2016 SISMID Module 16
Lecture 4: Disease Mapping

Jon Wakefield and Lance Waller

Departments of Statistics and Biostatistics
University of Washington

Outline

Disease mapping

Disease mapping has a long history in epidemiology, and may be defined as the estimation and presentation of summary measures of health outcomes.

The aims of disease mapping include

- simple description,
- hypothesis generation,
- allocation of health care resources, assessment of inequalities, and
- estimation of background variability in underlying risk in order to place epidemiological studies in context.

In this section we consider models for count data aggregated over areas at a single time point,in the situation in which we have (hopefully!) a complete enumeration of the cases (so not small area estimation).

Background reading, Wakefield *et al.* (2000).

The models we describe in this section, can also be used for regression modeling.

Motivation and Context

We begin by noting a number of non-statistical issues, for more background see Chapters 12 and 13 of Elliott *et al.* (2000):

- In broad-scale studies (in particular international endeavors), data comparability is a major issue.
- Precise disease definition (via ICD codes) is also extremely important.
- Mortality data tend to be more reliable than incidence data, but the latter are in general of greater epidemiological interest, because incident cases are closer in time to exposure.

Motivation and Context

There is a trade-off when a geographical scale is chosen:

- Larger geographical areas providing more stable rates and less problems of migration, but relative risk summaries may be distorted due to the large aggregation of individuals.
- If the relative risk shows marked variation within a particular area this information will be lost – if a particular subregion has a high relative risk then this will be diluted under aggregation.
- Larger study regions are likely to offer greater contrasts (range of covariates x) in relative risks and exposures.
- Localized effects can only be detected with data at a smaller level of aggregation.

Presentation

Chloropleth (areas shaded) are the most popular kind of maps, but isopleth (contours) and cartograms (size of areas proportional to denominator) have also been used.

Choice of color is important – multiple colors can be confusing, shading with a single color can work well.

Cut-points should be chosen to be epidemiologically meaningful and convey as much information as possible.

Use common cutpoints if multiple maps are to be compared/examined.

Instability of the Naive Relative Risk Estimate

We begin by concentrating on count (aggregated) data, since these are most common for mapping.

Unfortunately there are well-documented difficulties with the mapping of raw estimates since, for small areas and rare diseases in particular, these estimates will be dominated by sampling variability.

For the model $Y_i|\theta_i \sim \text{Poisson}(E_i\theta_i)$ the MLE is

$$\widehat{\theta}_i = \text{SMR}_i = \frac{Y_i}{E_i},$$

with variance

$$\text{var}(\widehat{\theta}_i) = \frac{\theta_i}{E_i},$$

which is estimated by

$$\widehat{\text{var}}(\widehat{\theta}_i) = \frac{\widehat{\theta}_i}{E_i}$$

so that areas with small E_i have high associated variance.

Surveillance

We imagine separate monthly surveillance for each of three areas over a 10-year period.

Data simulated from the model

$$Y_t|\theta \sim_{ind} \text{Poisson}(E\theta),$$

$t = 1, ..., 120$ months, where the relative risk is $\theta = 1$ in each case.

Recall that the MLE of the SMR in each time period is $\widehat{\theta}_t = Y_t/E$ with variance proportional to $1/E$ so that areas with small expected numbers have high variability.

The expected numbers differ in the three plots in Figure 1, and the resultant instability in the SMR is apparent.

For the $E = 0.2$ case there are a number of time periods with high estimates (and estimates of zero also!).

Surveillance

In general, maps showing p-values of exceedence of 1 are even less informative than maps of SMRs since areas with large populations may provide statistically significant SMRs, even for small deviations from a relative risk of 1.

Surveillance

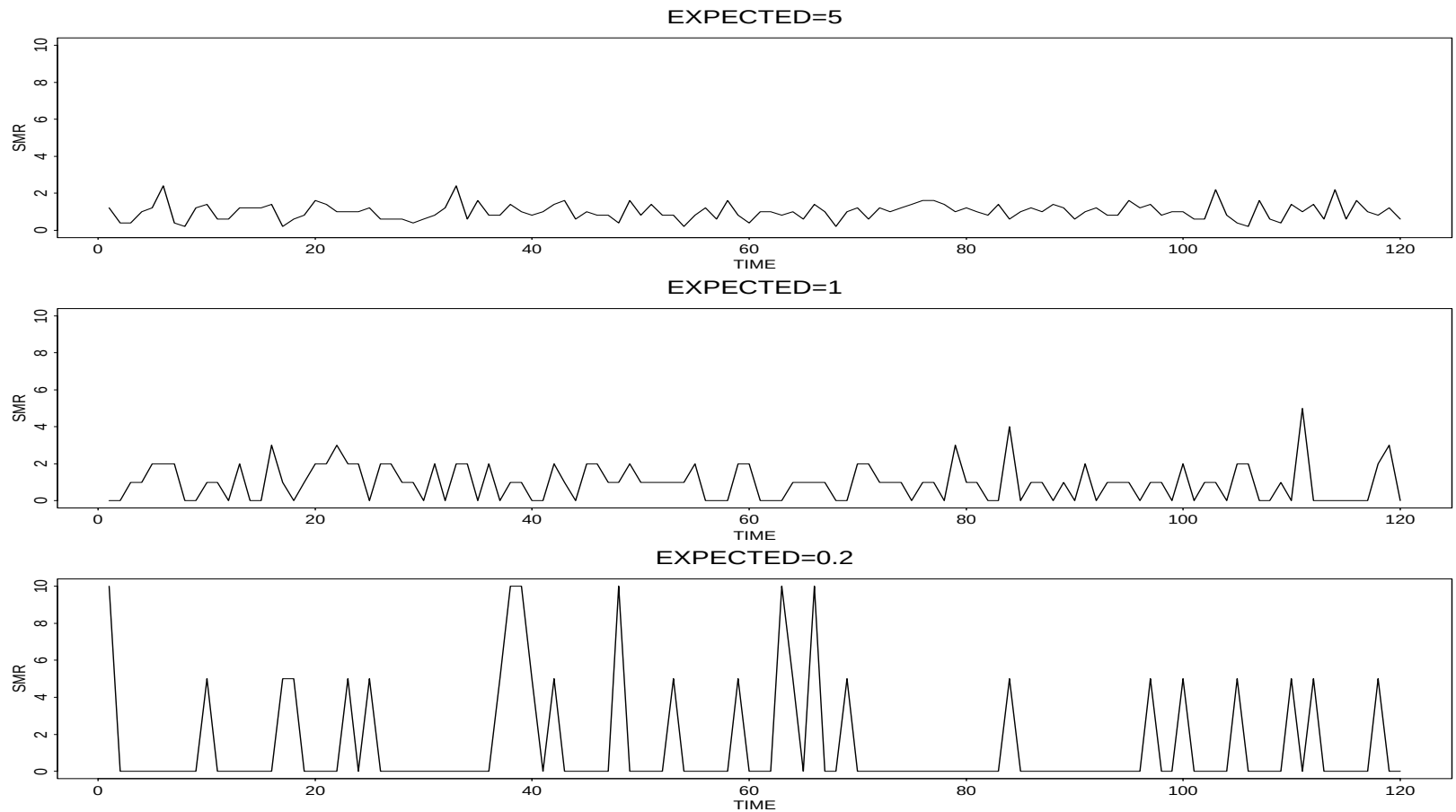


Figure 1 : Simulations from the Poisson distribution under different expected numbers.

Scottish Lip Cancer Data

Incidence rates of lip cancer in males in 56 counties of Scotland, registered in 1975–1980. These data were originally reported in the mapping atlas of Kemp *et al.* (1985).

The Scottish lip cancer data have been widely analyzed, because they have been around a long time, and the SIRs display a lot of spatial variability.

The form of the data is:

- Observed and expected number of cases (based on the county age populations, details shortly) – allows the calculation of the standardized morbidity ratio, the ratio of the observed to the expected cases (according to Clayton and Kaldor (1987) these are based on the MLEs for the age effects from a multiplicative model).
- A covariate measuring the proportion of the population engaged in agriculture, fishing, or forestry (AFF).
- The projections of the longitude and latitude of the area centroid, and the "position" of each county expressed as a list of adjacent counties.

