A Socio-Economic Mortality Index for Scotland

Andrew J.G. Cairns

Heriot-Watt University, Edinburgh

Joint work with Torsten Kleinow

Longevity 20 Conference, Singapore, September 2025

Outline

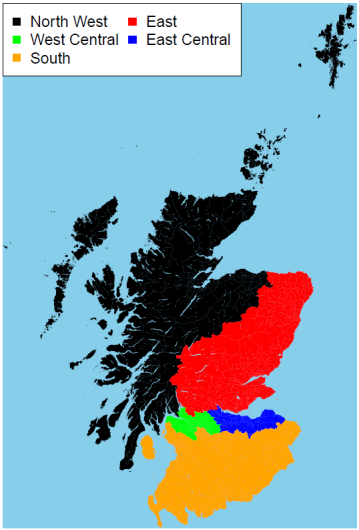
- Motivation and objectives
- Inequalities in mortality rates
- Data
- Methodology: Local linear regression
- Results
- Local sensitivity
- Summary

HERIOT WATT UNIVERSITY

Objectives and motivation

- To develop a customised mortality index for Scotland
- Mortality heterogeneity: mortality varies between subgroups.
  Main reason: prevalence of causal risk factors varies between groups (e.g. smoking)
- What are the most significant socio-economic factors that are associated with mortality rates?
  Note: relative significance varies with age
- How much inequality is there in mortality rates at different ages?
- Prior work: Longevity Index for England (LIFE)
  Objective: to extend this to the rest of the UK
  Focus here on Scottish mortality

Background: Relative mortality by region



Actual/Expected Deaths
Variation by Region
Ages 70-89, Years 2011-19

	Regional Relative To National Mortality (%)	
Region	Males	Females
Scotland	100	100
North West	**90.5**	**88.9**
East	95.0	95.0
West Central	**113.6**	**109.6**
East Central	96.2	96.7
South	98.9	102.6

Values show actual deaths (ages 70-89) by region as a percentage of expected deaths using national age-specific death rates

Background: Relative mortality by deprivation quintile (SIMD)

Actual/Expected Deaths
Variation by Region
Ages 70-89, Years 2011-19

Region	Males	Females
North West	90.5	88.9
East	95.0	95.0
West Central	113.6	109.6
East Central	96.2	96.7
South	98.9	102.6

Actual/Expected Deaths
Variation by Deprivation (SIMD) Quintile
Ages 70-89, Years 2011-19

Quintile	Males	Females
1	134.0	128.9
2	113.1	109.1
3	98.2	98.4
4	87.8	89.6
5	75.3	76.3

Regional variation $<$ variation by Scottish Index of Multiple Deprivation (SIMD)

- West Central has higher levels of deprivation than other regions
  $\Rightarrow$ 'naturally' higher mortality rates
- But is there an additional regional effect in addition to deprivation?
  Or does deprivation fully explain regional differences?

Does deprivation explain all variation in mortality by region

- West Central has higher levels of deprivation
  $\Rightarrow$ 'naturally' higher mortality rates
- But is there an additional regional effect in addition to deprivation?
  Or does deprivation fully explain regional differences?

- Four potential hypotheses:
  - H_0: year and age explains everything at the neighbourhood level
    $\Rightarrow m(t, x)$ national death rates (REJECT!)
  - H_1: year + age + deprivation (SIMD)
  - H_2: year + age + more general socio-economic data
  - H_3: none of the above!

Data: Neighbourhoods: "LSOA"s

- Small neighbourhoods:
  "Data Zones": DZ
- $L = 6976$ DZs
- Socio-economically homogeneous
- Average size ≈ 780 persons
- DZs $i = 1, \ldots, L$,
  single years (t =2011-2019),
  single or 5-year age groups, x:
  - Deaths: $D(i, t, x)$ (5-yr age groups) (by normal place of residence)
  - Exposures: $E(i, t, x)$ (single ages)

  - Many potential predictive variables, $\Rightarrow$ vector $X(i)$

Methodology

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

$$m(t,x) = \frac{D(t,x)}{E(t,x)} \qquad \text{(single ages)}$$

- Neighbourhood death rates ($i = 1, \ldots, L$)
  - local death rates: $m(\mathbf{i},t,x)$
  - how to model $m(i,t,x)$?

- Predictive variables $X(i)$ plus $m(t,x) \longrightarrow m(i,t,x)$
- Method: Local Linear Regression (LLR); Random Forest; XGboost; ...

