Environmental Health Impact Assessment using R

Mapping Risks

Prof. Gavin Shaddick
University of Exeter

23^{rd} November 2017

Outline

Disease Mapping

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.

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.
- There can be difficulties with the mapping of raw estimates since, for small areas and rare diseases in particular, these estimates will be dominated by sampling variability.

Standardised Mortality/Morbidity Ratios

- We can compare the observed expected number of deaths in an area with the expected number of deaths based on the population age-sex profile.

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

- The SMR is a ratio, not a rate or a percentage. An SMR of 100 means that the population of interest has the same number of deaths as we would expect from a reference population.
- If it is greater than 1, there are more deaths/disease cases than expected; if it is less than 1 there are less.

Lip Cancer in Scotland

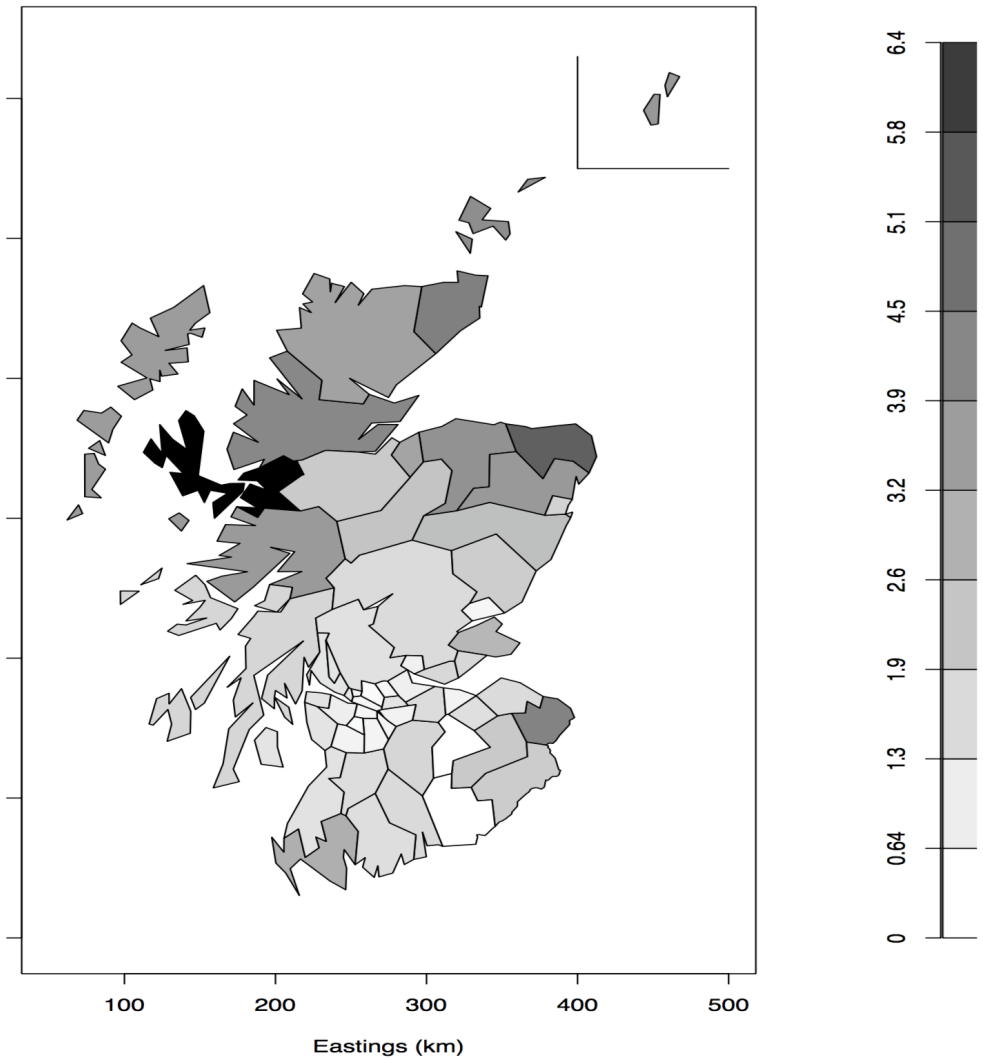
Example: Lip Cancer in Scotland

- Incidence rates of lip cancer in males in 56 counties of Scotland, registered in 1975–1980.
- Data
  - observed and 'expected' number of cases (based on the county age populations)
  - this allows the calculation of the standardised morbidity ratio
    - ratio of the observed to the expected cases.
  - 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.

Example: Lip Cancer in Scotland

County No. i	Obs Cases Y_i	Exp Cases 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
9	6	2.0	0.07	3.00	781.4	194.5	1,17,19,23,29
...							
47	2	5.6	0.01	0.36	640.8	277.0	24,31,46,48,49,53
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

Example: Lip Cancer in Scotland



Expected Numbers

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.

Cancer in North-West England

Example: Lung and Brain Cancer in North-West England

- Two tumors
  - one non-rare (lung)
  - and one rare (brain).
- Study period is 1981–1991.
- Analysis performed at ward level (144 wards)
  - incidence data by postcode.
- Brain cancer
  - the median number of cases per ward over the 11 year period is 6
  - range of 0 to 17.
- Lung cancer
  - the median number is 20
  - range 0–60.
- Expected counts were based on ward-level populations from the 1991 census, by 5-year age bands and sex.