Area i	Cases Y_i	Exp E_i	Prop AFF	SMR	Project N (km)	Project E (km)	Adjacent Counties
1	9	1.4	0.16	6.43	834.7	162.2	5,9,19
2	39	8.7	0.16	4.48	852.4	385.8	7,10
3	11	3.0	0.10	3.67	946.1	294.0	12
4	9	2.5	0.24	3.60	650.5	377.9	18,20,28
5	15	4.3	0.10	3.49	870.9	220.7	1,12,19
6	8	2.4	0.24	3.33	1015.2	340.2	Island
7	26	8.1	0.10	3.21	842.0	325.0	2,10,13,16,17
8	7	2.3	0.07	3.04	1168.9	442.2	Island
...							
48	3	9.3	0.01	0.32	654.7	282.0	24,44,47,49
49	28	88.7	0.00	0.32	666.7	267.8	38,41,44,47,48,52,53,54
50	6	19.6	0.01	0.31	736.5	342.2	21,29
51	1	3.4	0.01	0.29	678.9	274.9	34,38,42,54
52	1	3.6	0.00	0.28	683.7	257.8	34,40,49,54
53	1	5.7	0.01	0.18	646.6	265.6	41,46,47,49
54	1	7.0	0.01	0.14	682.3	267.9	34,38,49,51,52
55	0	4.2	0.16	0.00	640.1	321.5	18,24,30,33,45,56
56	0	1.8	0.10	0.00	589.9	322.2	18,20,24,27,55

Scottish Lip Cancer Data

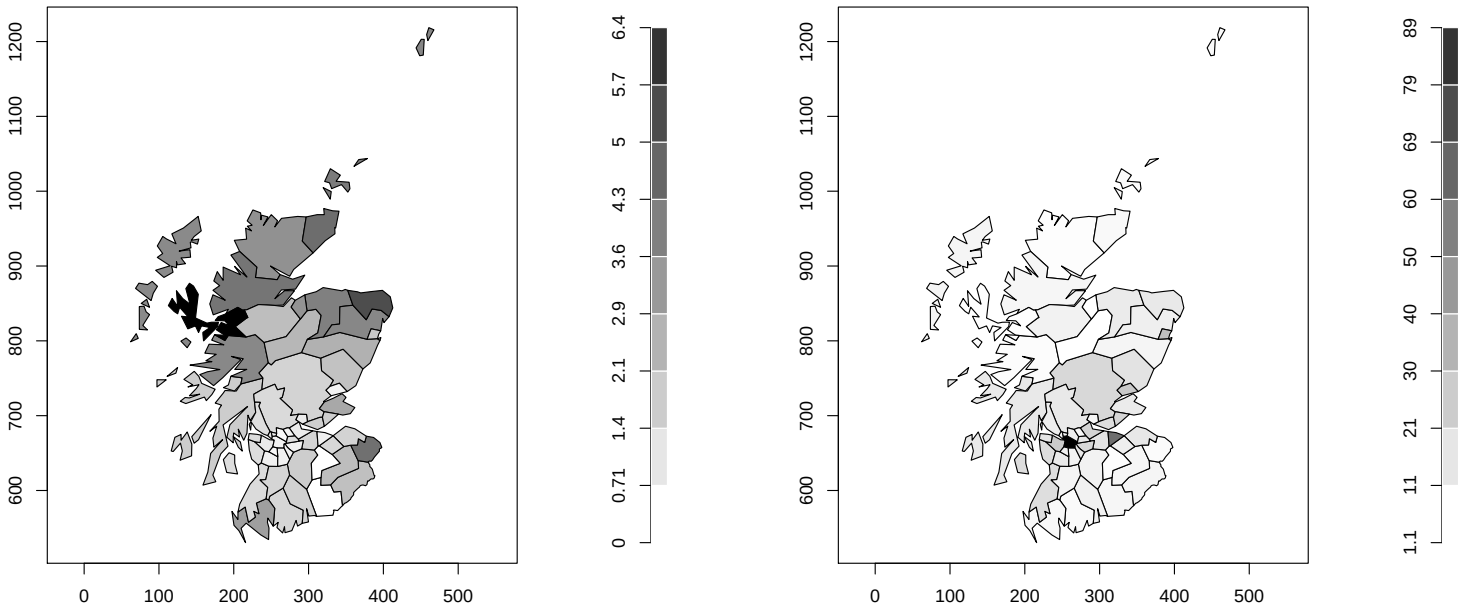
Figure 2 shows the SMRs for the Scottish lip cancer data, and indicates a large spread with an increasing trend in the south-north direction.

As just discussed, the variance of the estimate is $\text{var}(\text{SMR}_i) = \text{SMR}_i/E_i$, which will be large if E_i is small.

For the Scottish data the expected numbers are highly variable, with range 1.1–88.7.

This variability suggests that there is a good chance that the extreme SMRs are based on small expected numbers (many of the large, sparsely-populated rural areas in the north have high SMRs).

Figure 3 (left panel) shows the SMRs versus the estimated standard errors and clearly illustrates that the high SMRs have high associated standard error.



(a) (b)

Figure 2 : In 56 counties of Scotland: (a) SMRs and (b) expected numbers.

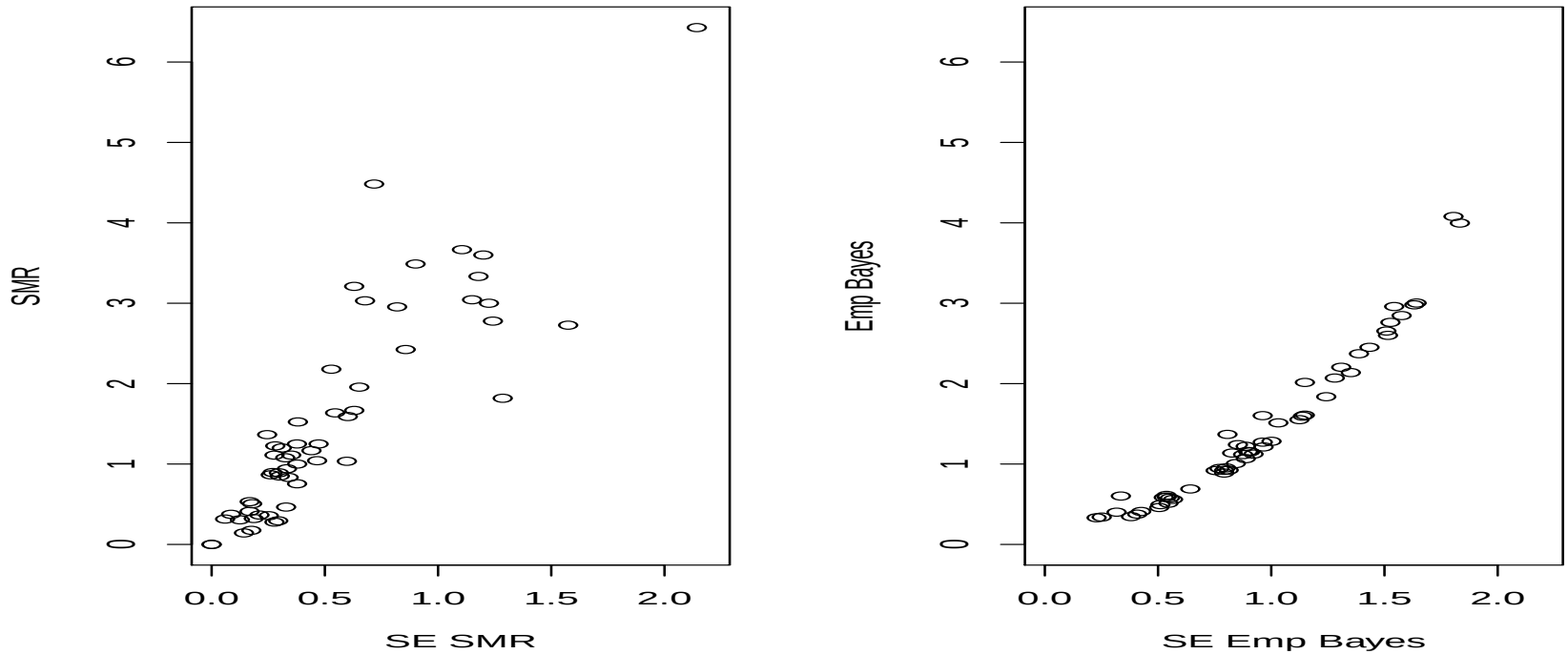


Figure 3 : Comparison of estimates and standard errors 56 counties of Scotland. On the left are the raw SMRs plotted against the standard errors on the right are estimates from a Bayesian smoothing model.

Smoothing Models

The above considerations of instability led to methods being developed to *smooth* the SMRs using hierarchical/random effects models that use the data from the totality of areas to provide more reliable estimates in each of the constituent areas.

Overview of Models:

- Basic Poisson Model: No smoothing.
- Random Effects Models:
  - Poisson-Gamma: Non-spatial smoothing.
  - Poisson-Lognormal: Non-spatial smoothing.
  - Poisson-Lognormal-Spatial: Spatial and non-spatial smoothing.
- Covariates may be added to each of these in order to smooth over covariate space.
- Estimation in these models is a separate issue.

Overview of Models

All of these models are of the form:

$$Y_i|\theta_i \sim \text{Poisson}\left(E_i\theta_i\right).$$

Having unconstrained θ_i and taking the MLE's leads to $\widehat{\theta}_i = Y_i/E_i$.[1]

Now suppose

$$\theta_i = \exp(\beta_0 + x_i\beta_1)\delta_i\eta_i,$$

with δ_i and η_i random effects.

Poisson-Gamma: δ_i independent gamma with $\eta_i = 1$.

Poisson-Lognormal: δ_i independent lognormal with $\eta_i = 1$.

Poisson-Lognormal-Spatial: δ_i independent lognormal with η_i dependent multivariate lognormal.

[1] Equivalent to taking the posterior mean in a Bayesian analysis with independent (improper) uniform prior on $\log\theta_i$ or, equivalently, $\pi(\boldsymbol{\theta}) \propto \prod_i \theta_i^{-1}$.

Poisson-Gamma Model Without Covariates

We begin by describing a simple Poisson-Gamma two-stage model that offers analytic tractability and ease of estimation.