Methodology: General Approach

- Over a limited age range (e.g. 70-89); and
- Over a limited range of years (e.g. 2011-2019):

$$\text{neighbourhood } i, \qquad m(i,t,x) = m(t,x)F_1(i)$$

- $F_1(i)$ = neighbourhood-specific *relative risk* due to socio-economic and other non-spatial characteristics, $X(i)$
  Estimate $F_1(i)$ using Local Linear Regression

Local Linear Regression (LLR): in brief

- i = neighbourhood index
- $D(i)$ = actual deaths
- $\hat{D}_0(i)$ = expected deaths based on national mortality, $m(t, x)$
- LLR: Locally regress
  - $Y_0(i) = D(i)/\hat{D}_0(i)$ response variable against
  - $\tilde{X}(i) = (X(i,1), \ldots, X(i,M))^T$ predictive variables
  - Local weights $w(i,j)$ depend on proximity of $\tilde{X}(j)$ to $\tilde{X}(i)$
- $F_1(i) = E[Y_0(i)|\tilde{X}(i)]$
- $\hat{F}_1(i) = \hat{a}(i) + \hat{b}(i)^T \tilde{X}(i)$

Feature Selection/Predictive Variables: what to include in $\tilde{X}(i)$

Sources of data available at the neighbourhood level:

- Scottish Index of Multiple Deprivation (SIMD)
  - Various domains (e.g. employment)
- Census 2022
  - Non-sensitive data publicly available
  - Based on census questions

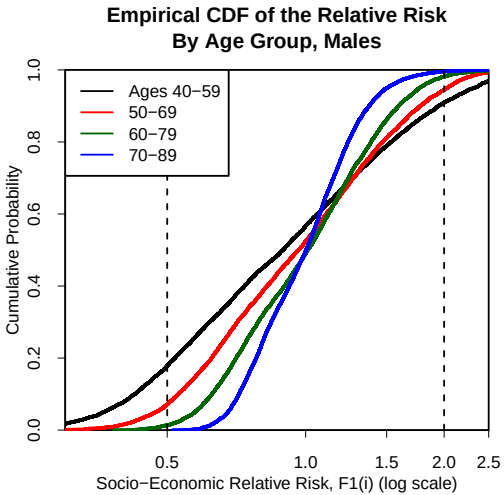
Advantage of LLR in feature selection process:

- During feature selection process: set weight $w(i, i) = 0$
  $\Rightarrow$ each $D(i, t, x)$ is "out of sample"
- In theory: maximise Poisson log-likelihood
- In practice: different age groups; males and females
  Want to choose the same set of predictive variables for each group

Choice of Predictive Variables: what to include in $\tilde{X}(i)$

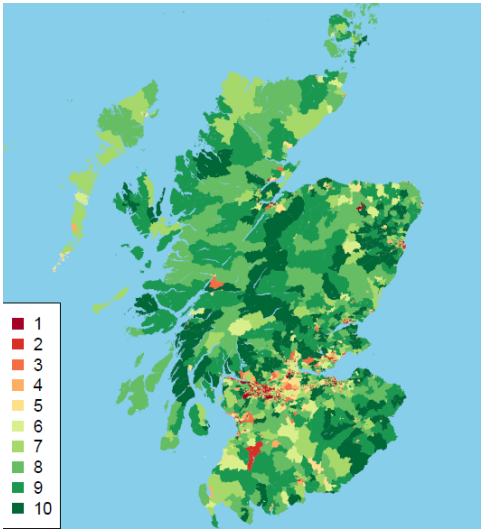
- **SIMD:**
  - Employment deprivation
  - Education deprivation: Proportion of the population with no qualfications
  - Crime rank (low rank $\Rightarrow$ high crime rate)
- **Census data:**
  - Communal establishments: proportion of the population aged 65+ in a communal establishment (mostly care homes in the 65+ group)
  - Proportion married: ages 35+
  - Proportion widowed or divorced: ages 35+
  - Proportion born in Scotland: all ages
- **2011-2019 population data:**
  - Old-age dependency ratio: # 65+ / # 20-64; males and females separately
  - Female to male ratio: ages 15+

Socio-Economic Relative Risk, $F_1(i)$: Significant Inequality



- High levels of inequality
  [Higher levels of inequality than England.]
- Inequality dimishes with age
- Age 40-59 main predictors:
  unemployment, married, education
- Age 70-89 main predictors:
  care homes, education, dependency ratio
- Males and females:
  Females slightly less spread out
  High correlation between male and female relative risks

