STATISTICA, anno 1, n. LXXVI, 2016

ON THE SPECIFICATION OF PRIOR DISTRIBUTIONS FOR VARIANCE COMPONENTS IN DISEASE MAPPING MODELS

Enrico Fabrizi
DISES, Università Cattolica, Piacenza, Italia.
Fedele Greco [1]
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia
Carlo Trivisano
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia.

1. Introduction

Maps of mortality or morbidity standardised rates at the *small areas* level are common tools used by epidemiologists to describe the geographical variation of diseases. For area i ($i = 1, \ldots, m$), the standardised mortality/morbidity rates (SMR_i), an estimate of the area's relative risk associated with a specific disease, is defined as the ratio of the observed number of cases (Y_i) to the expected count (E_i). In other words, $SMR_i = Y_i/E_i$. The expected incidence E_i would be observed if the disease risk was constant over the entire study region and if the spatial variations in incidence were caused exclusively by variations in population density and structure with respect to gender, age or other relevant variables. The expected incidence reflects the 'null-hypothesis spatial distribution' of the cases. The goal of disease mapping is to identify features of the geographical variation of the risks that are not captured by the null-hypothesis distribution (Lawson, 2001).

When the disease under study is rare, counts Y_i are heavily affected by random variability, and the estimates of the relative risk at the small-area level are unstable, particularly in areas with small populations. This instability leads to overdispersed maps, which are characterised by the presence of (possibly false) 'hot spots', and many estimates equal to 0. Using model-based techniques to obtain smoother maps is very popular, and a vast literature has been devoted to the topic (see Lawson (2009) for a recent review). The basic idea underpinning disease mapping models is 'borrowing strength', that is improving the precision of estimation, through the use of auxiliary area-level information or linking neighbouring areas through random effects. The latter strategy seeks to capture spatial and non-spatial structures that are unaccounted for by the auxiliary information.

[1] Corresponding Author. E-mail: fedele.greco@unibo.it

In this article we consider a popular disease mapping model known as the Besag-York-Mollié model (BYM, see Besag *et al.* (1991)). A distinguishing feature of this model is the use of two sets of random effects: one spatially structured to model spatial autocorrelation and the other spatially unstructured to describe residual unstructured heterogeneity. As the BYM model belongs to the class of generalised linear mixed models with complex random effects structures, it is often analysed with a Bayesian approach and typically assisted by MCMC or other computer-intensive integration techniques.

Bayesian analysis requires prior specification for all the unknown parameters in the model. This article examines prior specification for the variance components related to random effects in the BYM model. This topic has received attention in the literature (see Bernardinelli *et al.* (1995); Mollié (1999); Eberly and Carlin (2000) and Wakefield (2007)) because the choice of priors has a non-negligible influence on the posterior distribution of relevant parameters given the structure of the BYM model. Moreover, non-informative reference choices may lead to serious computational problems (Eberly and Carlin, 2000).

We propose using generalised inverse Gaussian (GIG) distributions for the priors on variance components. GIGs comprise many distributions used as priors for variance components as special cases, such as the gamma, inverse gamma and inverse Gaussian distributions. Using this general class of priors allows us to control the weights of the tails of the implied marginal priors on relative risks and to select hyperparameters such that these priors have finite moments up to a prespecified order. Priors on relative risks with finite expectations (and variances) make the inclusion of prior information straightforward and allows the implementation of our new proposal of balancing the prior weight of the spatially structured and unstructured components of variance. Moreover, GIG priors also provide an easy way to control the amount of shrinkage associated with posterior means of relative risks. The use of GIG priors in the estimation of the parameters of a log-normal distribution has been explored by Fabrizi and Trivisano (2012).

The rest of this article is organised as follows. In section 2, we review the Besag-York-Mollié model and the problems associated with prior specification on its variance components. In section 3, the GIG distribution is introduced, along with the related family of generalised hyperbolic distributions. Section 4 contains the description of our proposal about prior specification. In section 5, we use a simulation study to compare the performances of posterior estimates of relative risks obtained under the proposed priors with the performances of a wide range of priors on variance components now popular in the literature. Section 6 contains an application of the Besag-York-Mollié model to the analysis of two real data sets. Some conclusions are offered in section 7.

2. The Besag-York-Mollié model

Consider a study region subdivided into m contiguous areas. Let Y_i and E_i denote, respectively, the observed and expected counts for a disease in the i-th area ($i = 1, \ldots, m$). Expected counts for each area can be obtained by applying a standard table of group-specific sex and age rates to each area-specific background

population, subdivided by age and sex. The Besag-York-Mollié (BYM) model consists in the following set of assumptions:

$$Y_i|\theta_i \sim Poisson(\theta_i E_i) \quad (1)$$

$$\log\theta_i = \mu_i + v_i + u_i \quad (2)$$

in which $\mu_i = \mathbf{x}_i^t\beta$ describes the effect of area-level covariates $\mathbf{x}_i$, whereas vectors $\mathbf{v} = (v_1, \ldots, v_m)$ and $\mathbf{u} = (u_1, \ldots, u_m)$ represent unstructured and spatially structured random effects, respectively. Specifically, the following priors are assumed:

$$\mathbf{v}|\sigma_v^2 \sim MVN\Big(\mathbf{0}, \sigma_v^2\mathbf{I}\Big) \quad (3)$$

$$\mathbf{u}|\sigma_u^2 \propto \exp\Big(-\frac{\sigma_u^{-2}}{2}\mathbf{u^t Q u}\Big) \quad (4)$$

The implied prior on $v_i + u_i$ is termed a 'convolution prior' since it is the sum of two independent components. Prior (4) is typically obtained from a univariate specification in terms of conditional distributions $u_i|u_j$, $i \neq j$, for instance by means of Gaussian conditionally autoregressive models (CAR, see Besag (1968)), a special case of Gaussian Markov Random Fields. The basic idea of these models is that, conditional upon its neighbours, u_i is independent of all u_js at non-neighbouring areas. A neighbouring structure needs to be specified. Many structures have been proposed in the literature; in this paper, we follow the most popular approach: areas are considered to be neighbours if they share a boundary, and information about neighbouring is summarised in the $m \times m$ matrix $\mathbf{W}$, whose entries $w_{ij} = 1$ if areas i and j share a boundary and 0 otherwise ($w_{ii} = 0$ completes the specification). In the BYM model, the full conditional distributions are expressed as follows:

$$u_i|u_j, j \neq i, \sigma_u^2 \sim N\left(\sum_j \frac{u_j}{n_i}, \frac{\sigma_u^2}{n_i}\right) \quad (5)$$

in which n_i is the number of neighbours of the i-th area. As a consequence, $\mathbf{Q} = (\mathbf{D} - \mathbf{W})$, where $\mathbf{D} = diag(n_i)$. The matrix $\mathbf{Q}$, defined in this way, is of rank $m-1$. Therefore, according to prior (4), the intercepts u_i are only uniquely determined up to an additive constant. In line with Besag and Kooperberg (1995), we impose the constraint $\sum_{i=1}^m u_i = 0$. We may define $\mathbf{A} = [\mathbf{I}_{m-1}|\mathbf{1}]$ a matrix such that $\mathbf{Aw} = \mathbf{0}$, $\forall \mathbf{w} \in \Re^m$. The vector of random effects

$$\mathbf{z} = \mathbf{Au} \quad (6)$$

is characterised by an $(m-1)$-dimensional proper distribution $\mathbf{z} \sim MVN(\mathbf{0}, \sigma_u^2\Sigma_z)$ with $\Sigma_z = \big[\bar{\mathbf{A}}^t(\mathbf{D} - \mathbf{W})\bar{\mathbf{A}}\big]^{-1}$ where $\bar{\mathbf{A}}$ is the Moore-Penrose inverse of $\mathbf{A}$. Under the constraint $\sum_{i=1}^m u_i = 0$, we may write $\mathbf{u} = \bar{\mathbf{A}}\mathbf{z}$ whose positive-semidefinite covariance matrix is $\sigma_u^2\bar{\mathbf{A}}\big[\bar{\mathbf{A}}^t(\mathbf{D} - \mathbf{W})\bar{\mathbf{A}}\big]^{-1}\bar{\mathbf{A}}^t = \sigma_u^2\Sigma_u$. We may eventually write

$$\mathbf{u}|\sigma_u^2 \sim MVN(0, \sigma_u^2\Sigma_u) \quad (7)$$

i.e., $\mathbf{u}$ follows an Intrinsic Conditional Autoregressive Gaussian distribution.

