Big Data and Socio-Economic Mortality

Andrew J.G. Cairns

Heriot-Watt University, Edinburgh

Director, Actuarial Research Centre, IFoA

Joint work with J.Wen, T.Kleinow and G.Peters

Pensions Conference, Bristol
19 June 2019

Actuarial Research Centre

Institute and Faculty of Actuaries

Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries (IFoA). The IFoA does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation.

The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the IFoA.

Actuarial
Research Centre
Institute and Faculty
of Actuaries

The Actuarial Research Centre (ARC)

A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world. The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.

The 'Modelling, Measurement and Management of Longevity and Morbidity Risk' research programme is being funded by the ARC, the SoA and the CIA.

www.actuaries.org.uk/arc

SOCIETY OF ACTUARIES

Canadian Institute of Actuaries | Institut canadien des actuaires

Actuarial
Research Centre
Institute and Faculty
of Actuaries

Advert: Data Science MIG

Aims: To be a source of Data Science expertise, to champion the use of Data Science techniques in traditional and emerging actuarial fields, and to encourage collaboration and understanding of wider issues in Data Science; with a view to promoting debate, research and member awareness, and to assist the IFoA in reaching the objective of having members of the actuarial profession combine their business skills with a deep understanding of Data Science.

Workstreams:

- education
- regulation and ethics
- membership
- collaboration
- research

Actuarial Research Centre
Institute and Faculty of Actuaries

Outline

- Background
- Data – England male & female mortality
- Methodology
- Results
  - predictive variables
  - residual spatial variation
- Ongoing work

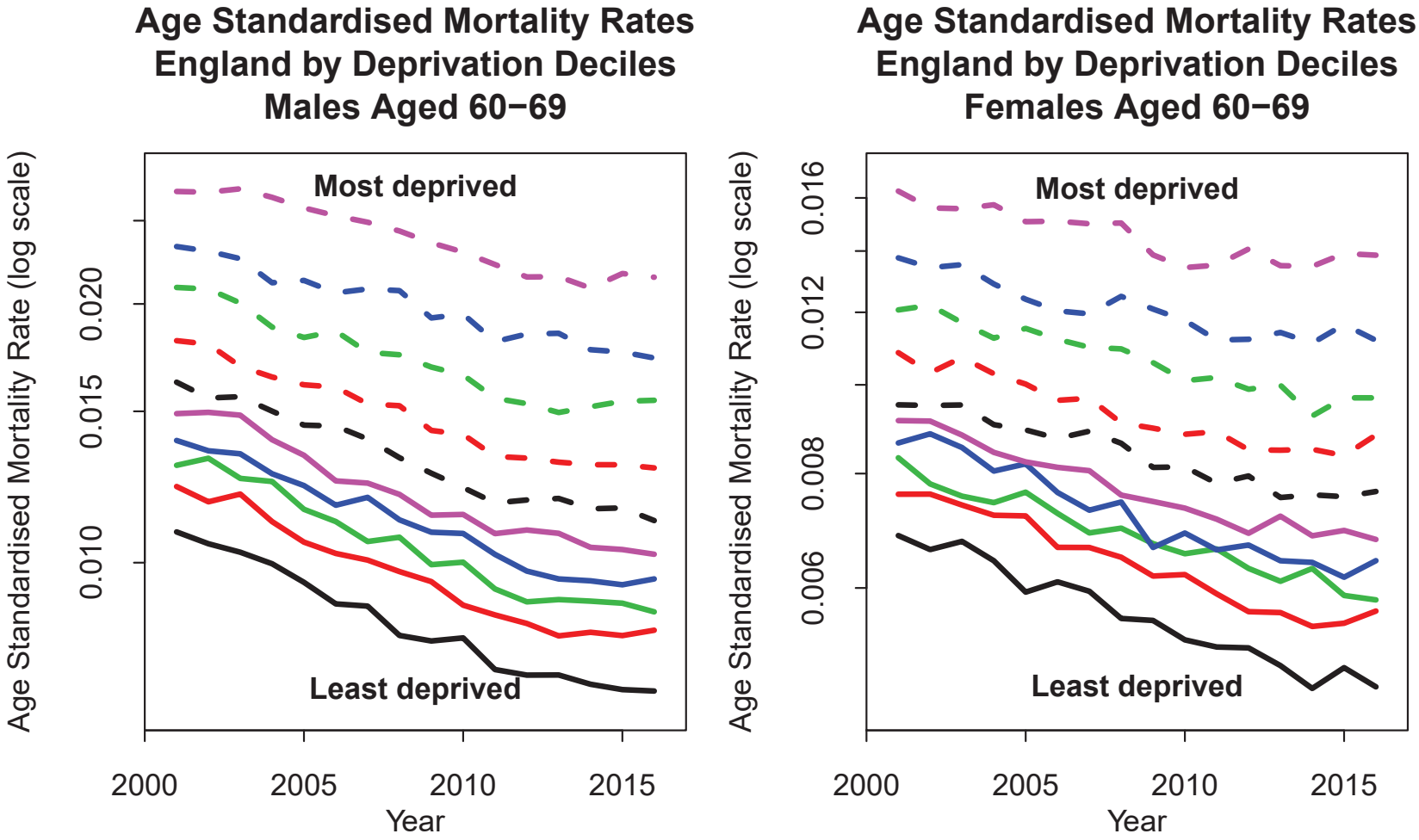
Background

- Considering here:
  male mortality in England
  (results for females similar and consistent)
- Stylised facts:
  - Mortality varies by socio-economic group
  - Mortality varies by region

Actuarial Research Centre
Institute and Faculty of Actuaries

Socio-Economic Differences in Mortality: England

England: mortality by *deprivation*



Background: Variation By Region

North East
North West
Yorkshire & Humber
East Midlands
West Midlands
East of England
London
South East
South West

Not in dataset:
Scotland, Wales,
Northern Ireland

Background: Relative mortality by region

England Variation by region (males 60-69)	
North East	118%
North West	116%
Yorkshire and The Humber	107%
East Midlands	98%
West Midlands	105%
East	88%
London	105%
South East	89%
South West	87%

Values show standardised mortality (ages 60-69) by region as a percentage of national standardised mortality

Regional variation $<$ variation by income deprivation

Background

- Mortality varies by socio-economic group
- Mortality in the north and in big cities is higher than mortality elsewhere
- *How much of this can be explained by underlying socio-economic differences?*
- *And how much variation is geographical?*

Actuarial Research Centre
Institute and Faculty of Actuaries

Data: LSOA's

- England only
- Lower Layer Super Output Areas: LSOA's
- $L = 32,844$ small geographical areas
- Socio-economically homogeneous
- Average size $\approx$ 1600 persons
- LSOA's $i = 1, \ldots, L$,
  single years (t =2001-2016), single ages, x:
  - Deaths: $D(i, t, x)$
  - Exposures: $E(i, t, x)$ (population)
- Plus many *static* predictive variables for each LSOA

Predictive variables by LSOA

- Indices of deprivation (2015) (single scores per LSOA)
  - income deprivation (benefits)
  - employment deprivation (unemployment)
  - education deprivation
  - crime
  - barriers to housing and services
    - geographical barriers (distance to services)
    - wider barriers (overcrowding; homelessness)
  - living environment (housing quality; unmodernised; air quality)
- Educational attainment (levels $\times$ age groups)
- Occupation groups (types $\times$ age groups)
- Average weekly income
- Average number of bedrooms
- # people in care homes with/without nursing
- Urban/rural classification (categorical)
- ....

Methodology

- $D(i,t,x)$, $E(i,t,x)$ deaths and exposures by LSOA
- National death rates (all t and x)

$$m(t,x) = \frac{\sum_{i=1}^{L} D(i,t,x)}{\sum_{i=1}^{L} E(i,t,x)}$$

- LSOA's ($i = 1, \ldots, L$) local death rates: $m(i,t,x)$

  General Model: $D(i,t,x) \sim \text{Poisson}\Big(m(i,t,x)E(i,t,x)\Big)$

Methodology (cont.)

General approach:

- Over a limited age range (e.g. 60-69); and
- Over a (potentially) limited range of years:

$$m(i,t,x) = m(t,x)F_1(i)F_2(i)$$

- $F_1(i)$ = relative risk due to socio-economic characteristics
  - GLM
  - kernel smoothing
  - local, weighted linear regression

- $F_2(i)$ = additional relative risk capturing spatial effects
  - kernel smoothing

Methodology (cont.)

- Years: $t = t_0, \ldots, t_1$
- Ages: $x = x_0, \ldots, x_1$
- Actual deaths by LSOA

$$D(i) = \sum_{t=t_0}^{t_1} \sum_{x=x_0}^{x_1} D(i,t,x)$$

- Expected deaths by LSOA (no modelled effects)

$$\hat{D}_0(i) = \sum_{t=t_0}^{t_1} \sum_{x=x_0}^{x_1} m(t,x) E(i,t,x)$$

- Actual-over-expected by LSOA

$$R_0(i) = D(i)/\hat{D}_0(i)$$

Actuarial Research Centre
Institute and Faculty of Actuaries

Stage 1: Introduce Predictive Variables

- LSOA's: $i = 1, \ldots, L$
- Predictive variables (PV): $j = 1, \ldots, n_P$
- $P(i,j) =$ unadjusted PV
- Different PVs are on different scales
  (e.g. $[0,1]$, $[0,100]$, $(-\infty, +\infty)$)
- Hence: standardise each PV

$$P(i,j) \longrightarrow X(i,j) \sim N(0,1)$$

- Purpose of standardisation:
  Simplifies the system of weighting later in Stage 1

- Vector: $X(i) = (X(i,1), \ldots, X(i,n_P))'$

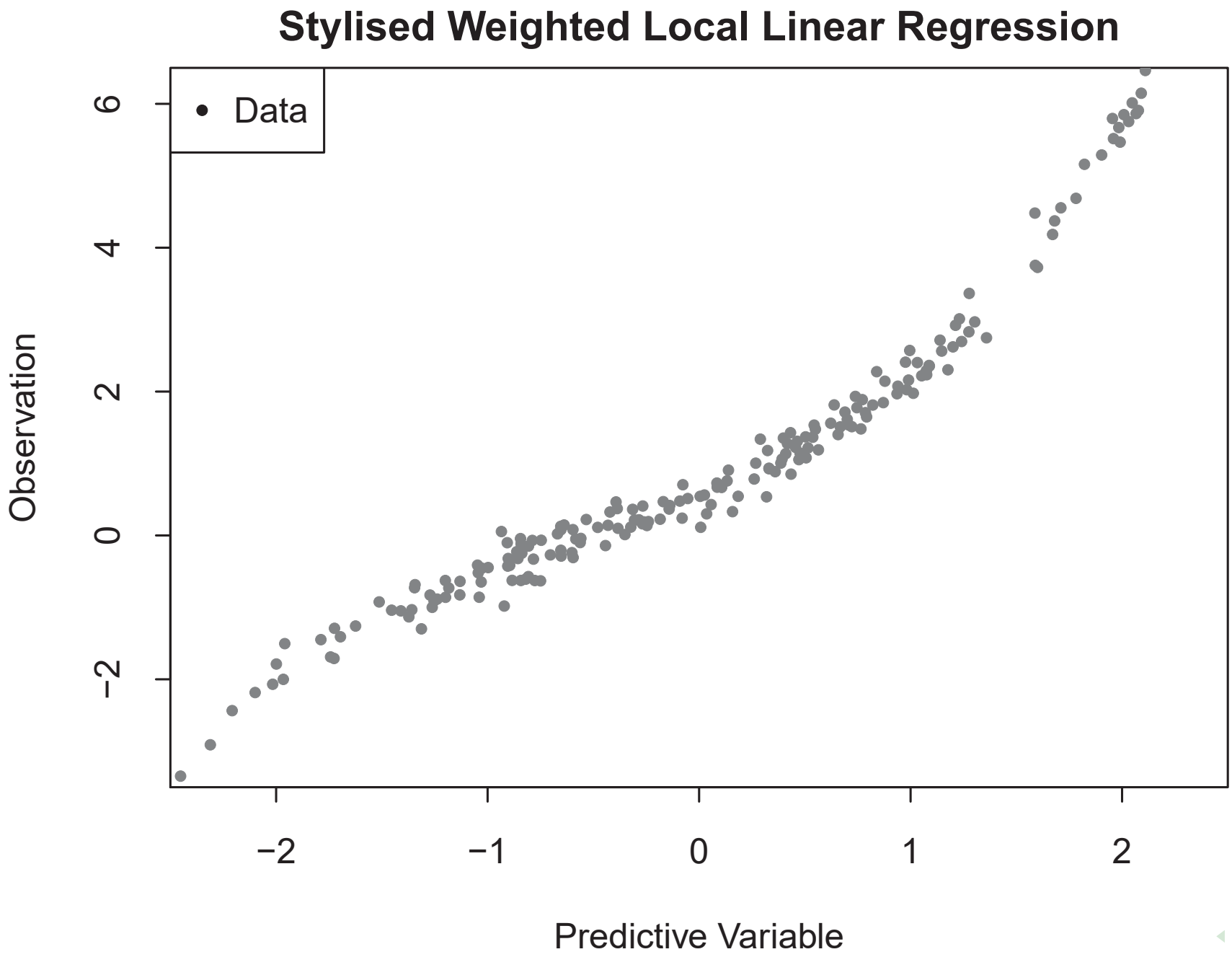
Actuarial Research Centre
Institute and Faculty of Actuaries