Geographic Spread of Inequality: Males 60-79



- Males 60-79
- Deciles by estimated $F_1(i)$
- Decile 1 (dark red): high mortality
- Decile 10 (dark green): low mortality

- Decile 1
  High mortality neighbourhoods
  Mostly concentrated in dense urban areas
- Rural areas:
  More green
  Lighter mortality

Regional Deaths Adjusting for Socio-Economic Variables

Actual / Expected Deaths
$m(i,t,x) = m(t,x)$
Variation by Region
Ages 70-89, Years 2011-19

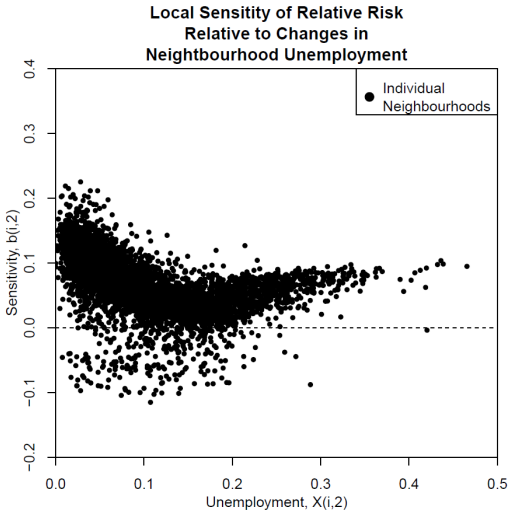
Region	Males	Females
North West	90.5	88.9
East	95.0	95.0
West Central	113.6	109.6
East Central	96.2	96.7
South	98.9	102.6

Actual / Adjusted Expected Deaths
$m(i,t,x) = m(t,x)F_1(i)$
Variation by Region
Ages 70-89, Years 2011-19

Region	Males	Females
North West	98.7	96.1
East	98.8	98.5
West Central	101.2	99.4
East Central	98.4	100.8
South	98.9	101.4

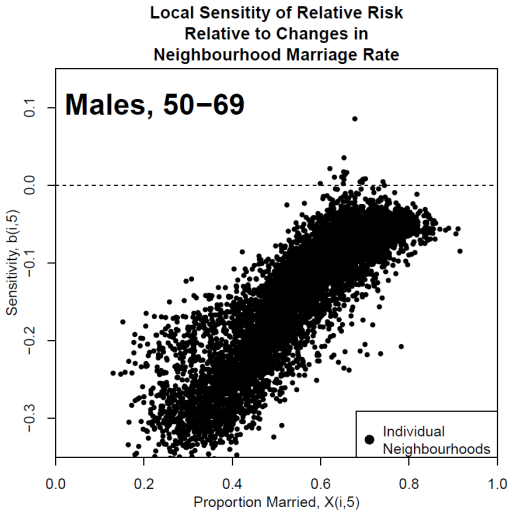
- Socio-economic modelling reduces significantly the variation between regions but it's not perfect

Local Sensitivity of Results to Specific Predictive Variables



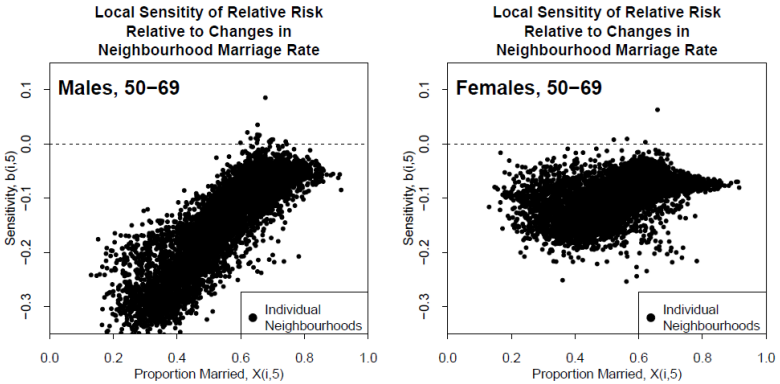
- $\hat{F}_1(i) = \hat{a}(i) + \sum_j b(i,j)X(i,j)$
- $b(i,j)$ measures the local sensitivity of $\hat{F}_1(i)$ to changes in $X(i,j)$
- Here: $j = 2$, unemployment
- Males, 60-79
- Consistent, mostly positive level
  $\Rightarrow$ relative risk rises with unemployment
- But also noisy
  - Sampling variation in death counts
  - Genuine variation across the range of predictive variables
    (unemployment and other variables)