2.1. Priors on the variance components

To complete the specification of the model, the priors for the variance components σ_v^2 and σ_u^2 are needed. Priors for these parameters should be chosen carefully, and a rich literature is devoted to this topic (see Bernardinelli *et al.* (1995); Mollié (1999); Eberly and Carlin (2000) and Wakefield (2007)). The problem in choosing these priors is with the marginal priors on u_i and v_i, which are induced from those on σ_v^2 and σ_v^2. Intuitively, a single data point (Y_i) cannot provide information about u_i and v_i individually, it can only provide information about their sum; u_i and v_i are said to be Bayesianly unidentified (Eberly and Carlin, 2000). Although Bayesian inference may be conducted on these parameters, improper priors on the random effects lead to improper posteriors that manifest themselves as convergence failures when the model is analysed using MCMC algorithms. Similarly, nearly improper (i.e., proper but very heavy-tailed) prior distributions may lead to near-improperty and very slowly converging MCMC chains. For computational convenience, proper inverse gamma priors on the variance components are popular. Some of the most popular choices of hyperparameters are as follows:

$$p(\sigma_v^2) = p(\sigma_u^2) = Inv - Gamma(0.5, 0.0005) \tag{8}$$

(Wakefield and Morris, 2001);

$$p(\sigma_v^2) = p(\sigma_u^2) = Inv - Gamma(0.001, 0.001) \tag{9}$$

(Best *et al.*, 1999).

If priors $\sigma_v^2 \sim Inv - Gamma(a_v, b_v)$ and $\sigma_u^2 \sim Inv - Gamma(a_u, b_u)$ are assumed, the implied distributions on $\mathbf{v}$ and $\mathbf{u}$ are multivariate t; in particular $v_i \sim t_{2a}(0, b_v/a_v)$ is a heavy-tailed distribution. The first moment of v_i does not exist in any of the previously mentioned parameter choices. More generally, inversa gamma priors give rise to a multivariate $\log - t$ distribution on the relative risk scale, i.e. considering $\exp(\mathbf{v})$ and $\exp(\mathbf{u})$; this distribution has no moments of all orders.

In this weakly informative setting, the priors keep their influence on the posteriors for θ_i (particularly when E_i s are small), and heavy-tailed priors, inconsistent with small amounts of variability observed in many practical applications, lead to inefficient "Bayes estimators" of the relative risks (i.e. $E(\theta_i|data)$ with large frequentist mean square error).

Another problem in prior specification is balancing the priors of the two random effects, which is related to inference on parameters such as

$$\phi = \frac{SD(\mathbf{u})}{SD(\mathbf{v}) + SD(\mathbf{u})} \tag{10}$$

(see Best *et al.* (1999)) in which $SD(.)$ is the marginal standard deviation of a vector. This parameter is relevant to analysts, as it expresses the relative weights of spatial clustering and unstructured heterogeneity characterising the residual variation in log-relative risks (i.e., unaccounted for by the area-level covariates $\mathbf{x}_i$). Posterior inference on ϕ is sensitive to the priors chosen for σ_v^2 and σ_v^2.

Moreover, $p(\sigma_v^2) = p(\sigma_u^2)$ is not a neutral choice because σ_v^2 is a marginal variance whereas σ_u^2 is a conditional variance.

Mollié (1999) notes that $V(u_i + v_i | u_j, j \neq i, \sigma_u^2, \sigma_v^2) = \frac{\sigma_u^2}{n_i} + \sigma_v^2$, so she proposes to approximately balance a priori the two set of random effects using the variance of the log-SMRs as a guess on the total variability of the log relative risks and scaling the prior mean of σ_u^2 with the average number of neighbours $\bar{n} = m^{-1} \sum_i n_i$. An interesting balancing procedure is suggested by Wakefield (2007). He considers the set of random effects z_i, $i, \ldots, m-1$ defined in (6). The procedure starts by defining the average marginal variance $\bar{\sigma}_z^2 = \bar{a}\sigma_u^2$, in which $\bar{a}$, a constant depending on the neighbouring structure, is the average of the marginal variances of z_i (i.e., the average of the diagonal entries of the matrix $\mathbf{\Sigma_z}$). Next, Wakefield (2007) specifies a prior on $\tau_T = (\sigma_v^2 + \bar{\sigma}_z^2)^{-1}$, which is approximately the marginal precision of the residual relative risk $u_i + v_i$. He assumes $\tau_T \sim Gamma(a, b)$, introduces the parameter $p = \sigma_v^2/(\sigma_v^2 + \bar{a}\bar{\sigma}_z^2)$ to govern the relative weights of the random effects and suggests $p \sim Beta(c, d)$. Priors on the variance components are induced from the identities $\sigma_v^2 = (1-p)\tau_T^{-1}$ and $\sigma_u^2 = p\tau_T^{-1}/\bar{a}$. Specifically, in the empirical analysis of Scottish lip cancer data discussed in the paper, Wakefield (2007) specifies $\tau_T \sim Gamma(1, 0.0260)$ and $p \sim Beta(1, 1)$.

3. The generalised inverse Gaussian and the multivariate generalised hyperbolic distributions

In this section, we briefly introduce the generalised inverse Gaussian (GIG) and generalised hyperbolic (GH) distributions, establish the notation and mention some key properties that will be used later. For more details on these distributions, see Bibby and Sørensen (2003) and Bibby and Sørensen (2004), among others.

The density of the GIG distribution may be written as follows:

$$f(x) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{1}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\Big\{ -\frac{1}{2}\big(\delta^2 x^{-1} + \gamma^2 x\big)\Big\} \mathbf{1}_{\Re^+} \tag{11}$$

in which $K_\lambda(.)$ is the modified Bessel function of the third kind (see Bibby and Sørensen (2004) for more details). If $\delta > 0$, the permissible values for the other parameters are $\gamma \geq 0$ if $\lambda < 0$ and $\gamma > 0$ if $\lambda = 0$. If $\delta \geq 0$, then γ, λ should be strictly positive.

The moments of the GIG can be expressed as functions of the Bessel-K functions by

$$E(X^j) = \left(\frac{\delta}{\gamma}\right)^j \frac{K_{\lambda+j}(\delta\gamma)}{K_\lambda(\delta\gamma)}. \tag{12}$$

Many important distributions may be obtained as special cases of the GIG. For $\lambda > 0$ and $\gamma > 0$, the gamma distribution emerges as the limit when $\delta \to 0$. The inverse gamma is obtained when $\lambda < 0$, $\delta > 0$ and $\gamma \to 0$, whereas an inverse Gaussian distribution results when $\lambda = -\frac{1}{2}$.

