ANALYSIS OF SPATIAL DATA IN EPIDEMIOLOGY

Prof. Dr. Maria A Barceló and Prof. Dr. Marc Saez

September 8, 10, 14 and 16, 2021

Research Group on Statistics, Econometrics and Health (GRECS), University of Girona
CIBER of Epidemiology and Public Health (CIBERESP)

GRECS
GRUP DE RECERCA EN ESTADÍSTICA
ECONOMETRIA I SALUT
UdG
ciberesp
Centro de Investigación Biomédica en Red
Epidemiología y Salud Pública

COURSE INTRODUCTION

1. Course introduction
2. Introduction to epidemiology and spatial statistics
3. Overview of mixed models
4. Overview of mixed models - Practicals
5. Introduction to INLA and R INLA
6. R INLA - Practicals

Wednesday 8

Friday 10

Marc Saez Maria Antònia Barceló

COURSE INTRODUCTION

7. Disease mapping. Standardisation of incidence and mortality rates
8. **Disease mapping. Smoothing standardised incidence and mortality rates**
9. Disease mapping – Practicals
10. Geographical association studies. Spatial ecological regression
11. Spatial ecological regression - Practicals

Tuesday 14

Marc Saez Maria Antònia Barceló

COURSE INTRODUCTION

12. Clustering
13. Extensions: BYM2, point processes, leaflet, pc priors
14. Extensions – Practicals

Thursday 16

Marc Saez Maria Antònia Barceló

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Remember that we said that we are interested in considering the spatial component:

- Because we are explicitly interested in the spatial pattern of the risk factor: **disease maps**
- Because it contains a large part of the non-observed confounding: **spatial regre**ssion
- Because we observe agglomerations in the space: **cluster detection**
- Because we are interested in the effects of a pollutant source on the health of the residents in the surrounding area: **source identification**

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Remember that:

- **Disease maps** provide an initial overview of the spatial distribution of the disease, health event or its risk factors. They are a visual summary of the geographical risk..
- The indicators of disease occurrence most often represented on these maps are morbidity indicators (prevalence and incidence) and mortality indicators.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

- it is incorrect to use absolute numbers or crude rates. At the least, standardized rates, SIRs or SMRs, should be used.
- to estimate ***relative risk***, a ***Poisson regression*** should be used, or a ***binomial regression*** in the case of more frequent diseases.
- even if we use a regression to estimate the relative risk, we must not forget that the regression models still estimate SIRs (or SMRs).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Nonetheless, we have seen that standardized rates have a series of problems: :

- ***In the case of rare diseases and/or small areas, SMRs (or SIRs) are very inaccurate*** because the variance is proportional to the square of the denominator (in other words, it is proportional to the number of expected cases squared)
- ***The variance*** associated with areas with small expected cases ***will be very high***.
- ***The SMRs (or SIRs) are estimated independently in each area***.
- SMRs (or SIRs) ***do not take into account the very likely spatial dependence***.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Possible solution:

- Use ***smoothed standardized incidence ratio*** (smoothed SIR) and/or ***smoothed standardized mortality ratio*** (smoothed SMR).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

In summary:

- The ***problems*** we encounter when estimating standardized rates (both mortality and incidence rates) are the ***instability of the estimators*** and ***the effects of not taking into account the very likely spatial dependence***.
- These problems can be partly solved by ***spatial smoothing of the rates***.
- There are several methods to smooth these rates.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Methods to smooth maps

- Local smoothing rates, ad hoc, for example, spatial moving averages, headbanging algorithms
- Trend surface analysis, for example, kriging, smoothing by polynomials/splines
- Random effects models

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Methods to smooth maps

- ***Local smoothing rates, ad hoc***, for example, spatial moving averages, headbanging algorithms
  - Advantages: quick and simple to implement
  - Disadvantages: can be very sensitive to ad hoc choice of weights, etc. and there are no estimates for uncertainty (standard errors).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Methods to smooth maps

- ***Trend surface analysis***, for example, kriging, smoothing by polynomials/splines
  - Advantages: the estimation of the "smoothing parameters" is based on the balance between the fit and the smoothing, and standard errors are generally available.
  - Disadvantages: can be sensitive to the choice of the penalisation parameter for balancing.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Methods to smooth maps

- ***Random effects models***: the most popular alternatives to the methods discussed previously are random effects models.

- Two types of random models:
    - ***Empirical Bayes***: include only the spatial heterogeneity as a random effect.
    - ***Bayesian hierarchical models (mixed models)***: include two random effects, spatial heterogeneity and spatial dependence.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Methods to smooth maps

➢ ***Random effects models***

Prior to the vast IT boom from 2000 onwards, empirical Bayes models were mainly used, since the other models were computationally complex. These are now obsolete and ***Bayesian hierarchical models are used***.