Stage 1: Urban versus Rural

- Urban-rural classification
  - 1: Conurbation; London (4810 LSOA's)
  - 2: Conurbation: not London (7921)
  - 3: City or town (14515)
  - 4: Rural town (3056)
  - 5: Rural village and dispersed (2542)
- Preliminary experiments $\Rightarrow$
  contribution and importance of specific predictive variables varies significantly between urban and rural LSOA's
- Hence: incorporate urban/rural classification into the process.

Actuarial Research Centre
Institute and Faculty of Actuaries

Stage 1: Local Linear Regression



Stage 1: Local Linear Regression

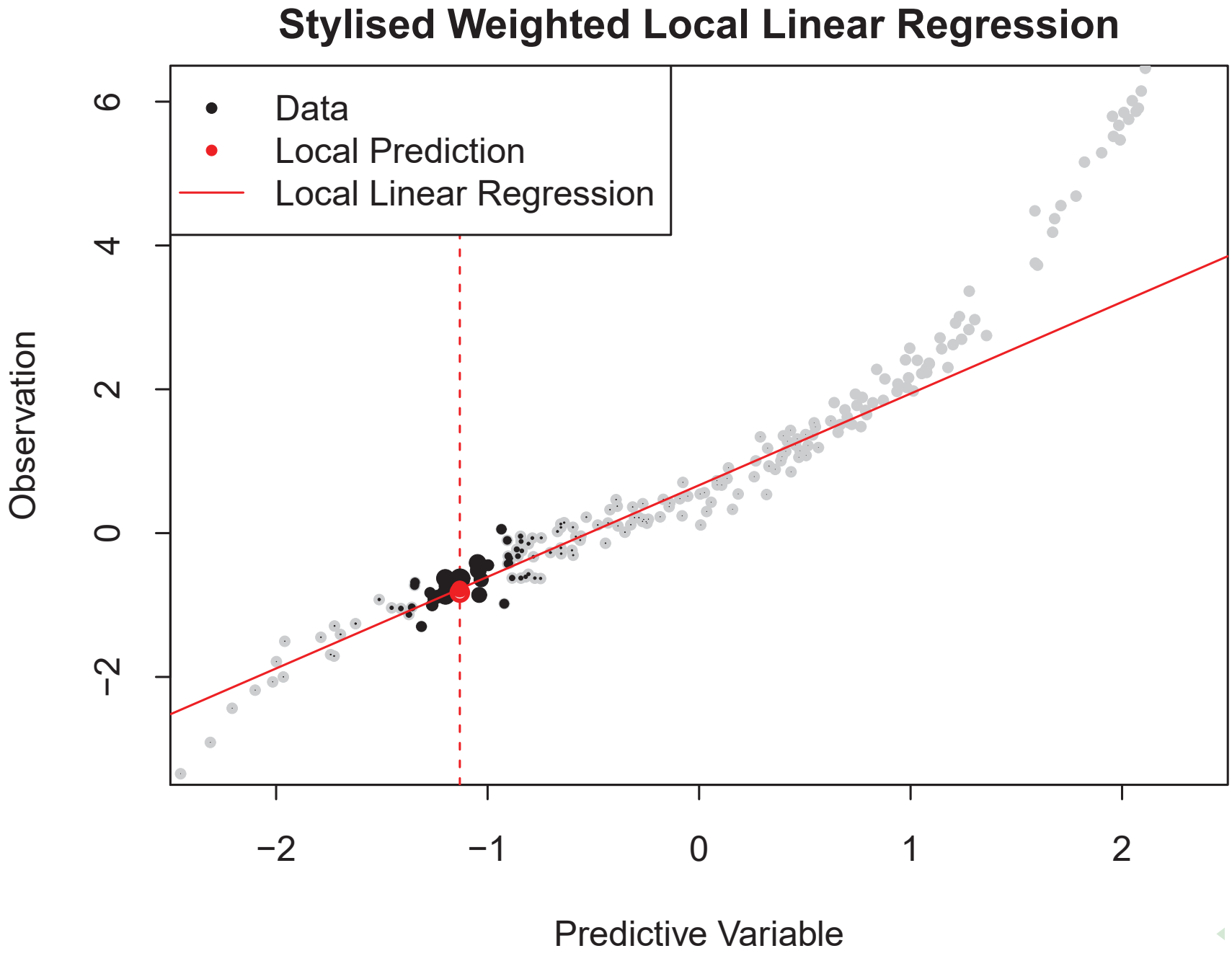
Stylised Example: X one dimensional

- Observe $(X(i), Y(i)), \quad i = 1, \ldots, n$
- What is $\hat{Y}(i) = E[Y(i)|X(i)]$?
- Weighted least squares:

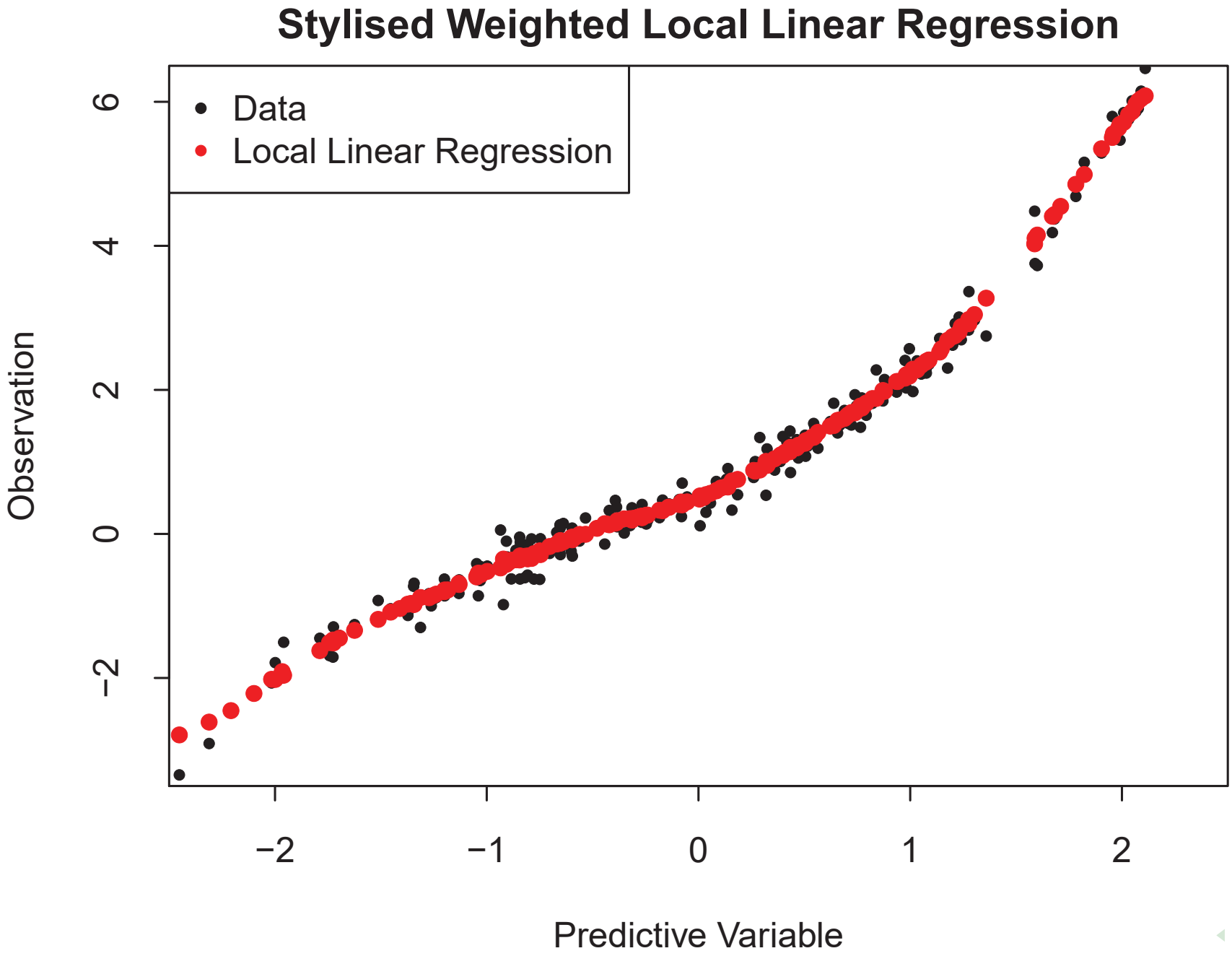
$$\text{minimise} \quad S_i = \sum_{j=1}^{n} w(i,j) \left(Y(j) - (a + bX(j))\right)^2$$

- Weights, $w(i,j) \to 0$ as $X(j)$ gets further from $X(i)$
  $\Rightarrow$ fit a straight line through points near $X(i)$
- Minimisation $\Rightarrow \hat{a}(i), \hat{b}(i)$
- $\hat{Y}(i) = \hat{a}(i) + \hat{b}(i)X(i)$
- Could also use e.g. B-splines
  But might not be practical if X has several dimensions.