Barndorff-Nielsen (1977) shows that the marginal distribution of a random vector $\mathbf{X}$ (of size d) for which we have that $\mathbf{X}|W = w \sim MVN(\mathbf{0}, w\Sigma)$ with

$W \sim GIG(\lambda, \delta, \gamma)$ is a multivariate generalised hyperbolic distribution that is symmetric around 0, i.e., $\mathbf{X} \sim MGH(\lambda, \delta, \gamma, \mathbf{0}, \Sigma, \mathbf{0})$, a distribution characterised by the density

$$f_d(\mathbf{x}) = \frac{(\delta/\gamma)^\lambda (\gamma^2)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} K_\lambda(\delta\gamma)} \frac{K_{\lambda-\frac{d}{2}}\big(\sqrt{(\delta^2 + \mathbf{x}^T\Sigma^{-1}\mathbf{x})\gamma}\big)}{\big(\sqrt{(\delta^2 + \mathbf{x}^T\Sigma^{-1}\mathbf{x})\gamma}\big)^{\frac{d}{2}-\lambda}} \tag{13}$$

Many important multivariate distributions may be obtained as special cases: the Gaussian, for $\lambda \to \pm\infty$ or for $\gamma \to +\infty$; the Student's t for $-\infty \leq \lambda \leq -2$ and $\gamma = 0$; and the multivariate Normal inverse Gaussian (NIG) distribution for $\lambda = -\frac{1}{2}$. This distribution is in the class of elliptically contoured distributions, so it is closed under passing to marginal distributions. Diagonal Σ does not imply independence.

For the purposes of this paper, the moment generating function of the multivariate generalised hyperbolic distribution is important and is as follows:

$$M_{MGH}(\mathbf{t}) = E(e^{\mathbf{t}^T X}) = \left(\frac{\gamma^2}{\gamma^2 - \mathbf{t}^T\Sigma\mathbf{t}}\right)^{\frac{\lambda}{2}} \frac{K_\lambda\big(\delta\sqrt{\gamma^2 - \mathbf{t}^T\Sigma\mathbf{t}}\big)}{K_\lambda(\delta\gamma)} \tag{14}$$

provided that $\gamma^2 > \mathbf{t}^T\Sigma\mathbf{t}$. Note that if $\lambda = -.5$, the moment generating function simplifies to $M_{MGH}(\mathbf{t}) = \exp\big(-\delta\sqrt{\gamma^2 - \mathbf{t}^T\Sigma\mathbf{t}} + \delta\gamma\big)$.

The reason for considering the moment generating function is that the resulting marginal distributions for $\mathbf{v}$, $\mathbf{u}$ will be MGH, assuming a prior for the variance components in the GIG class; moments of $\exp(\mathbf{v})$, $\exp(\mathbf{u})$ may be studied directly using the moment generating functions.

4. Generalised inverse Gaussian priors on variance components

In this section, we propose a strategy for choosing the hyperparameters for the priors $p(\sigma_v^2)$ and $p(\sigma_u^2)$ within the GIG class. Our aims are the following: *i*) to properly balance the relative weights of spatially structured and unstructured effects;

ii) to specify priors that induce light tails on $p(\mathbf{v})$ and $p(\mathbf{u})$ in the sense that $E\{\exp(\mathbf{v})\}$, $E\{\exp(\mathbf{v}^t\mathbf{v})\}$, $E\{\exp(\mathbf{u})\}$ and $E\{\exp(\mathbf{u}^t\mathbf{u})\} < +\infty$; and *iii*) to make prior information easy to include when it is available.

We assume GIG priors for σ_v^2 and σ_u^2. The GIG class of distributions includes many special cases that are already popular as priors for variance components in epidemiology.

i) We first consider the problem of balancing the two sets of random effects. We specifically aim to balance their marginal variances, i.e. the expected variance of a sample drawn from the joint distribution of the two random effect vectors. It

may be shown that

$$E\Big\{\frac{1}{m-1}\sum_{i=1}^{m}\big(v_i-\bar{v}\big)^2|\sigma_v^2\Big\}=\sigma_v^2 \tag{15}$$

$$E\Big\{\frac{1}{m-1}\sum_{i=1}^{m}\big(u_i-\bar{u}\big)^2|\sigma_u^2\Big\}=\sigma_u^2\frac{tr(\Sigma_u)}{m-1}=\sigma_u^2 t_u \tag{16}$$

with $\bar{v}=m^{-1}\sum_i v_i$, $\bar{u}=m^{-1}\sum_i u_i$. Note that because v_is are independent and identically distributed, identity (15) is straightforward. To prove (16), note that because $\sum_{i=1}^{m} u_i = 0$ by construction, the left hand side of (16) may be written as $(m-1)^{-1}E(\mathbf{u}^t\mathbf{u}|\sigma_u^2)$. Identity (16) follows from general results on the expectation of quadratic forms.

It is worth noting that the conditional variance σ_u^2 is not necessarily greater than the marginal variance of the spatially structured random effects. In fact, in graph theory, the number t_u is related to the Kirchhoff index (KF) by the relationship

$$t_u=\frac{KF}{m(m-1)}$$

If we consider a cycle graph C_m of size m, then $KF(P_m)=\frac{m(m-1)(m+1)}{12}$ (Palacios, 2001), thus $t_u(C_m)>1$ for $m>11$. In contrast, for a complete graph $t_u=m^{-1}$; therefore, $t_u\leq 1\ \forall m$.

We base the balancing of (15) and (16) on overall measures of the influence of random effects; this overcomes the well-known problem of comparing an unconditional and a conditional variance. To consider two extreme situations, if, on one hand, the total amount of variability is a result of unstructured heterogeneity, then we denote the variance parameter with ς_v^2. If, on the other hand, the total amount of variability is a result of structured heterogeneity, we denote the variance parameter with (ς_u^2). Given the independence between us and vs, the marginal variance of the residual log-relative risks $\xi_i=\log\theta_i-\mu_i$, conditional upon ς_v and ς_u, can be expressed as a linear combination of the variance parameters in the two extreme situations posited above:

$$E\Big\{\frac{1}{m-1}\sum_{i=1}^{m}\big(\xi_i-\bar{\xi}\big)^2|\varsigma_v^2,\varsigma_u^2,\psi\Big\}=(1-\psi)\varsigma_v^2+\psi\varsigma_u^2 t_u$$

If we set $\sigma_v^2=(1-\psi)\varsigma_v^2$, $\sigma_u^2 t_u=\psi\varsigma_u^2 t_u$, this expression may be useful for managing the balance between structured and unstructured components. The parameter $\psi\in(0,1)$ weights the contributions that the two random effects give to the empirical variance of the residual log relative risks; $\psi=0.5$ implies an equal a priori weight.

Thus, assuming $p(\varsigma_v^2)\sim GIG(\lambda_v,\delta_v,\gamma_v)$ and $p(t_u\varsigma_u^2)\sim GIG(\lambda_u,\delta_u,\gamma_u)$ in the purely spatial and non-spatial scenarios and using the properties of the GIG distribution, it follows that the priors on the variance components are given by $p(\sigma_v^2|\psi)\sim GIG(\lambda_v,\delta_v\sqrt{1-\psi},\frac{\gamma_v}{\sqrt{1-\psi}})$ and $p(\sigma_u^2|\psi)\sim GIG(\lambda_u,\delta_u\sqrt{\psi},\frac{\gamma u}{\sqrt{\psi}})$.

