Quantifying the Health Impacts of Air Pollution

Day 3: Mapping Risks

Prof. Gavin Shaddick
University of Exeter

10^{th} August 2017

Outline

Spatial Epidemiology

What is Spatial Epidemiology?

- Epidemiology is the study of the distribution of diseases in populations.
- Disease risk depends on the person (genetics/behaviour), place and time.
- Spatial epidemiology focuses on the second of these.
- Place is a surrogate for exposures present at that location
  - environmental exposures in water/air/soil
  - lifestyle characteristics of those living in particular areas.

Growing interest in spatial epidemiology

- Public interest in effects of environmental hazards/pollution.
- Epidemiological interest in differences in disease rates across different areas.
- Data availability: collection of health data at different geographical scales.
- Increase in computing power and methods
  - Geographical Informations Systems (GIS).
- Development of statistical/epidemiological methods for investigating disease 'clusters'.

The Need for Spatial Methods

- Many epidemiological studies are spatial
  - many are spatio-temporal!
- When do we need to 'worry'?
  - are we explicitly interested in the spatial pattern of disease incidence?
    - disease mapping
    - cluster detection.
  - is the clustering a nuisance quantity that we wish to acknowledge, but are not explicitly interested in?
    - spatial regression.

Types of Spatial Data

- Point data
  - 'exact' residential locations exist, e.g. for cases and controls.
- Count data
  - aggregation
  - typically over administrative units.

Disease Mapping

Overview of Disease Mapping

- 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
  - estimation of background variability in underlying risk in order to place epidemiological studies in context.
- 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.

Example: COPD in England

- Incidence rates of hospital admissions for Chronic Pulmonary Obstructive Disease (COPD) in 324 local authorities of England, between in 2001–2010.
- Data:
  - observed and 'expected' number of cases (based on the local authority age-sex profiles)
  - this allows the calculation of the standardised morbidity ratio (ratio of the observed to the expected cases).

Expected Numbers

- The expected number of deaths/disease are calculated using *indirect* standardisation.
- Rates from a reference population are applied to the population of interest.
- The expected number

$$E = \sum_k N_k r_k$$

where r_k is the rate in the reference population and N_k is the population in the study. Commonly, k would denote age–sex categories.

Standardised Mortality/Morbidity Ratios

- The observed number of deaths can be compared to the expected number using the standardised mortality/morbidity ratio (SMR).
- Let O be the observed number of deaths/disease cases in the population of interest, and E be the expected number.

$$\text{SMR} = \frac{O}{E}$$

- An SMR of 1 means that the rates of death/disease in the population of interest are the same as in reference population.
- If it is greater than 1, we have more deaths/disease cases than expected; if it is less than 1 we have less.