- A very simple model (when there are no covariates) is

$$Y_i|\beta_0 \sim_{ind} \text{Poisson}\left(e^{\beta_0} E_i\right),$$

  so that the relative risk is constant across all areas, and equal to e^{β_0}. The latter is the overall relative risk, and reflects differences between the reference rates and the rates in the study region.
- We would like a model between the above form with one parameter and the n distinct, unrelated relative risks model, which leads to the SMRs, Y_i/E_i.
- We assume there are no covariates and assume the first stage likelihood is given by

$$Y_i|\delta_i, \beta_0 \sim_{ind} \text{Poisson}\left(e^{\beta_0} E_i \delta_i\right), \tag{1}$$

  where e^{β_0} is again the overall relative risk.

Poisson-Gamma Model Without Covariates

At the second stage the random effects δ_i are assigned a distribution.

The δ_i are the deviations of the relative risk in area i from the level across the whole region (e^{β_0}).

We initially assume that across the map the deviations of the relative risks from the mean, e^{β_0}, are modeled by

$$\delta_i|\alpha \sim_{iid} \text{Ga}(\alpha, \alpha), \tag{2}$$

a gamma distribution with mean 1, and variance $1/\alpha$.

If α is small we have a wide gamma distribution, and we would expect little shrinkage of an area's estimate to the overall level, but if α is large we have a narrow distribution and large shrinkage is anticipated.

δ_i are the residual relative risks, the relative risks are

$$\text{RR}_i = e^{\beta_0} \times \delta_i.$$

Poisson-Gamma Model Without Covariates

The rationale here is that we expect some similarity of residual relative risks δ_i across the map.

How do we decide upon a value for α, which determines the spread of the δ_i?

- We might hope that the totality of data might aid in estimating the δ_i in each area.
- One possibility would be to simply fix α, based on the context/historical data.
- However, estimating *a* from the data will often lead to an appropriate measure of the spread of the distribution.
- Estimation may be carried out using empirical Bayes or full Bayes methods.

Before we discuss estimation of α we see how we would proceed, if it were known.

Poisson-Gamma Model Without Covariates

The model is

$$\begin{aligned} Y_i|\delta_i, \beta_0 &\sim_{ind} \text{Poisson}\left(e^{\beta_0} E_i \delta_i\right) \\ \delta_i|\alpha &\sim_{iid} \text{Ga}(\alpha, \alpha) \end{aligned}$$

This leads to a gamma posterior for δ_i:

$$\delta_i|y_i, \alpha, \beta_0 \sim \text{Ga}(\alpha + y_i, \alpha + E_i e^{\beta_0}).$$

Hence, the posterior mean residual relative risk estimate is

$$\begin{aligned} \widehat{\delta}_i &= \frac{\alpha + y_i}{\alpha + E_i e^{\beta_0}} = \frac{\alpha}{\alpha} \frac{\alpha}{\alpha + E_i e^{\beta_0}} + \frac{y_i}{E_i} \frac{E_i}{\alpha + E_i e^{\beta_0}} \\ &= 1 \times \frac{\alpha}{\alpha + E_i e^{\beta_0}} + \text{SMR}_i \times \frac{E_i}{\alpha + E_i e^{\beta_0}}. \end{aligned}$$

Poisson-Gamma Model Without Covariates

The residual relative risk estimate is:

$$\widehat{\delta}_i \quad = \quad 1 \times \frac{\alpha}{\alpha + E_i e^{\beta_0}} + \text{SMR}_i \times \frac{E_i}{\alpha + E_i e^{\beta_0}}.$$

If α is large then the random effects have a tight spread, and there is more shrinkage towards the prior mean of 1, since SMRs that are far from 1 are inconsistent with the total collection of estimates.

However, an outlying estimate that is not based on a large expected number, will be shrunk, and we may miss an important excess.

Poisson-Gamma Model Without Covariates

We have obtained the form of the smoothed estimate for the residual relative risk.

The posterior mean estimate of the relative risk is

$$\begin{aligned}
\widehat{\text{RR}}_i &= e^{\beta_0} \times \widehat{\delta}_i = e^{\beta_0} \times \frac{\alpha + y_i}{\alpha + E_i e^{\beta_0}} \\
&= e^{\beta_0} \frac{\alpha}{\alpha + E_i e^{\beta_0}} + \frac{y_i}{E_i} \frac{E_i e^{\beta_0}}{\alpha + E_i e^{\beta_0}} \\
&= e^{\beta_0} \times (1 - W_i) + \text{SMR}_i \times W_i
\end{aligned}$$

where

$$W_i = \frac{E_i e^{\beta_0}}{\alpha + E_i e^{\beta_0}}$$

is the weight on the SMR in area i.

The weight on the observed SMR increases as E_i increases so for areas with large populations the estimate is dominated by the data.

Poisson-Gamma Model Without Covariates

Since

$$\delta_i | y_i, \alpha, \beta_0 \sim \text{Ga}(\alpha + y_i, \alpha + E_i e^{\beta_0})$$

then

$$\text{RR}_i | y_i, \alpha, \beta_0 \sim \text{Ga}[\alpha + y_i, (\alpha + E_i e^{\beta_0})/e^{\beta_0}].$$

Let RR_i^0 is the true relative risk of area i; this is not Y_i/E_i but the hypothetical parameter that gave rise to this observed relative risk.

Note that the SMR is unbiased:

$$\text{E}[\text{SMR}_i] = \text{E}\left[\frac{Y_i}{E_i}\right] = \frac{\text{E}[Y_i]}{E_i} = \frac{E_i \theta_i}{E_i} = \theta_i.$$

An important aspect of this shrinkage model is that it introduces finite-sample bias, e.g. for the posteror mean:

$$\text{E}[\text{RR}_i | y_i, \alpha, \beta_0] = \text{E}[\widehat{\text{RR}}_i] = e^{\beta_0} \times (1 - W_i) + \text{E}[\text{SMR}_i] \times W_i \neq \text{RR}_i^0.$$

Notice that

$$\text{E}[\text{RR}_i | y_i, \alpha, \beta_0] \to \text{SMR}_i,$$

as $E_i \to \infty$, as we would hope.

Poisson-Gamma Model Without Covariates

The estimated variance of the SMR in area i is estimated as

$$\frac{y_i}{E_i^2},$$

so the variance of the SMR can grow without bound as E_i decreases.

For the smoothed estimate the variance is obtained from the gamma posterior – recall the variance of a $\text{Ga}(a, b)$ is a/b^2.

We have

$$\text{var}(\text{e}^{\beta_0}\delta_i|y_i, \alpha, \beta_0) = \frac{(\alpha + y_i)\text{e}^{2\beta_0}}{(\alpha + E_i\text{e}^{\beta_0})^2}$$

showing that the posterior variances are bounded above.

We now turn to the question of how we estimate α and β_0.

Empirical Bayes Estimation in the Poisson-Gamma Model Without Covariates

In an empirical Bayes approach the random effects δ_i are eliminated from the model to give a negative binomial likelihood that depends on β_0 and α only:

$$\begin{aligned}\Pr(Y_i|\beta_0,\alpha) &= \int \Pr(Y_i|\beta_0,\delta_i) \times p(\delta_i|\alpha) d\delta_i \\ &= \frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)} \left(\frac{E_i e^{\beta_0}}{E_i e^{\beta_0}+\alpha}\right)^{y_i} \left(\frac{\alpha}{E_i e^{\beta_0}+\alpha}\right)^{\alpha}.\end{aligned}$$

The likelihood is

$$L(\beta_0,\alpha) = \prod_{i=1}^{n} \Pr(Y_i|\beta_0,\alpha),$$

which is maximized as a function of β_0 and α – `R` can do this for us using the `glm.nb()` function in the `MASS` library.

We then proceed as if α and β_0 are known, i.e. the estimates are $\mathrm{E}[\delta_i|y_i,\widehat{\alpha},\widehat{\beta_0}]$.

Full Bayes Estimation in the Poisson-Gamma Model Without Covariates

The full Bayes approach assigns a (hyper) prior to the (hyper) parameters α, β_0 to give the three stage hierarchical model:

Stage 1: $Y_i|\delta_i, \beta_0 \sim_{ind} \text{Poisson}(e^{\beta_0} E_i \delta_i)$, $i = 1, \ldots, n$.

Stage 2: $\delta_i|\alpha \sim_{iid} \text{Ga}(\alpha, \alpha)$, $i = 1, \ldots, n$.

Stage 3: Priors for α, β_0.

The posterior is

$$p(\delta_1, ..., \delta_n, \alpha, \beta_0|\mathbf{y}) \propto \left[\prod_{i=1}^{n} p(y_i|\delta_i, \beta_0) p(\delta_i|\alpha)\right] p(\alpha, \beta_0).$$

This model is not analytically tractable and we do not discuss further (including the issue of prior choice), since the Poisson-Lognormal model we describe shortly is more flexible.

Full Bayes Estimation in the Poisson-Gamma Model Without Covariates

What do we gain by full Bayes? Uncertainty in α, β_0 can be acknowledged.

The posterior distribution is analytically intractable but can be implemented using

- Markov chain Monte Carlo (MCMC). `WinBUGS` and more specifically the `GeoBUGS` module is a convenient way to do this. Other (generic) MCMC environments include `JAGS` (very similar to `WinBUGS`) and `Stan`.
- This is the method that has been used since the early 1990s (Besag *et al.*, 1991).
- More recently (Rue *et al.*, 2009) the integrated nested Laplace approximation (INLA) has been developed — can't be used for this model, but for the lognormal models we will see later.

