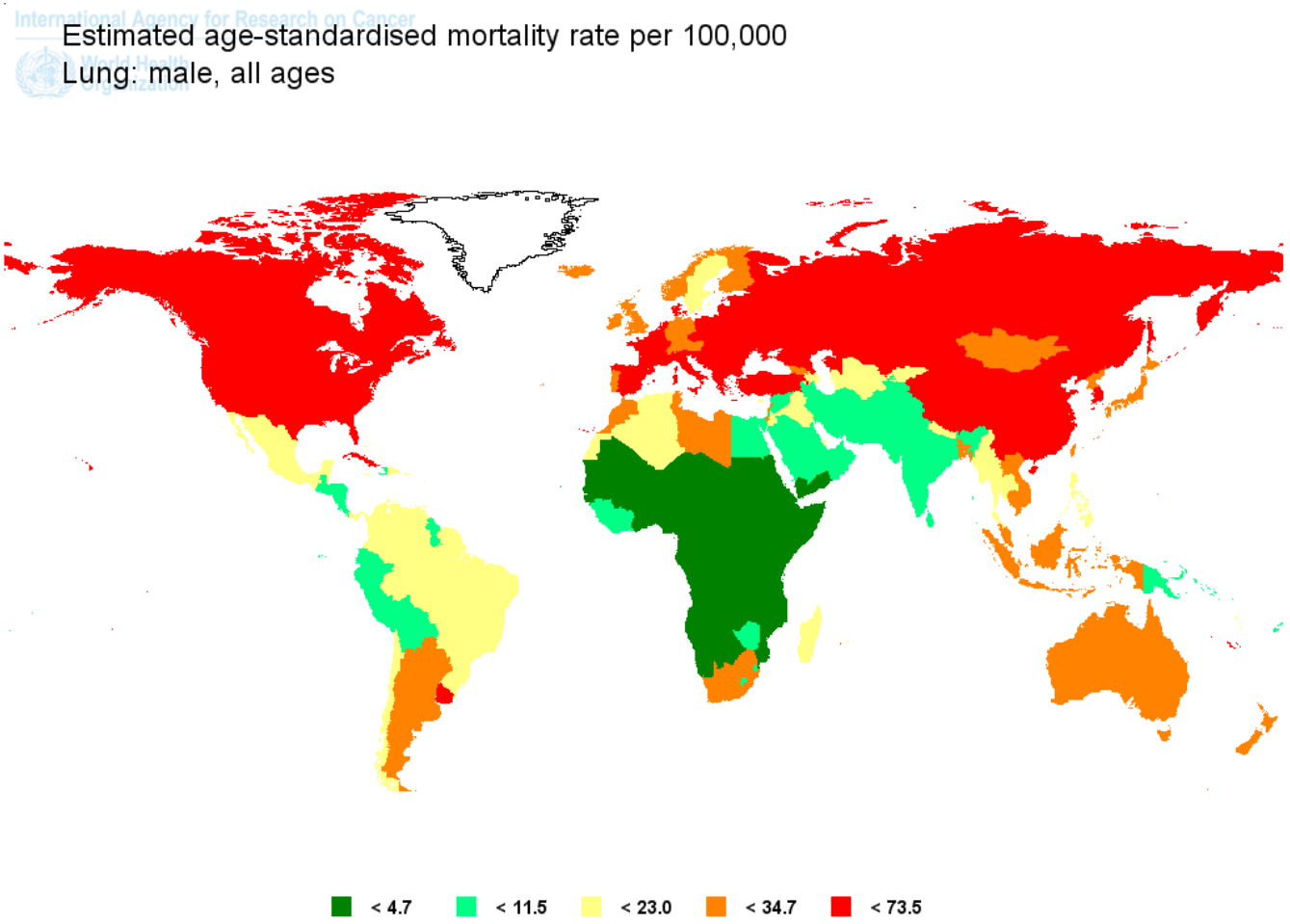
Small area studies

Lea Fortunato

Small areas studies

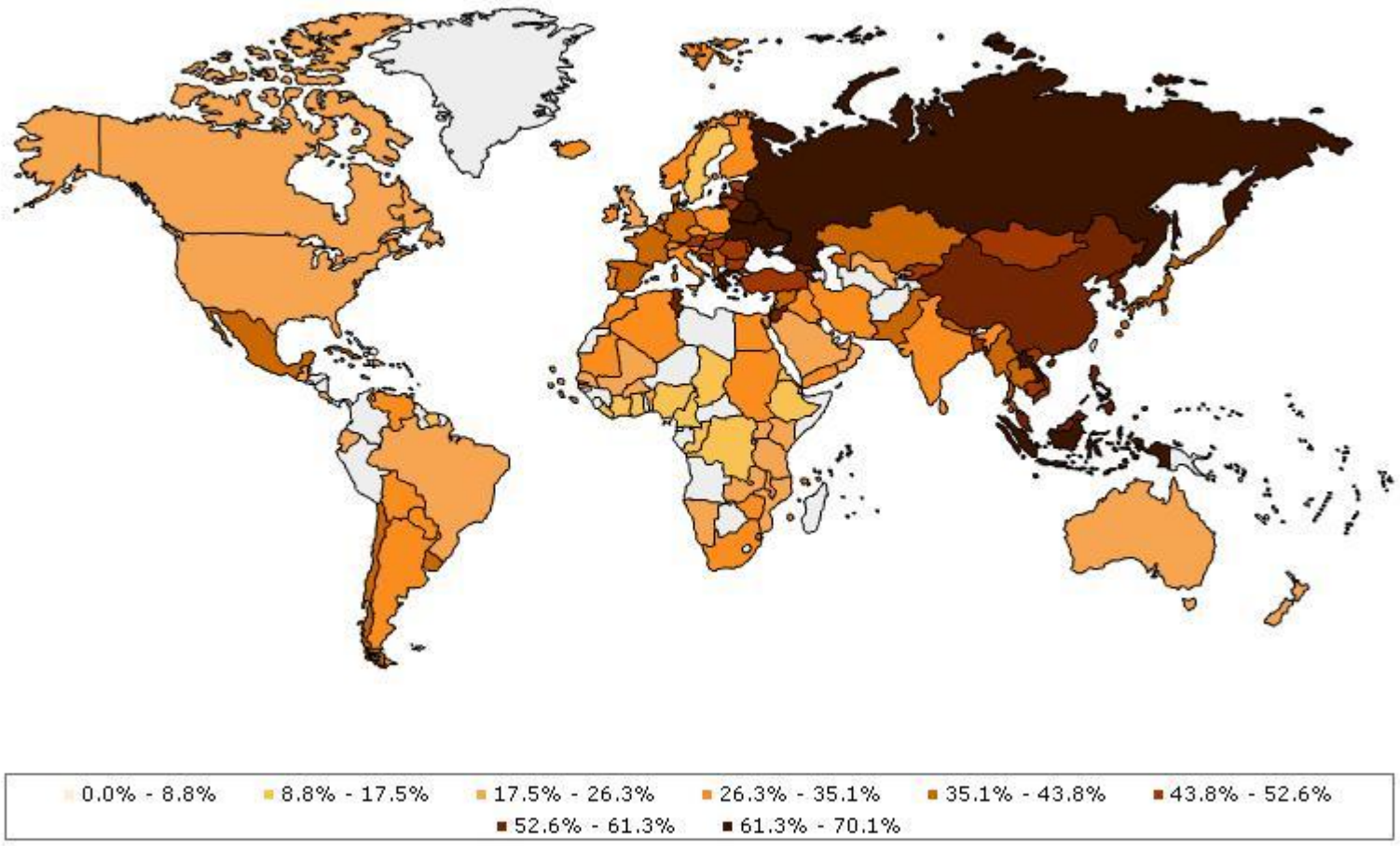
- Geographical variations of chronic diseases have been historically used as important sources of knowledge in epidemiology
  - Highlight sources of heterogeneity and spatial patterns
  - Suggest public health determinants or aetiological clues
  - e.g. international and national atlases
- No individual data
- Use of routinely collected data: registries (national and regional), surveys, censuses,...
  - counts and population by area, age group and gender
  - easy to obtain
  - greatly improved in quality

Lung cancer mortality worldwide in 2008 - males
(IARC)



Large difference between developed and developing countries
Pattern explained by differences in the prevalence of smoking
Source: http://globocan.iarc.fr/