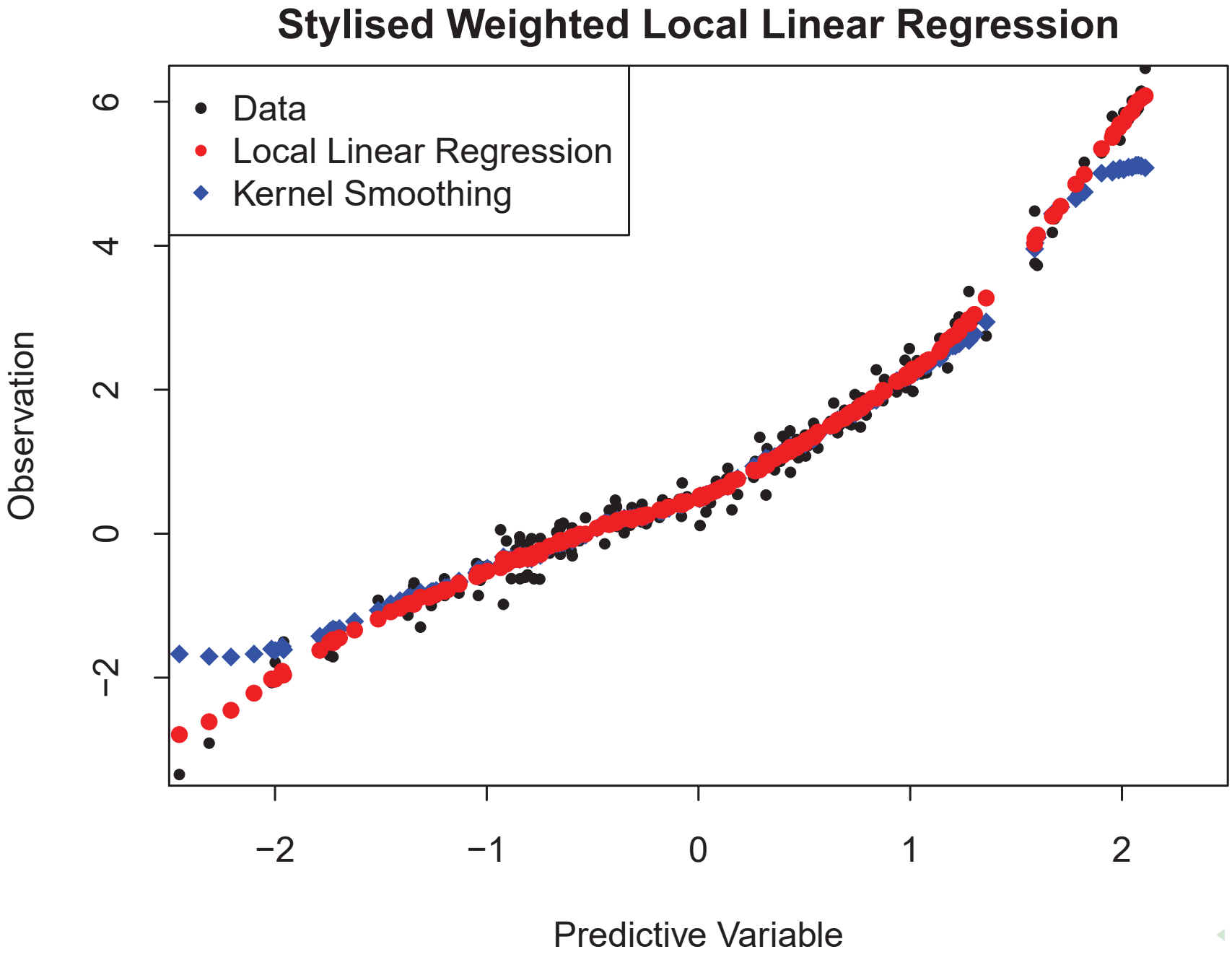
Stage 1: Local Linear Regression



Local Linear Regression



Stage 1: Local Linear Regression



Stage 1: Local Linear Regression

- LSOA i
- Estimate the socio-economic-specific Relative Risk, $F_1(i)$
- For each i, fit an n_P-dimensional sheet around $X(i)$

$$F(i, \mathbf{x}) = a(i) + \mathbf{b}(i)^T \mathbf{x}$$

- n_P predictive variables exclude urban-rural classification urban-rural handled in the weights, $w_1(i, j)$
- Minimise

$$S(a(i), b(i)) = \sum_j w_1(i, j) \left(R_0(j) - a(i) - b(i)^T X(j) \right)^2$$

over $a(i)$ and $b(i)$

Stage 1: Local Linear Regression (cont.)

- Then set

$$F_1(i) \quad = \quad a(i) + b(i)^T X(i)$$

$\Rightarrow$ relative risk accounting for socio-economic factors

- Update estimated deaths:

$$\hat{D}_1(i) \quad = \quad \hat{D}_0(i) F_1(i)$$

Stage 1: Local Linear Regression (cont.)

How to calculate the weights?

- $w(i,j)$ depends on the "distance" between predictive variables $X(i)$ and $X(j)$
- $w(i,j) \to 0$ as the distance gets larger
- $w(i,i) = 0$
- $w(i,j) = 0$ if LSOA's i and j are in different urban-rural groups

Stage 1 → Stage 2

$$
\begin{aligned}
D(i) &= \text{LSOA actual deaths} \\
\hat{D}_0(i) &= \text{LSOA expected deaths with no predictive variables} \\
\hat{D}_1(i) &= \text{LSOA expected deaths with predictive variables} \\
R_1(i) &= \frac{D(i)}{\hat{D}_1(i)} = \text{updated actual-over-expected}
\end{aligned}
$$

Stage 2: Add location data:

$$
\begin{aligned}
Y(i) &= \text{LSOA location co-ordinates} \\
&= \text{(latitude, longitude)}
\end{aligned}
$$

Kernel smooth the $R_1(i)$ using location data.

Actuarial Research Centre
Institute and Faculty of Actuaries

Stage 2: Smooth A/E by Location

Estimate the *additional* location-specific relative risk

$$F_2(i) = \frac{\sum_j w_2(i,j) R_1(i)}{\sum_j w_2(i,j)}$$

Then the fitted expected deaths are

$$\hat{D}_2(i) = \hat{D}_0(i) F_1(i) F_2(i)$$

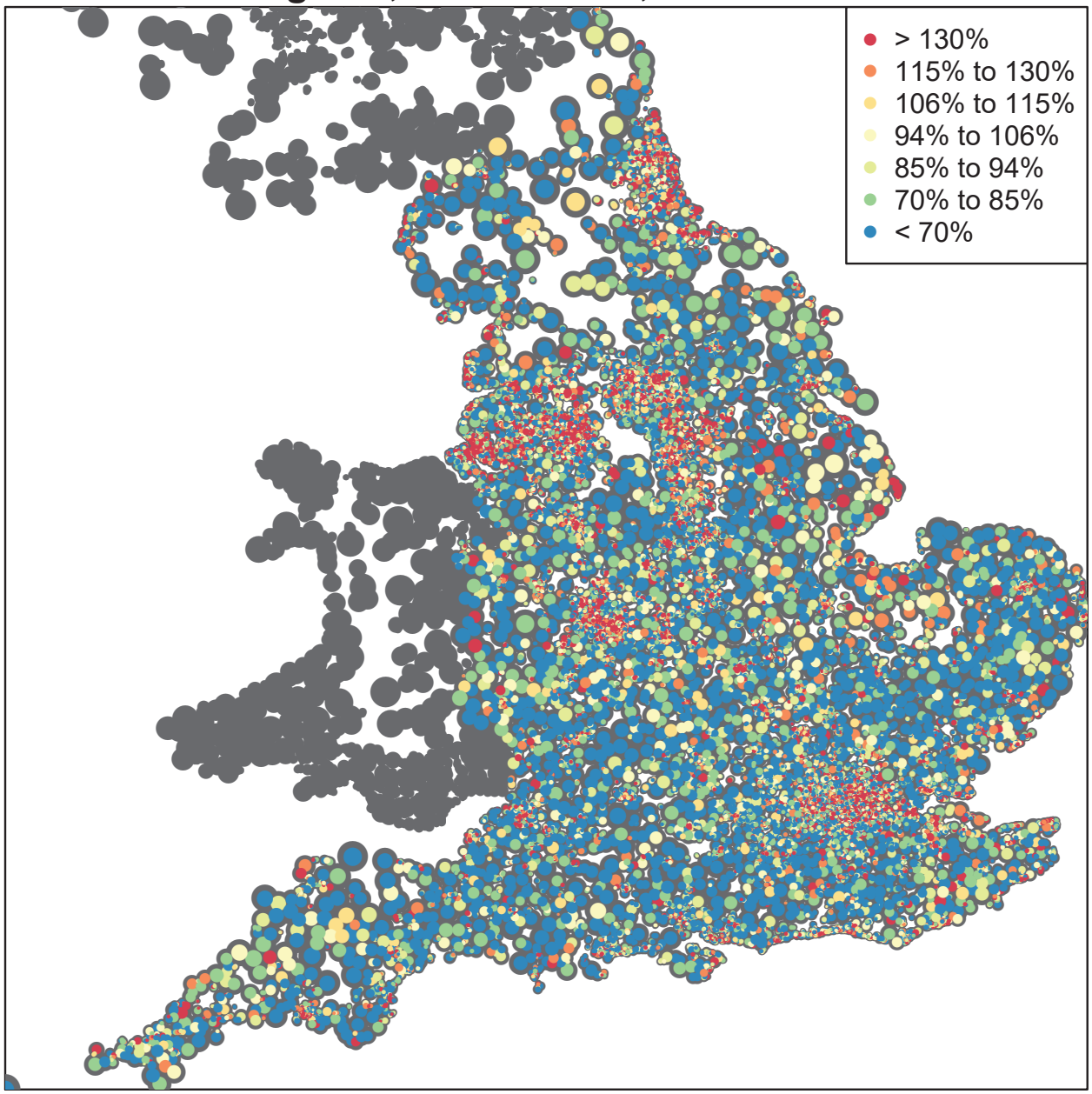
Weights, $w_2(i,j)$, depend on the physical distance between the two LSOA's

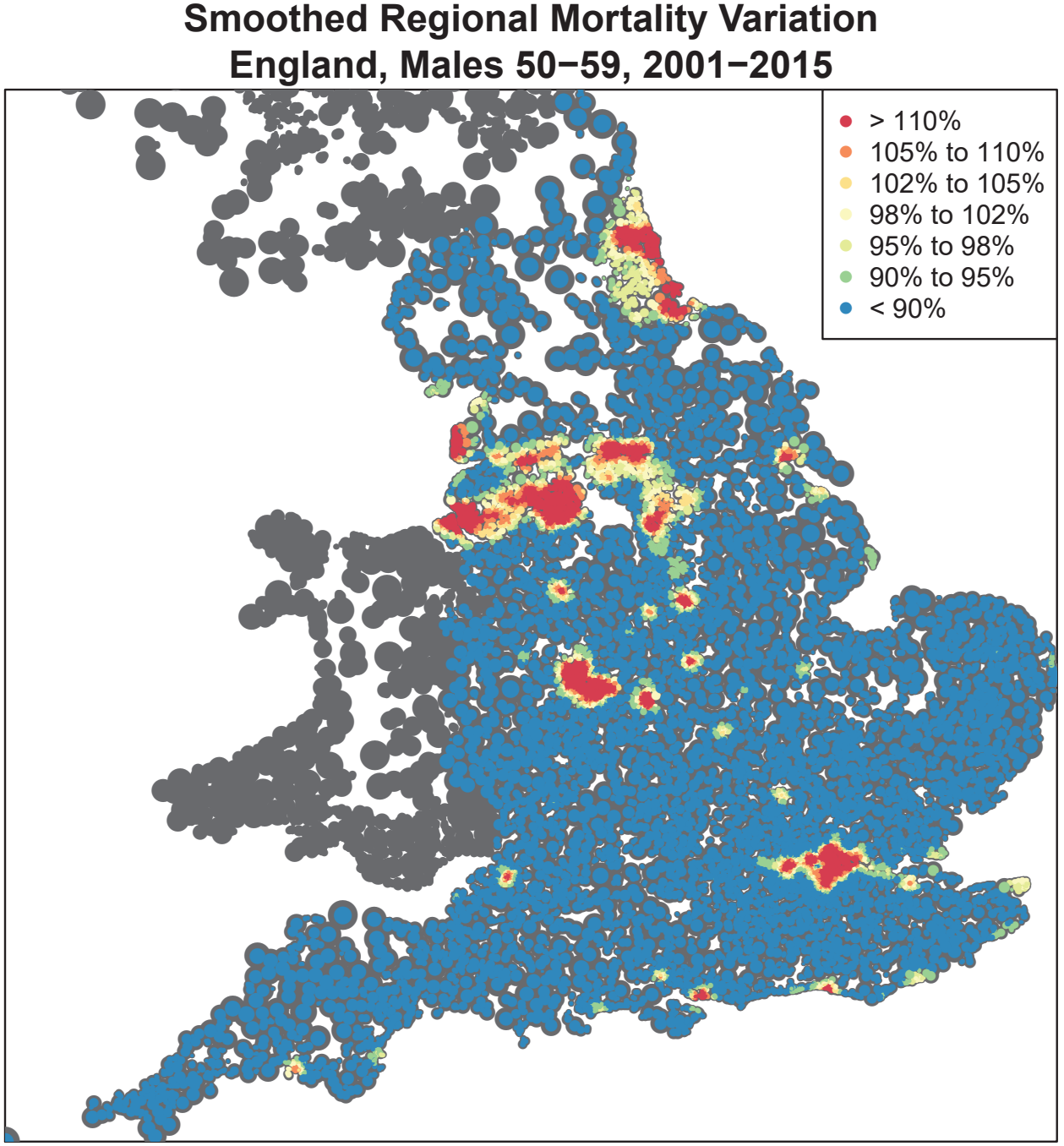
Actuarial Research Centre
Institute and Faculty of Actuaries