Note: the Poisson-Gamma model is useful to introduce the smoothing concept and for non-spatially dependent random effects, but cannot be extended easily.

Assumptions in the Poisson-Gamma Model Without Covariates

We have a Poisson model for the data with random effects assumed to follow a gamma distribution.

We therefore need to check whether this gamma distribution appears reasonable.

One way to assess this is to form a so-called QQ plot in which

- The observed ordered residual relative risks are plotted against the expected ordered residual relative risks.
- The latter are given (approximately) by $f_i = F^{-1}_{\text{GAMMA}}\left(\frac{i-0.5}{n}|\alpha\right)$ where $F^{-1}_{\text{GAMMA}}(\cdot|\alpha)$ is the inverse cumulative distribution function of a $\text{Ga}(\alpha, \alpha)$ distribution.

Poisson-Gamma Model with Covariates

With area-level covariates we have the model

$$Y_i|\delta_i, \beta_0, \beta_1 \sim_{ind} \text{Poisson}\left(e^{\beta_0+\beta_1 x_i} E_i \delta_i\right),$$

where we have assumed a loglinear regression model for area-level covariates x_i.

The random effect δ_i now describes the deviation of area i's relative risk from the loglinear mean model.

Letting $\mu_i = \mathrm{E}[Y_i|\delta_i, \beta_0, \beta_1]$ notice that on the log scale we have

$$\log \mu_i = [\beta_0 + \beta_1 x_i + \log(E_i)] + \underbrace{\log(\delta_i)}_{\text{Residual}},$$

so that the log random effects, $\log \delta_i$, may be viewed as residuals.

Poisson-Gamma Model with Covariates

At the second stage the random effects δ_i are assigned a distribution.

We assume that across the map the deviations of the relative risks from the mean, $e^{\beta_0+\beta_1 x_i}$, are again modelled by

$$\delta_i|\alpha \sim_{iid} \mathsf{Ga}(\alpha, \alpha),$$

a gamma distribution with mean 1, and variance $1/\alpha$.

Everything follows through as before.

Poisson-Lognormal Models

The gamma model is computationally convenient but cannot easily be extended to allow for residual spatial dependence.

Hence, we turn our attention to Poisson-Lognormal models.

A Poisson-lognormal non-spatial random effect model is given by:

$$Y_i|\beta_0, \beta_1, \epsilon_i \sim_{ind} \text{Poisson}(E_i e^{\beta_0 + x_i\beta_1} e^{\epsilon_i}) \quad (3)$$

$$\epsilon_i|\sigma^2_\epsilon \sim_{iid} N(0, \sigma^2_\epsilon) \quad (4)$$

where ϵ_i are area-specific random effects that capture the residual or unexplained (log) relative risk of disease in area i, $i = 1, \ldots, n$.

Whereas in the Poisson-Gamma model we have $\delta_i \sim \text{Ga}(\alpha, \alpha)$, here we have

$$\delta_i = e^{\epsilon_i} \sim \text{LogNormal}(0, \sigma^2_\epsilon).$$

We still have a single parameter controlling the spread of the random effects, now σ^2_ϵ, rather than α.

Poisson-Lognormal Models

The model given by (3) and (4) does not give a marginal distribution for the data of known form (in contrast to the gamma model), but does naturally lead to the addition of spatial random effects.

Empirical Bayes is not so convenient for this model, though see the function `lognormalEB()` within the DCluster() package.

Hence, we resort to a fully Bayesian approach for which we need to specify prior distributions.

Prior Choice for Poisson-Lognormal Models

We need to specify priors for:

- The intercept β_0 and regression coefficient $\beta_!$.
- The variance of the normal random effects σ^2_ϵ.

An improper prior[2]

$$p(\beta_0, \beta_1) \propto 1$$

may often be used, but in some circumstances such a choice may lead to an improper posterior.

If there are a large numbers of covariates, or high dependence amongst multiple covariates then more informative priors will be beneficial.

If an informative prior is required, then a multivariate normal distribution is the natural choice.

This is equivalent to a multivariate lognormal distribution for the relative risks.

[2] This means that it doesn't integrate to 1

Prior Choice for Poisson-Lognormal Models

It is convenient to specify lognormal priors for a positive parameter $\exp(\beta)$, since one may specify two quantiles of the distribution, and directly solve for the two parameters of the lognormal.

Denote by LogNormal(μ, σ) the lognormal distribution for a generic parameter θ with

$$\mathsf{E}[\log(\theta)] = \mu, \qquad \mathsf{var}(\log(\theta)) = \sigma^2,$$

and let θ_1 and θ_2 be the q_1 and q_2 quantiles of this prior.

In our example, $\theta = \exp(\beta)$.

Then it is straightforward to show that

$$\begin{aligned} \mu &= \log(\theta_1)\left(\frac{z_{q_2}}{z_{q_2} - z_{q_1}}\right) - \log(\theta_2)\left(\frac{z_{q_1}}{z_{q_2} - z_{q_1}}\right), \\ \sigma &= \frac{\log(\theta_1) - \log(\theta_2)}{z_{q_1} - z_{q_2}}. \end{aligned}$$

Prior Choice for Poisson-Lognormal Models

As an example, suppose that for the ecological relative risk

$$\theta = \exp(\beta)$$

we believe there is a 50% chance that the relative risk is less than 1 and a 95% chance that it is less than 5.

This gives

$$q_1 = 0.5, \qquad \theta_1 = 1.0, \qquad q_2 = 0.95, \qquad \theta_2 = 5.0,$$

we obtain lognormal parameters

$$\mu = 0, \qquad \sigma = \frac{\log 5}{1.645} = 0.98.$$

The density is shown in Figure 4.

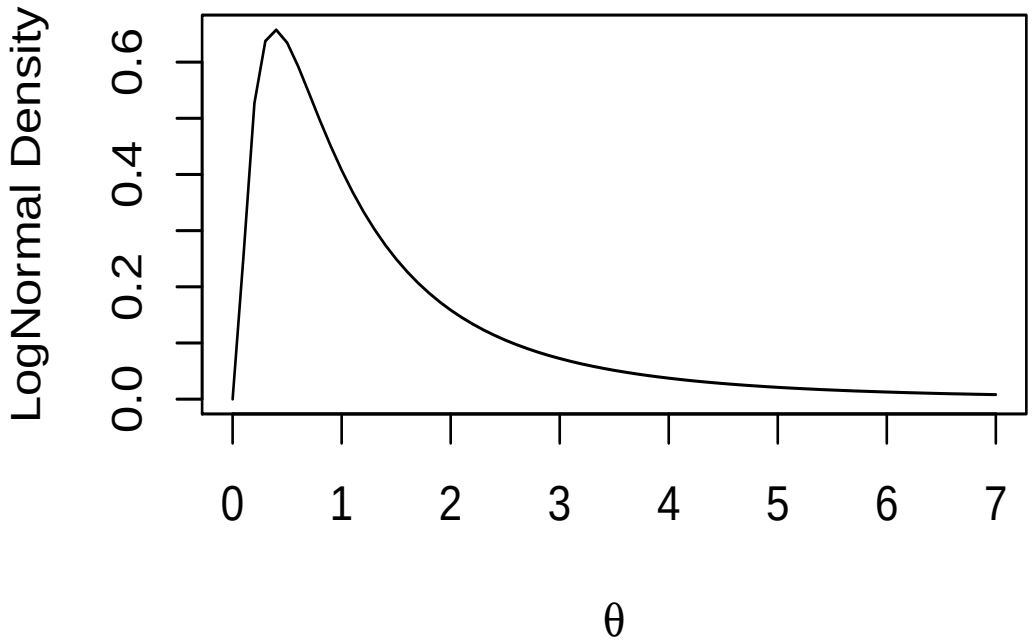


Figure 4 : Lognormal density with 50% point 1 and 95% point 5.

Prior Choice for Poisson-Lognormal Models

The priors $\tau_\epsilon = \sigma_\epsilon^{-2} \sim \text{Ga}(1, 0.0260)$ or $\tau_\epsilon = \sigma_\epsilon^{-2} \sim \text{Ga}(0.5, 0.0005)$ will often be suitable in a mapping context.

τ_ϵ is the precision, i.e., the reciprocal variance.

For the Ga(1,0.026) prior the 2.5%, 50% (median) and 97.5% quantiles for σ_ϵ are:

$$(0.014, \quad 0.047, \quad 1.01).$$

For the Ga(0.5,0.0005) prior the 2.5%, 50% (median) and 97.5% quantiles for σ_ϵ are:

$$(0.084, \quad 0.194, \quad 1.01).$$

So the Ga(1,0.026) prior favors smaller values, i.e. more shrinkage is anticipated.

Prior Choice for Poisson-Lognormal Models

Interpretation is helped by approximation of the residual relative risk

$$\exp(\epsilon) \approx 1 + \epsilon$$

for small ϵ and so

$$\text{s.d}(e^{\epsilon}) = \sigma_{\epsilon}$$

is approximately the standard deviation of the residual relative risks.

Sensitivity of the results to the specification should be carried out, particularly if the number of areas is not large.

Poisson-Lognormal Spatial Model

In general we might expect residual relative risks in areas that are "close" to be more similar than in areas that are not "close".