Male Prevalence of Smoking (Percent of Adults) 2006



Source: WHO, World Health Statistics 2010: http://www.who.int/whosis/whostat/2010/en/index.htm

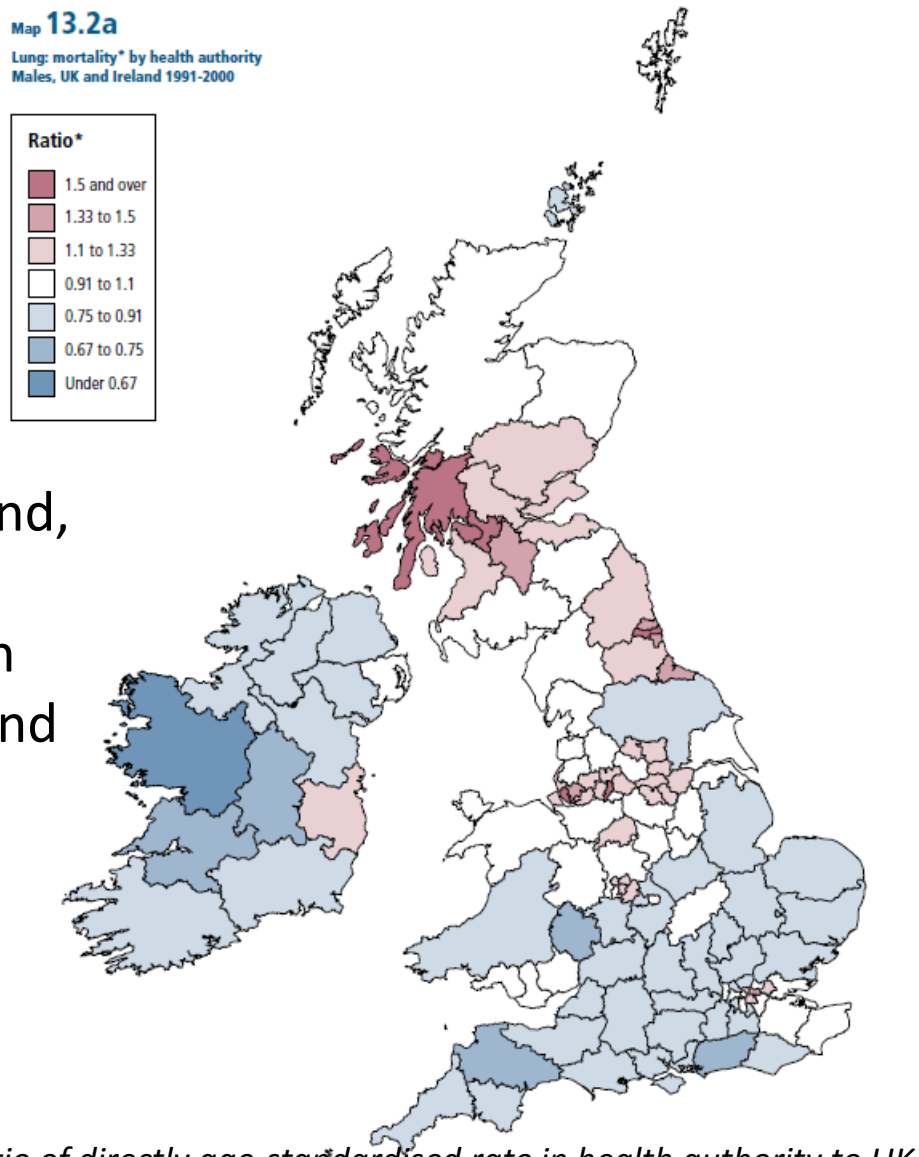
Implications for global health

- Smoking related cancers:
  - National policies for reducing smoking
  - WHO Tobacco Free Initiative (TFI): http://www.who.int/topics/tobacco/en/

- Liver and cervical cancers: prevent by vaccination (Jemal et al, 2011)
- Infant mortality and mothers' education (paper to discuss later)

=> Same analyses at smaller level

Lung cancer mortality in UK and Ireland (1991-2000)



- Higher mortality in northern England, and Scotland
- Lower mortality in Wales, southern and eastern England, Northern Ireland and Ireland

**Ratio of directly age-standardised rate in health authority to UK and Ireland average*

Small area studies in practice

- Relative risk estimated by SMR: for each area i (i=1,...N)

  $SMR_i = O_i/E_i$ = Standardised Mortality/Morbidity Ratio

  O_i Observed number of cases of disease D

  E_i Expected number of cases of disease D

- Calculation of the expected numbers of cases
  - Under null hypothesis that risk of contracting D is the same in area i as in a reference area:

$$E_i = N_i \; r$$

where N_i=population at risk in area i

r = risk of disease D in reference area (e.g. the whole country)

- published data (e.g. ONS)
- total number of cases/total population at risk

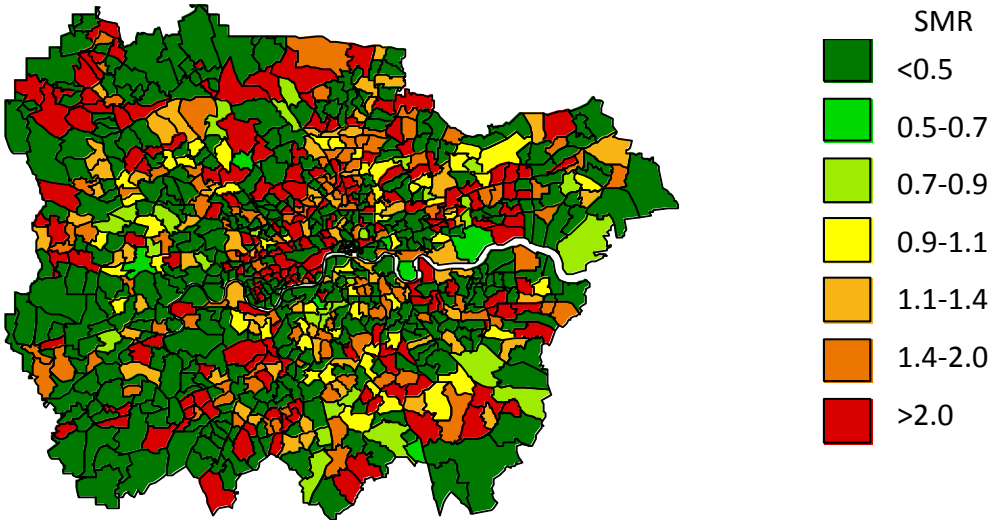
  - Adjustment to take into differences in the population
    - e.g. Mortality rates increase with age

$$E_{ij} = \sum_j N_{ij} \; r_j$$

where N_{ij} =population at risk in area i, strata j

r_j = mortality rate for age-sex strata j in reference area

Childhood leukaemia incidence in London 1985-1998



(from Best et al, JRSSA, 2001)

Issues

- Common practice is to map SMRs
  - SMR represents estimate of the true (underlying) risk in an area
  - Statistical uncertainty about estimate based on variance

$$\text{var}(SMR_i) \propto 1/E_i$$

  - $\Rightarrow$ SMR_i very imprecise for rare diseases and/or areas with small populations

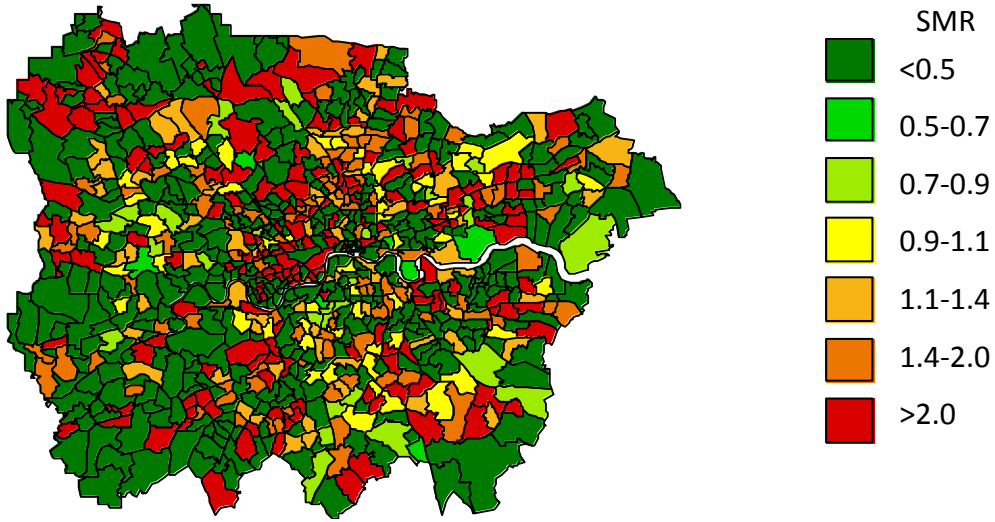
- SMR in each area is estimated independently
  - makes no use of risk estimates in other areas of the map, even though these are likely to be similar

$\Rightarrow$ Highlights extreme risk estimates based on small numbers

$\Rightarrow$ Ignores possible spatial correlation between disease risk in nearby areas due to possible dependence on spatially varying risk factors

Childhood leukaemia incidence in London

Map of SMR of childhood leukaemia



- Is the variability real or simply reflecting unequal E_is?
- Have the highlighted areas truly a raised relative risk?