Regarding the parameters of these distributions, we suggest a convenience choice for the λ paramters, namely $\lambda_v=\lambda_u=-0.5$. This choice dramatically

simplifies the expression of the residual relative risks' prior moments. Note that this choice implies that the priors being considered in this section are actually inverse Gaussians.

In choosing the remaining parameters, we want the prior to hold the first two moments equal when the variability of $\log \xi_i$ is entirely due to spatially unstructured (structured) random effects; this allows for an adequate prior balance of the two types of heterogeneity. This is achieved if

$$\gamma_v = \frac{\gamma_u}{\sqrt{t_u}} \tag{17}$$

and

$$\delta_u = \frac{\delta_v}{\sqrt{t_u}} \tag{18}$$

In fact, because $E(\varsigma_v^2) = \frac{\delta_v}{\gamma_v}$, $E(t_u\varsigma_u^2) = \frac{\delta_u t_u}{\gamma_u}$, and $V(\varsigma_v^2) = \frac{\delta_v}{\gamma_v^3}$, $V(t_u\varsigma_u^2) = \frac{\delta_u t_u^2}{\gamma_u^3}$, the equalities $E(\varsigma_v^2) = E(\varsigma_u^2 t_u)$ and $V(\varsigma_v^2) = V(\varsigma_u^2 t_u)$ follow from (17) and (18).

As for ψ, it may specified by the user as a constant in the $(0, 1)$ interval. Alternatively, the model can be made more flexible by adding a further level to the hierarchy, i.e. specifying

$$\psi \sim Beta(a, b). \tag{19}$$

The choice $a = 1$, $b = 1$ leads to a prior on the parameter ϕ, defined in (10), quite peaked around 0.5. An alternative choice, leading to approximately $p(\phi) \sim Unif(0, 1)$, may be more advisable. It can be shown numerically that this may be approximately achieved specifying

$$\psi \sim Beta(0.4, 0.4) \tag{20}$$

that is the prior that will be used in the subsequent sections of this paper.

Our approach to the a priori balancing of the two sets of random effects is quite similar to that of Wakefield (2007) in the sense that it depends on the neighbouring structure. Nonetheless, note that our strategy is based on $\sigma_u^2 t_u$, which is a measure of the variability of structured random effects, whereas the approach proposed in Wakefield (2007) is based on $\bar{a}\sigma_u^2$ (i.e., an average marginal variance). Calibration on the basis of $\bar{a}$ downweights the contribution of the structured component because $\bar{a} \geq t_u$. To see this, first note that $tr(\boldsymbol{\Sigma}_u) = tr\left(\overline{\mathbf{A}}\boldsymbol{\Sigma}_z\overline{\mathbf{A}}^t\right) = tr\left(\overline{\mathbf{A}}^t\overline{\mathbf{A}}\boldsymbol{\Sigma}_z\right)$. Because the $i-$th diagonal element of the matrix $\overline{\mathbf{A}}^t\overline{\mathbf{A}}$ is equal to $\frac{m-1}{m}$ by construction, and the off-diagonal elements of this matrix are equal to $-\frac{1}{m}$, we find that

$$\begin{aligned} tr\left(\overline{\mathbf{A}}^t\overline{\mathbf{A}}\boldsymbol{\Sigma}_z\right) &= \frac{m-1}{m}tr(\boldsymbol{\Sigma}_z) - \frac{1}{m}\left[\sum_{ij}\boldsymbol{\Sigma}_z(i,j) - tr(\boldsymbol{\Sigma}_z)\right] \\ &= tr(\boldsymbol{\Sigma}_z) - m^{-1}\sum_{ij}\boldsymbol{\Sigma}_z(i,j) \end{aligned}$$

Because $\sum_{ij}\boldsymbol{\Sigma}_z(i,j) > 0$, it follows that $\overline{a} = \frac{tr(\boldsymbol{\Sigma}_z)}{m-1} > \frac{tr(\boldsymbol{\Sigma}_u)}{m-1} = t_u$. Moreover, Wakefield (2007) specifies a single prior on the marginal precision of the

residual relative risk, allocated to two sets of random effects according to a parameter similar to our ψ, whereas we specify two independent priors with parameters chosen to guarantee the a priori balance. The assumption of GIG priors on σ_v^2 and σ_u^2 implies that $\mathbf{v} \sim MGH(\lambda_v, \delta_v\sqrt{1-\psi}, \gamma_v/\sqrt{1-\psi}, \mathbf{0}, \mathbf{I}_m, \mathbf{0})$ and $\mathbf{u} \sim MGH(\lambda_u, \delta_u\sqrt{\psi}, \gamma_u/\sqrt{\psi}, \mathbf{0}, \Sigma_u, \mathbf{0})$. As a consequence, prior moments of the residual relative risks are available in closed form and can be expressed as follows:

$$E(\xi_i|\psi) = \exp\Big(-\delta_v\sqrt{\frac{\gamma_u^2}{t_u} - (1-\psi)} - \frac{\delta_v}{\sqrt{t_u}}\sqrt{\gamma_u^2 - \psi\Sigma_u(i,i)} + 2\frac{\delta_v\gamma_u}{\sqrt{t_u}}\Big) \tag{21}$$

$$V(\xi_i|\psi) = \exp\Big(-\delta_v\sqrt{\frac{\gamma_u^2}{t_u} - 4(1-\psi)} - \frac{\delta_v}{\sqrt{t_u}}\sqrt{\gamma_u^2 - 4\psi\Sigma_u(i,i)} + 2\frac{\delta_v\gamma_u}{\sqrt{t_u}}\Big) - \tag{22}$$
$$- \big\{E(\xi_i)\big\}^2$$

ii) We set $\gamma_u = \sqrt{4\max(diag(\Sigma_u))} + \varepsilon$ in order to have $V\{\exp(u_i)\} < +\infty$ $\forall i$ for some positive (typically small) ε. Parameter γ_v is set according to (17). When $t_u < 1$, as is the case with 'regular maps' met in most applications in which the number of neighbours is relatively high with respect to m, (17) and (18) imply $V(exp(v_i)) < +\infty$ $\forall i$. The only parameter left to specify is δ_v. For fixed γ_v, parameter δ_v characterizes the prior distribution of ς_v^2, the variance of the unstructured heterogeneity when $\psi = 0$. More precisely, the value chosen for δ_v will reflect prior assumptions on the variability of the risk in the study region. We use a heuristic specification based on assuming that the log residual relative risks coincide with $\mathbf{v}$.

iii) Suppose that we dispose of prior information that enables us to assume that residual relative risks are included in an interval in the form (c^{-1}, c), $c > 1$, with probability $1 - \alpha$. For $\psi = 0$, this is equivalent to say that each v_i lies in an interval $(-\log c, \log c)$ with probability $1 - \alpha$. Because for $\psi = 0$, $\mathbf{v} \sim MGH(\lambda_v, \delta_v, \gamma_v, \mathbf{0}, \mathbf{I}, \mathbf{0})$ - a symmetric distribution around $\mathbf{0}$ whose distribution will depend on δ_v - we can use numerical methods (for instance, the `ghyp` package in `R` (W. Breymann, 2010)) to obtain a value of δ_v such that $\Pr(-log(c) \leq v_i \leq log(c)) = 1 - \alpha$. Figure 1 plots the suggested value for δ_v based on a wide range of values of c and $\alpha = 0.1$.

If prior information about the variation of the residual relative risks is missing a choice of δ_v in the interval $(0.2, 0.4)$ may be reasonable for most applications. Note that although the heuristic argument is assuming only spatially unstructured heterogeneity, the balance between the priors on the two sets of random effects is guaranteed by (17) and (18). The suggested argument for the choice of δ_v is not based on an 'average' or 'typical' area but on the joint distribution of $\mathbf{v}$; using a parallel argument, we could, in principle, base our choice on the distribution of $\mathbf{u}$.

