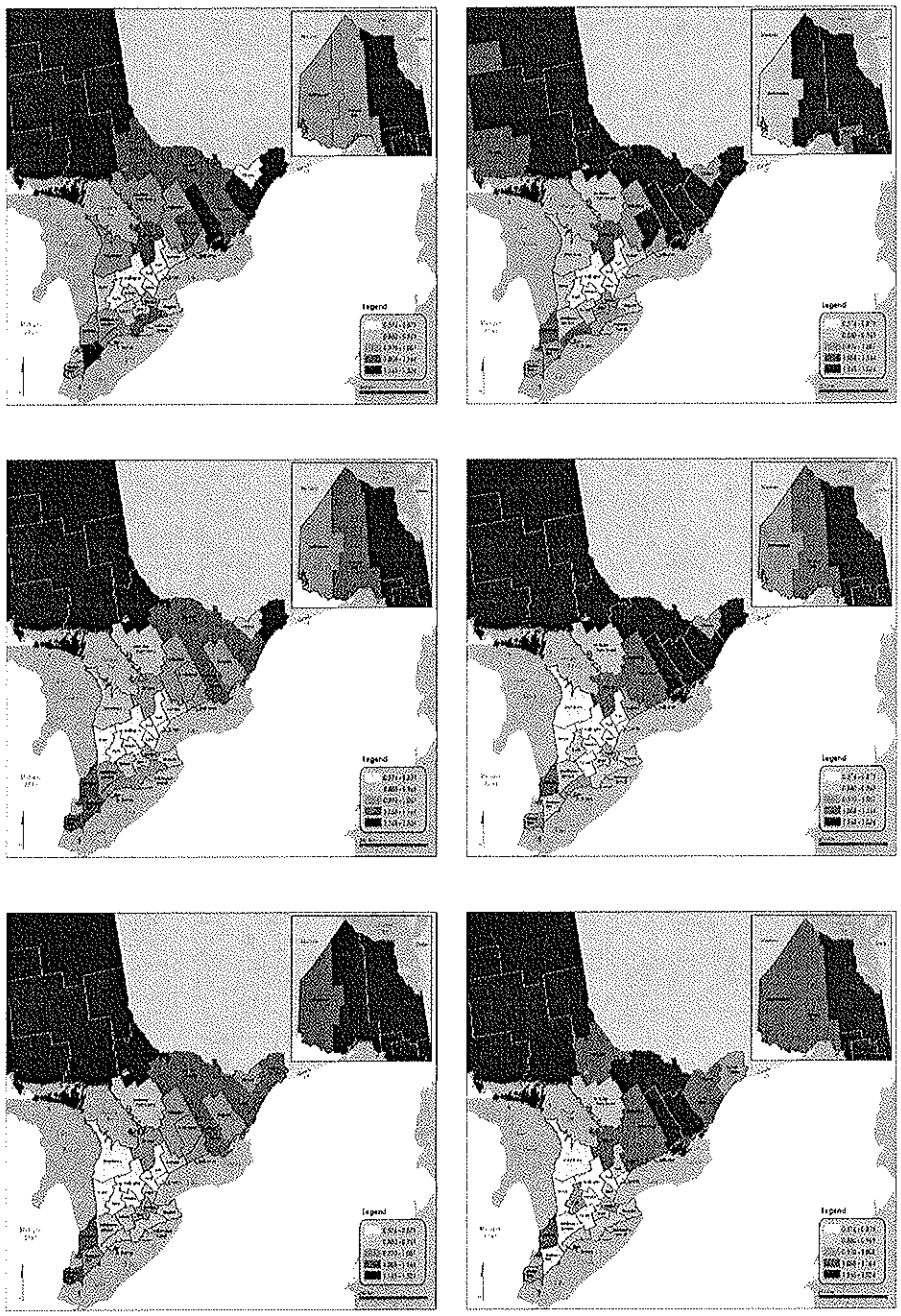

Figure 7: Maps of the spatial age-specific relative risk of lung cancer incidence for males (left) and females (right) and age-group 50-54 (top), 65-69 (middle) and 80-84 (bottom) based on model M_{3C}.