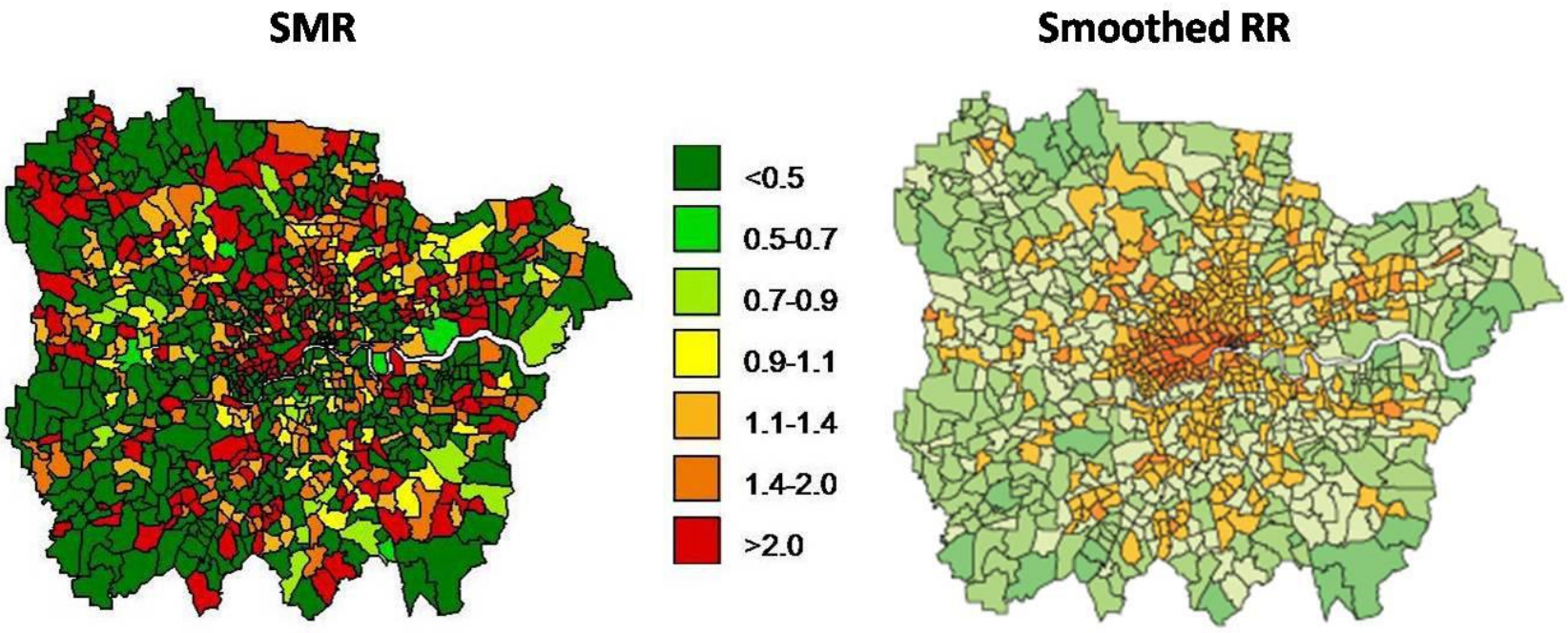
Smoothing disease maps

- These problems may be addressed by spatial smoothing of the raw data

- Idea is to “borrow information” for neighbouring areas to produce “better” estimates of the risk of each area (e.g. more stable, less noisy)

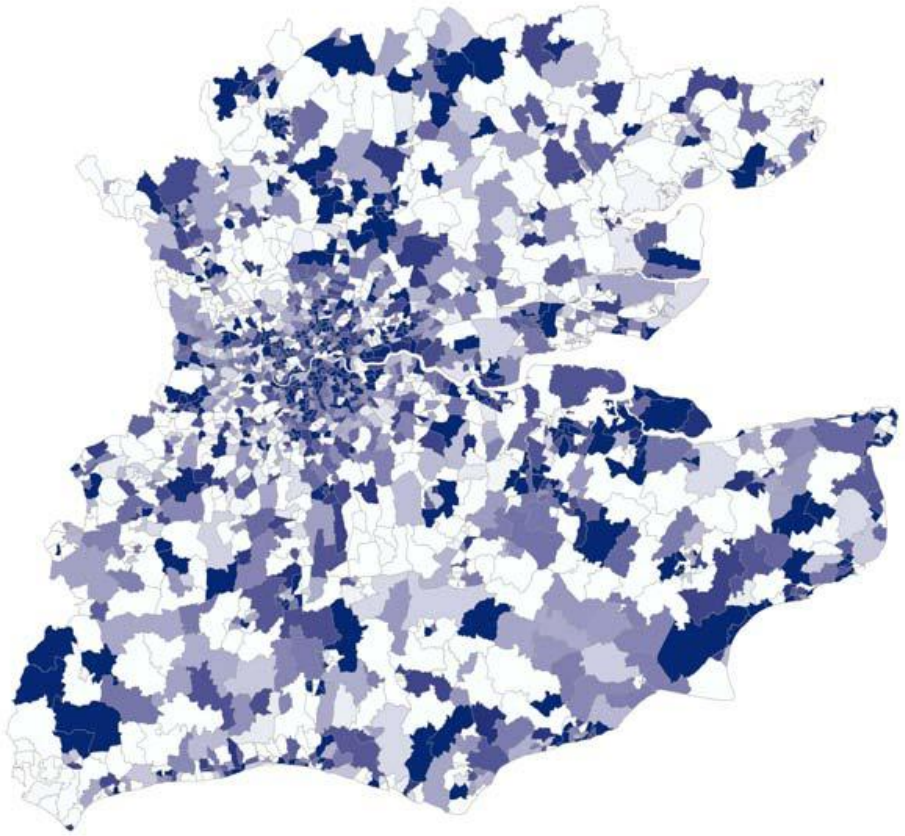
- Many methods available
  - Bayesian disease mapping models (course on spatial statistics)