Data and Results So Far

- 2001-2015; 2001-2008; 2009-2016
- Ages: 40-49, 50-59, 60-69, 70-79, 80-89
- Predictive variables:
  - income deprivation (elderly; receiving government benefits)
  - employment deprivation (unemployment)
  - average number of bedrooms
  - living environment deprivation (housing quality and air quality)
  - wider barriers (overcrowding)
  - % in care home (60+ with nursing)
  - % in care home (60+ without nursing)
  - urban-rural classification

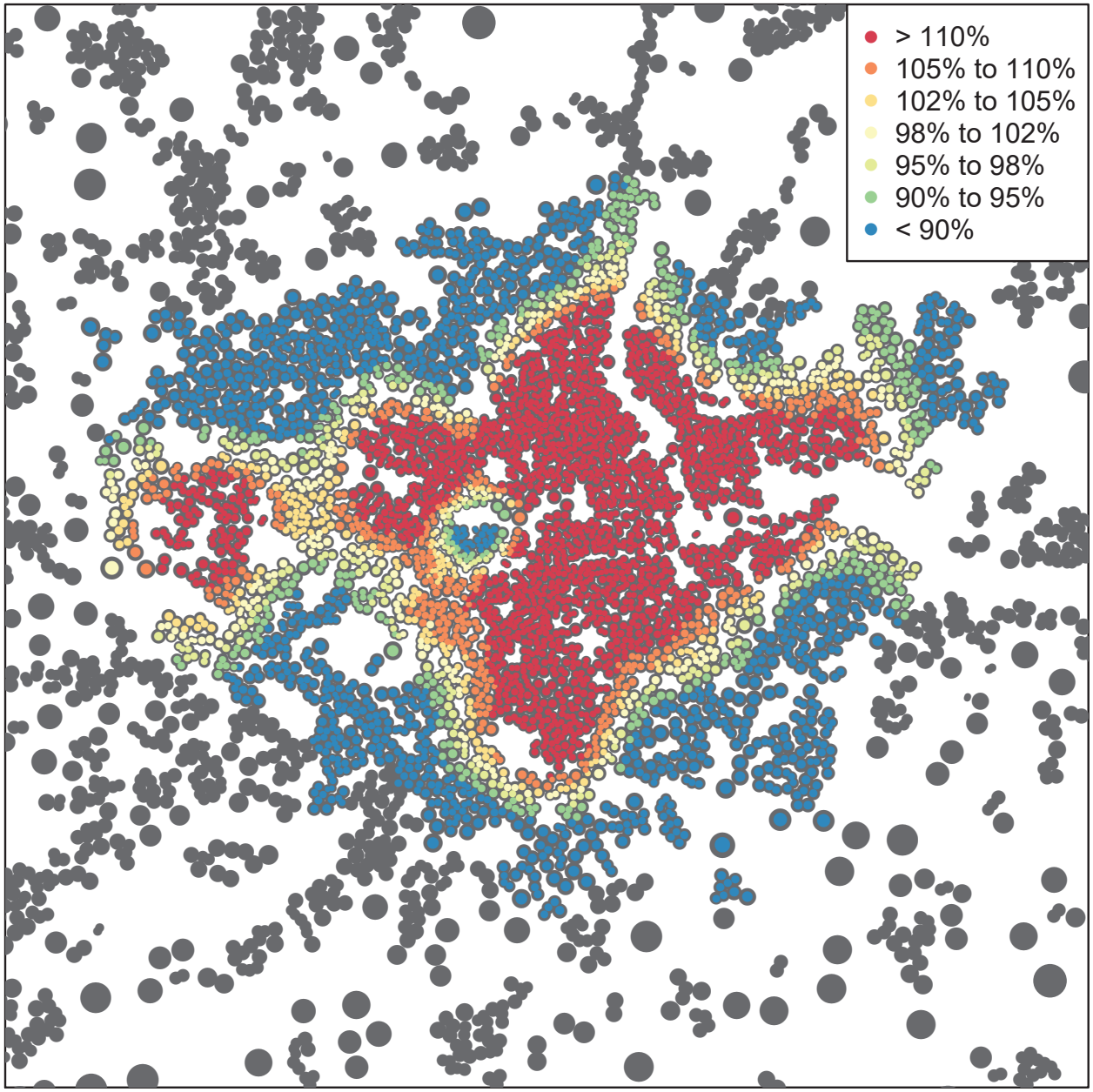
Actuarial Research Centre
Institute and Faculty of Actuaries


R0(i) Actual−over−Expected
England, Males 50−59, 2001−2015
> 130%
115% to 130%
106% to 115%
94% to 106%
85% to 94%
70% to 85%
< 70%



Fix $F_1(i) \equiv 1$; estimate $F_2(i)$.

Smoothed Regional Mortality Variation
London, Males 50−59, 2001−2015



Socio-Economic-Specific Relative Risk

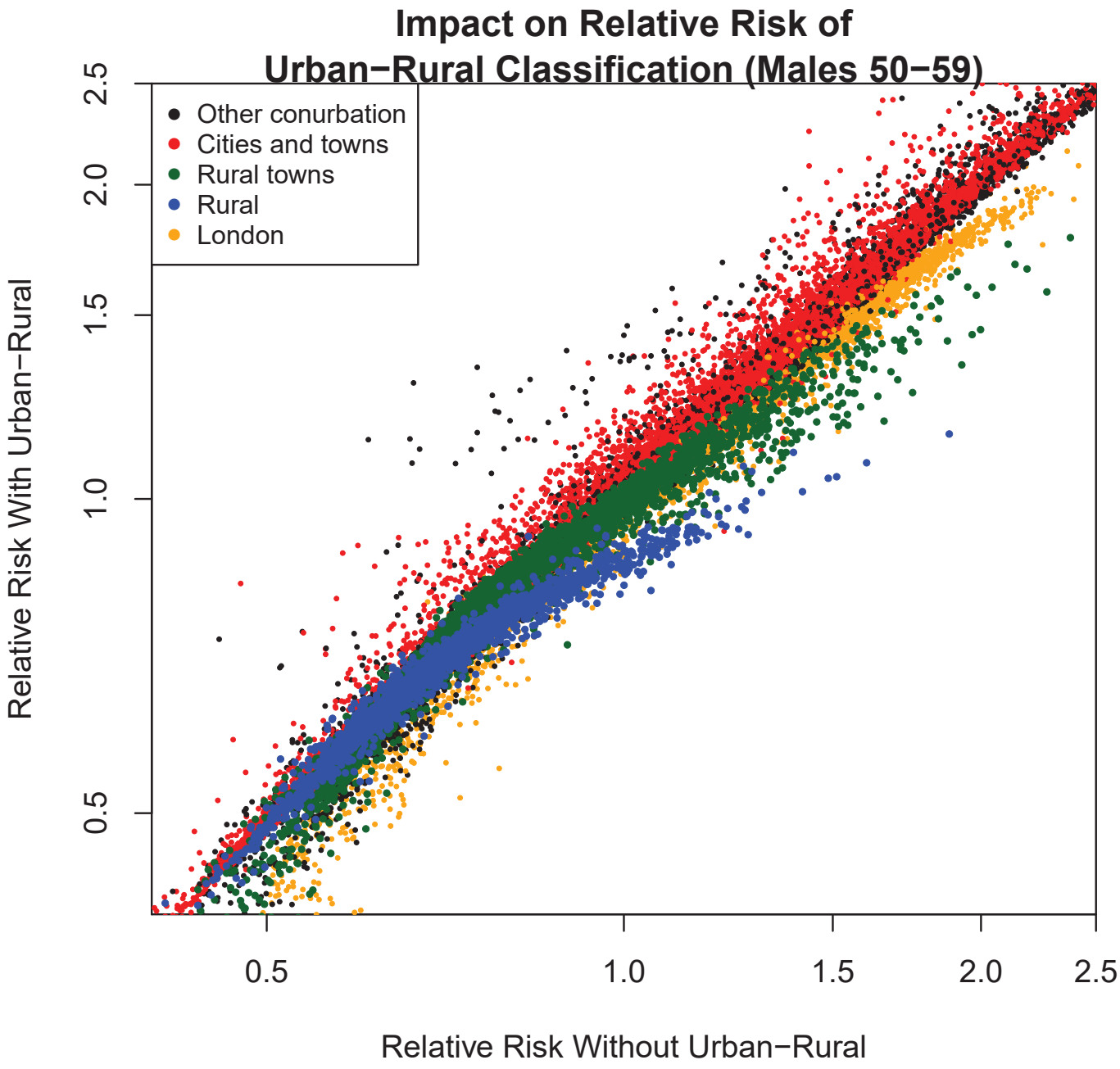
$$\hat{D}_0(i) = \sum_{t=t_0}^{t_1} \sum_{x=x_0}^{x_1} m(t,x) E(i,t,x)$$

$$F_1(i) = a(i) + b(i)^T X(i) \quad \text{(weighted local linear regression)}$$

$$\hat{D}_1(i) = F_1(i) \hat{D}_0(i)$$

- How dependent is $F_1(i)$ on urban-rural classification?
- Does the inclusion of urban-rural classification make a difference?