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models

In order to solve the problems we have seen presented by the SMR and/or SIR, we use smoothed Bayesian estimators of the SMR and/or SIR obtained by means of:

- ***The Poisson-logNormal-spatial model*** (Bayesian hierarchical model with a random effect).
- ***Besag, York and Mollié model, BYM*** (Bayesian hierarchical model with two random effects).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: The Poisson-logNormal-spatial model

Poisson-logNormal model

$$
\begin{aligned}
O_i &\sim \text{Poisson}(\lambda_i E_i) \\
\log \lambda_i &= \alpha + V_i \\
V_i &\sim \text{Normal}(0, \sigma_v^2)
\end{aligned}
$$

Priors (vague, non informative):

- between-area variance σ_v^2:
  $\sigma_v^2 \sim$ Inverse Gamma$(0.5, 0.0005) \Leftrightarrow \tau_v \sim$ Gamma$(0.5, 0.0005)$
  $\sigma_v \sim$ Truncated Normal$(0, 100)_{[0, Inf)}$
- mean log relative risk: $\alpha \sim N(0, 100)$

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: The Poisson-logNormal-spatial model

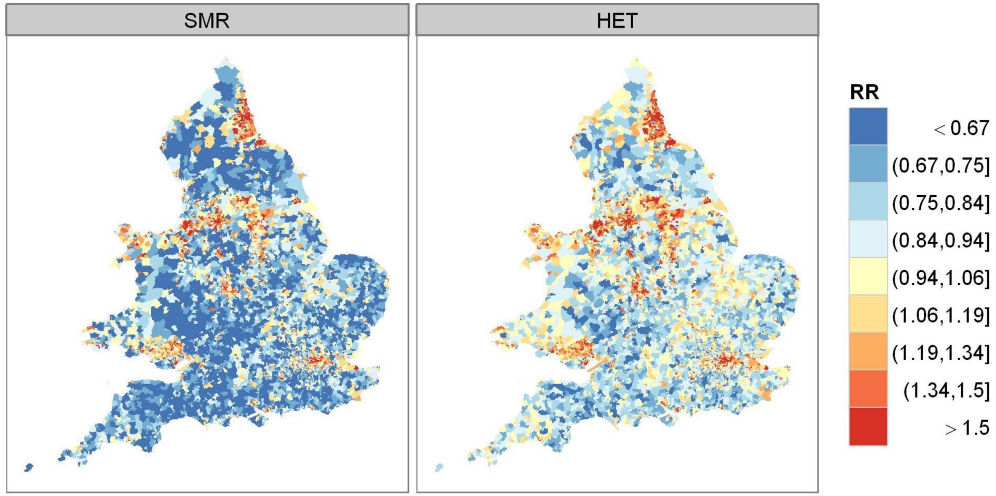
- $\boldsymbol{O_i}$, $\boldsymbol{E_i}$ are the number of observed and expected cases in area i, respectively
- $\boldsymbol{\lambda_i = exp(\alpha + V_i)}$ is the relative risk (RR) in area i compared with the expected risk based on the age and gender of the population
- Parameters $\boldsymbol{V_i}$ are specific random effects for each area to control the heterogeneity (overdispersion)

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: The Poisson-logNormal-spatial model

Lung cancer incidence in males, 1985-2009, England and Wales (I)