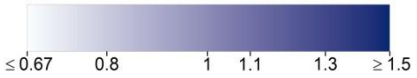
Childhood leukaemia incidence in London



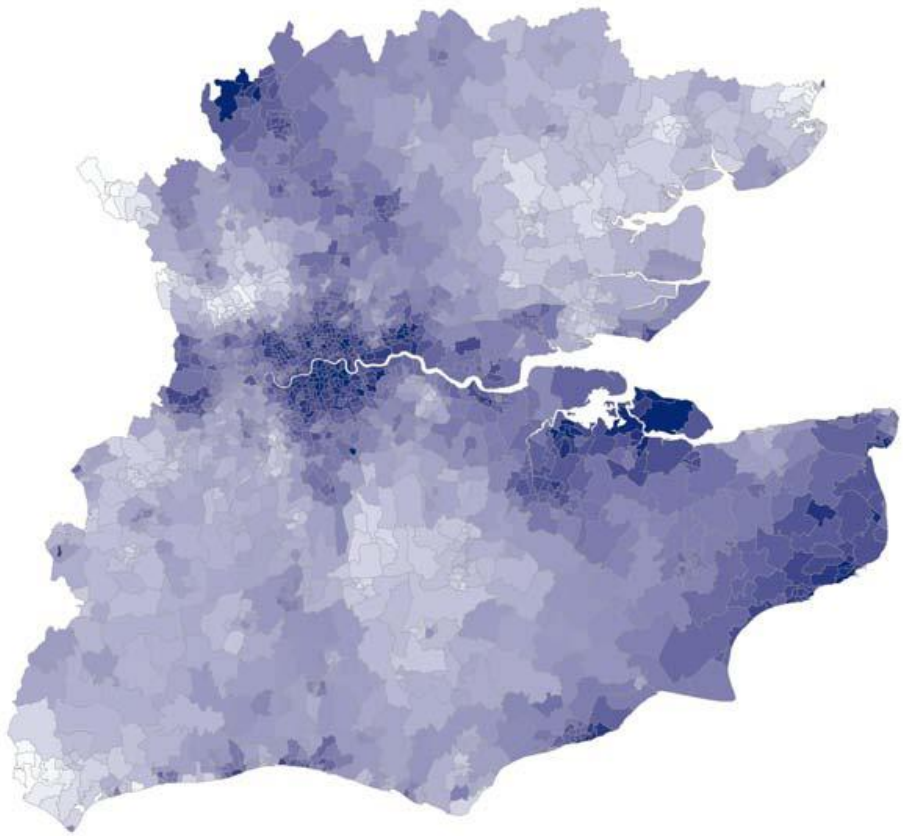
Cervix cancer incidence, South East England 1989 - 2003

SMR





Smoothed RR



Space-time analyses

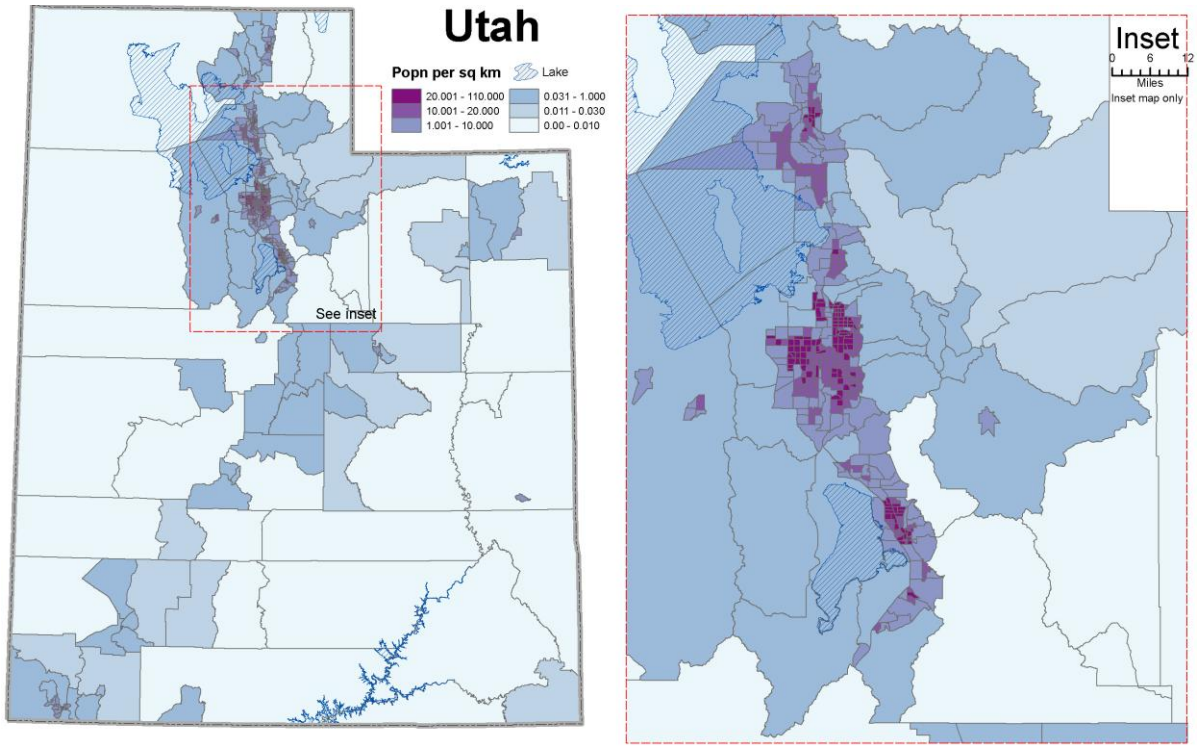
- Disease mapping is usually carried out on aggregated data over a time period
- Rather than suppressing the time dimension, it can be interesting to use models that combine the space and time dimension
- The stability (or not) of the spatial pattern can aid interpretation:
  - Stability interpreted as associated with stable risk factors, environmental effects, distribution of health care access
  - Unstability can pinpoint unusual/emerging hazards

$\Rightarrow$Increased epidemiological interpretability

$\Rightarrow$Potential tool for surveillance

Bladder cancer – Utah

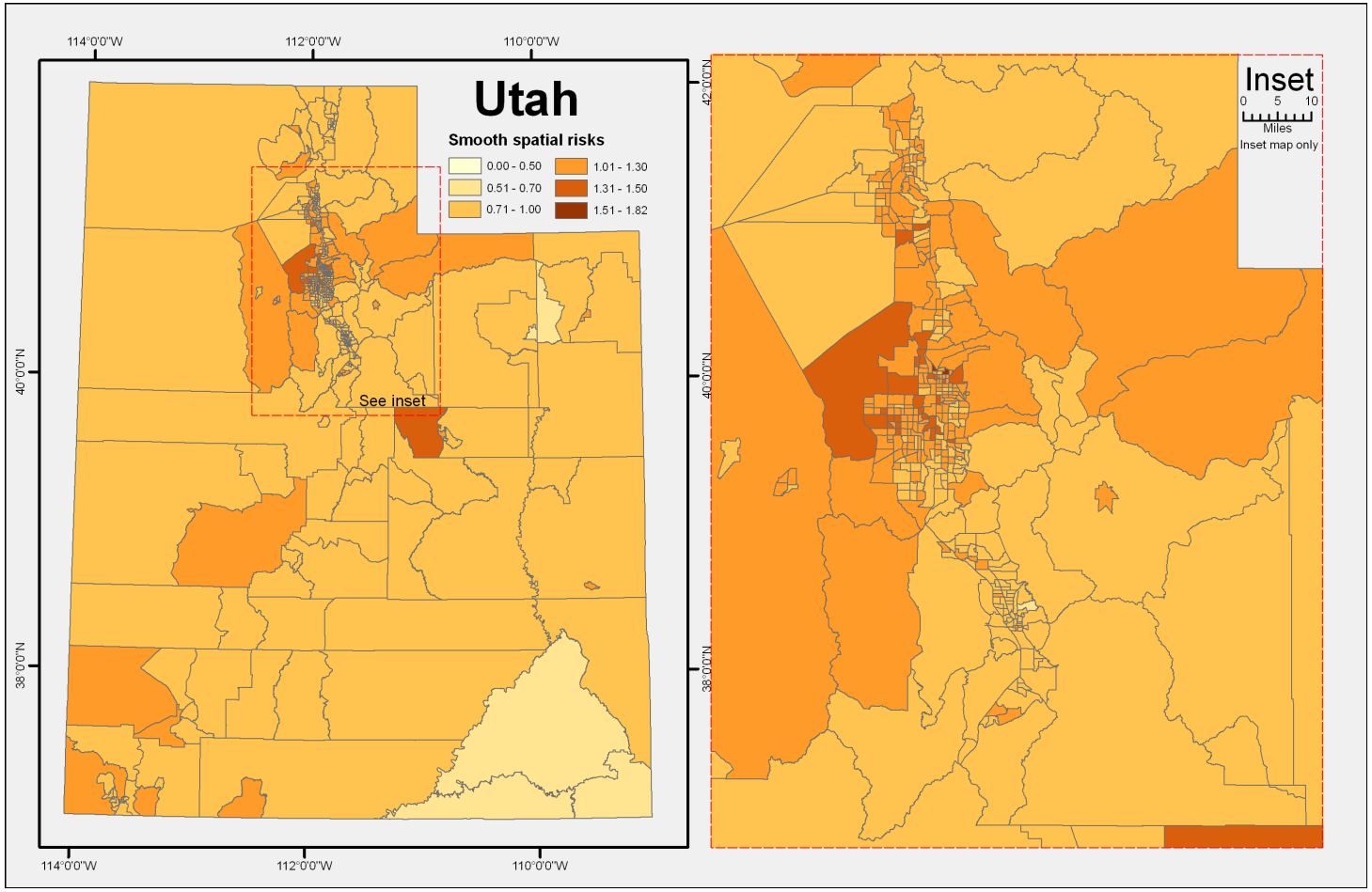
- Spatio-temporal variations of baldder cancer incidence, at census tract level (496 areas), 1973 – 2004
- Main risk factors: tobacco smoking and occupational exposure to aromatic amines
- Data from Utah registry and 2000 US census



Utah's population was mainly concentrated in the counties of Weber, Davis, Salt Lake and Utah.