Example: Lung and Brain Cancer in North-West England

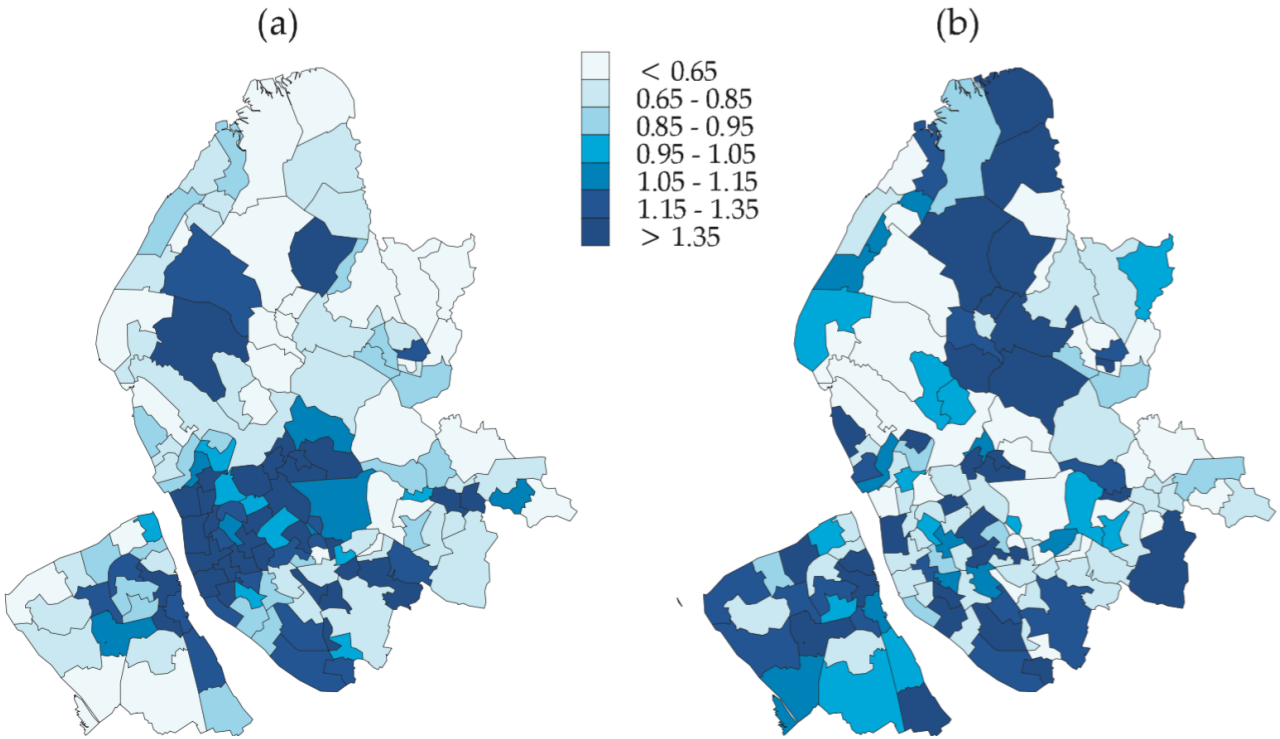


Figure: SIRs for (a) lung cancer, and (b) brain cancer in the North-West of England.

COPD in England

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).

Example: COPD in England

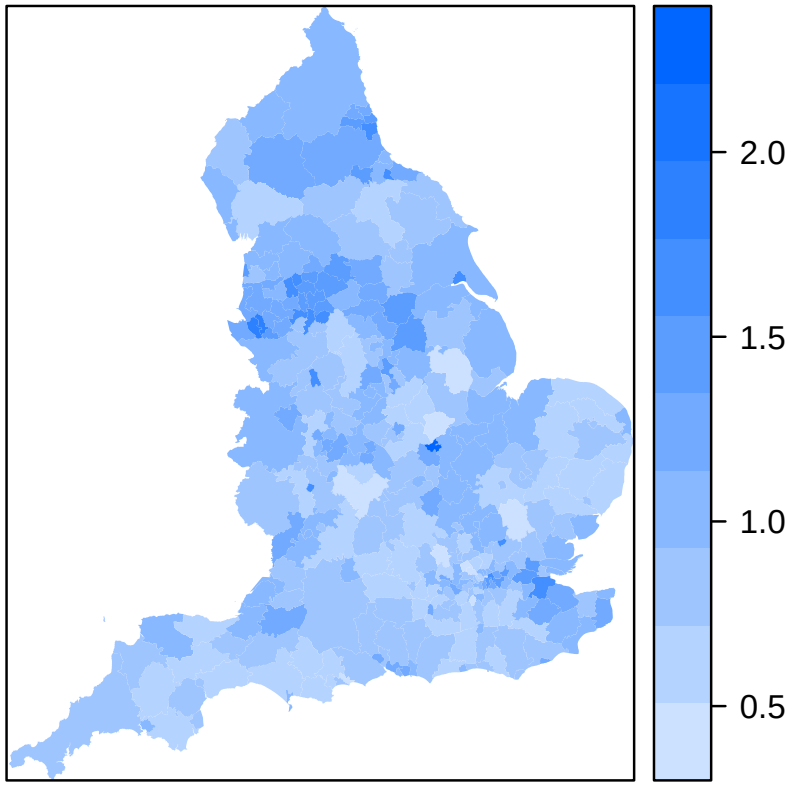


Figure: SMRs for hospital admissions of COPD in 324 local authorities in England.

Data Quality Issues

Data Quality Issues

- In routinely carried out investigations the constituent data are often subject to errors.
- *Population data*
  - Population registers are the gold standard but counts from the census are those that are typically routinely-available.
  - Census counts should be treated as estimates, however, since inaccuracies, in particular underenumeration, are common.
  - For inter-censual years, as well as births and deaths, migration must also be considered.
  - The geography, that is, the geographical areas of the study variables, may also change across censuses which causes complications.

Data Quality Issues

- In routinely carried out investigations the constituent data are often subject to errors.
- *Health data*
  - For any health event there is always the possibility of diagnostic error or misclassification.
  - For other events such as cancers, case registrations may be subject to double counting and under registration.
- In both instances *local knowledge* is invaluable.

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