In other words we would like to smooth in space.

We would like to exploit this information in order to provide more reliable relative risk estimates in each area.

This is analogous to the use of a covariate x, in that areas with similar x values are likely to have similar relative risks.

Unfortunately the modelling of spatial dependence is much more difficult since spatial location is acting as a surrogate for unobserved covariates.

We need to choose an appropriate spatial model, but do not directly observe the covariates whose effect we are trying to mimic.

Poisson-Lognormal Spatial Model

We first consider the model

$$Y_i|\theta_i \sim_{ind} \text{Poisson}(E_i\theta_i) \tag{5}$$

with

$$\log\theta_i = \beta_0 + x_i\beta_1 + s_{i1}\gamma_1 + s_{i2}\gamma_2 + S_i + \epsilon_i, \tag{6}$$

where

- $\mathbf{s}_i = (s_{i1}, s_{i2})$ denotes spatial location, the centroid of area i,
- the large-scale spatial trend is captured by

$$s_{i1}\gamma_1 + s_{i2}\gamma_2. \tag{7}$$

- the random effects $\epsilon_i|\sigma^2_\epsilon \sim_{iid} \text{N}(0, \sigma^2_\epsilon)$ represent non-spatial overdispersion,
- S_i are random effects with local spatial structure.
- We describe two possible forms for the spatial random effects.

Spatial Models Overview

In general, there have been two approaches to modeling spatial dependence:

- Local conditional modeling.
- Multivariate (geostatistical) modeling.

The local approach, an early reference to which is Besag (1974), is based on conditional specifications $S_i|S_{-i}$, where

$$S_{-i} = (S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_n).$$

In general, the only variables in S_{-i} that are relevant are the neighbors (suitably defined, which we write as $S_i|S_j, j \in \text{ne}(i)$.

In words, what is the distribution of S_i, given we know the values taken by the neighboring random variables.

The multivariate approach, see for example Stein (1999), is based on the specification of the full multivariate distribution of $\mathbf{S} = (S_1, \ldots, S_n)$.

A Multivariate Spatial Model

Assume that

$$\mathbf{S} = (S_1, \ldots, S_n)$$

arise from a zero mean multivariate normal distribution with variances

$$\text{var}(S_i) = \sigma_s^2$$

and correlations $\text{corr}(S_i, S_j)$.

The obvious approach in a spatial setting is to assume a form such that the correlation between S_i and S_j decreases as d_{ij}, the distance between the locations at which S_i and S_j are measured, decreases.

A model in which the correlations are a function of distance only between the points is known as isotropic.

Care must be taken with specification of the function relating correlations to distance, as the resulting variance-covariance matrix must be invertible.

A Multivariate Spatial Model

A simple form is

$$\text{corr}(S_i, S_j) = \rho^{d_{ij}}$$

where

- $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ is the distance between the centroids of areas i and j, and
- $\rho > 0$ is a parameter that determines the extent of the correlation; ρ is the correlation between the residual spatial variability in two locations that are one unit of distance apart.

The correlation above is the marginal correlation between the random variables S_i and S_j.

This multivariate spatial model has two parameters, σ_s^2, which determines the scale of the spatial variability, and ρ, which determines the extent of the spatial variability.

A Multivariate Spatial Model

More generally the correlations can be modeled as a Matérn correlation function (Stein, 1999):

$$\text{corr}(S_i, S_j) = \frac{1}{\Gamma(\upsilon + 1/2)(4\pi)^{1/2}\kappa^{2\upsilon}2^{\upsilon-1}}(\kappa d_{ij})^{\upsilon} K_{\upsilon}(\kappa d_{ij})$$

where $K_{\upsilon}(\cdot)$ is a modified Bessel function of the second kind, $\kappa > 0$ is a scale parameter and $\upsilon > 0$ is a smoothness parameter.

In general, difficult to estimate many parameters in a spatial model and often υ is fixed.

A Multivariate Spatial Model

The multivariate model with correlations of this form is computationally expensive to fit, because one has to carry out operations on the $n \times n$ covariance matrix, which we call $\mathbf{\Sigma}$.

The multivariate normal distribution $\mathbf{s}|\mathbf{\Sigma} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$ is given by

$$p(\mathbf{s}) = (2\pi|\mathbf{\Sigma}|)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{s}^{\mathrm{T}}\Sigma^{-1}\mathbf{s}\right),$$

so to evaluate the density we need to calculate a determinant and an inverse.

We consider the multivariate normal model no more in a mapping context, but return to it when we consider exposure surface modeling and Kriging.

A Conditional Spatial Model

An alternative approach is to specify the distribution of each S_i as if we knew the values of the spatial random effects S_j in neighboring areas.

Hence, we have a conditional specification since we are conditioning on knowing the neighbors.

A Conditional Spatial Model

We need to specify a rule for determining the neighbors of each area.

In an epidemiological context the areas are not regular in shape.

This is in contrast to image processing applications in which the data are collected on a regular grid.

Hence, there is an arbitrariness in specification of the neighborhood structure.

A Conditional Spatial Model

To define neighbors, a number of authors have taken the neighborhood scheme to be such that two areas are taken to be neighbors if they share a common boundary.

This is reasonable if all regions are (at least roughly) of similar size and arranged in a regular pattern (as is the case for pixels in image analysis where these models originated), but is not particularly attractive otherwise (but reasonable practical alternatives are not available).

Various other neighborhood/weighting schemes are possible:

- One can take the neighborhood structure to depend on the distance between area centroids and determine the extent of the spatial correlation (i.e. the distance within which regions are considered neighbors).
- One could also define neighbors in terms of cultural similarity.

In typical applications it is difficult to assess whether the spatial model chosen is appropriate, which argues for a simple form, and to assess the sensitivity of conclusions to different choices.

A Conditional Spatial Model

A common model, due to Besag *et al.* (1991), is to assign the spatial random effects an intrinsic conditional autorgressive (ICAR) prior.

Under this specification it is assumed that the spatial random effect is drawn from a normal distribution whose mean is the mean of the neighbors random effects, with variance proportional to one over the number of neighbors (so more neighbors, less variability).

In math-speak:

$$S_i | S_j, j \in \text{ne}(i) \sim \text{N}\left(\overline{S}_i, \frac{\sigma_s^2}{m_i}\right),$$

where $\text{ne}(i)$ is the set of neighbors of area i, m_i is the number of neighbours, and

$$\overline{S}_i = \frac{1}{m_i} \sum_{j \in \text{ne}(i)} S_j$$

is the mean of the spatial random effects of these neighbors.

A Conditional Spatial Model

The parameter σ_s^2 is a conditional variance and its magnitude determines the amount of spatial variation.

The variance parameters σ_ϵ^2 and σ_s^2 have different interpretations.

Both are defined with respect to the log relative risk scale, but σ_ϵ has a marginal interpretation while σ_s has a *conditional* interpretation.

Specifically, for area i, the variance of S_i is conditional on $S_j, j \in \text{ne}(i)$.

Hence the variances are not directly comparable (in contrast to the joint model in which σ_s is on the same scale as σ_ϵ).

Notice that if σ_s^2 is "small" then although the residual is strongly dependent on the neighboring value the overall contribution to the residual relative risk is small.

It is not necessary to include the $S_{i1}\gamma_1 + S_{i2}\gamma_2$ term in the model with the ICAR formulation, since local linear trends are accommodated.

A Conditional Spatial Model

This is a little counterintuitive but stems from spatial models having two aspects, the strength of dependence and the magnitude of spatial dependence, and in the ICAR model there is only a single parameter which controls both aspects.

In the joint model (with covariance $\sigma_s^2 \rho^{d_{ij}}$ for example) the strength is determined by ρ and the total amount by σ_s^2.

A non-spatial random effect should always be included along with the ICAR random effect since this model cannot take a limiting form that allows non-spatial variability.

In the joint model with S_i only, this is achieved as $\rho \to 0$.

If the majority of the variability is non-spatial, inference for this model might incorrectly suggest that spatial dependence was present.

Prior specification is difficult for the conditional variance is difficult because it has a conditional rather than a marginal interpretation.

Computation for the Conditional Model

Let $\mathbf{Q}/\sigma_s^2$ denote the precision matrix of the ICAR model.

For simplicity, suppose all areas are connected to at least one other area.

The elements $Q_{ij} = 0$ if S_i and S_j are conditionally independent, i.e., not neighbors.

The elements $Q_{ij} = -1$ if S_i and S_j are conditionally dependent, i.e., neighbors.

The elements $Q_{ii} = m_i$, where m_i is the number of neighbors of area i.

Hence, most of the elements of $\mathbf{Q}$ are zero (so the matrix is sparse) and this aids greatly in computation, see Rue and Held (2005) for details.

Computation for the Conditional Model

The form of the joint 'density' is

$$\begin{aligned} p(\mathbf{s}|\mathbf{Q},\sigma_s^2) &= (2\pi)^{-1/2}|\mathbf{Q}|^{1/2}\sigma_s^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_s^2}\mathbf{s}^\top\mathbf{Q}\mathbf{s}\right) \\ &= (2\pi)^{-1/2}|\mathbf{Q}|^{1/2}\sigma_s^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_s^2}\sum_{i\sim j}(s_i-s_j)^2\right) \end{aligned}$$

where $i \sim j$ means i and j are neighbors.