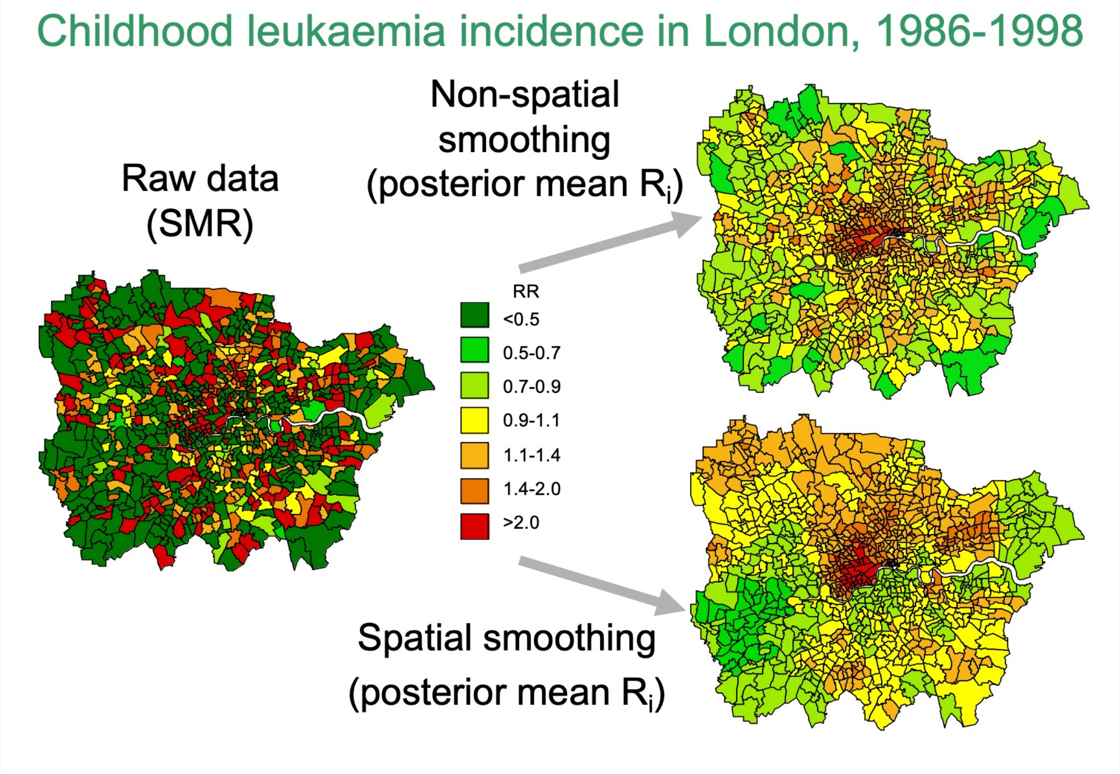
RR estimates using 2 methods



SMRs and non-spatially smoothed RRs

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: The Poisson-logNormal-spatial model



SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: BYM model

- ***The random effect used to capture the spatial heterogeneity*** is the same in all the models (iid normal).
- ***The random effect used to capture the spatial dependency*** can be approximated using a CAR (Conditional Autoregressive) (used when having areal data)

The spatial dependency is currently approximated using a Matern (**the so called log-Gaussian-Cox**, used for all types of spatial data).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects models: BYM model

The λ_i are usuallly spatially correlated because they reflect, in part, spatially varying risk factors.

To correct for this we must:

- incorporate the spatial dependence in the distribution of the λ_i.
- in the BYM model, this dependence is modelled as a CAR (conditional autoregressive model).

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

- **BYM model. CAR model**

$\log(R_i) \sim \text{Normal}(m_i, v_i)$

$m_i = \Sigma_k R_k / n_i$ = average risk in neighbouring areas

$v_i = v / n_i \rightarrow$ variance inversely proportional to number of neighbours

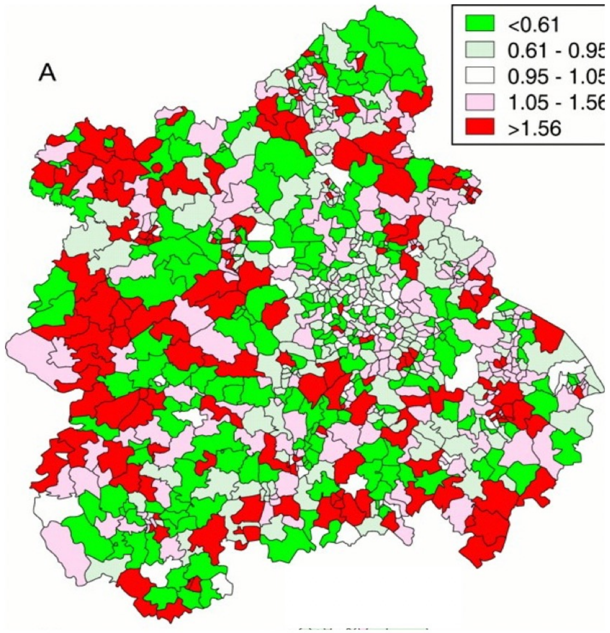
Besag, York, Mollie (1991) *Annals of the Institute of Statistics and Mathematics*, 43: 1-59

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Example:

Map of SMR of adult leukaemia in West Midlands Region, England 1974-86

(Olsen, Martuzzi and Elliott, *BMJ* 1996;313:863-866).



Is the variability real or simply reflecting unequal expected counts ?

Have the red highlighted areas truly got a raised relative risk?

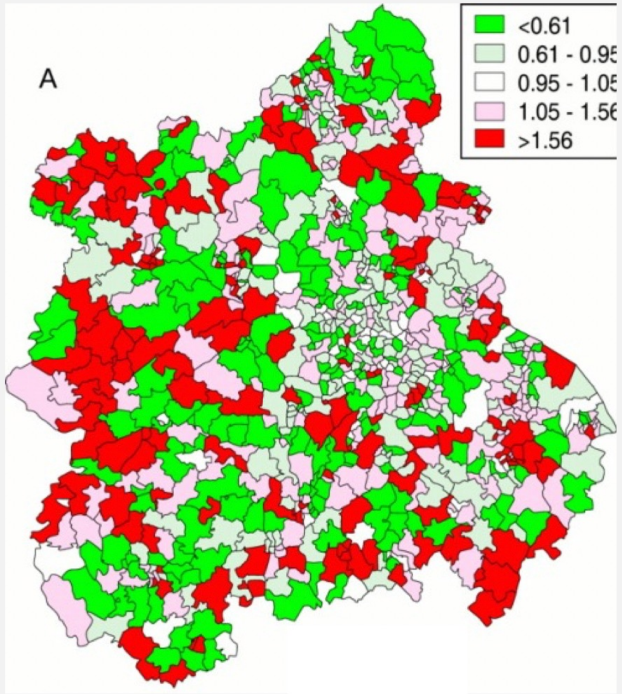
SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Example:

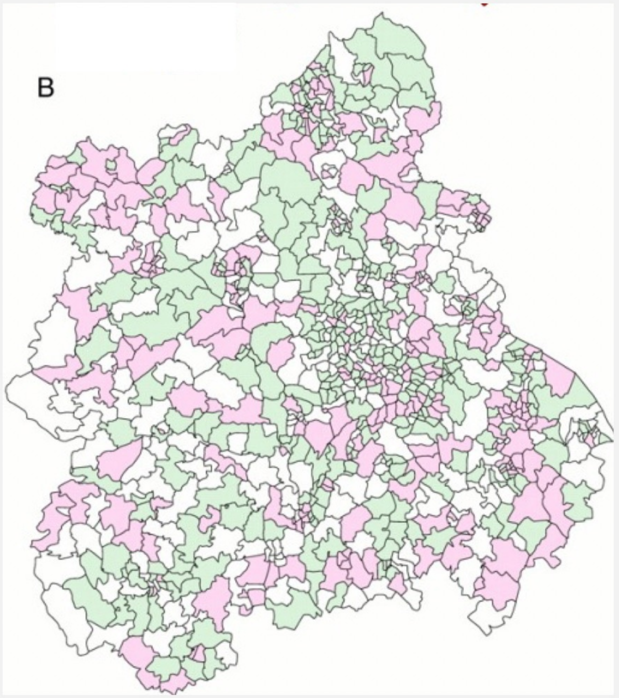
Map of SMR of adult leukaemia in West Midlands Region, England 1974-86

(Olsen, Martuzzi and Elliott, *BMJ* 1996;313:863-866).