5. A simulation exercise

In this section, we present the results of a simulation exercise aimed at evaluating the impact of the prior specification suggested in section 4 on the posterior distribution of relative risks. Although the approach to estimation is Bayesian,

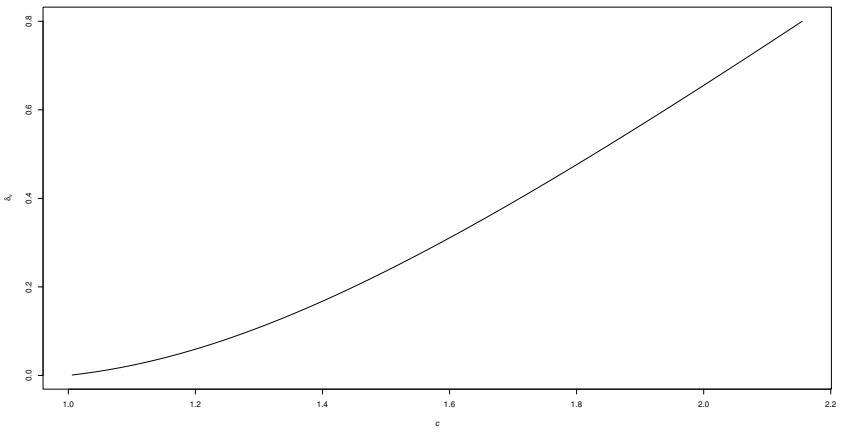

Figure 1 – Suggested δ_v as a function of c at the probability level 0.90

we focus on frequentist properties, namely the average mean square error and the amount of shrinkage. To this end, the posterior means $E(\theta_i|Y_i, E_i)$ are assumed to be the estimators of the unknown θ_i.

We compare these posterior summaries to those obtained using the $Inv - Gamma$ priors (8) and (9) mentioned in section 2.1. We also consider prior balancing of the two variance components according to Mollié (1999), as well as the prior specified by Wakefield (2007).

We consider a region partitioned into $m = 48$ areas. For the purpose of comparison, we also consider a region partitioned into $m = 95$ areas; simulation results obtained on this region are only partially shown for the sake of brevity. For each simulation setting, a set of $R = 500$ relative risks vector $\boldsymbol{\theta}_r = (\theta_{1r}, \ldots, \theta_{mr})$ is generated from the BYM model, assuming $\log(\theta_{ir}) = u_{ir} + v_{ir}$. Expected counts E_i and the proximity structure of the region are taken from a real mortality map (ischemic heart disease in the municipalities of the provinces of Piacenza (48 areas) and Parma (47 areas) in Italy) analysed in Greco and C.trivisano (2009).

Counts Y_{ir} are simulated from a multinomial distribution in order to have $\sum_i Y_{ir} = \sum_i E_i$. To study the effect of various rates of disease incidence on the inference we will consider a multiplying factor k for E_i, Y_{ir}, specifically $k = 0.25, 0.5, 1, 2.5$; the implied average expected counts $\bar{E} = m^{-1} \sum E_i$ are equal to 3.5, 7.1, 14.2, 35.5, respectively.

The marginal variance $V(\boldsymbol{\xi})$ is a function of σ_u^2, σ_v^2 and ϕ. We set $V(\boldsymbol{\xi}) = 0.07$ in order to have approximately 90% of the relative risks in the interval $(0.66, 1.5)$, set specific values of ϕ and choose σ_u^2 and σ_v^2 accordingly. Here, we consider $\phi = 0.5$. Comments on alternative values for ϕ will be discussed later.

To summarise, the simulation study is characterised by several simulation settings in which we allow the number of areas m, the disease incidence (controlled by k), the contribution of spatial and non-spatial heterogeneity (controlled by ϕ) and the total variability of the 'true' relative risk distribution to all vary. Simulation results are only partially shown for the sake of brevity.

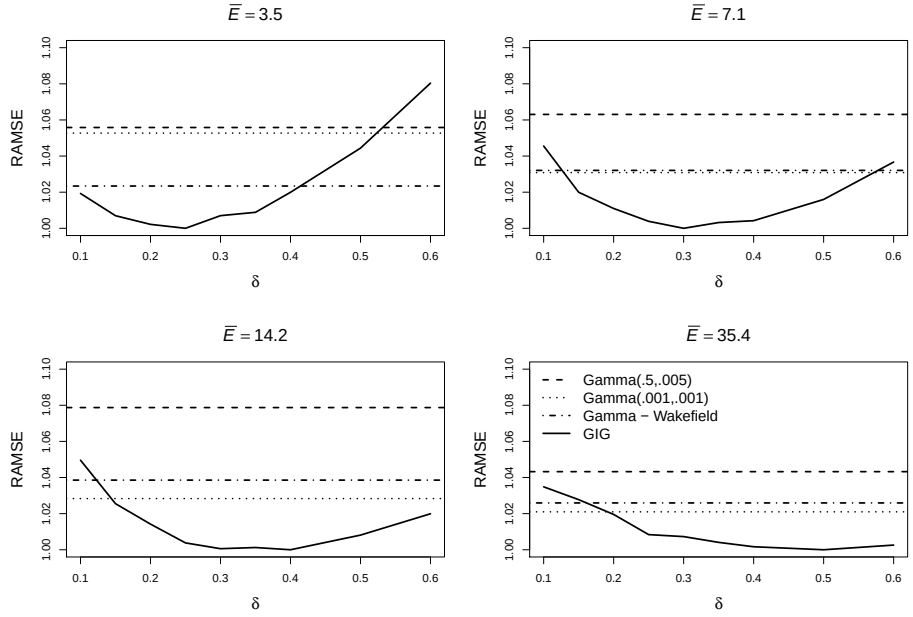

Figure 2 – Different values of δ_v vs $RAMSE(\delta_v)/\min RAMSE(\delta_v)$

For each simulated data set and for each model, inference is based on 10000 samples from the MCMC algorithm, after a burn-in of 15.000 iterations. All the estimated models can be implemented in the Openbugs software (Thomas *et al.*, 2006). The code for estimating the BYM model with GIG priors is available upon request from the authors.

The summary measures that we consider in the analysis of the simulation are defined as follows:

$$RAMSE = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{m}\sum_{i=1}^{m}\left(\frac{\hat{\theta}_{ir}}{\theta_{ir}}-1\right)^2 \tag{23}$$

$$SHR = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{\sum_{i=1}^{m}(\hat{\theta}_{ir}-\hat{\bar{\theta}}_r)^2}{\sum_{i=1}^{m}(\theta_{ir}-\bar{\theta}_r)^2}} \tag{24}$$

where $\hat{\theta}_{ir} = E(\theta_{ir}|Y_{ir}, E_{ir})$, $\hat{\bar{\theta}}_r = m^{-1}\sum_{i=1}^{m}\hat{\theta}_{ir}$ to simplify notation.