Bladder cancer – Utah

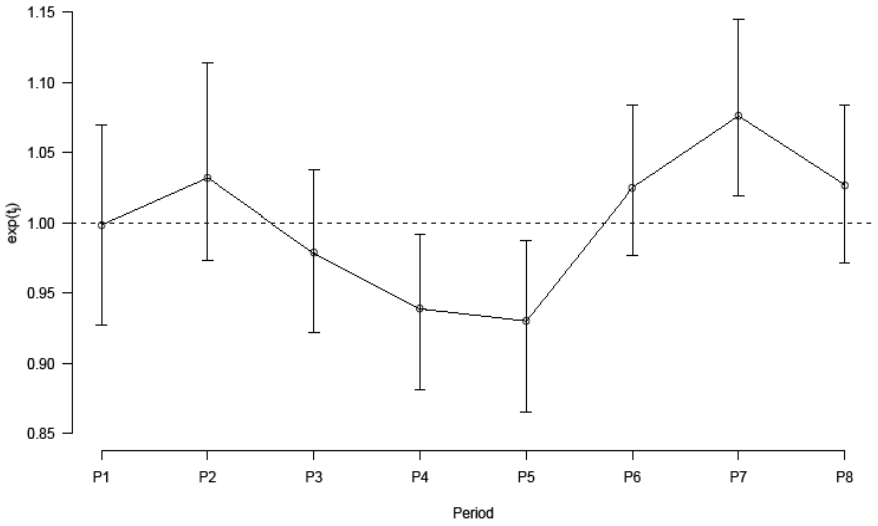
Smoothed RRs by census tract



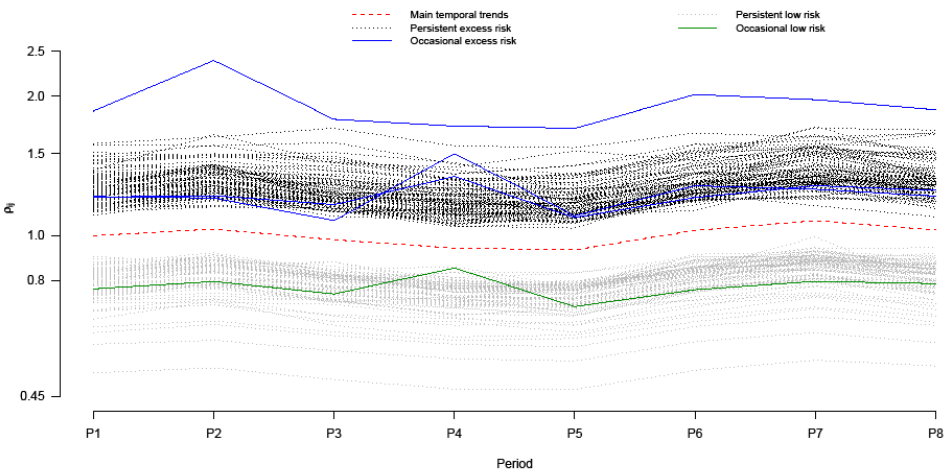
Spatial heterogeneity with higher risks in central areas around Salt Lake City

Bladder cancer – Utah

Smoothed temporal trends by period



Stable versus unusual patterns



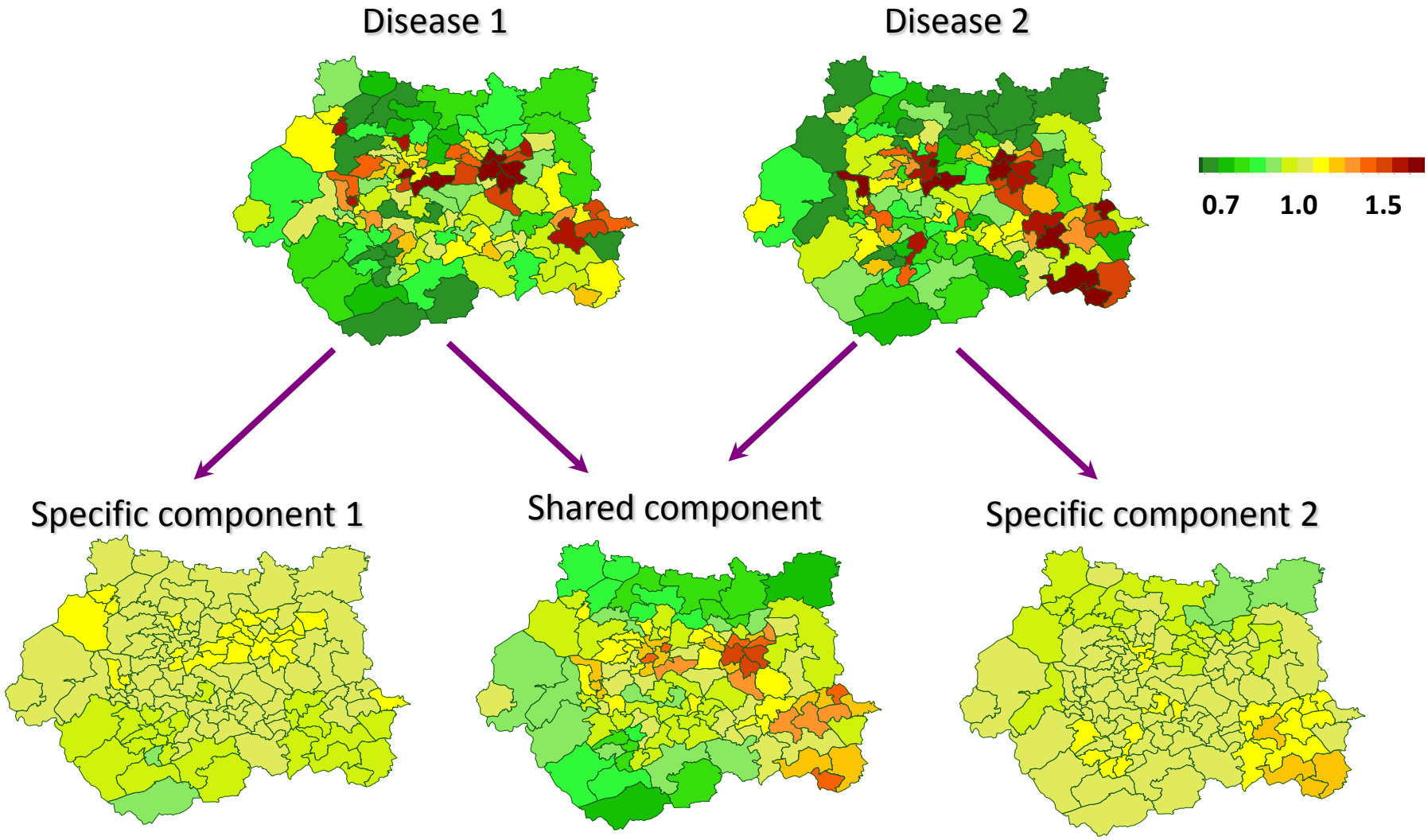
Slow but continuous decrease of risk of bladder cancer between the periods 77-80 (P2) and 89-92 (P5), followed by a steep increase in the period 1997-2000 (P7).