Relative risk with or without urban-rural (50-59)



Actuarial Research Centre
Institute and Faculty of Actuaries

Role of Predictive Variables

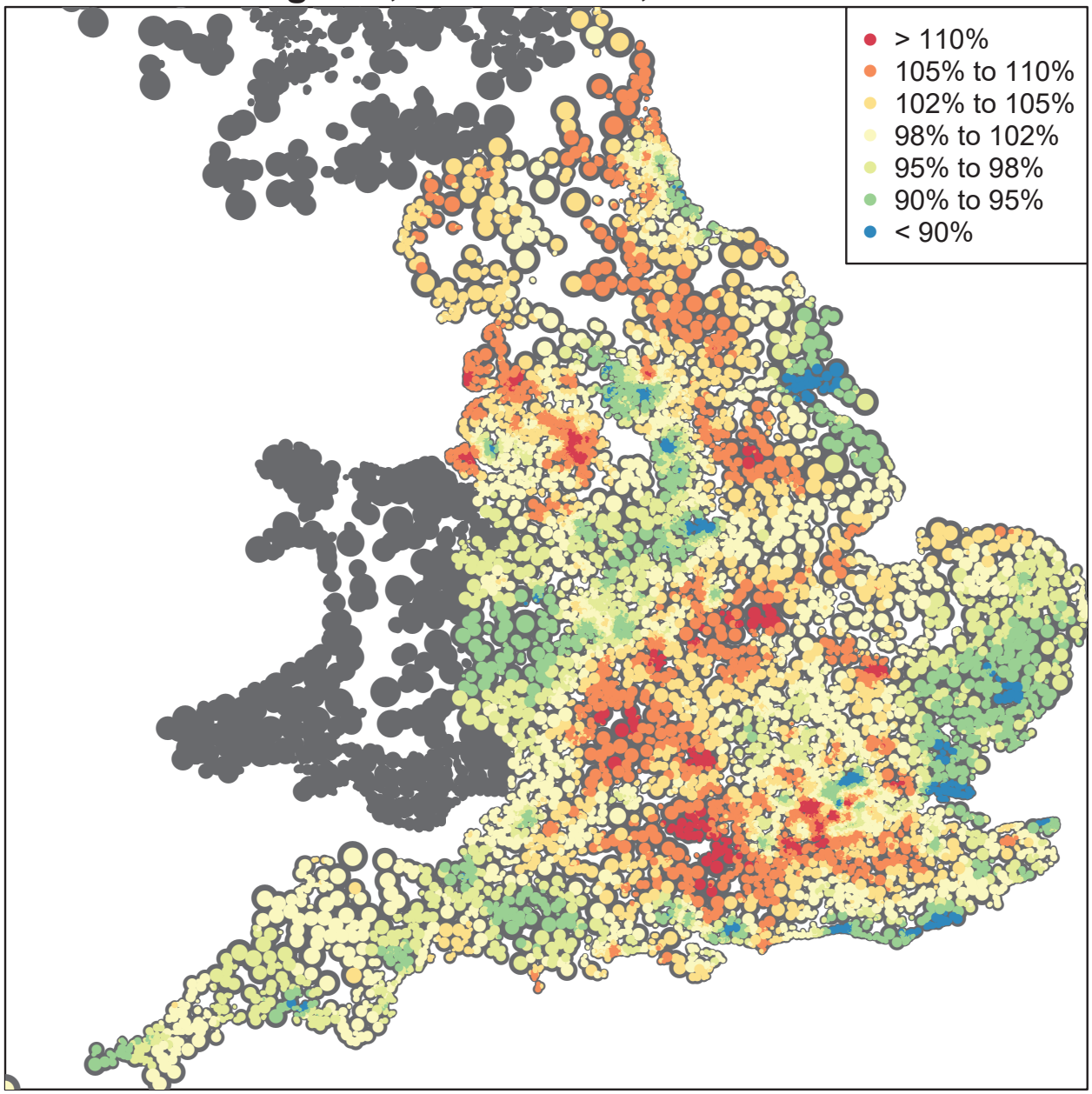
- Income deprivation (elderly) and employment deprivation are the main drivers
- Employment deprivation is the main driver for younger age groups
- Income deprivation (elderly) is the main driver for older age groups
- Urban-rural classification is also an important driver
- Bedrooms, living environment and wider barriers are second order but significant
- Care homes:
  - "nuisance" variables when considering socio-economic effects
  - but including these predictive variables is very important
  - methodology allows us to filter out the impact of care homes on individual LSOA mortality
  - E.g. males 80-89 in a care home with nursing: mortality is 3x to 6x higher than not in a care home

Actuarial Research Centre
Institute and Faculty of Actuaries

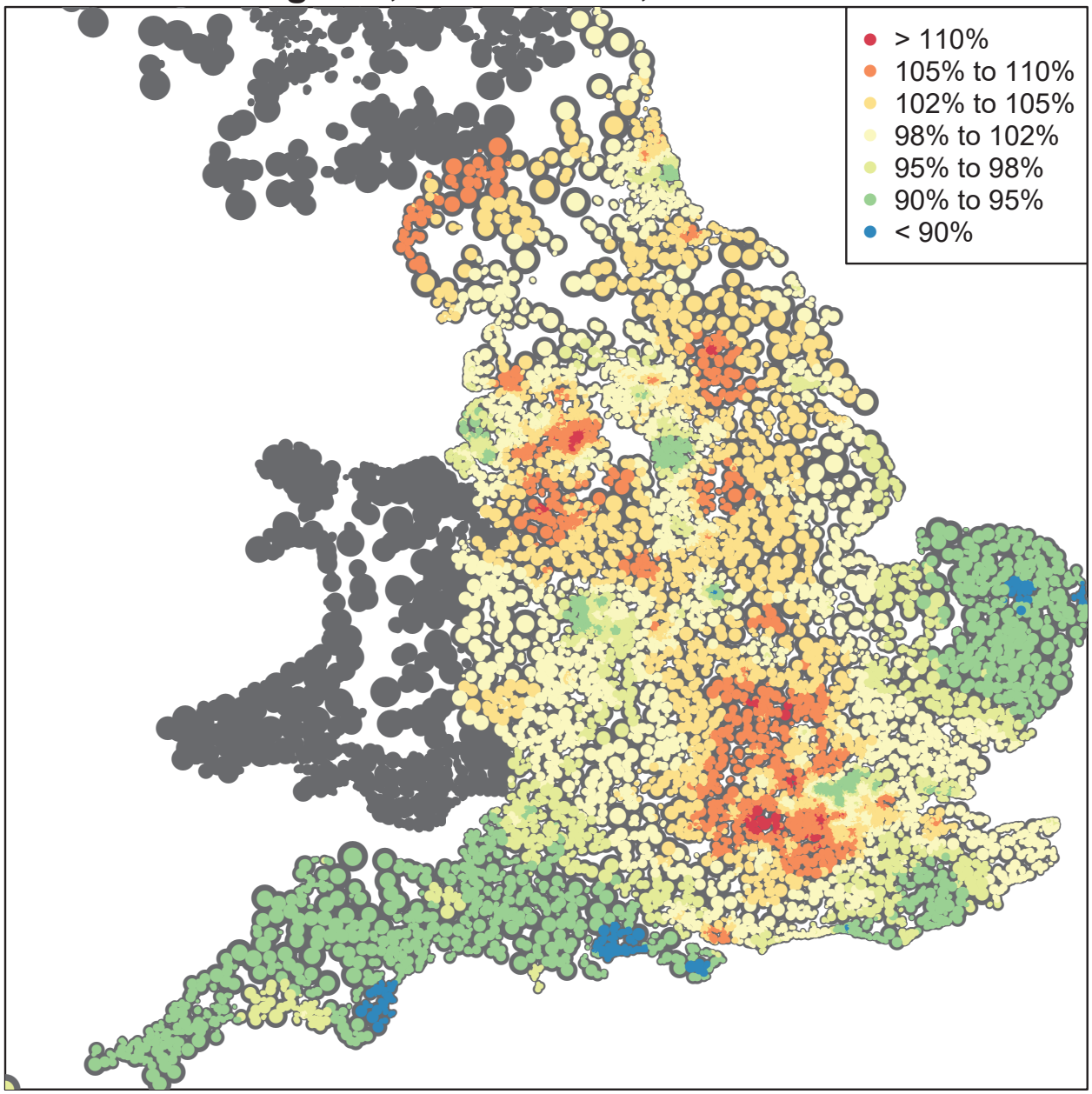
Location-Specific Relative Risk

$$F_2(i) = \frac{\sum_j w_2(i,j) R_1(i)}{\sum_j w_2(i,j)}$$

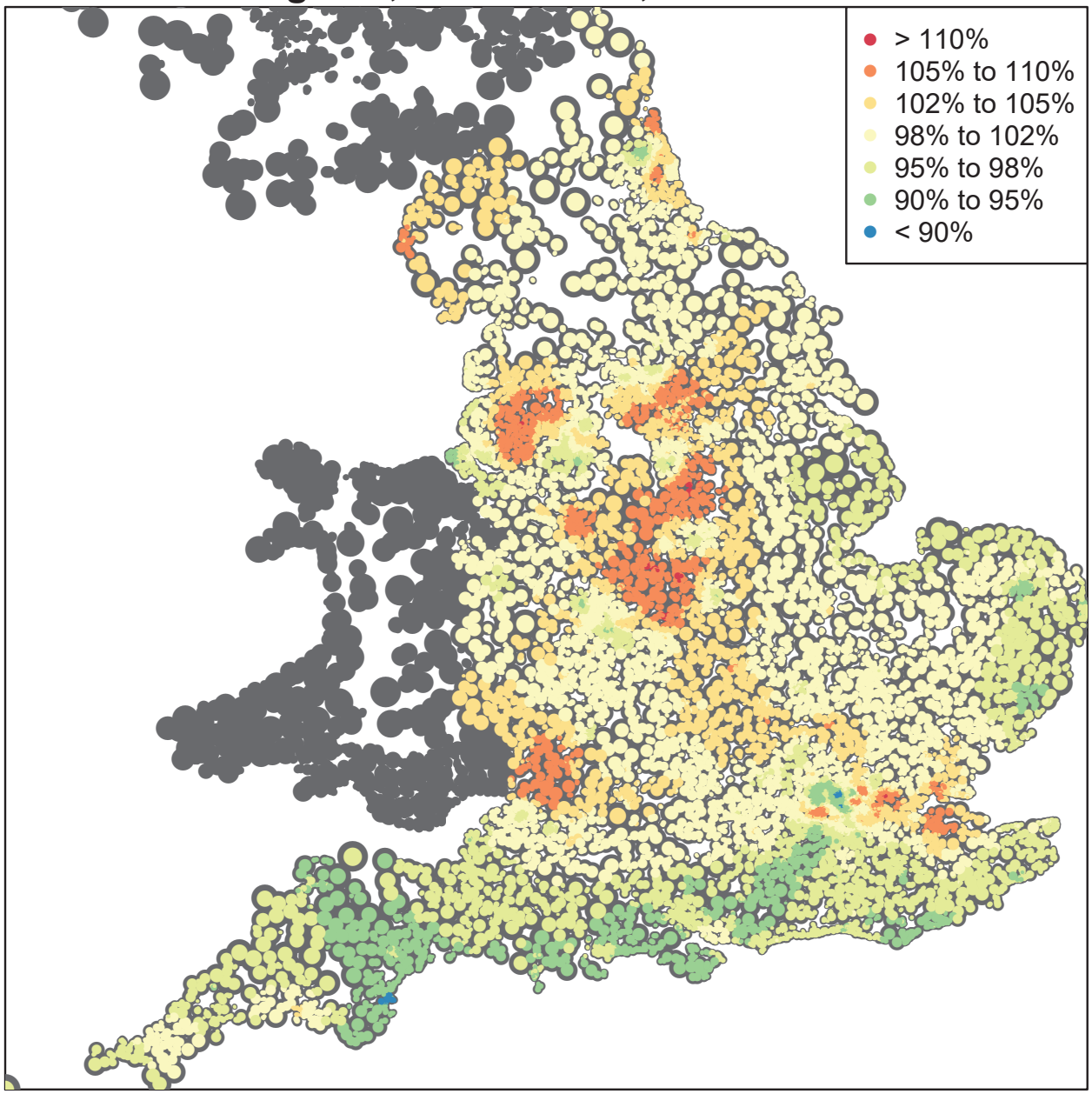
Location−Specific Relative Risk
England, Males 40−49, 2001−2015