Sensitivity to Proportion Married



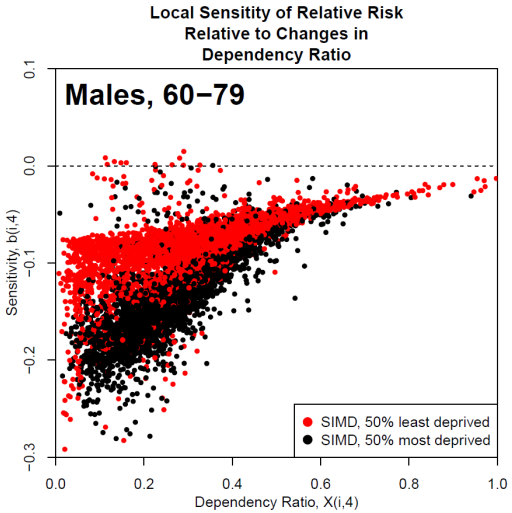
- Sensitivity to the proportion married (or in a civil partnership)
- Males 50-69
- Strong negative sensitivity
  $\Rightarrow$ strong correlation between marriage and lower mortality
- Stronger influence in neighbourhoods with lower marriage rates
- Potential inference:
  marriage is a possible predictor of a healthier lifestyle

Sensitivity to Proportion Married: Males vs Females



- Relative risk for females much less sensitive than males
- Inference:
  Female mortality much less dependent on marital status
  Male lifestyles more heavily influenced by marital status

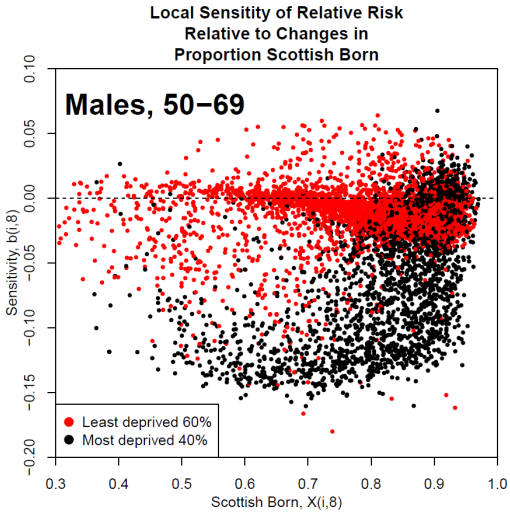
Sensitivity to Dependency Ratio, by Deprivation



$$\text{Dependency Ratio, DR} = \frac{\text{population 65+}}{\text{population 20-64}}$$

- Dependency ratio depends on:
  - high/low death rates
  - migration into/out of the neighbourhood as individuals get older
- Strong negative sensitivity
  $\Rightarrow$ migration in is a strong predictor of lower mortality
- SIMD deciles (black/red):
  The strength of the DR effect is very dependent on the level of general deprivation

Sensitivity to Proportion Born in Scotland



- **Least deprived 60%:**
  Limited sensitivity to Scottish birth
- **Most deprived 40%:**
  Significant effect
  Higher proportion $\Rightarrow$ lower mortality
- This is the opposite of a
  healthy immigrant effect!
- But for females at higher ages
  sensitivity is positive

Discussion: Neighbourhood Effects and Identifiability

- Dependency ratio effect $\Rightarrow$ two explanations
  - Healthy people are more likely to move to "nice" or "healthy" areas on retirement $\Rightarrow$ bf individual "migrant" status
  - Age-related immigration and a larger population of 65+ creates a positive **environment** for retirees
- Similarly: proportion married
  - Is an **individual's marital status** a strong predictor?
  - Or is the **environment** in which they live a strong predictor?

- With neighbourhood data only:
  we cannot differentiate between the two explanations
  $\Rightarrow$ we have an identifiability problem
- Reality (?):
  Both individual status and environment are likely to be influential.

Summary

- Customised mortality index is much better at explaining mortality inequalities than the generic Scottish Index of Multiple Deprivation (SIMD)
- Inequality between different socio-economic groups is very significant: greatest at younger ages; and narrows significantly with age

- Key predictive variables depend on age group
- Local Linear Regression $\Rightarrow$ complex dependencies between variables that need careful analysis and interpretation

  Sensitivities vary significantly within the predictive variable space

- Annuities: implications for pricing and reserving?

Thank You!

E: A.J.G.Cairns@hw.ac.uk
W: www.macs.hw.ac.uk/~andrewc/ARCresources

Coming soon: THe Index of Scottish Life Expectancy – THISLE

HERIOT WATT UNIVERSITY