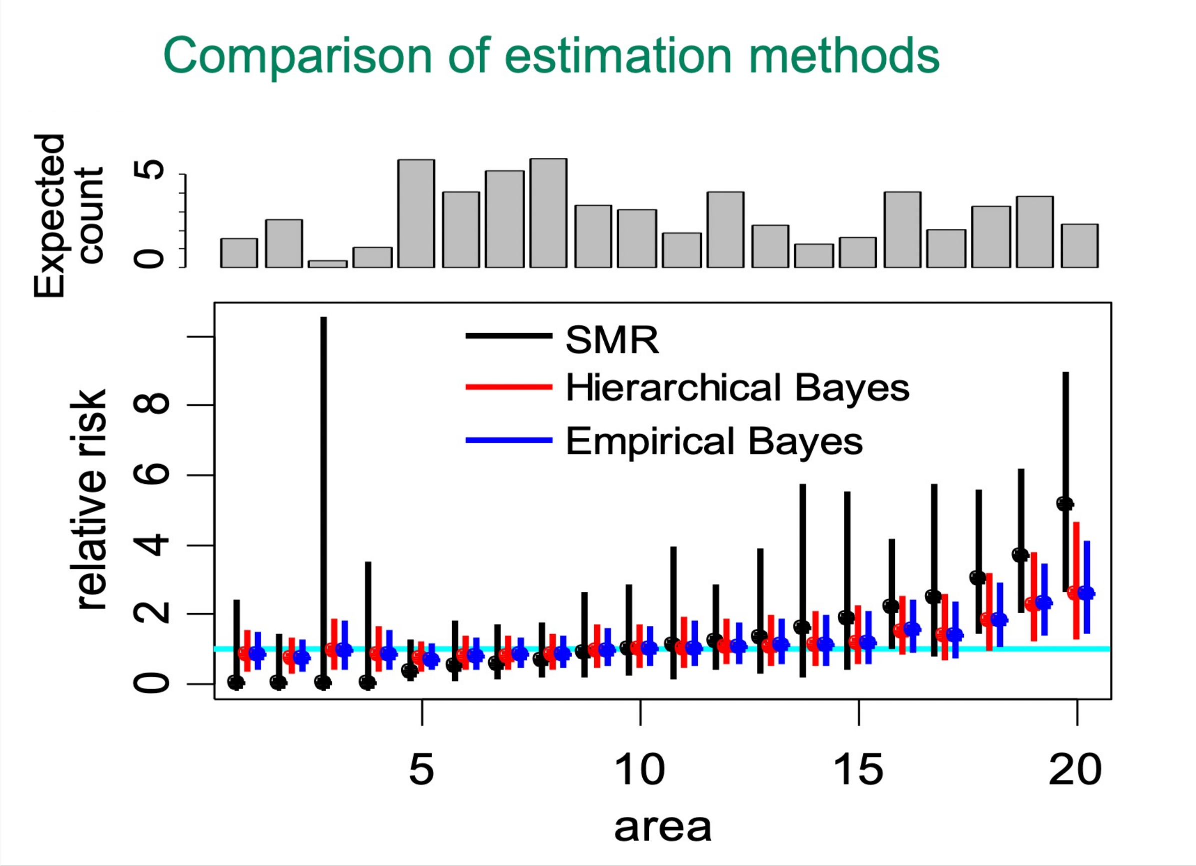
(A) unsmoothed SMR



(B) smoothed by Bayesian methods

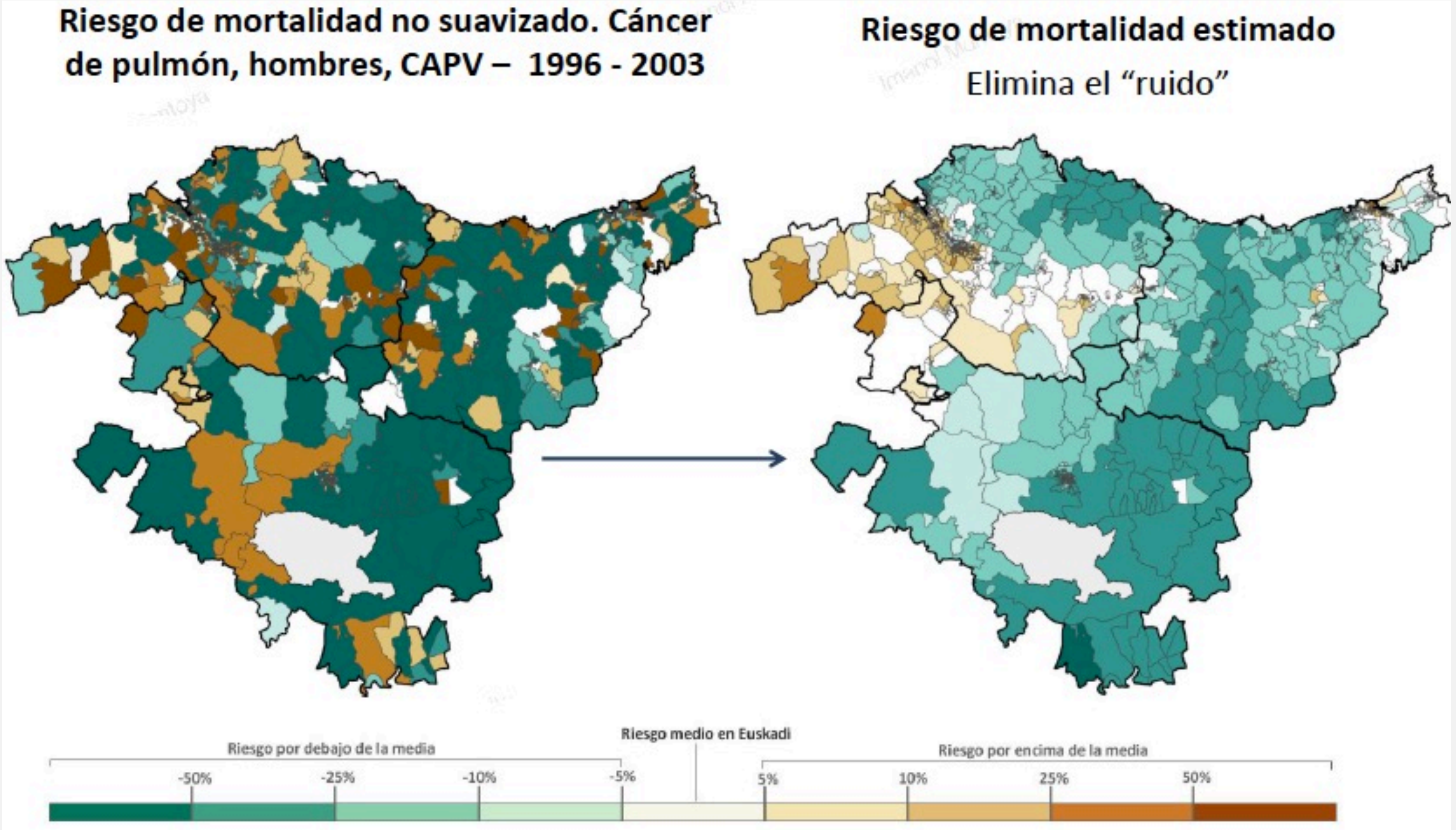


SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES



SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Ejemplo:

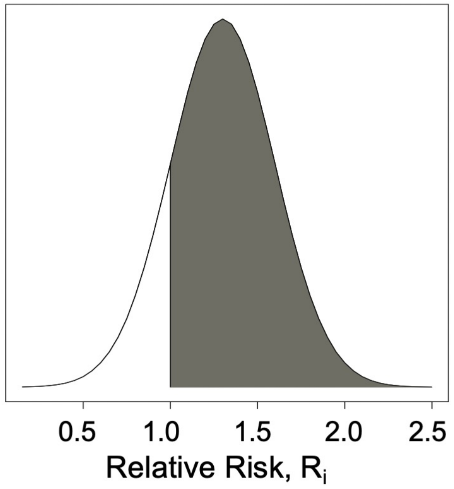


SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Random effects model: Poisson-logNormal-spatial model

Mapping uncertainty

- Mapping the mean posterior value of R_i does not make full use of the posterior distribution



- Map posterior SD
- Map
  Probability (R_i > 1)

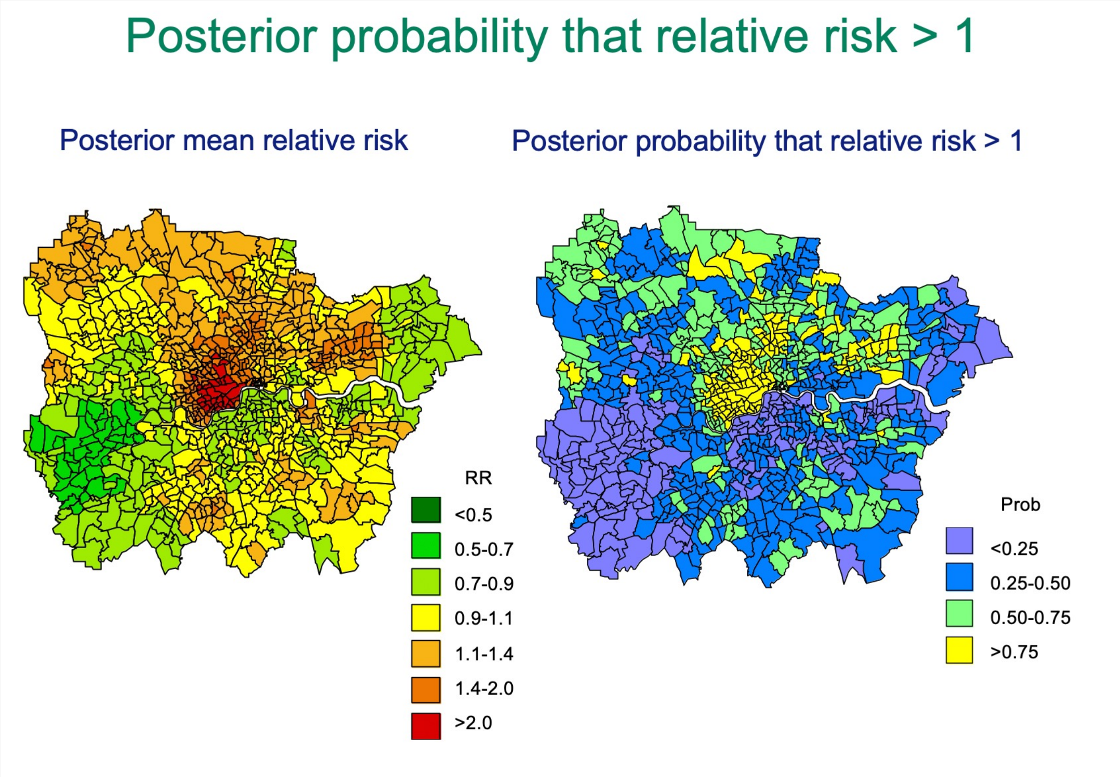
Note – this is not the same as a classical p-value

SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

- Richardson et al. (2004): simulation study that investigates the use of a posteriori probabilities de in studies focused on the representation of disease on maps
  - An area is classified as elevated risk if $[Prob(R_i > 1)] > 0.8$
  - An area is classified as elevated specificity if $false\ detection < 10\%$
  - Sensitivity 60%-95% for E_i of 5-20 and true R_i of 1.5-3.0