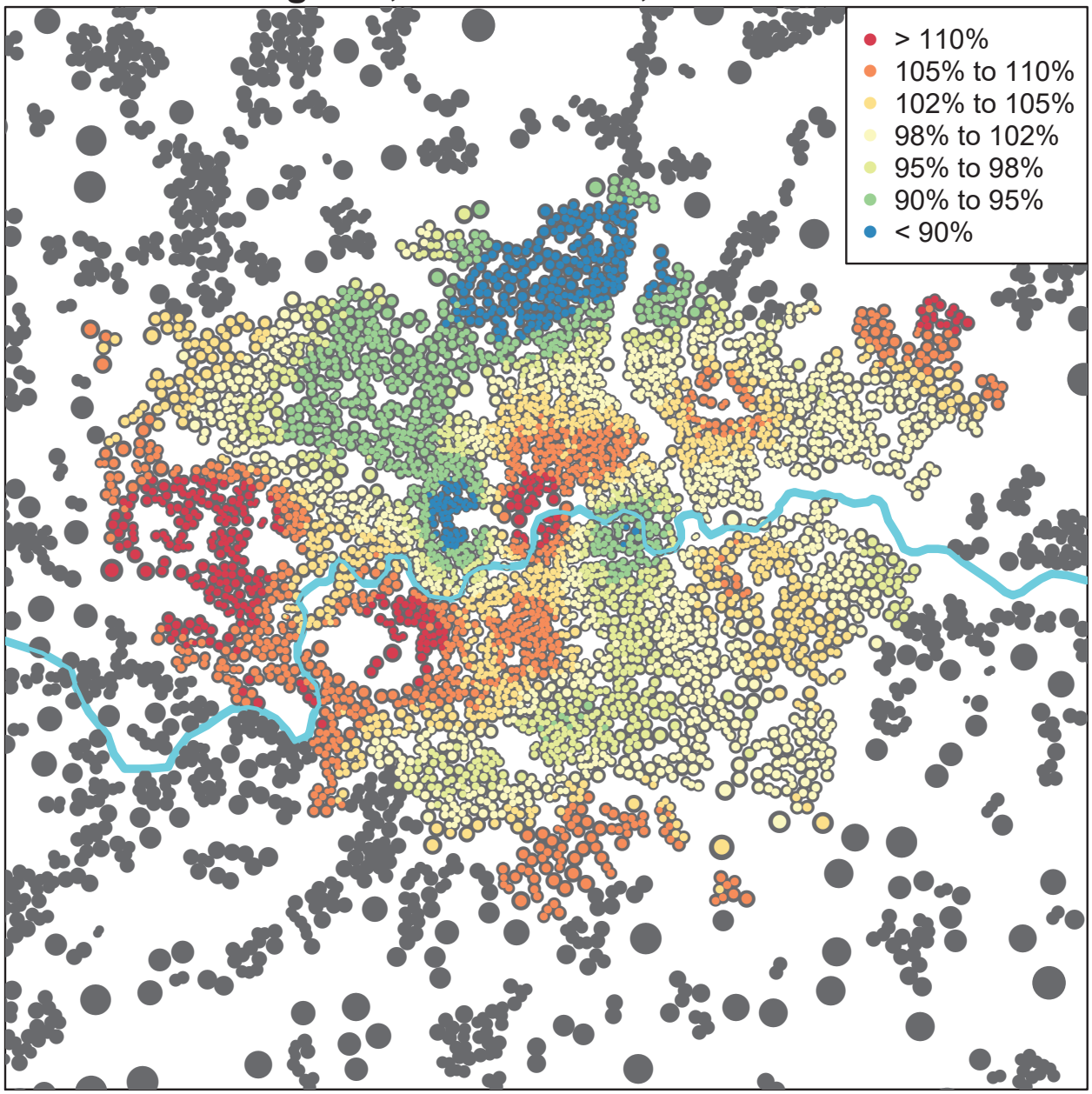
Location−Specific Relative Risk
England, Males 60−69, 2001−2015



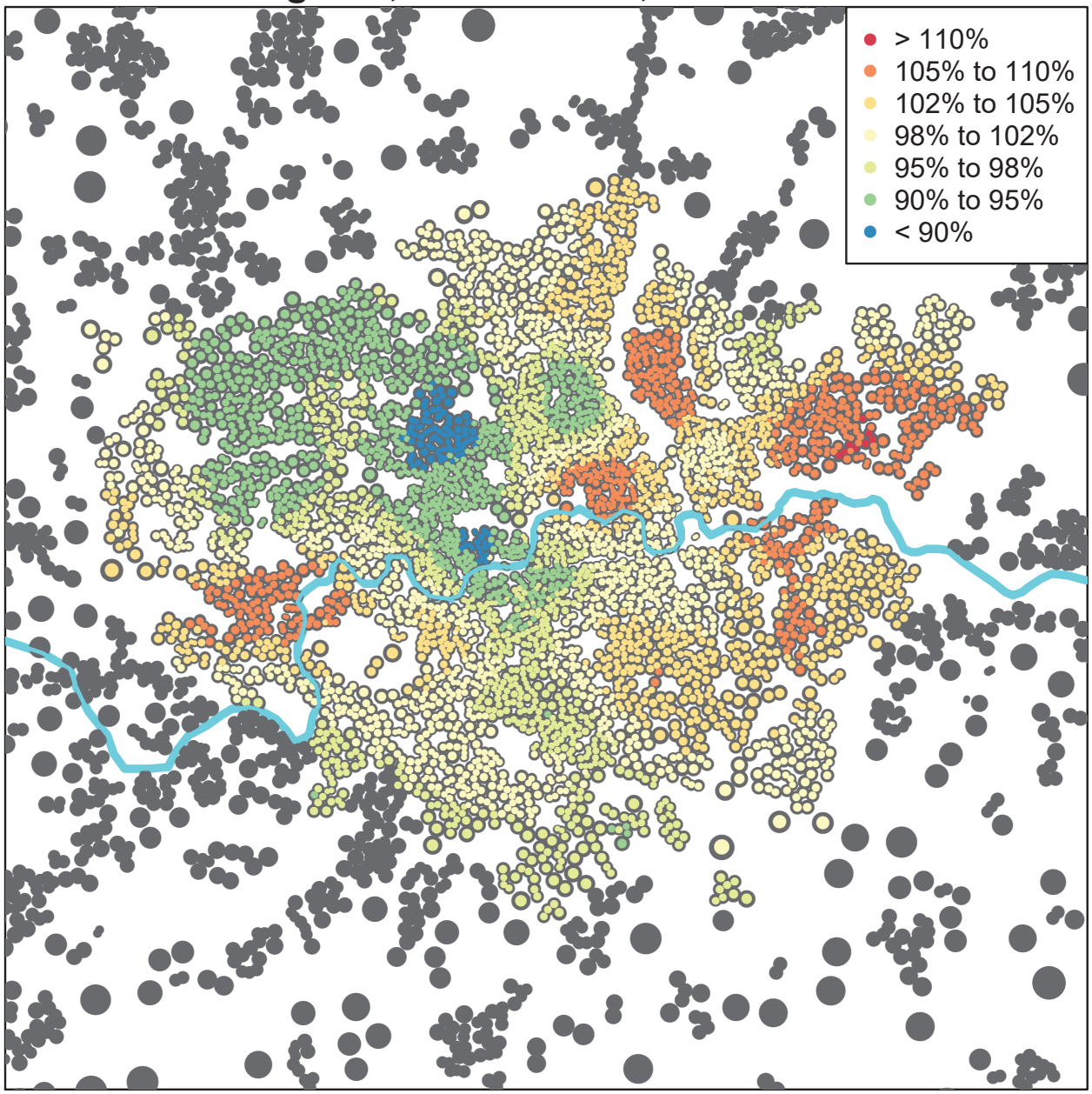
Location−Specific Relative Risk
England, Males 80−89, 2001−2015



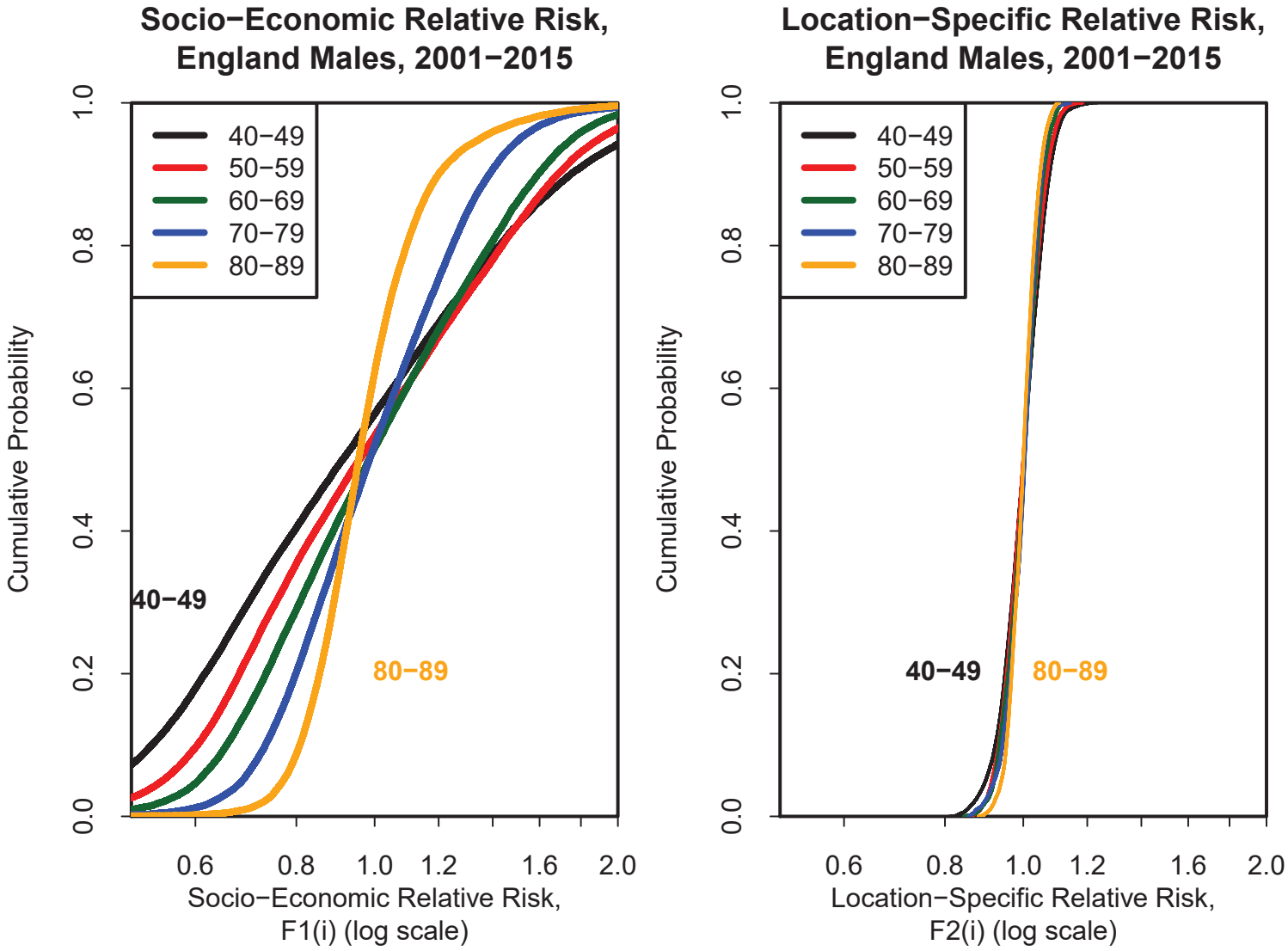
Location−Specific Relative Risk
England, Males 40−49, 2001−2015