Example: COPD in England

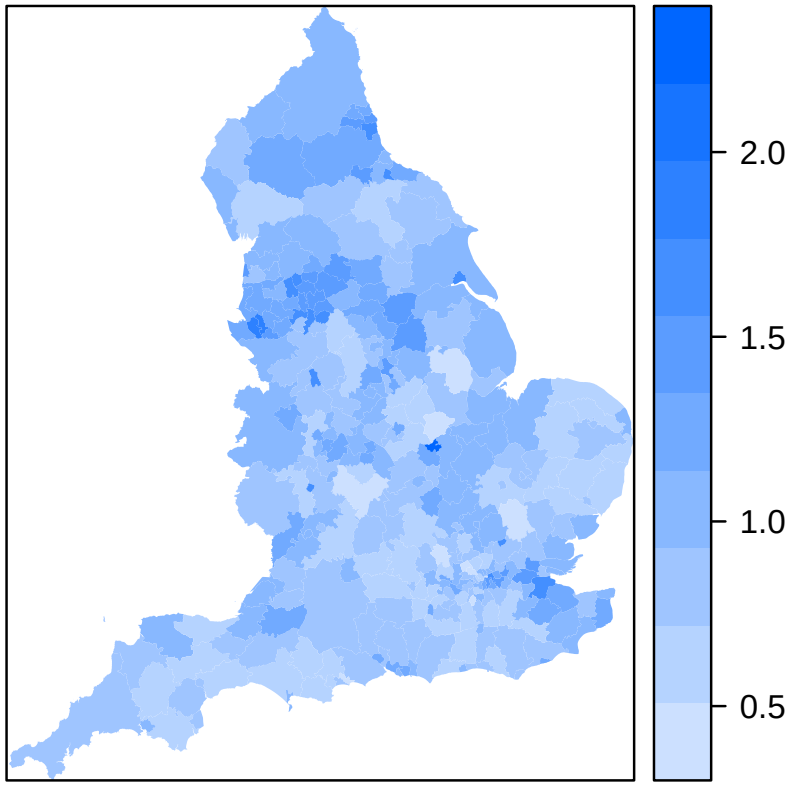


Figure: SMRs for hospital admissions of COPD in 324 local authorities in England.

Non-statistical issues

- There is a trade-off when a geographical scale is chosen
  - larger geographical areas provide more stable rates
  - risks may be distorted due to the large aggregation of individuals.
- The size of the areas chosen also determines the sort of questions that can be posed
  - smaller areas often mean greater contrasts in relative risks and exposures
  - localized effects can only be detected with data at a smaller level of aggregation
  - high (and low) relative risks will be diluted under aggregation
  - if the relative risk shows marked variation within a particular area this information may be lost under aggregation.

Statistical Issues

- Unfortunately there are well-documented difficulties with the mapping of raw estimates of risk.
- For small areas and rare diseases in particular, estimates will be dominated by sampling variability.
- The SMR is given by