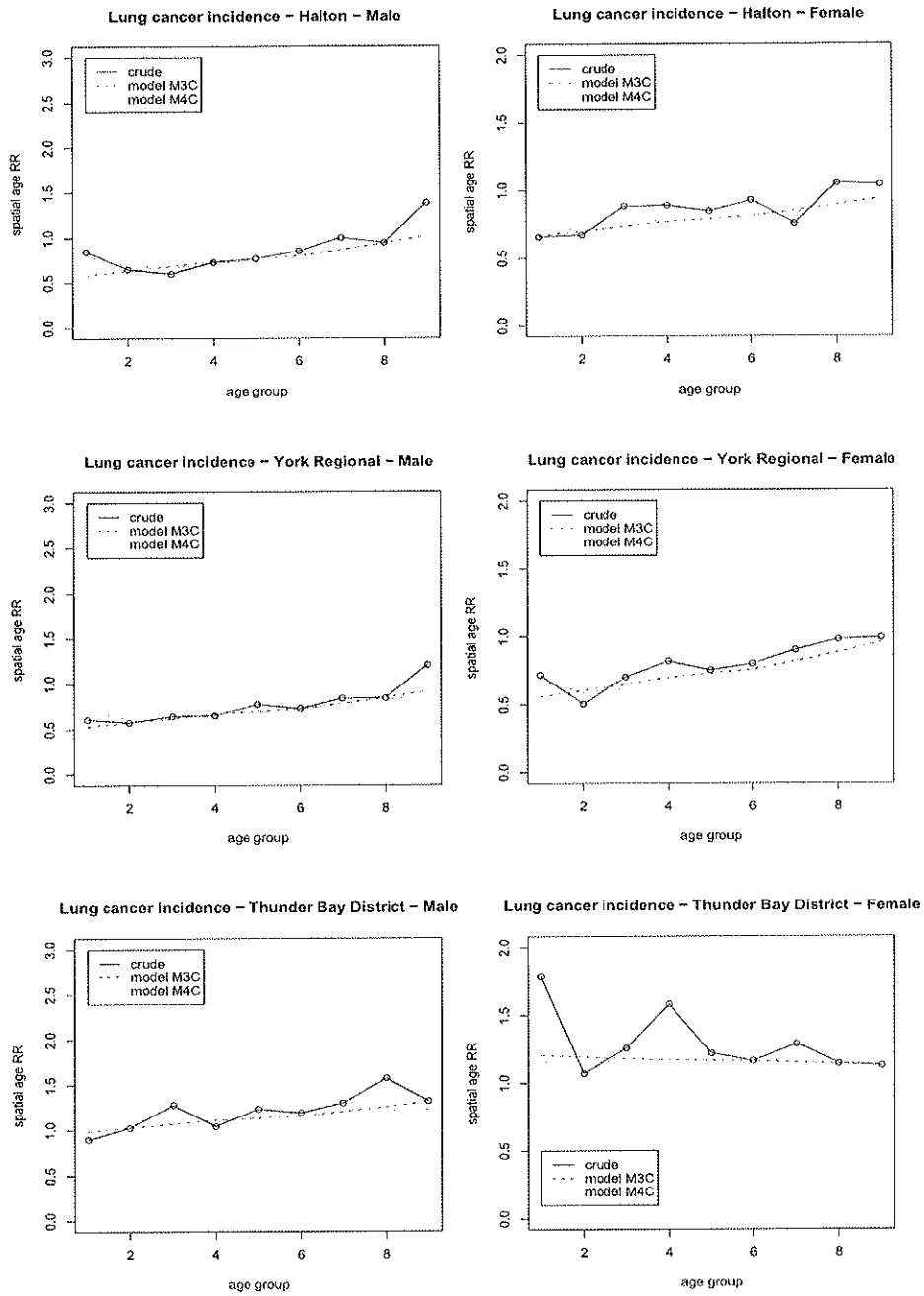

Figure 8: Plots of the standardized crude rate (——) for specific PHUs by age and gender (males/left and females/right) and estimated lung cancer incidence based on model M_{3C} (- - -) and model M_{4C} (· · ·) - Part 1.