This is not a true density since it is not proper; $\mathbf{Q}$ is singular and has rank $n-1$.

The ICAR model is an example of a Gaussian Markov Random Field.

Note the contrast with the multivariate model in which $\Sigma_{ij} = 0$ if the marginal covariance between S_i and S_j is zero.

Discussion

Often SMRs are unstable because of small denominators.

More reliable estimates can be obtained by using the totality of data to inform on the distribution, both locally and globally, of the relative risks across the study region.

The gamma model is mathematically convenient, but is not well suited to modeling spatial dependence.

A lognormal model can be extended to allow spatial dependence relatively easily, with the ICAR model being particularly popular.

Spatio-Temporal Disease Mapping

We now consider space-time modeling of disease counts.

Suppose now that we are in the situation where we have population counts N_{itj} in area i, time period t and stratum j, with associated disease counts Y_{itj}, $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, J$.

We first form expected numbers

$$E_{it} = \sum_{j=1}^{J} q_j N_{itj}.$$

Note that the reference rates are for stratum only and not time; evaluating over time also would lead to the loss of the temporal component.

We allow for the possibility of counts changing over time, however.

The SMRs are, as usual, $\text{SMR}_{it} = Y_{it}/E_{it}$.

Spatio-Temporal Disease Mapping

Recall that in a mapping context we wish to obtain best guesses at risk over space and time, along with assessment of trends.

An overall aim in spatio-temporal disease mapping is to apportion the variability in the data to:

- space,
- time,
- space-time (i.e., the interaction) and
- covariates.

Ohio lung cancer mortality data

We examine data on lung cancer deaths in 88 counties of Ohio over the years 1968–1988.

We adjust for gender and race and for age, the latter via 5 age bands: 0–44, 45–54, 55–64, 65–74, 75+.

These data have been analyzed by a number of authors including Waller *et al.* (1997), Xia and Carlin (1998), Knorr-Held and Besag (1998).

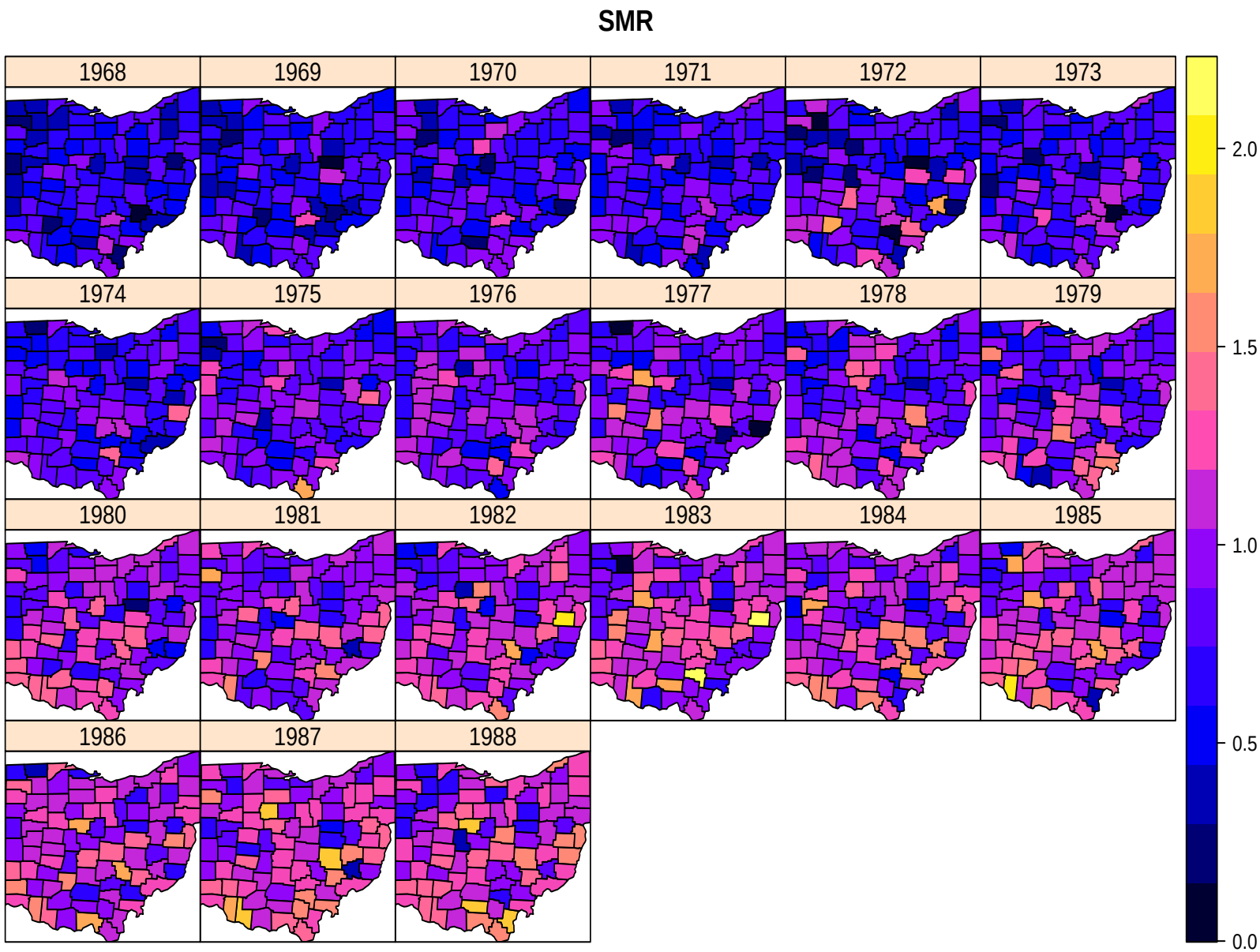


Figure 5 : SMRs over time for Ohio lung cancer mortality.

A Quick Tangent on Time Series Smoothing Models

Suppose we have counts Y_t of disease from denominators N (constant across time) and with equally-spaced time intervals $t = 1, \ldots, T$.

The model is $Y_t|\theta_t \sim \text{Poisson}(N\theta_t)$ with

$$\log \theta_t = \beta_0 + \omega_t + \tau_t$$

with

- independent terms

$$\omega_t \sim_{iid} \text{N}(0, \sigma_\omega^2)$$

  and

- first-order (random walk) smoothing

$$\tau_t|\tau_{t-1}, \tau_{t+1} \sim \text{N}\left(\frac{1}{2}(\tau_{t-1} + \tau_{t+1}), \frac{\sigma_\tau^2}{2}\right).$$

A Quick Tangent on Time Series Smoothing Models

A second-order (random walk) smoothing model would have

$$\tau_t | \tau_{t-1}, \tau_{t+1}, \tau_{t-2}, \tau_{t+2} \sim \mathsf{N}\left(\frac{4}{6}(\tau_{t-1} + \tau_{t+1}) - \frac{1}{6}(\tau_{t+2} - \tau_{t-2}), \frac{\sigma_\tau^2}{6}\right).$$

These two models are often abbreviated to RW1 and RW2 and the RW1 is the 1D analog of the ICAR model we have used for spatial smoothing.

Both RW1 and RW2 are local smoothing models.

Rue and Held (2005) is the definitive text on Gaussian Markov random field (GMRF) models.

We will now describe the combination of spatial and temporal models.