$$\text{SMR}_i = \frac{O_i}{E_i}$$

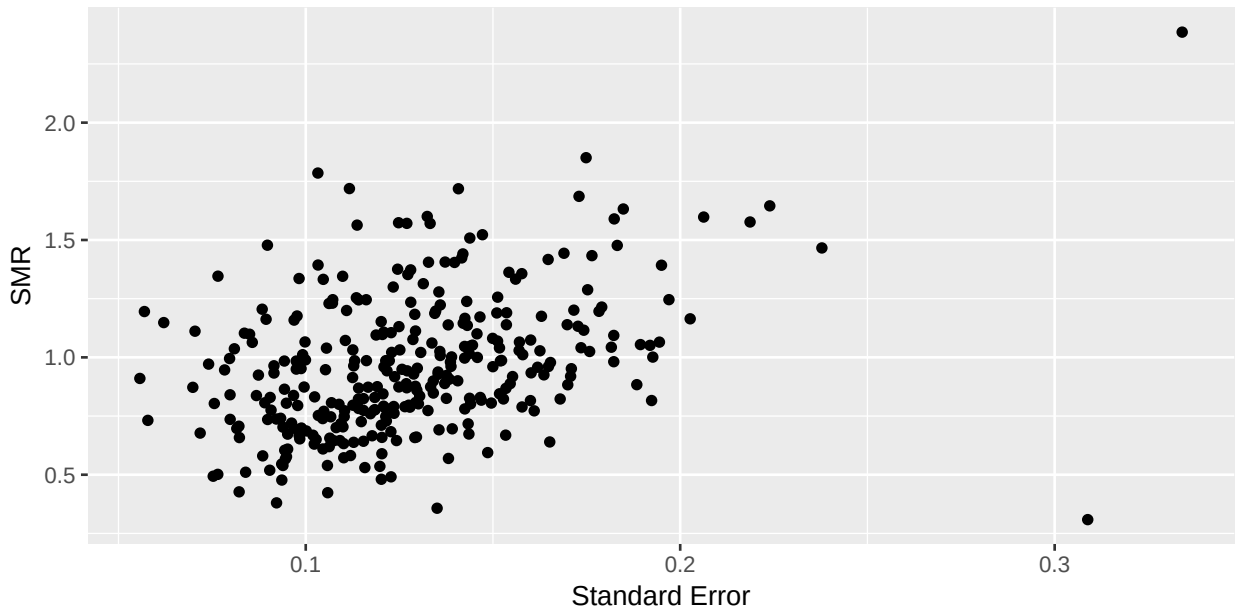
and it has variance

$$\text{var}(\text{SMR}_i) = \frac{\text{SMR}_i}{E_i} = \frac{O_i}{E_i^2}$$

therefore areas with smaller expected numbers, E_i have higher variance.

EXAMPLE: SCOTTISH LIP CANCER

- For COPD in England, the expected numbers are highly variable, with range 3.2–368.9.
- Are the extreme SMRs based on small expected numbers?
  - many of the large, sparsely populated, rural areas in the north have high SMRs.



Smoothing Models

- The above considerations have led to the development of models to *smooth* SMRs.
- These approaches use data from surrounding areas to improve the stability of estimates.