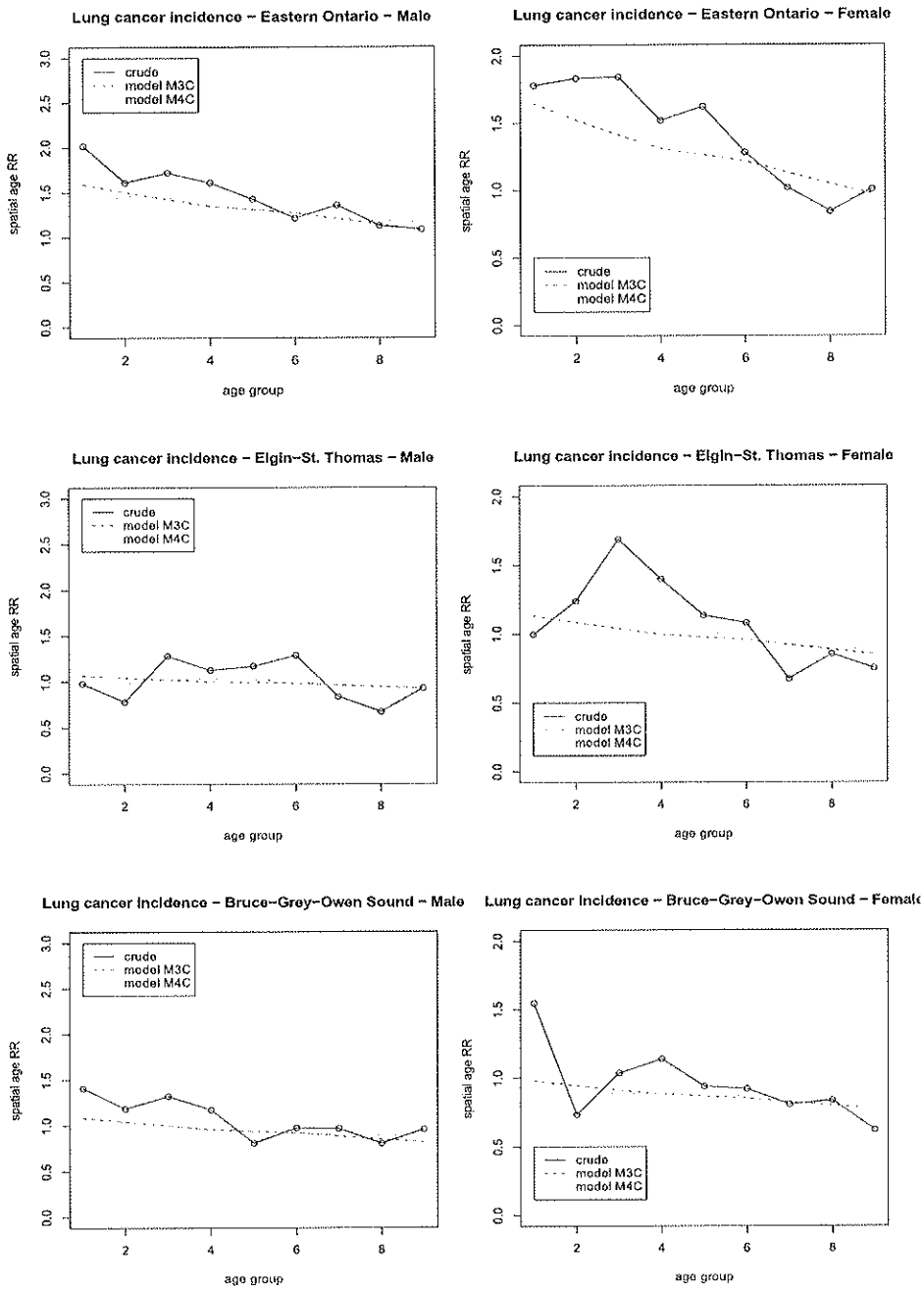

Figure 9: Plots of the standardized crude rate (——) for specific PHUs by age and gender (males/left and females/right) and estimated lung cancer incidence based on model M_{3C} (- - -) and model M_{4C} ($\cdots$) - Part 2.