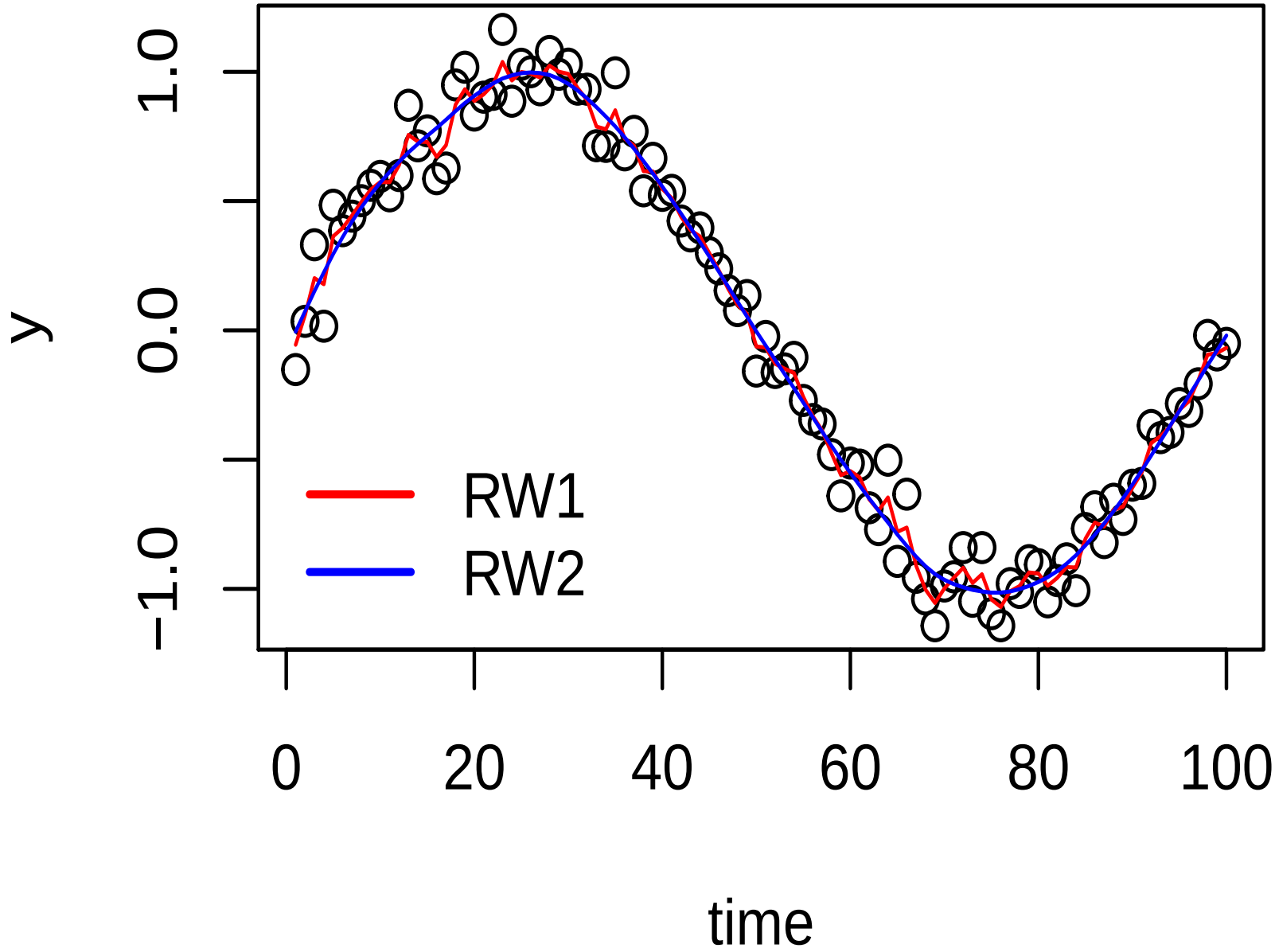


Figure 6 : RW1 and RW2 fits to simulated data; note the greater smoothness of the RW2 model.

Simple Space-Time Models

As a starting point we assume $Y_{it}|\theta_{it} \sim \text{Poisson}(E_{it}\theta_{it})$

We can first fit separate models

$$\log \theta_{it} = \beta_{0i} + \beta_{1i} t,$$

with quasi-likelihood used for inference.

We can then informally examine the variability in the area-specific relative risk slopes $\exp(\beta_{1i})$.

Mapping these slopes may give indication of space-time interaction.

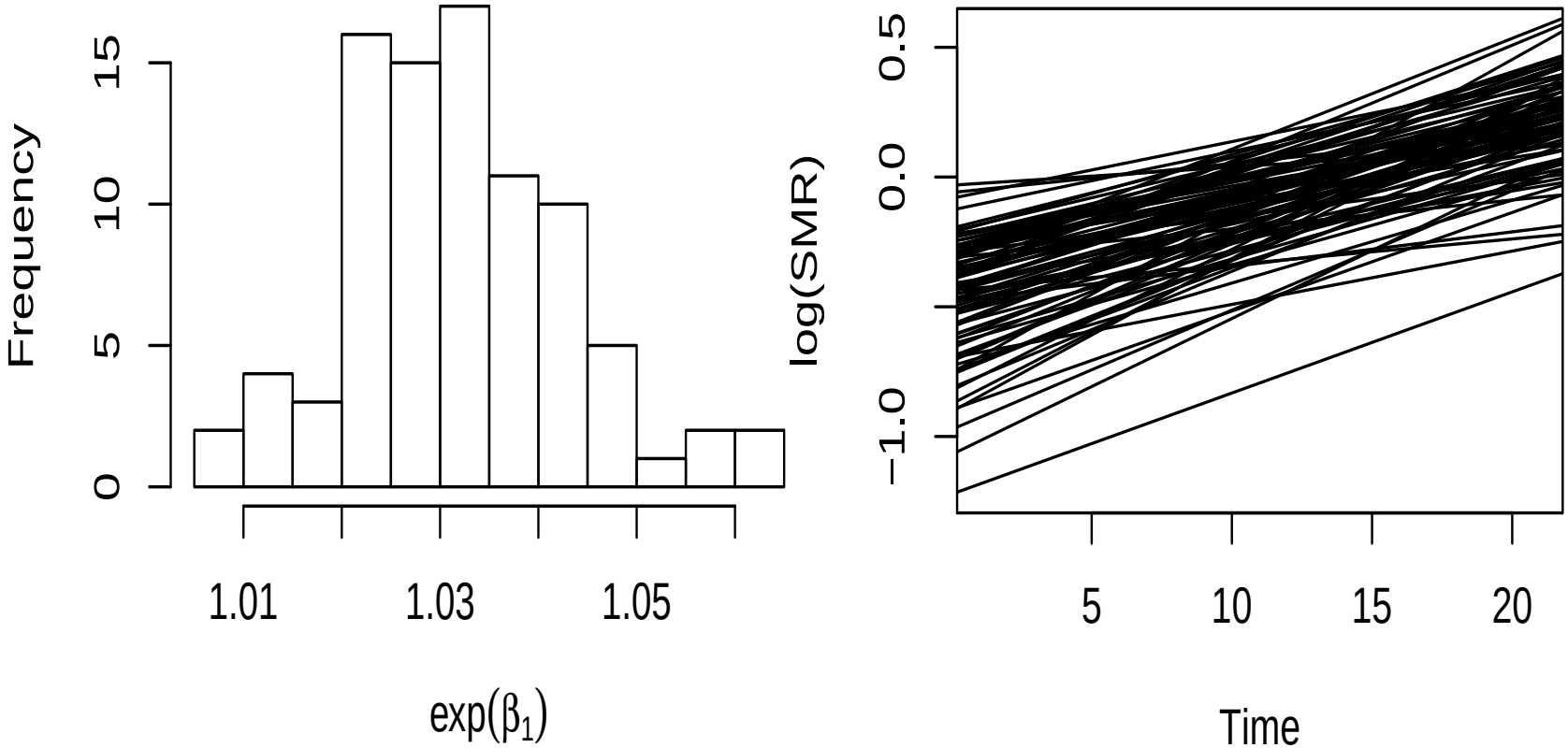


Figure 7 : Ohio lung cancer data: fitted slopes (on log relative risk scale) and histogram of slopes. Clearly the trend in relative risk is increasing across all areas.

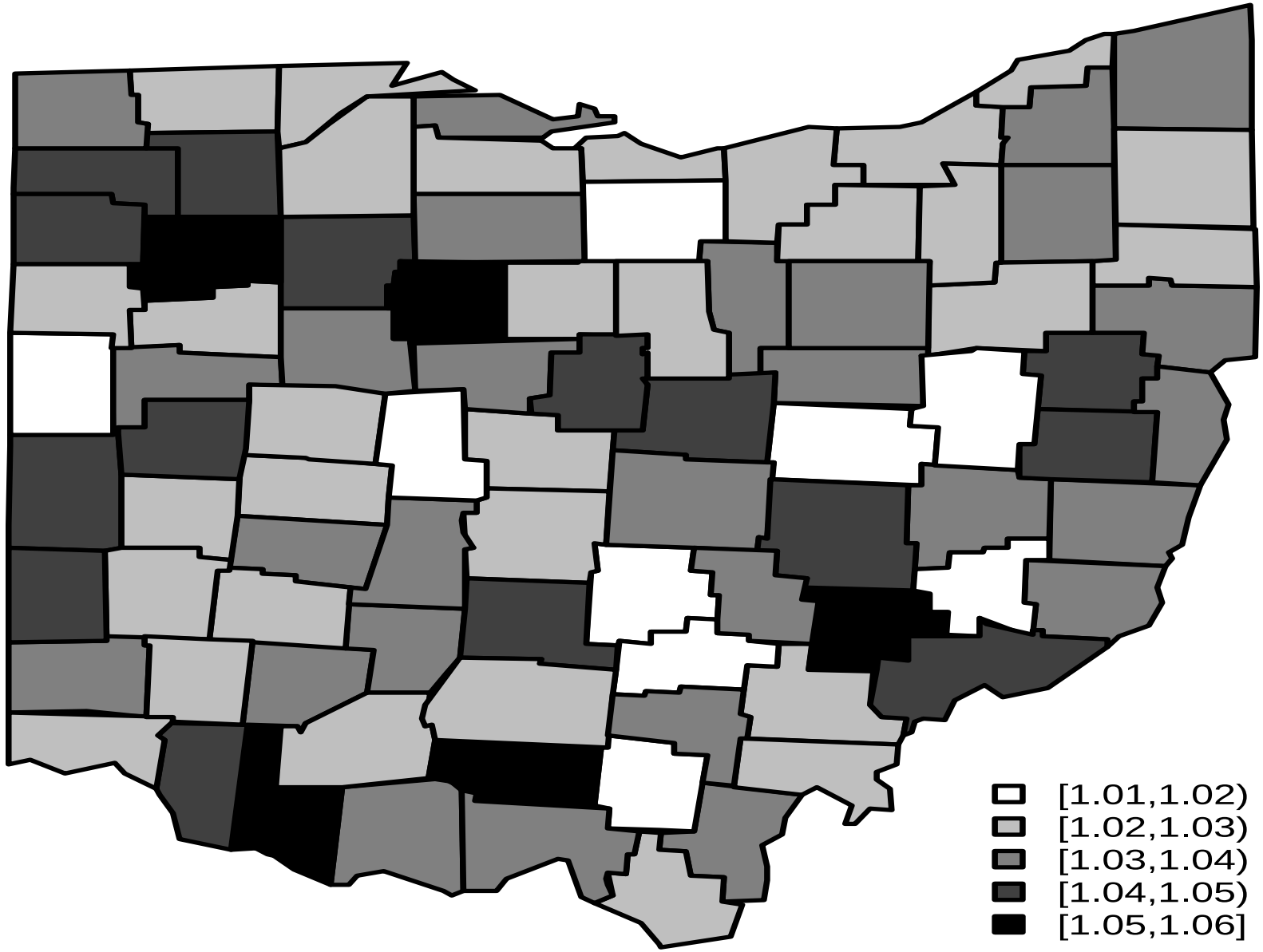


Figure 8 : Map of exponentiated slopes for Ohio lung cancer data; there is no obvious pattern.