- 93 areas with sustained increased risk
- 81 with sustained low risk (grey lines)
- 3 high risk areas (blue)
- 1 low risk area (green) had unusual temporal patterns

Joint spatial variation in risk of multiple diseases

- Spatial modelling of disease risk almost exclusively for a single disease
- But: many diseases share common risk factors, e.g. Smoking
- Joint formulation seems appropriate
  - Improved precision of risk estimates
  - Greater aetiological insight by identification of geographical variations in shared and disease-specific risk

Joint analysis of 2 diseases

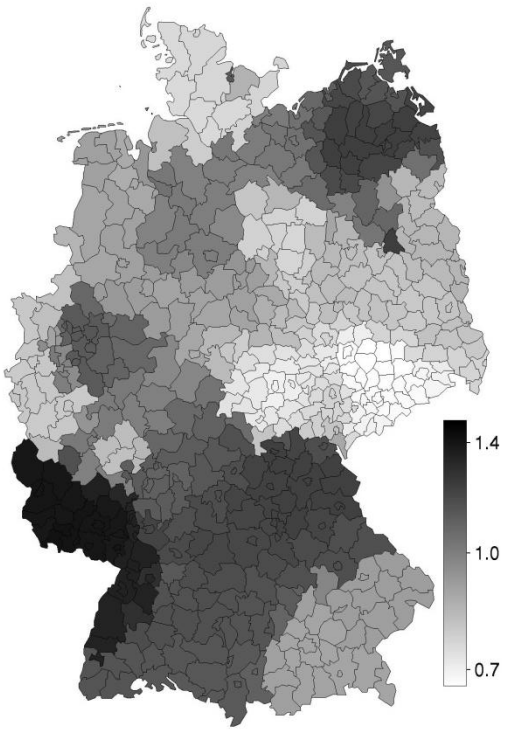


Knorr-Held and Best (2001)

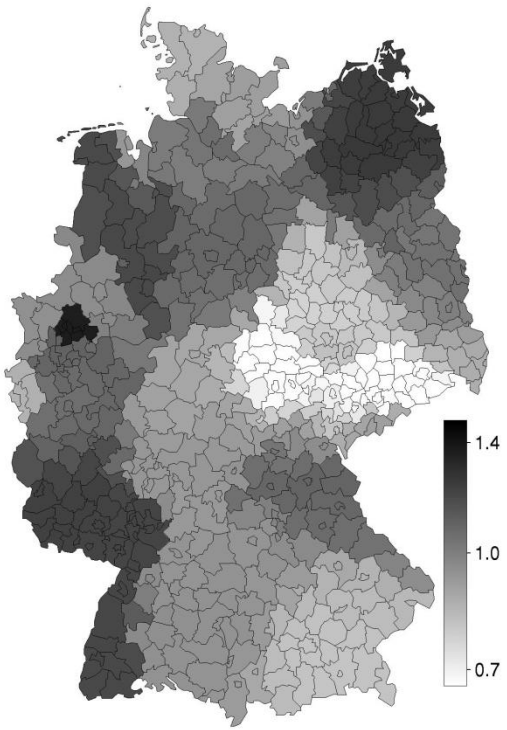
Oral cavity and oesophagus cancers

- Oral cavity and oesophagus cancer mortality, 544 districts in Germany, 1986-1990 (Knorr-Held and Best, 2001)
- Established risk factors: tobacco and alcohol
- Separate analysis: map of the smoothed RR

Oral cavity



Oesophagus

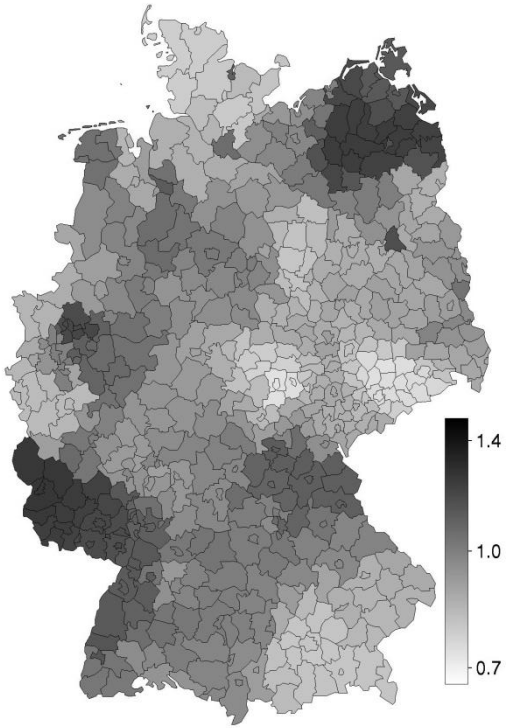


Spatial structure similar for both cancers:
High values in the North-East and South West

$\Rightarrow$ Joint analysis

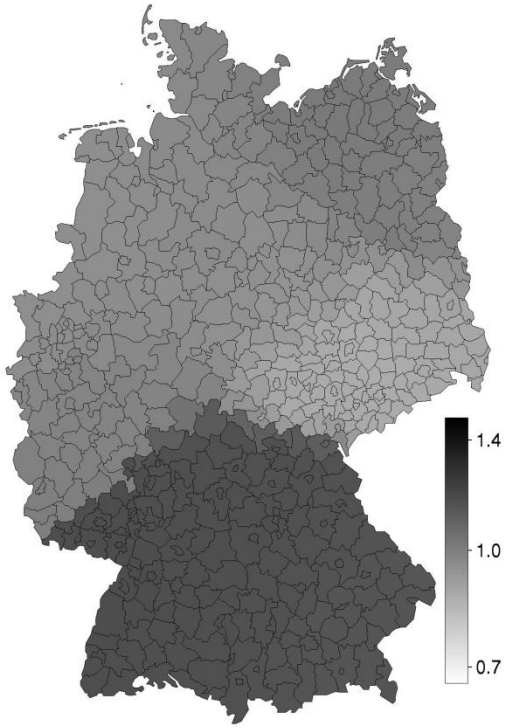
Oral cavity and oesophagus cancers

Shared RR



2 large clusters in the NE and SW (regions where alcohol or tobacco consumptions are high,respectively)

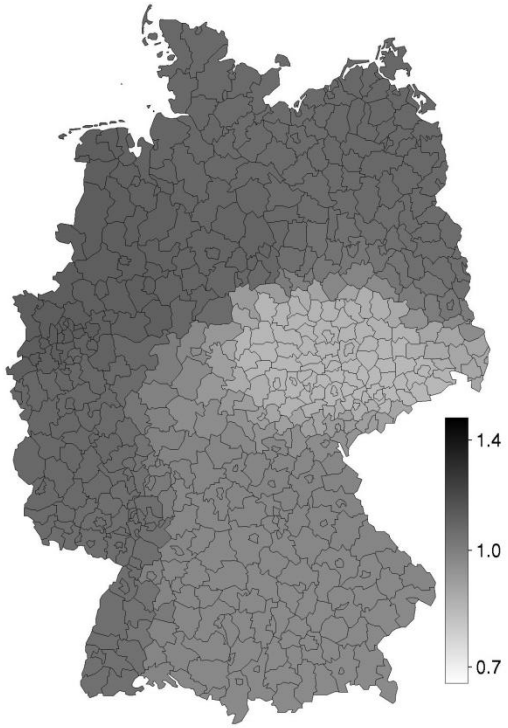
Oral cavity-specific RR



Clear spatial pattern: higher RRs in the S and lower RRs in the N

$\Rightarrow$ existence of additional risk factors that are relevant only to oral cavity but not to oesophageal cancer

Oesophageal-specific RR



Different spatial pattern with less variation and slightly higher RRs in the W and N.

Ecological regression studies

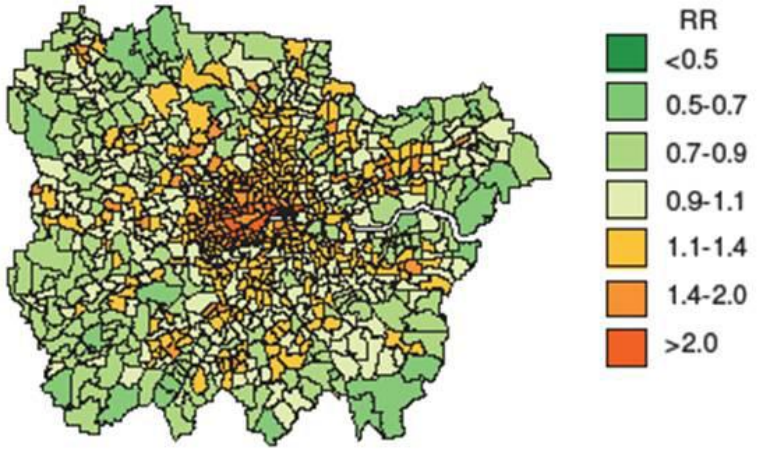
- Disease mapping:
  - Focus is on description
  - Level of inference is at the aggregate (small area) level
- Ecological regression studies:
  - Focus is on `explanation'
  - Used for investigating specific aetiological hypothesis at small-area scale
  - Typically aim to transfer level of inference from aggregate to individual level
- Idea is to regress area level measure of outcome (i.e. number of cases of disease) on possible explanatory variables also measured at the area level (e.g. mean pollution, mean income, proportion of population who smoke)
- Typically aim to transfer level of inference from aggregate to individual level