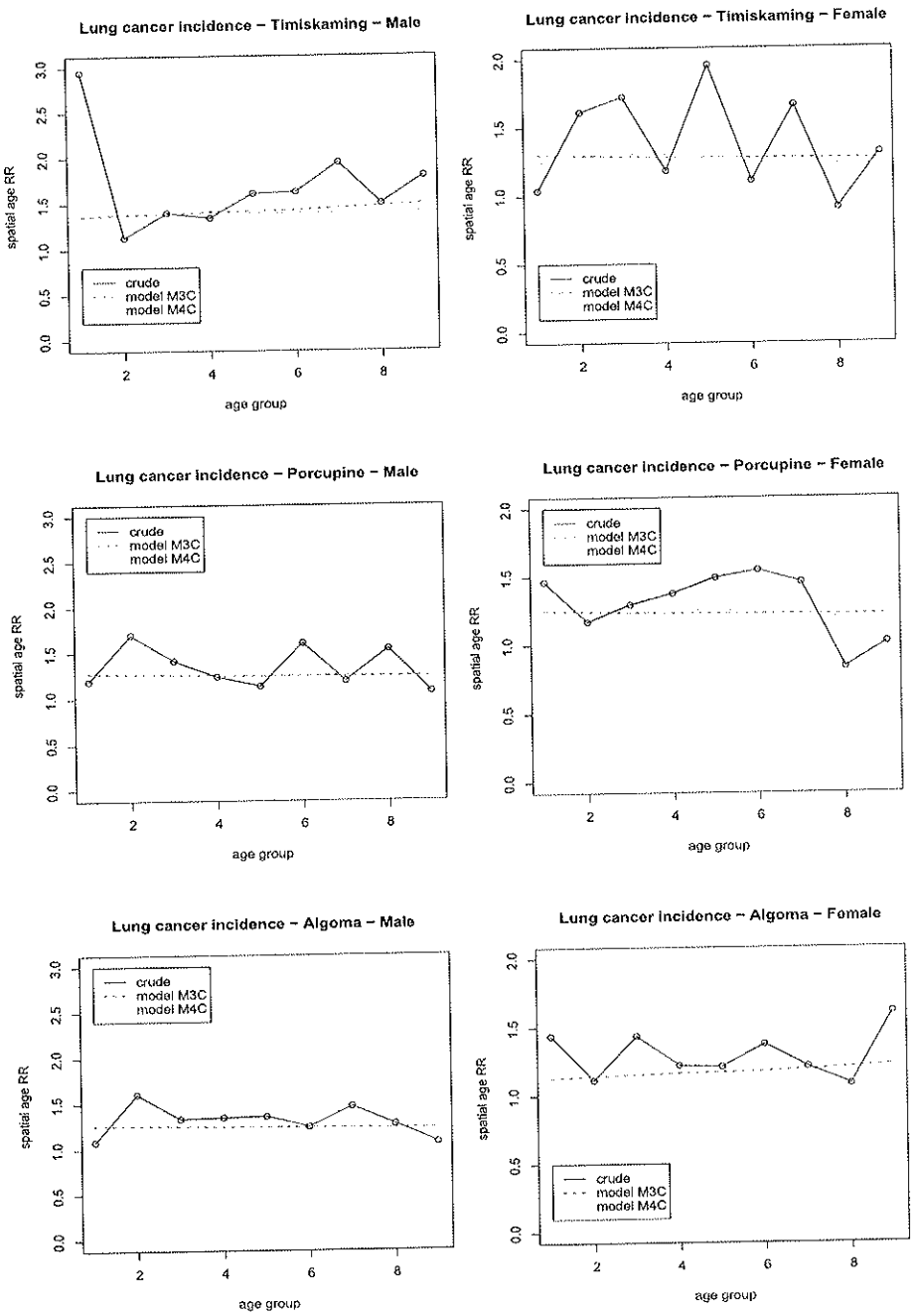

Figure 10: Plots of the standardized crude rate (——) for specific PHUs by age and gender (males/left and females/right) and estimated lung cancer incidence based on model M_{3C} (- - -) and model M_{4C} (· · ·) - Part 3.