Simple Space-Time Models

Bernardinelli *et al.* (1995) suggested the model

$$\log \theta_{it} = \beta_0 + \epsilon_i + S_i + (\beta_1 + \eta_i)t$$

where

- β_0 is the intercept,
- ϵ_i and S_i are non-spatial and spatial random effects and
- η_i are area-specific interaction parameters that adjust the average slope β_1; the η_i may or may not have spatial structure.

Hence, this model assumes temporal trends are linear (on the log scale) but that the slopes differ between areas.

Simple Space-Time Models

Waller *et al.* (1997) assume that

$$\log \theta_{it} = \beta_0 + \epsilon_{it} + S_{it}$$

with non-spatial and spatial random effects ϵ_{it} and S_{it}, respectively.

These random effects are assumed independent across time (though we may allow the variances of the distributions to depend on time).

As pointed out by Knorr-Held and Besag (1998) there is no structure (smoothing) across time in this model which is not realistic.

Main Effects and Interactions

Suppose we have a univariate continuous response Y.

Suppose we have two factors with levels, A and B, with $i = 1, \ldots, I$ and $j = 1, \ldots, J$ indexing the levels.

A main effects only model takes the form

$$\mathsf{E}[Y|\beta_0, \eta_i, \phi_j] = \beta_0 + \eta_i + \phi_j.$$

Interpretation: η_i is the effect of being at level i for factor A, regardless of the level assumed by B, i.e. there is no interaction.

Main Effects and Interactions

An interaction model adds a set of interaction parameters

$$\mathsf{E}[Y|\beta_0, \eta_i, \phi_j, \delta_{ij}] = \beta_0 + \eta_i + \phi_j + \delta_{ij}.$$

Interpretation: δ_{ij} is the additional effect, beyond $\eta_i + \phi_j$ of being simultaneously at levels i and j of factors A and B.

If the factor correspond to nominal levels (e.g., a factor for color with 2 levels: "red", "blue") then we would not expect similarity between adjacent levels.

In a space-time context the "factors" space and time have structure and we would expect similarity.

Separable Main Effects Model

First, consider a separable space-time model

$$\begin{aligned} Y_{it}|\theta_{it} &\sim \text{Poisson}(E_{it}\theta_{it}) \\ \log\theta_{it} &= \beta_0 + \epsilon_i + S_i + \omega_t + \tau_t \end{aligned}$$

Components:

- Unstructured spatial term $\epsilon_i \sim_{iid} \text{N}(0, \sigma_v^2), i = 1, \ldots, n$.
- Smooth spatial term $(S_1, \ldots, S_n)$ smooth in space (e.g. ICAR model).
- Smooth temporal term $(\tau_1, \ldots, \tau_T)$ smooth in time (e.g. follows a random walk of first or second order).
- Unstructured temporal term $\omega_t \sim_{iid} \text{N}(0, \sigma_\omega^2), t = 1, \ldots, T$.

Notice there is no interaction between space and time. The spatial effects are constant across time and temporal trends are constant across space.

Inseparable Space-Time Interaction Models

Knorr-Held (2000) considered the model:

$$\theta_{it} = \beta_0 + \epsilon_i + S_i + \omega_t + \tau_t + \delta_{it}$$

with ϵ_i, S_i, ω_t, η_t are as in the separable model.

Four different models for the interaction δ_{it}:

- Type I: Independent interaction.
- Type II: Temporal trends differ between areas but don't have spatial structure.
- Type III: Spatial patterns differ between time points but don't have temporal structure.
- Type IV: Temporal trends differ between areas but more likely to be similar for adjacent areas.

Inseparable Space-Time Interaction Models

Type II: Temporal trends differ between areas but don't have spatial structure. For example, an RW(2) model in each area has conditional distribution:

$$\delta_{it}|\delta_{i,t-1},\delta_{i,t+1},\delta_{i,t-2},\delta_{i,t+2} \sim \mathrm{N}\left(\frac{4}{6}(\delta_{i,t-1}+\delta_{i,t+1})-\frac{1}{6}(\delta_{i,t+2}-\delta_{i,t-2}),\frac{\sigma_\delta^2}{6}\right).$$

The joint distribution for this model can also be written

$$f(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{i=1}^{I}\sum_{t=3}^{T}(\delta_{it}-2\delta_{i,t-1}+\delta_{i,t-2})^2\right).$$

Realistic to assume that time trends have no spatial structure?

Inseparable Space-Time Interaction Models

Type III: Spatial patterns differ between time points but without temporal structure:

$$f(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{t=1}^{T}\sum_{i\sim j}(\delta_{it}-\delta_{jt})^2\right).$$

So this model says we have independent ICAR models at each time point (though with the same variance, σ_δ^2).

Realistic to assume that spatial structure changes at every time point without smooth patterns in space?

Inseparable Space-Time Interaction Models

Type IV: Temporal trends differ between areas but more likely to be similar for adjacent areas.

This will often be the most realistic model if interactions are present.

In the case of a RW1 temporal model and an ICAR spatial model, the joint distribution can be written:

$$f(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{t=3}^{T}\sum_{i\sim j}(\delta_{it}-\delta_{jt}-2\delta_{i,t-1}+2\delta_{j,t-1}+\delta_{i,t-2}-\delta_{j,t-2})^2\right)$$

The `inla` implementation of Type II, III and IV interaction models is complex.

On the class website is code to fit Type II–IV interaction models.

Conclusions

Raw estimates of area relative risks can be unstable when population sizes/expected numbers are small.

Random effects shrinkage models can stabilize rates by jointly estimating the complete set of relative risks over the complete study region.

These models have good properties to describe the complete collection of areas estimates, but any one area can be poorly estimated, since the accuracy depends on the appropriateness of the model for that area.

One common model includes a set of independent random effects and a set of spatial (ICAR) random effects.

Bayesian modeling is the most common approach so priors must be specified, and the sensitivity to the priors on the variances is important to examine.

Examination of model assumptions is important, though very difficult to asses whether the spatial model is reasonable.

Conclusions

For chronic diseases such as cancer the time trends are often slow, though may change more abruptly if large-scale screening is implemented.

Often there are more areas than time points (though counts in areas may be small), and so the sophistication of the temporal model may be restricted.

Age-period-cohort (APC) models may be used if there are sufficient periods, spatial-APC models have been developed.

References

Bernardinelli, L., Clayton, D. G., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, **14**, 2433–2444.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, **36**, 192–236.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*, **43**, 1–59.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–682.

Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. John Wiley and Sons.

Dominici, F., McDermott, and Hastie, T. (2004). Improved semi-parametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, **468**, 938–948.

Elliott, P., Wakefield, J. C., Best, N. G., and Briggs, D. J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press, Oxford.

Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, **11**, 397–412.

Gabriel, E., Rowlingson, B., and Diggle, P. (2013). stpp: an r package for plotting, simulating and analyzing spatio-temporal point patterns. *Journal of Statistical Software*, **53**, 1–29.

Jones, S. and Kulldorff, M. (2012). Influence of spatial resolution on space-time disease cluster detection. *PLoS One*, page e48036.

Kemp, I., Boyle, P., Smans, M., and Muir, C. (1985). *Atlas of Cancer in Scotland, 1975–1980: Incidence and Epidemiologic Persepctive*. IARC Scientific Publication 72, Lyon, France, International Agency for Research on Cancer.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**, 2555–2567.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, **17**, 2045–60.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A*, **164**, 61–72.

Le, N. and Zidek, J. (2006). *Statistical Analysis of Environmental Space-Time Processes*. Springer, New York.

Mantel, N. (1967). The detection of disease clustering and a generalised regression approach. *Cancer Research*, **27**, 209–220.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Application*. Chapman and Hall/CRC Press, Boca Raton.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.

Taylor, B., Davies, T., Rowlingson, B., and Diggle, P. (2013). lgcp: an r package for inference with spatial and spatio-temporal log-Gaussian Cox processes. *Journal of Statistical Software*, **53**.

Wakefield, J. C., Best, N. G., and Waller, L. A. (2000). Bayesian approaches to disease mapping. In P. Elliott, J. C. Wakefield, N. G. Best, and D. Briggs, editors, *Spatial Epidemiology: Methods and Applications*, pages 104–27. Oxford University Press, Oxford.

Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607–617.

Xia, H. and Carlin, B. (1998). Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in Medicine*, **18**, 2025–2043.