Example: COPD in England

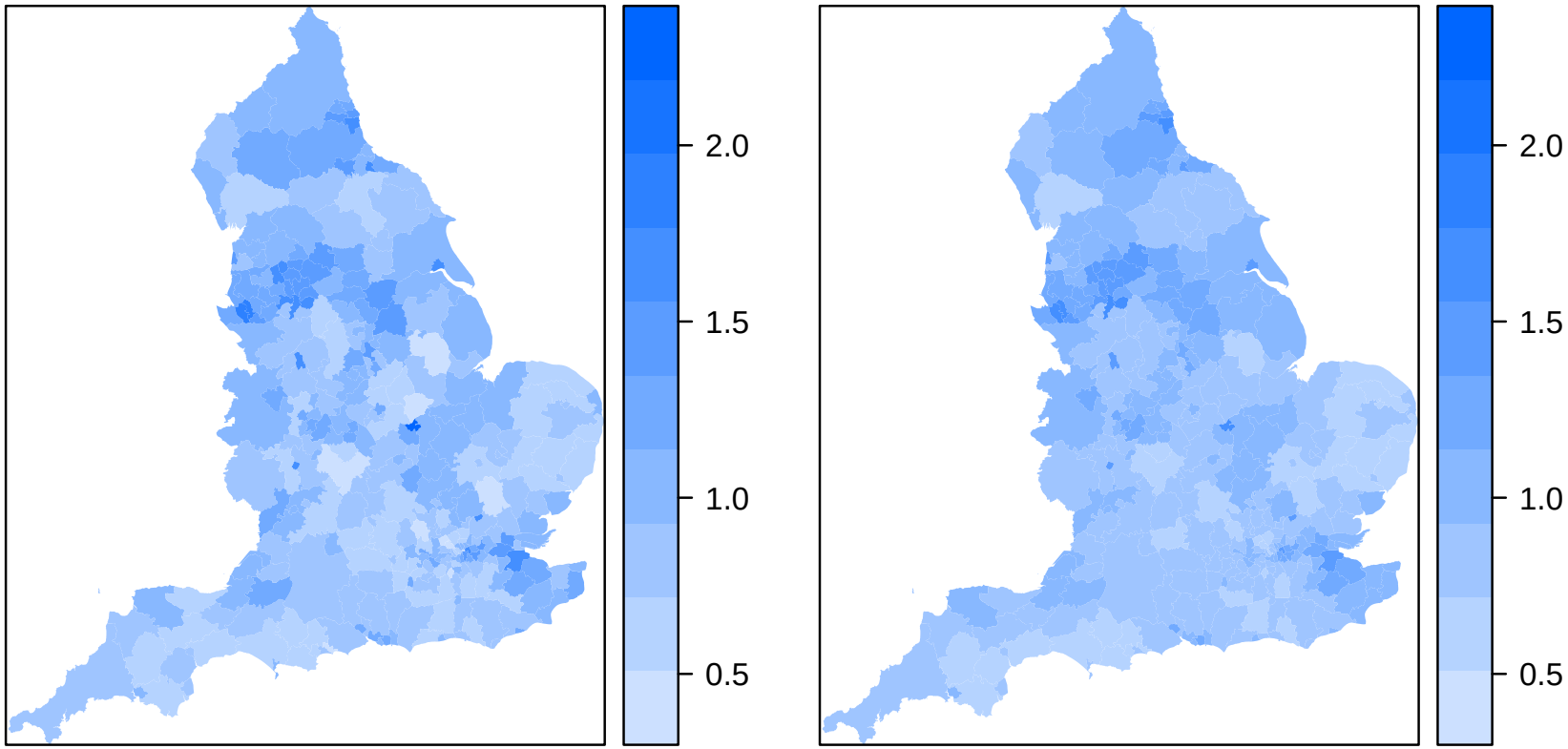


Figure: (Left) Unsmoothed SMRs and (Right) Smoothed SMRs for hospital admissions of COPD in England.

Statistical Methods for Smoothing Risks

SMOOTHING

- Defined in terms of 'neighbouring areas'.
- Often defined as when areas share a *common boundary*.
- May use an intrinsic conditional autoregressive (ICAR) distribution.

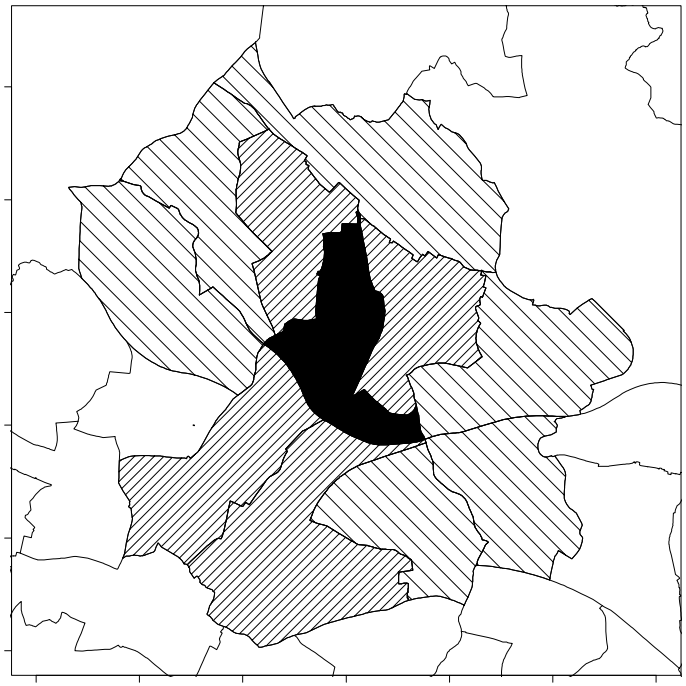


Figure: Close-up of a region within Birmingham, UK.

Bayesian Analysis

- Spatial smoothing models are not easily fit using traditional techniques, and so Bayesian methods are commonly used.
- Unfortunately, Bayesian models are not available in most standard software packages.
- Markov chain Monte Carlo (MCMC)
  - `Win/OpenBUGS` – A package that allows very general Bayesian modelling
  - `GeoBUGS` – Module of WinBUGS that contains spatial models and mapping facilities
  - `R` – Packages for example `CARBayes`.
- For fast computation on big datasets
  - `R-INLA`.