Ecological study of childhood leukaemia and Benzene exposure related to traffic

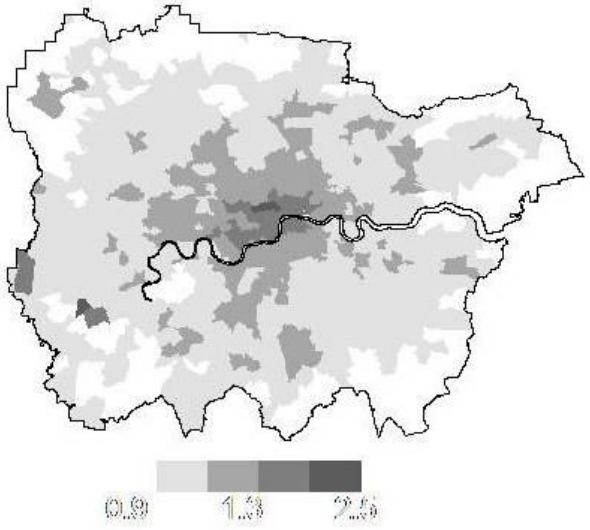
- Benzene is a recognized carcinogen at moderate to high doses (IARC, 1981)
- Recent concerns that low dose exposure to environmental benzene may increase risk of leukaemia in vulnerable groups, e.g. Children
- Geographical study by Best et al (2001):
  - 872 wards in London
  - 734 leukaemia cases in children 0-15 yrs, 1985-96
  - Benzene emissions from outdoor sources on 1km grid, aggregated at ward level

Childhood leukaemia incidence and benzene

Smoothed RR



Cube root benzene emissions (tonnes/year)



RR of leukaemia associated with a unit increase in cube root benzene emissions in the area of residence

= 2.23, 95% interval = (1.64, 2.96)

Important issues of interpretation

- Measurement errors may occur in:
  - Exposure: Inaccurate sampling of contaminant concentrations, use of proxy measures for true exposure
  - Disease counts: Missing cases / duplicates in cancer registries, differential use of ICD codes
  - Population at risk: Census underenumeration
- Confounding:
  - Area-level socio-economic deprivation is the major confounder because correlated with disease and coincides with e.g. Industrial sites, busy roads, smoking
- Ecological bias: difference between individual and group-level estimate of disease risk

Summary

- Smoothing of small area risks is important to help to separate spatial pattern from “noise”
- Achieved by borrowing information from neighbouring regions
- Many methods available in the literature
- Natural extensions to
  - Joint mapping of 2 or more diseases
  - Joint modelling of spatial and temporal variation
  - Adjustments for covariates
- Methods allowing to deal with some issues, e.g. uncertainty, ecological bias

Mothers' education and child mortality

Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis.

Gakidou E, Cowling K, Lozano R, Murray CJ. Lancet (2010)

- Education is a major determinant of health
- Strong association between mothers' education and child mortality: use of health services, economic advantages, reductions in fertility...
- Aims:
  - estimate the time series of the mean number of years of education
  - Investigate the association between child mortality and education

- Data:
  - 915 censuses and national surveys of respondents' educational attainment between 1958 and 2008 for 175 countries
  - Estimates of child mortality from Rajaratnam et al (2010)

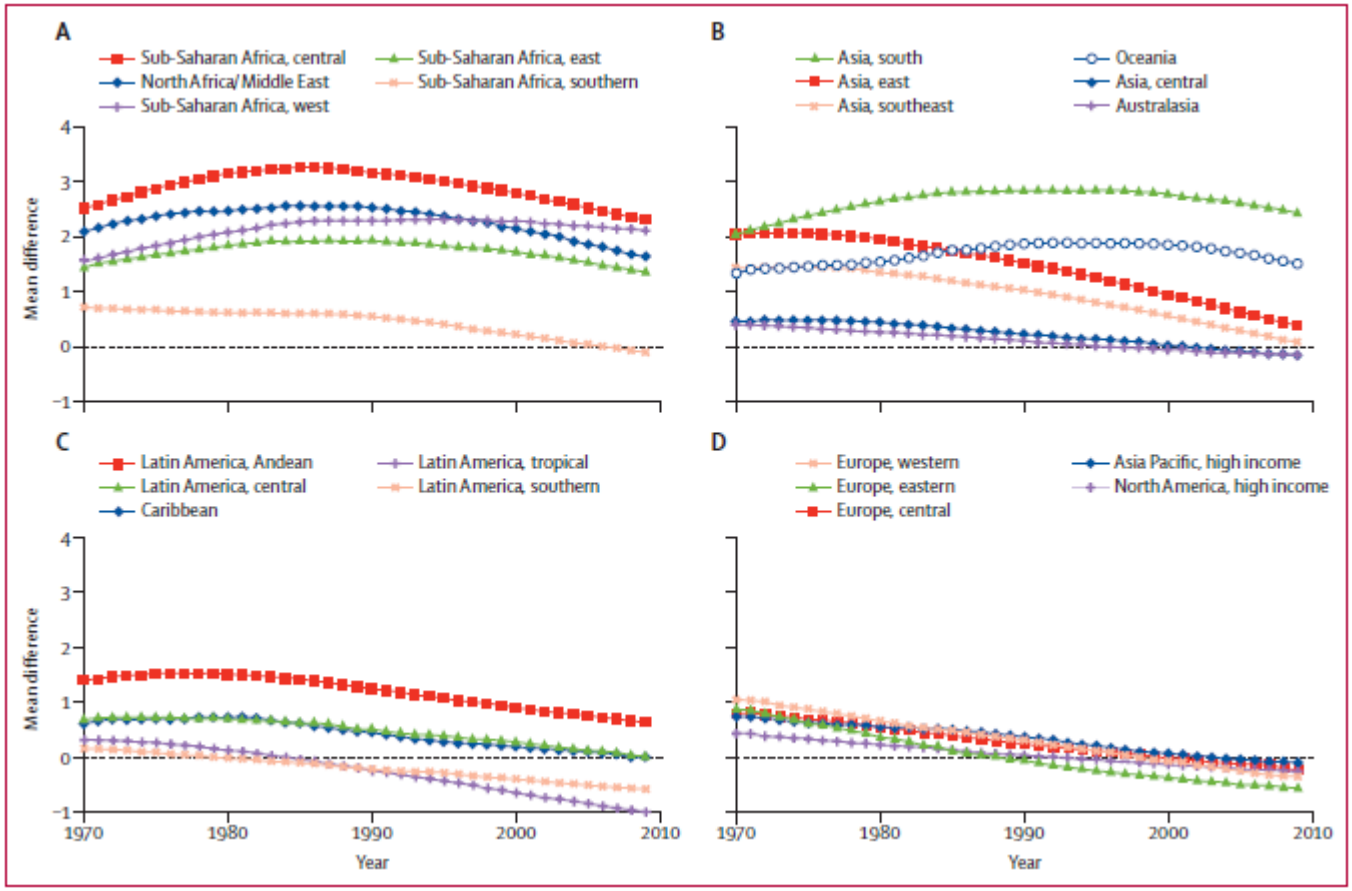
Difference in mean years of education between men and women aged 25–34 years with time for 21 regions