Figure 2 shows that the RAMSE is sensitive to the choice of the prior, particularly for small values of $\bar{E}$. For a wide range of choices of δ_v, estimators based on the GIG prior are characterised by lower RAMSE than the considered alternatives. On one hand, if we focus only on the priors (8) and (9), then this is true for almost the entire range of δ_v that we study in the simulation. On the other hand, the prior suggested in Wakefield (2007) is a more serious competitor, especially for $\bar{E} = 3.5$, when it may lead to more efficient estimators (when δ_v is large), although it leads to (almost) uniformly larger RAMSE in the remaining

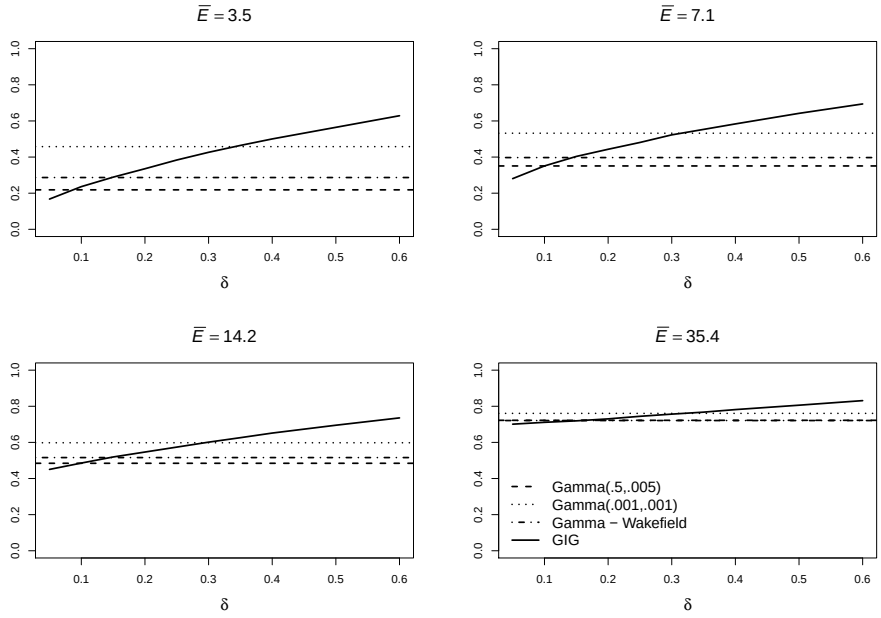

Figure 3 – Different values of δ_v vs $SHR(\delta_v)$

scenarios. The simulation study has been performed by varying the variability of the log-relative risks to $V(\boldsymbol{\xi}) = 0.17$. Results show that, as expected, the optimal value of δ_v moves right in order to capture this higher variability. Nonetheless, loss in RAMSE due to a choice of δ_v in the interval $(0.2, 0.4)$ is not dramatic.

The priors (8) and (9) perform similarly when $\bar{E} = 3.5$ whereas, for larger $\bar{E}$, the prior (9) improves its relative performances. For $\bar{E} = 35.4$, the prior (9) is the closest to the priors that we propose. The balancing of gamma priors according to Mollié (1999) does not yield any improvement in terms of RAMSE; although not reported in the plot, we observed that the performances of these balanced specifications are always close to, although slightly worse than their equivalents that are based on exactly the same prior distributions for the two variance components.

Regarding GIG priors, the RAMSE is an approximately U-shaped function of δ_v. In line with expectation, the curve becomes less pronounced as $\bar{E}$ grows. The minimum seems to move to the right as $\bar{E}$ grows, but it should be noted that the plot is rather flat around the minimum in all cases.

Figure 3 suggests an interpretation of δ_v as a shrinkage parameter, since the shrinkage (as measured by (24)) decreases almost linearly with δ_v. As expected, the slope of this line decreases as $\bar{E}$ grows. For $\delta_v \geq 0.3$, the posterior means obtained under the prior specification that we suggest shrink less than all the alternatives considered; the difference is rather large unless we consider the prior (9) that leads to more moderate shrinkage with respect to the other two.

If evidence from figures 2 and 3 are combined, we see that small δ_vs, (i.e.

$\delta_v \leq 0.2$) are suboptimal in terms of both efficiency (RAMSE) and the amount of shrinkage. Large values (i.e. $\delta_v \geq 0.4$) imply less shrinkage with respect to the alternatives, but may be not as efficient when the disease under study is very rare. Intermediate choices of δ_v (i.e $0.2 < \delta_v < 0.4$), lead to improved efficiency and comparable amount of shrinkage with respect to inverse gamma priors.

In table 1, we present Monte Carlo expectations of posterior means $E_{MC}\{E(\phi|\mathbf{d})\}$ and standard deviations $E_{MC}\{SD(\phi|\mathbf{d})\}$ for different population values of the parameter ϕ defined in (10), different levels of k and different number of areas ($m = 48$, $m = 95$). Note that $\mathbf{d} = (Y_i, E_i)$, $i = 1, \ldots, m$ is shortcut notation for the data on which the posterior is conditioned. The GIG priors considered in this comparison are those obtained when setting $\delta_v = 0.25$.

Alternative prior specifications have a moderate, although non-negligible, impact on the posterior distributions of the parameter ϕ. All of the considered priors seem to allow for a relevant learning about the parameter; posterior means get closer to the actual ϕ population value as the information contained in the data gets larger in terms of the average expected cases $\bar{E}$ and the number of areas (with the exception of the Wakefield prior). The posterior means associated with the GIG priors and with the $Inv - Gamma(0.001, 0.001)$ specified for both variance components are very close; nonetheless, GIG priors lead, on average, to smaller posterior standard deviations, and thus to a better identification of the parameter ϕ (which is notoriously difficult to estimate). Underestimation of the weight of the spatial component is expected when using Wakefield priors, because $\bar{a}$ is not suited for prior balancing of the variances and because this prior is quite informative on the total random effect variance. Moreover, without carefully balancing the moments of $p(\sigma_v^2)$ and $p(\sigma_u^2)$, the GIG priors may also lead to poor identification of ϕ, which is a consequence of the fact that these priors have lighter tails.

6. An application on real data

In this section two case studies are examined. Data refer to death counts observed from 1998 to 2001 in the municipalities of the Italian provinces of Parma and Piacenza, regions previously considered in section 5. We consider two different diseases: stomach cancer (STC) and genitourinary system diseases (GSD). In the study period, stomach cancer is characterised by an average municipality count of 14.2, whereas the average count for genitourinary disease is equal to 6.2. For each disease, expected counts are obtained via external standardisation by applying the sex-age specific disease rates for the region of Emilia Romagna (which includes the two provinces along with seven others) to the municipal populations. In panels [SMR] of figures 4 and 5, the SMRs for STC and GSD are mapped.

We analyse these data using the BYM model with four different prior specifications for the variance components. The considered priors are as follows:

- prior [1]: $p\left(\sigma_v^2\right) = p\left(\sigma_u^2\right) = Inv - Gamma\left(0.5, 0.0005\right)$ i.e., prior (8) of section 2.1;
- prior [2]: $p\left(\sigma_v^2\right) = p\left(\sigma_u^2\right) = Inv - Gamma\left(0.001, 0.001\right)$ i.e., prior (9) of

TABLE 1
$E_{MC}\{SD(\phi|\mathbf{d})\}$ and $E_{MC}\{SD(\phi|\mathbf{d})\}$ for various ϕ in the population, different levels of k and prior specifications. For simplicity the notation $m_\phi = E_{MC}\{SD(\phi|\mathbf{d})\}$ and $sd_\phi = E_{MC}\{SD(\phi|\mathbf{d})\}$ is adopted