Example: COPD in England

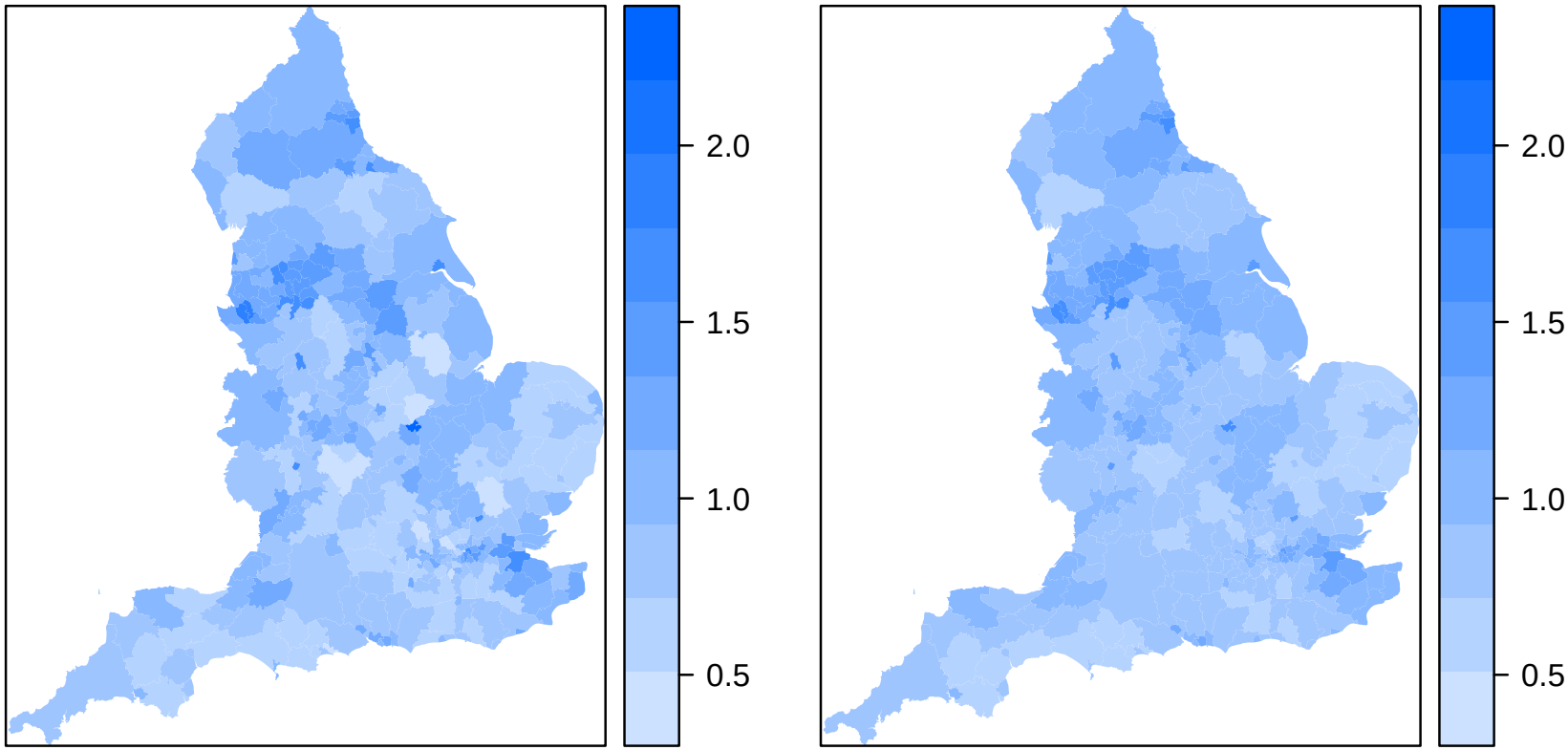


Figure: (Left) Unsmoothed SMRs and (Right) Smoothed SMRs for hospital admissions of COPD in England.

Summary of Smoothing in Disease Mapping

- The aim is to reduce the inherent instability in SMRs that are based on small expected numbers (from small populations).
- This is achieved by fitting a smoothing model.
- Comparing raw and smoothed SMRs is important.
- If there are big changes, are they appropriate?
  - were the expected numbers small?
  - can inform whether the statistical model was sensible or not.

And you will be able to do all of this by the end of the course!