Africa

Asia

Latin America

High-income countries



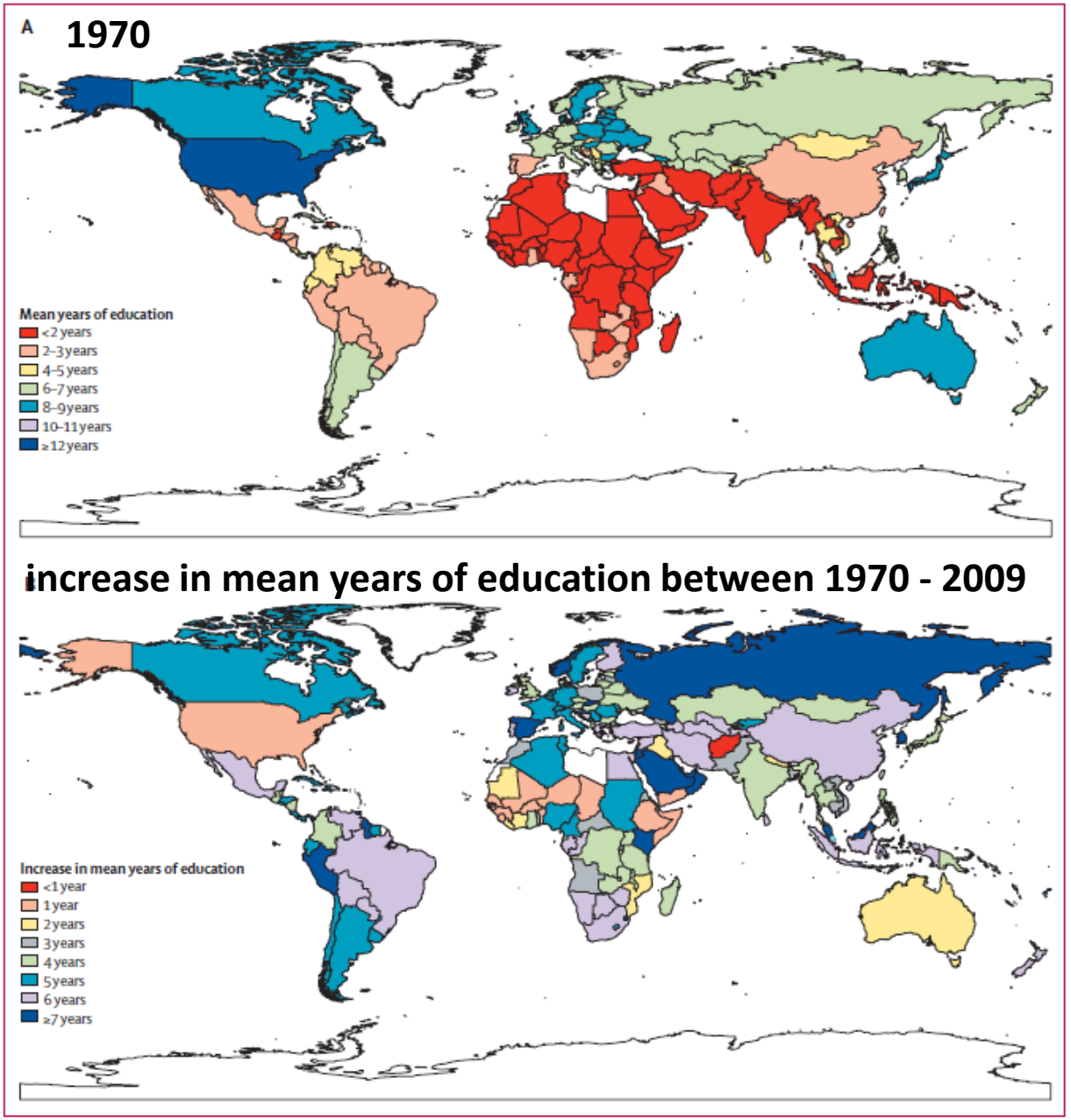
Comment the graphs:

- Difference between men and women
- Temporal trends
- Geographical differences

Mean years of education in women aged 25–34

How could you interpret these maps?

What else should you take into account?



Association between child mortality and education

- Coefficient associated with education = -0.1, se=0.007, p-value<0.0001

How do you interpret this coefficient?

- “Of 8.2 million fewer deaths in children younger than 5 years between 1970 and 2009, we estimated that 4.2 million (51.2%) could be attributed to increased educational attainment in women of reproductive age”

As policy maker, what would you do?

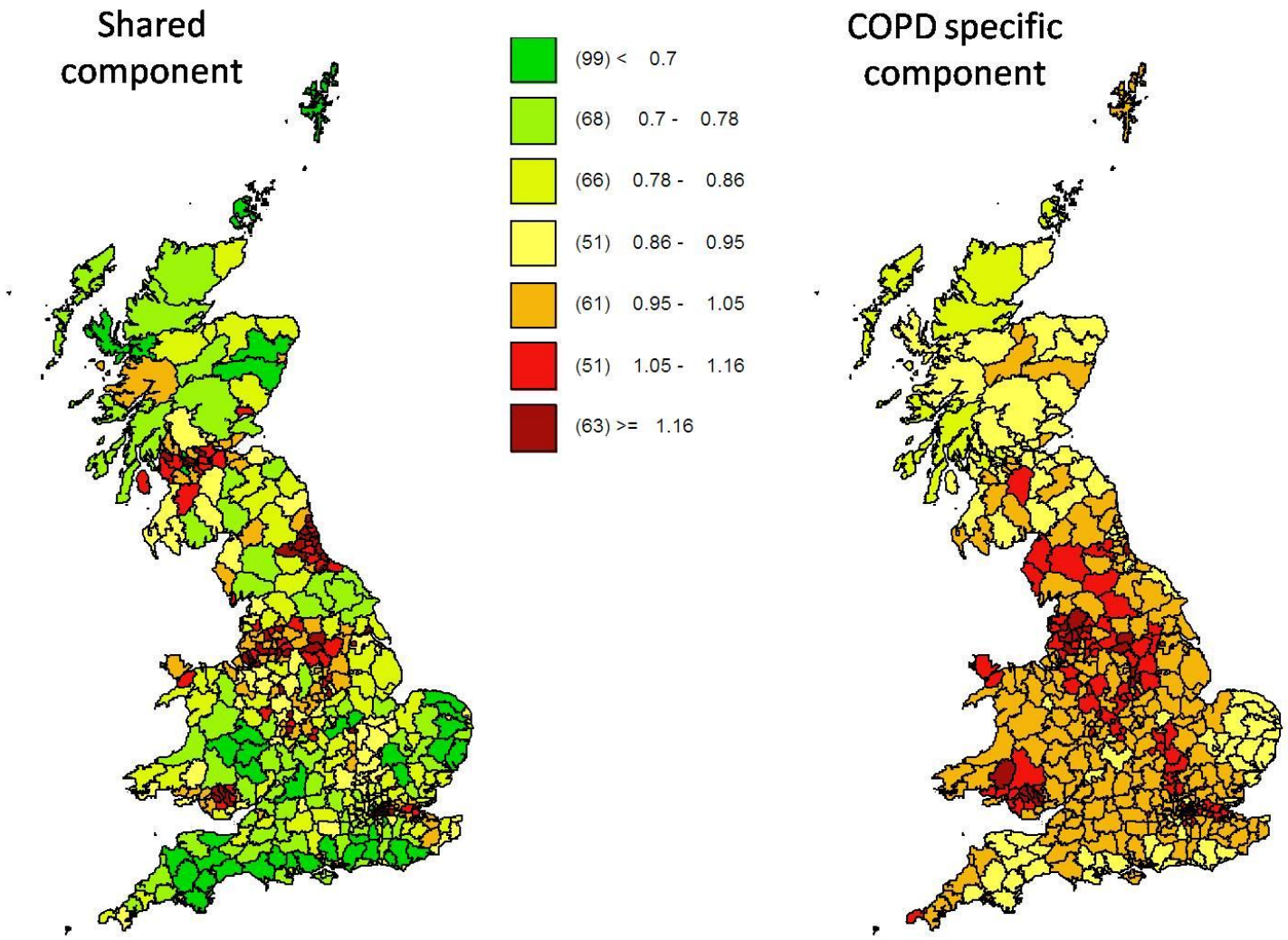
Joint analysis of COPD and lung cancer

Geographic Variations in Risk: Adjusting for Unmeasured Confounders Through Joint Modeling of Multiple Diseases.

Best N, Hansell AL. *Epidemiology 2009*

- Lung cancer and Chronic obstructive pulmonary disease (COPD) are leading causes of death in England and Wales
- Smoking is major risk factor for both diseases
- Interest in non-smoking related risk factors for COPD (particularly air pollution)
- Aim: carry out joint spatial analysis of COPD and lung cancer
  - Shared spatial effects primarily reflect smoking
  - Interest in spatial pattern of residual (non-smoking related) risk for COPD
- Data: Deaths from COPD and lung cancer in males over 45, 1981-1999, by district in GB

Joint analysis of COPD and lung cancer



Comment the 2 maps