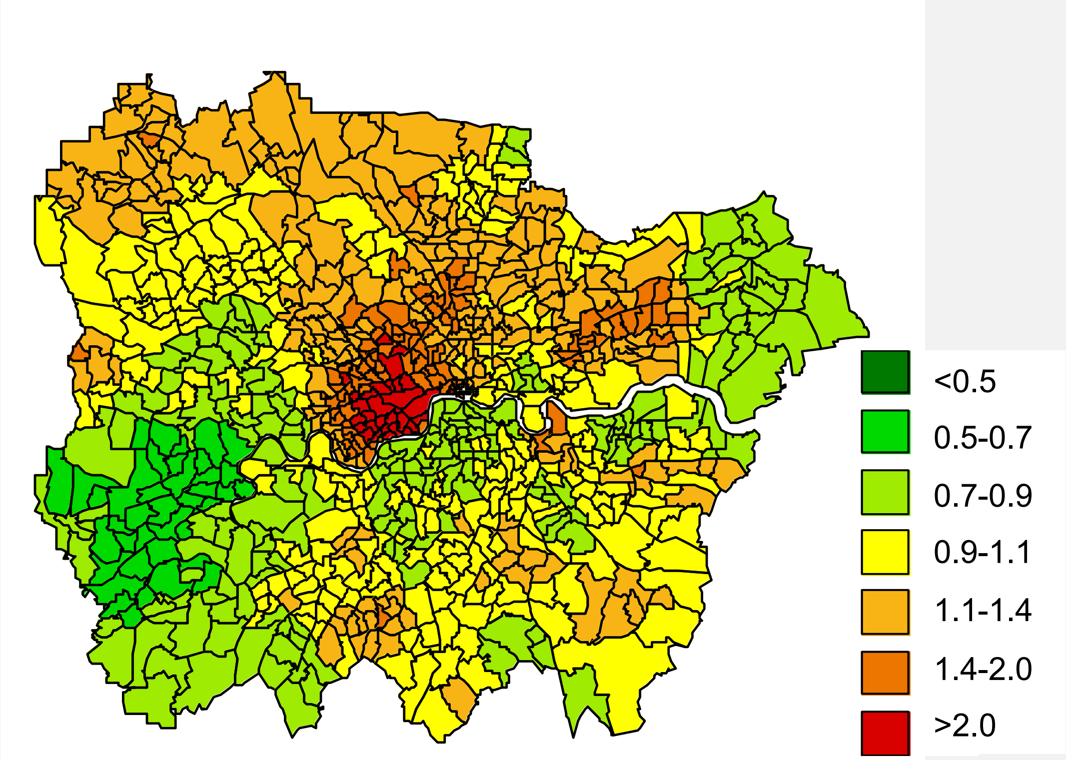
SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

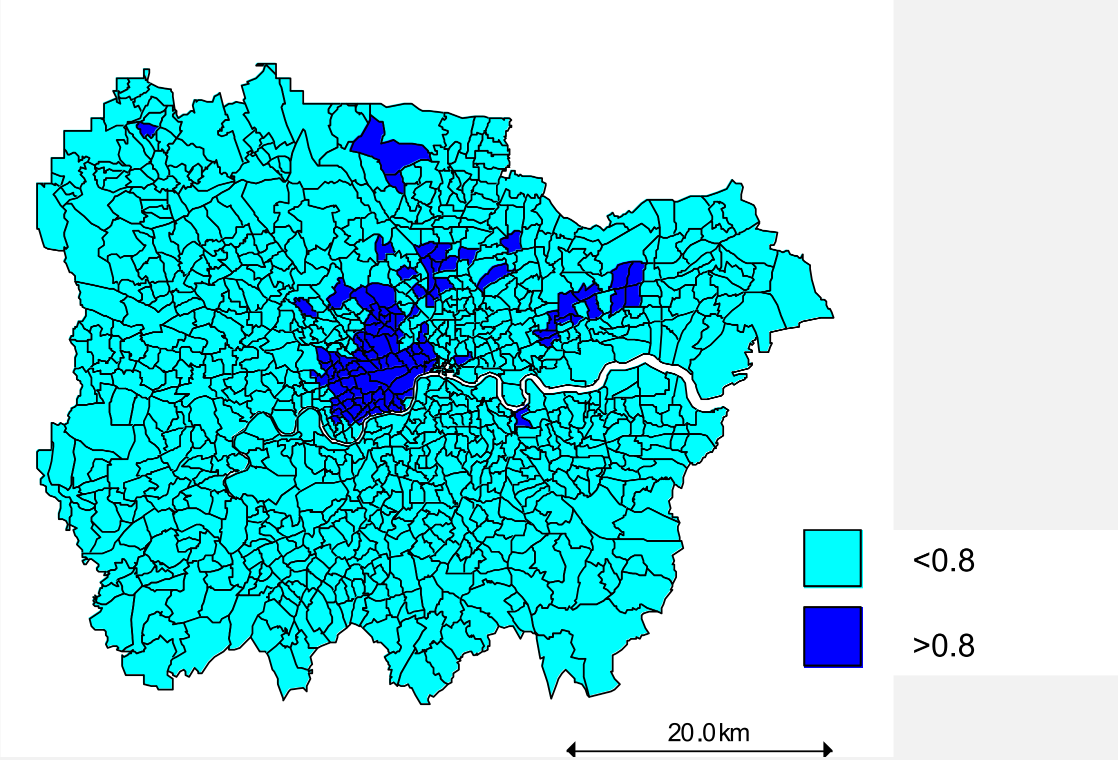
Example: Childhood leukaemia incidence in London, 1986-1998



SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Example: Childhood leukaemia incidence in London, 1986-1998

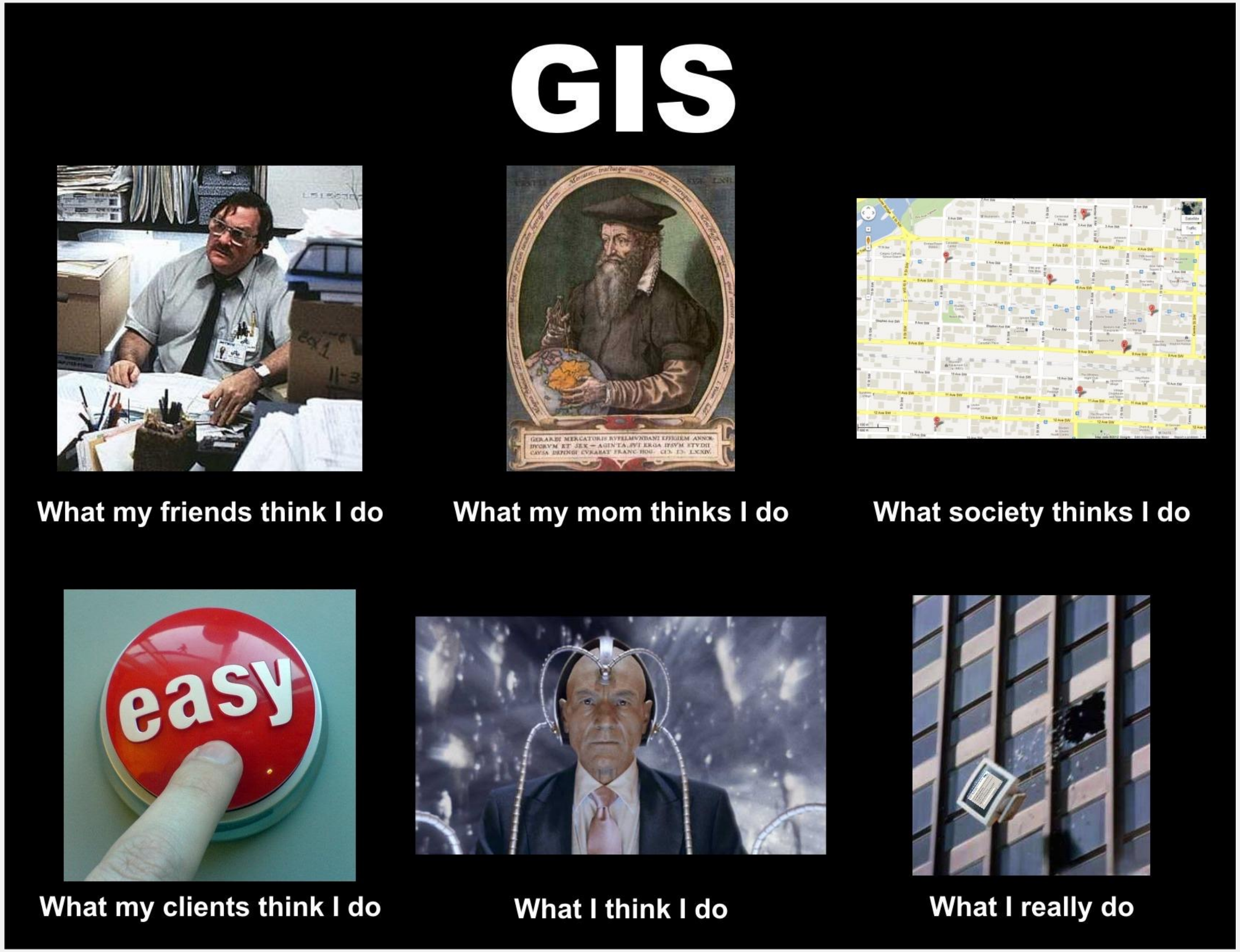




SMOOTHING STANDARDISED INCIDENCE AND MORTALITY RATES

Software

- For smoothing using Empirical Bayes, the Rapid Inquiry Facility (RIF) of SAHSU from Imperial College can be used.
- Estimation of Bayesian hierarchical models is done using computationally complex simulation methods (MCMC) implemented in the free WinBUGS and GeoBUGS software.
- Since 2008, the free software INLA (Rue et al, 2008) has implemented a fast and more efficient approach to MCMCs to estimate these models.


GIS
What my friends think I do
What my mom thinks I do
What society thinks I do
easy
What my clients think I do
What I think I do
What I really do