Location−Specific Relative Risk
England, Males 80−89, 2001−2015

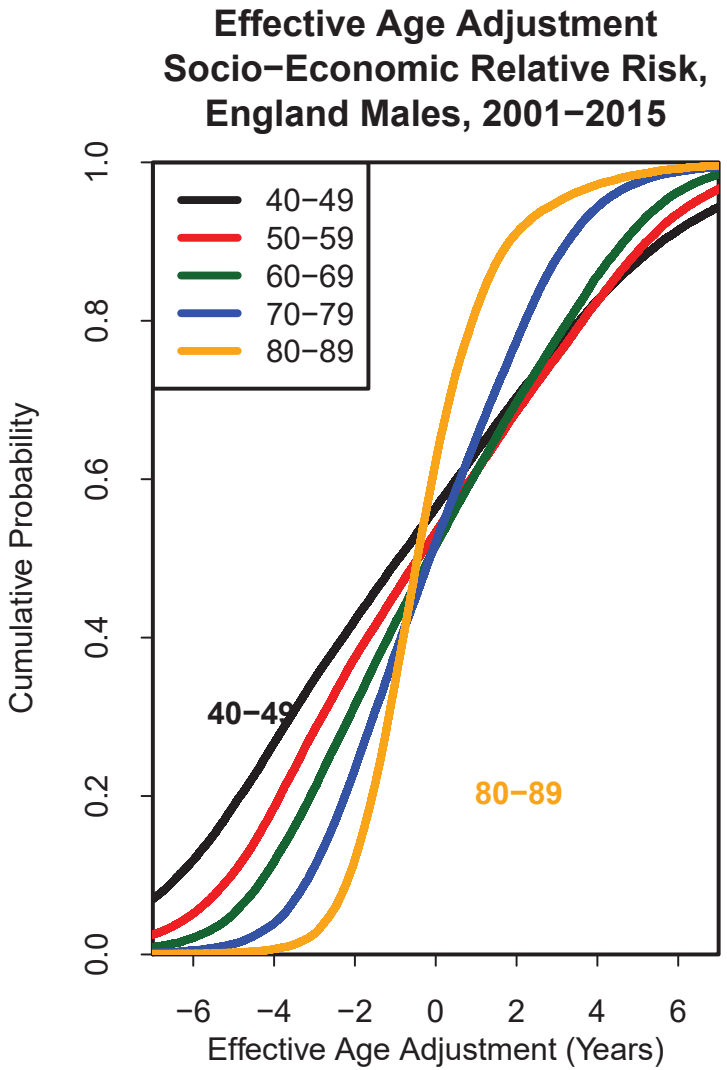


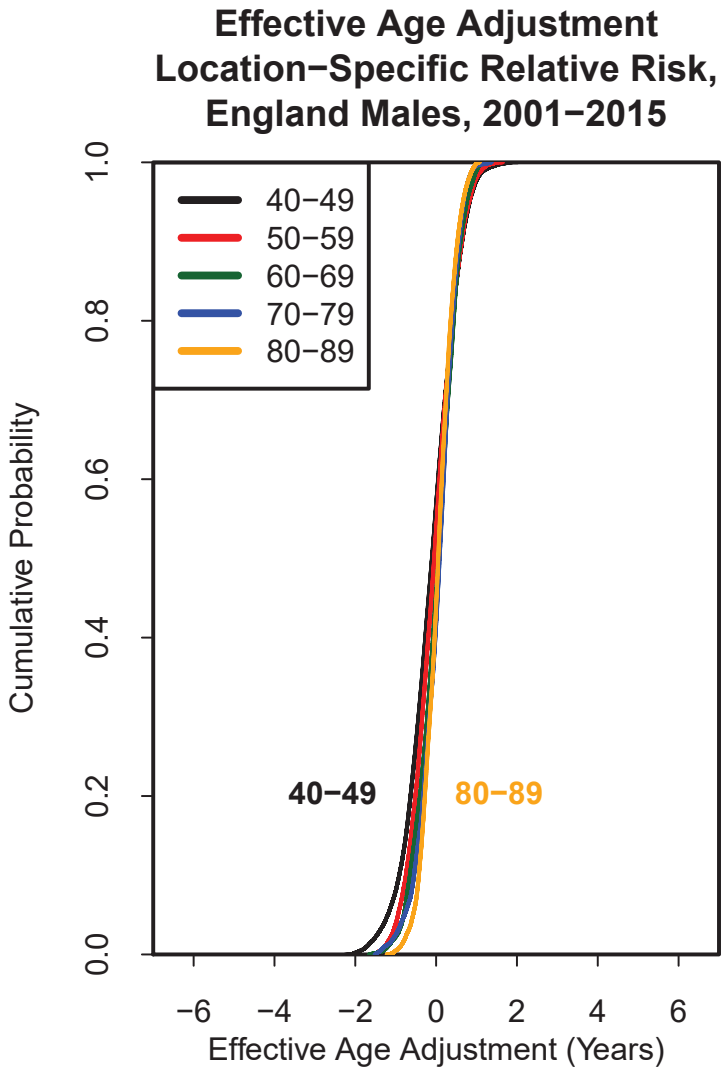
Socio-Economic vs Spatial Effects



- Location contributes 1.3% to 3.5% of the variance in the relative risk

Socio-Economic vs Spatial Effects





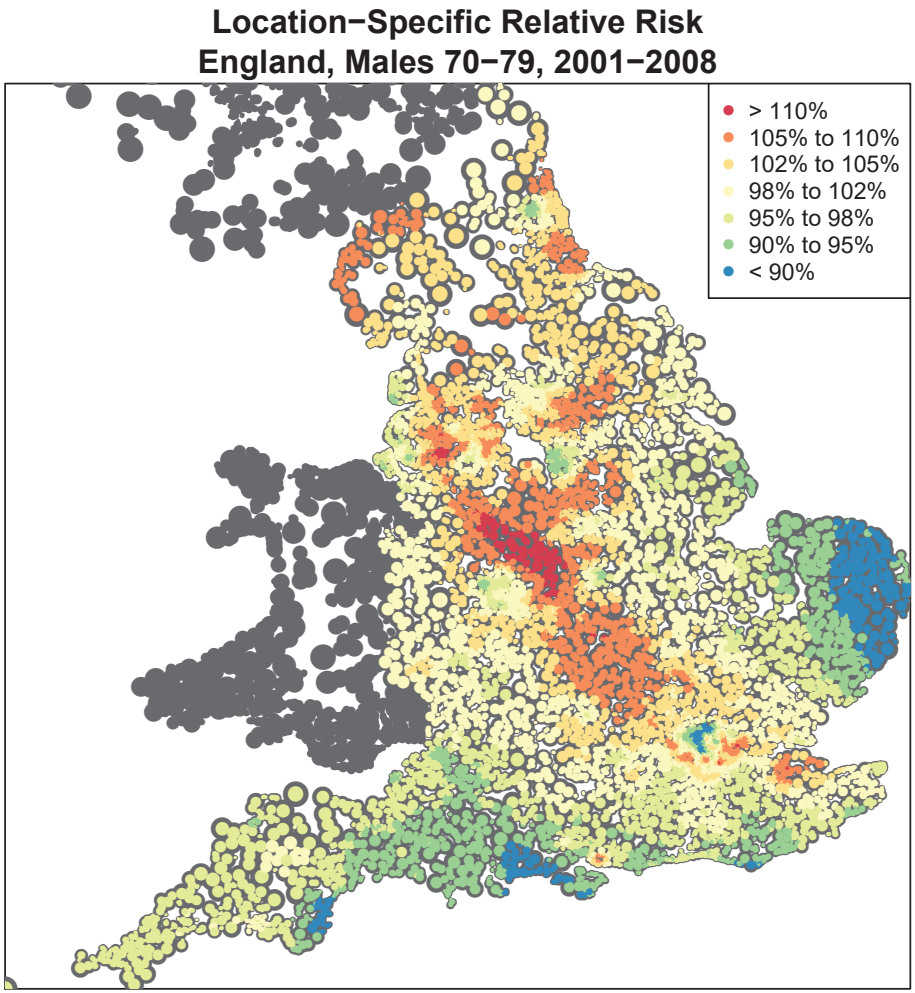
e.g. Effective age adjustment$= -4 \Rightarrow$ mortality is as if 4 years older

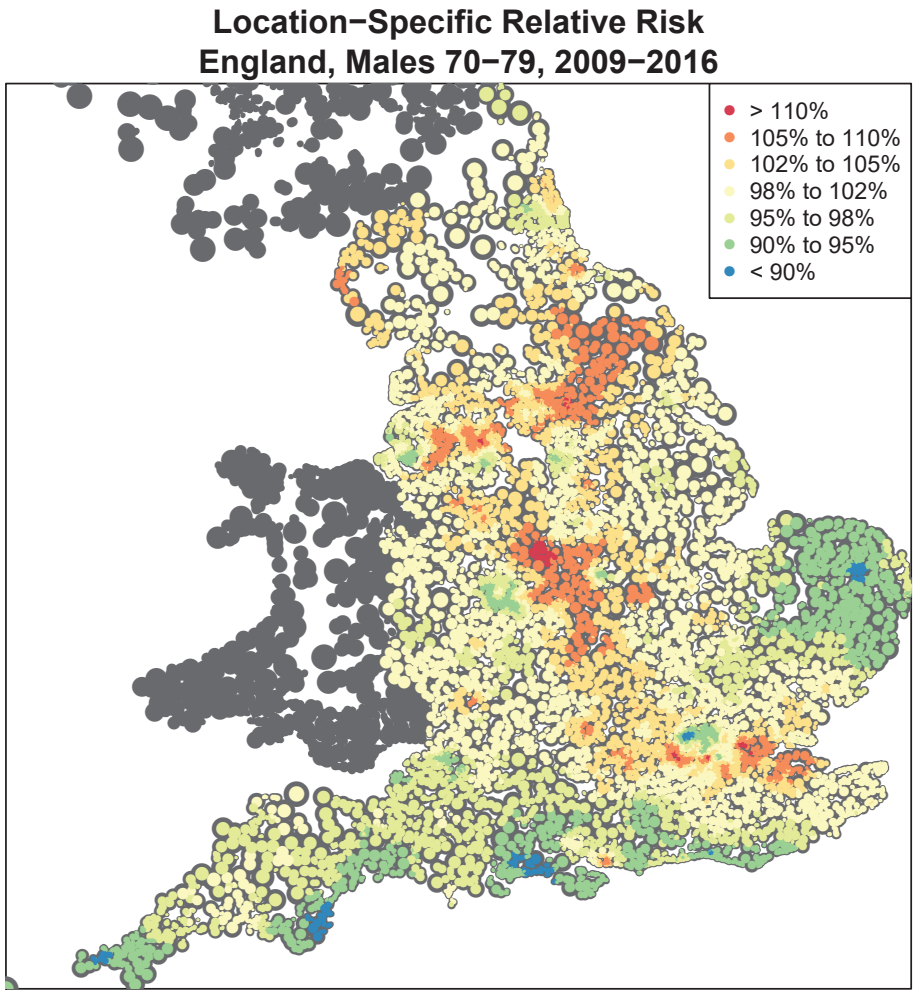
Actual-over-expected: Ages 60-69

Region	No effect	Socio-economic only	Full Model
North East	118	100	99
North West	116	102	100
Yorkshire and The Humber	107	100	100
East Midlands	98	100	99
West Midlands	105	99	100
East	88	96	98
London	105	100	99
South East	89	101	100
South West	87	94	99