$m = 48$									
k	ϕ	GIG priors		Wakefield priors		$IGamma(.5,.0005)$		$IGamma(.001,.001)$	
		m_ϕ	sd_ϕ	m_ϕ	sd_ϕ	m_ϕ	sd_ϕ	m_ϕ	sd_ϕ
0.25	0.33	0.45	0.19	0.39	0.16	0.41	0.23	0.44	0.23
0.5	0.33	0.44	0.18	0.39	0.15	0.40	0.23	0.43	0.22
1	0.33	0.42	0.16	0.38	0.15	0.36	0.22	0.43	0.21
2.5	0.33	0.37	0.14	0.35	0.13	0.30	0.19	0.37	0.18
0.25	0.5	0.48	0.18	0.39	0.16	0.43	0.23	0.46	0.23
0.5	0.5	0.48	0.17	0.41	0.16	0.43	0.24	0.48	0.21
1	0.5	0.50	0.17	0.41	0.15	0.44	0.23	0.48	0.21
2.5	0.5	0.49	0.14	0.43	0.14	0.44	0.20	0.49	0.18
0.25	0.66	0.51	0.18	0.43	0.17	0.44	0.23	0.49	0.23
0.5	0.66	0.52	0.17	0.43	0.16	0.49	0.23	0.51	0.22
1	0.66	0.54	0.16	0.45	0.16	0.52	0.23	0.53	0.19
2.5	0.66	0.57	0.14	0.48	0.14	0.57	0.20	0.57	0.17
$m = 95$									
k	ϕ	GIG priors		Wakefield priors		$IGamma(.5,.0005)$		$IGamma(.001,.001)$	
		m_ϕ	sd_ϕ	m_ϕ	sd_ϕ	m_ϕ	sd_ϕ	m_ϕ	sd_ϕ
0.25	0.33	0.43	0.15	0.40	0.14	0.38	0.20	0.42	0.20
0.5	0.33	0.39	0.14	0.36	0.13	0.32	0.17	0.38	0.18
1	0.33	0.34	0.12	0.34	0.12	0.28	0.15	0.36	0.16
2.5	0.33	0.32	0.11	0.33	0.10	0.24	0.13	0.33	0.13
0.25	0.5	0.49	0.16	0.44	0.15	0.47	0.21	0.48	0.20
0.5	0.5	0.49	0.15	0.43	0.14	0.48	0.18	0.49	0.18
1	0.5	0.50	0.13	0.44	0.13	0.50	0.18	0.51	0.16
2.5	0.5	0.47	0.12	0.45	0.11	0.44	0.15	0.48	0.14
0.25	0.66	0.52	0.16	0.45	0.15	0.50	0.20	0.52	0.20
0.5	0.66	0.56	0.14	0.48	0.14	0.55	0.18	0.55	0.17
1	0.66	0.59	0.13	0.53	0.13	0.62	0.16	0.60	0.15
2.5	0.66	0.608	0.11	0.54	0.10	0.62	0.13	0.61	0.12

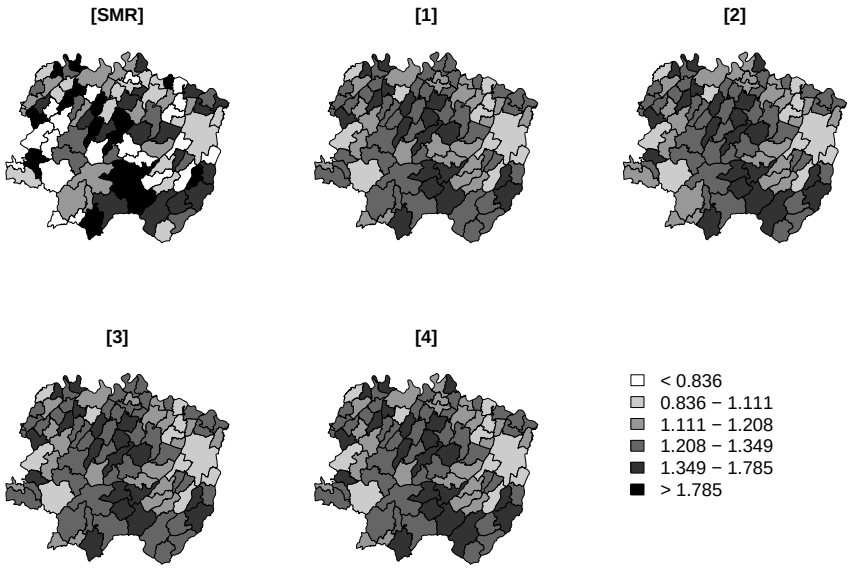

Figure 4 – Maps of standardized mortality rates and posterior means of relative risks under alternative prior specifications for stomach cancer

section 2.1;

- prior [3]: $p(\tau_T) = Gamma(1, .026)$, $p \sim Unif(0,1)$ and $\bar{a} = 0.917$. This follows the prior specification strategy proposed in Wakefield (2007) and discussed in section 2.1;
- prior [4]: $p\left(\sigma_v^2 \,|\psi\right) = GIG\left(-0.5, 0.25\sqrt{1-\psi}, \frac{3.7}{\sqrt{1-\psi}}\right)$, $p\left(\sigma_u^2 \,|\psi\right) = GIG\left(-0.5, 0.373\sqrt{\psi}, \frac{2.5}{\sqrt{\psi}}\right)$, $p(\psi) = Beta(0.4, 0.4)$. Hyperparameters are chosen according to the discussion of section 4 and $t_u = 0.45$.

Constants $\bar{a}$ and t_u are calculated from the adjacency matrix of the map under study.

For each considered model, inference is based on 50.000 samples from the MCMC algorithm, after 50.000 burn-in iterations for all the estimated models. Convergence has been checked via the graphical examination of the trace plots of sample values against iteration, and of the autocorrelation plot in each chain.

The maps for STC and GSD are plotted in figure 4 and 5 respectively.

Spatial features of the distribution of relative risks are not evident from the SMR maps because of sampling variation characterizing counts. Smoothed maps, no matter what prior is chosen, show a clear spatial pattern for GSD, whereas a less pronounced pattern is apparent for STC.

With reference to GSD, we note that the posterior mean of ϕ is 0.695 when prior [4] is used, and that the posterior mean takes intermediate values (0.61 and

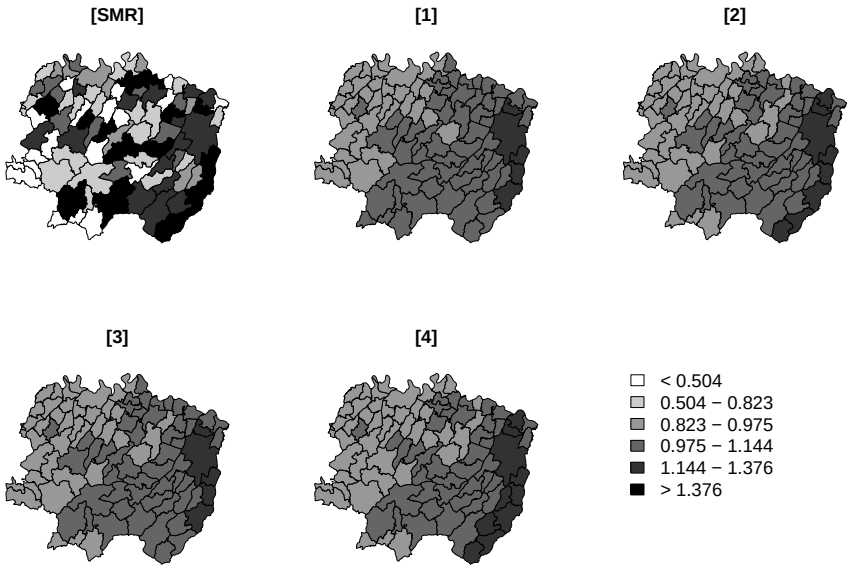

Figure 5 – Maps of standardized mortality rates and posterior means of relative risks under alternative prior specifications for genitourinary system diseases

0.56) when priors [1] and [2] are considered, whereas it equals only 0.51 under prior [3]. Thus, prior [4] to give a result that is more consistent with map appearance. Regarding prior [3], underestimation of ϕ is expected because the balancing of the priors favours the spatially unstructured effect. For STC, we find that all posterior means of ϕ are lower than 0.5, which highlights a predominance of the spatial component, consistent with the map appearance. More precisely, posterior means of ϕ range from 0.29 under prior [1] to 0.40 under prior [2], whereas prior [4] gives an intermediate value equal to .34.