- Similar patterns for other age groups and for females

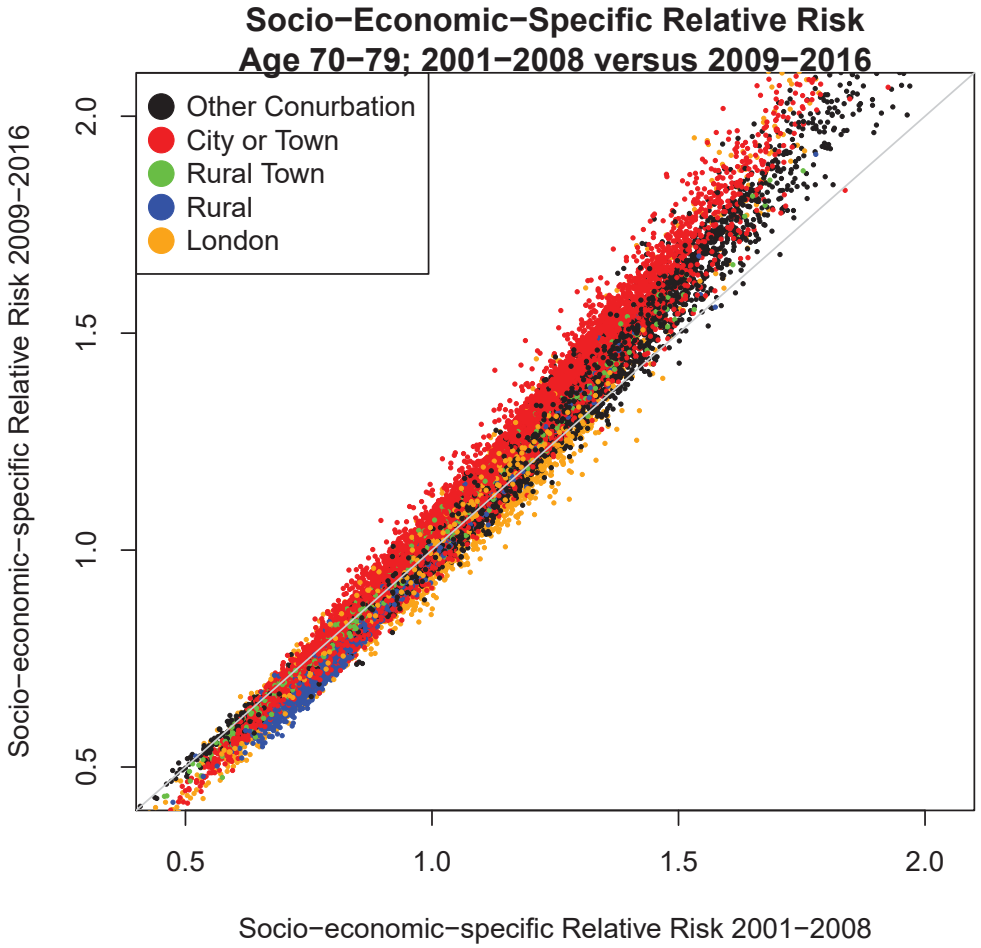
2001-2008 versus 2009-2016

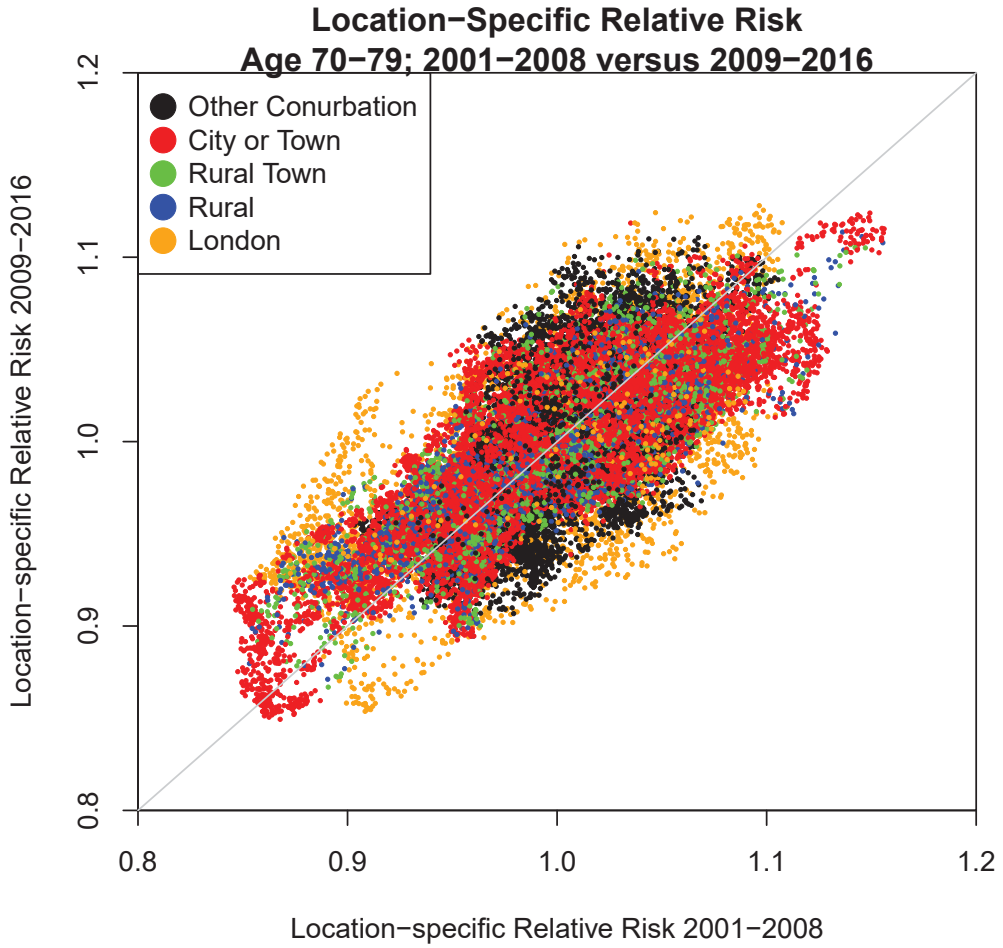




Some variation over time.

2001-2008 versus 2009-2016: Ages 70-79



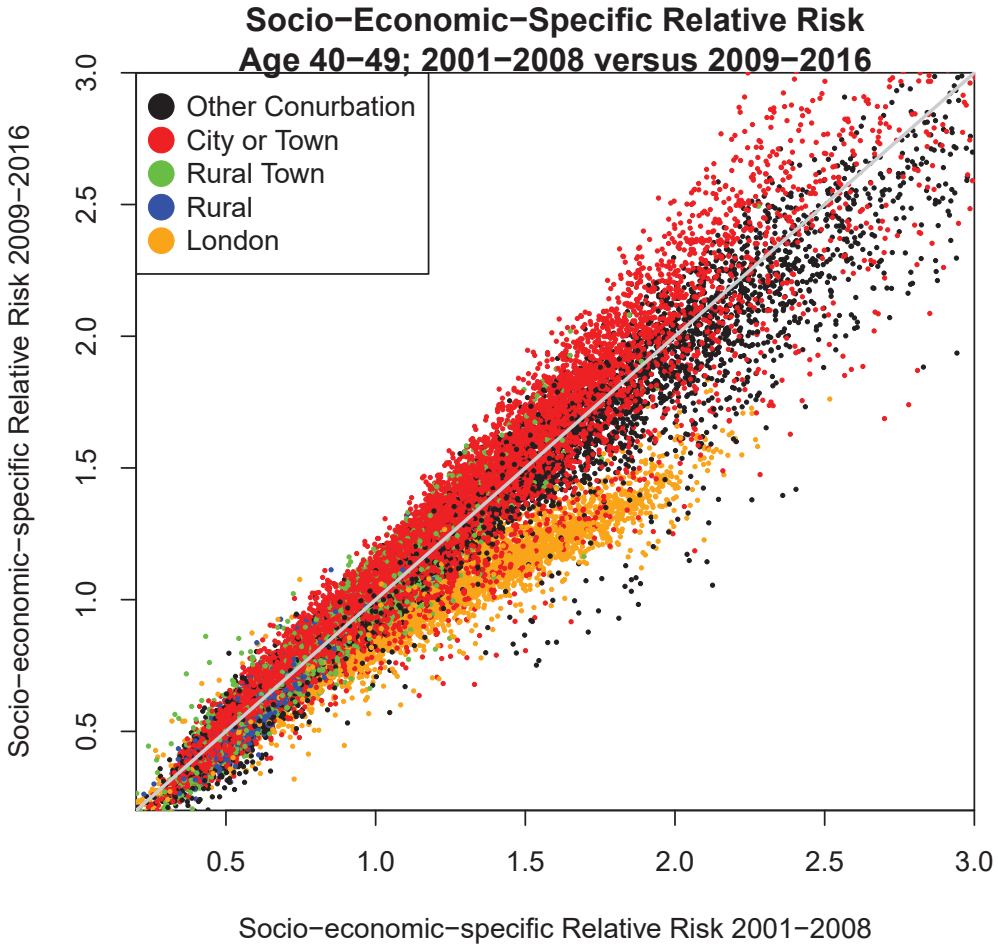


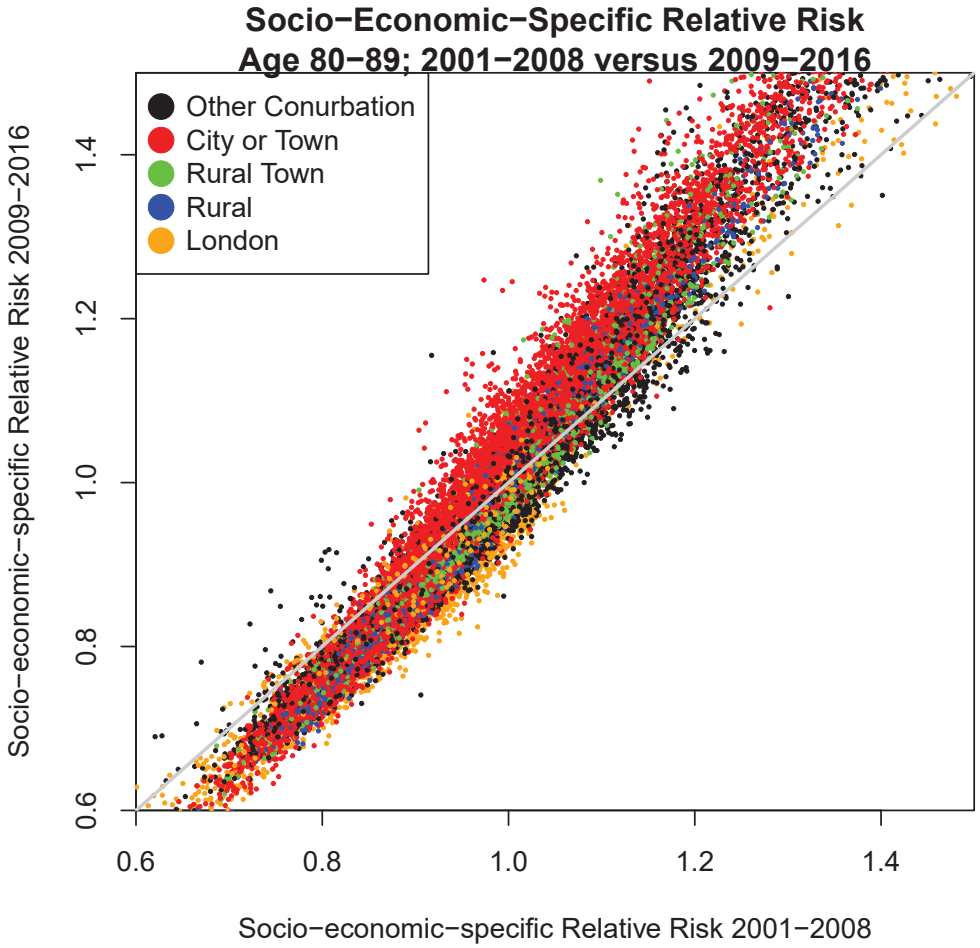
Comments on next slide ...

2001-2008 versus 2009-2016: Ages 70-79

- Socio-economic-specific relative risk:
  - consistent rankings from one period to the next
  - the inequality gap has widened
- Location-specific relative risk
  - Sampling variation is significant
  - Location-specific relative risk has narrowed
  - Sampling variation $\Rightarrow$
    - Mostly, no strong evidence for a change over time in the regional patterns

2001-2008 versus 2009-2016: Ages 40-49 and 80-89





- Widening inequality gap at 80-89
- Stable gap at 40-49, except London: narrowing gap

Conclusions

- Spatial/regional effects are significant
- But much less important than socio-economic (non-regional) effects
- Both effects: can these be used to improve predictions of insurance and pensions mortality?
- Longer term objective:
  Can we form e.g. 10 clusters of LSOA's with similar mortality experience over the period of observation?
- Work in progress

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

Thank You!

Questions?

E: A.J.G.Cairns@hw.ac.uk

W: www.macs.hw.ac.uk/∼andrewc/ARCresources

HERIOT WATT UNIVERSITY

SOCIETY OF ACTUARIES

Canadian Institute of Actuaries | Institut canadien des actuaires

Actuarial Research Centre
Institute and Faculty of Actuaries