From both sets of maps, we note that the BYM model shrinks the distribution of the SMRs considerably. As expected, the shrinkage of the SMRs toward the regional mean is far higher for GSD because of its rarity. Nonetheless, different priors lead to estimates characterised by different empirical variances. Specifically, the variance of the groups of estimates is maximised under prior [4] for both STC and GSD. If we compare the models associated with the four priors in terms of DIC , a popular model comparison tool, the model based on prior [4] is characterised by the minimum DIC value for both STC and GSD. The reduced shrinkage caused by this prior is therefore associated with the best fit.

7. Conclusions

In this paper, we discuss the use of generalised inverse Gaussian distributions as priors for the variance components in a Besag-York-Mollié model. The main aim

of this work is to propose a simple strategy for this specification that allows one to highlight the meaning of prior specification in terms of a hypothesis on the amount and structure of the variability of the disease under study. This problem has only been partially addressed in a variety of papers on this topic, sometimes with controversial results. The generalised inverse Gaussian distribution includes the inverse-gamma, a popular choice for these parameters, as a special case. Working with these more general distributions allows us to keep the weights of the tails of the implied priors on random effects under control imposing the existence of prior moments of relative risks up to a pre-specified order and employ a parameter that can be used to incorporate prior information on the variability of residual relative risks when it is available. In a simulation study, it is shown that this parameter is also strongly related to the amount of shrinkage that characterises posterior estimates. Moreover, we build our prior specification strategy in order to properly balance the a priori weight attributed to each of the two sets of random effects (the spatially structured and unstructured heterogeneity). Generalised inverse Gaussian distributions thus provide a generalisation of the inverse gammas for which the parameters may be easily and sensibly chosen. Both the simulation study and the analysis of real case studies highlight that our prior specification strategy seems to produce better results in terms of fit and similar (or lower) amount of shrinkage when compared with results obtained under more popular priors. We did not fully explore the potential of generalised inverse Gaussian priors in the Bayesian analysis of disease mapping models; we illustrated only one possible specification strategy and others could be considered. The same priors may be applied to the analysis of different models, especially those based on a log-normal likelihood or involving a log-normal prior, as in the case of count data models, which are widely popular in epidemiology and in other fields of scientific enquiry.

Acknowledgements

The research activity of Fedele Greco and Carlo Trivisano was partially funded by a FIRB 2012 grant (title: Statistical modeling of environmental phenomena: pollution, meteorology, health and their interactions) for research projects by the Italian Ministry of Education, Universities and Research.

References

O. Barndorff-Nielsen (1977). *Exponentially decreasing distributions for the logarithm of particle size.* Proceedings of the Royal Statistical Society, series A, 353, pp. 401–419.

L. Bernardinelli, D. Clayton, C. Montomoli, M. G. M, M. Songini (1995). *Bayesian estimates of disease maps: how important are priors.* Statistics in Medicine, 14, pp. 2411–2431.

J. Besag (1968). *Spatial interaction and the statistical analysis of lattice systems*

(with discussion). Journal of the Royal Statistical Society, series B, 36, pp. 192–236.

J. Besag, C. Kooperberg (1995). *On conditional and instrinsic autoregressions.* Biometrika, 82, pp. 733–746.

J. Besag, J. York, S. Mollié (1991). *A bayesian image restoration, with two applications in spatial statistics (with discussion).* Annals of the Institute of Statistical Mathematics, 43, pp. 1–59.

N. Best, L. Waller, A. Thomas, E. Conlon, R. Arnold (1999). *Bayesian models for spatially correlated disease and exposure data.* In J. Berardo, J. Berger, A. Dawid, A. Smith (eds.), *Sixth Valencia International Meeting on Bayesian Statistics*, Oxford University Press, London, pp. 131–156.

B. Bibby, M. Sørensen (2003). *Generalized hyperbolic and inverse gaussian distributions: limiting cases and approximation of processes.* In S. Rachev (ed.), *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, New York, pp. 211–248.

B. Bibby, M. Sørensen (2004). *Hyperbolic processes in finance.* In R. Dalang, M. Dozzi, F. Russo (eds.), *Seminar on Stochastic Analysis, Random Fields and Applications IV, Progress in Probability*, Birkhäuser Verlag, Berlin, pp. 221–264.

L. Eberly, B. Carlin (2000). *Identifiability problems and convergence issues for markov chain monte carlo fitting of spatial models.* Statistics in Medicine, 19, pp. 2279–2294.

E. Fabrizi, C. Trivisano (2012). *Bayesian estimation of log-normal means with finite quadratic expected loss.* Bayesian Analysis, 4, pp. 975 – 996.

F. Greco, C.trivisano (2009). *A multivariate car model for improving the estimation of relative risks.* Statistics in Medicine, 28, pp. 1707–1724.

A. Lawson (2001). *Statistical methods in spatial epidemiology.* Wiley, Chichester.

A. Lawson (2009). *Bayesian disease mapping: hierarchical modelling in spatial epidemiology.* Chapman and Hall, New York.

A. Mollié (1999). *Bayesian and empirical bayes approaches to disease mapping.* In A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J. Viel, R. Bertollini (eds.), *Disease mapping and risk assessment for public health*, Wiley, New York, pp. 15–29.

J. Palacios (2001). *Closed-form formulas for kirchhoff index.* International Journal of Quantitative Chemistry, 81, pp. 135–140.

A. Thomas, B. O'Hara, U. Ligges, S. Sturz (2006). *Making bugs open.* R News, 6, pp. 12–17.

D. L. W. Breymann (2010). *R: A Language and Environment for Statistical Computing*. URL `cran.r-project.org/web/packages/ghyp/index.html`.

J. Wakefield (2007). *Disease mapping and spatial regression with count data*. Biostatistics, 8, pp. 158–183.

J. Wakefield, S. Morris (2001). *The bayesian modeling of disease risk in relation to a point source*. Journal of the American Statistical Association, 96, pp. 77–91.

Summary

In this paper, we consider the problem of specifying priors for the variance components in the Bayesian analysis of the Besag-York-Mollié model, a model that is popular among epidemiologists for disease mapping. The model encompasses two sets of random effects: one spatially structured to model spatial autocorrelation and the other spatially unstructured to describe residual heterogeneity. In this model, prior specification for variance components is an important problem because these priors maintain their influence on the posterior distributions of relative risks when mapping rare diseases. We propose using generalised inverse Gaussian priors, a broad class of distributions that includes many distributions commonly used as priors in this context, such as inverse gammas. We discuss the prior parameter choice with the aim of balancing the prior weight of the two sets of random effects on total variation and controlling the amount of shrinkage. The suggested prior specification strategy is compared to popular alternatives using a simulation exercise and applications to real data sets.

Keywords: Hierarchical models; Spatial epidemiology; Generalised inverse Gaussian distribution; Intrinsic conditional autoregressive models.