Introducing the New Longevity Index for England (LIFE) App

Andrew J.G. Cairns and Torsten Kleinow
joint work with Jie Wen

Heriot-Watt University, Edinburgh

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

Actuarial Research Centre
Institute and Faculty of Actuaries

SOCIETY OF ACTUARIES

Canadian Institute of Actuaries / Institut canadien des actuaires

Outline

- Introduction
  - Motivation and background
  - Mortality data and predictive variables
  - LIFE index objectives
- Construction of the LIFE index
- Results
- The LIFE App

Background: Socio-economic variation in mortality

Motivation for development of the Longevity Index For England (LIFE):

- mortality varies significantly between sub-populations
- how much variation is there?
- what is the best way to sub-divide the population?

Applications

- mortality base tables
- mortality improvement rates for sub-populations linked to past sub-population improvement rates
- contribute to the wider debate on mortality inequalities

Common Practice: The Index of Multiple Deprivation (IMD) As a Predictive Variable

- Motivation
  - IMD deciles often used to segregate the population
  - IMD is a reasonable mortality predictor, but ...
    it is designed for a wide range of purposes
    $\Rightarrow$ not optimised as a mortality predictor
  - Can we develop a better, customised mortality predictor?

- Known problem: analysis of subgroups within an IMD decile
  - Construct a "standard table" for each IMD decile
  - Decile $d = 1, \ldots, 10$ death rates, year t, age a:

$$\hat{m}(d,t,a) = \frac{\text{actual deaths in decile } d\text{, year } t\text{, age } a}{\text{exposures in decile } d\text{, year } t\text{, age } a} = \frac{\sum_{i \in \mathcal{L}(d)} D(i,t,a)}{\sum_{i \in \mathcal{L}(d)} E(i,t,a)}$$

where $\mathcal{L}(d)$ is the set of all Lower Layer Super Output Areas (LSOAs) in IMD decile d

- What is an LSOA? What are $D(i,t,a)$ and $E(i,t,a)$?

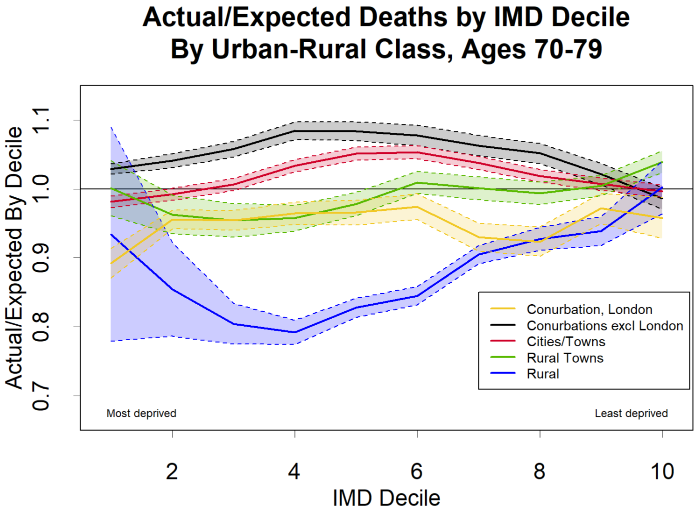
Digression: Lower Layer Super Output Areas (LSOAs)

- What is an LSOA?
  - small neighbourhoods (32,844 in England)
  - 1,000 to 3,000 population; average 1,600
  - Socially homogeneous (tenure of household and dwelling type)
- A wide variety of data are available at the LSOA level: deaths, exposures, socio-economic variables, ...
- $D(i, t, a)$ = deaths in LSOA i, year t, age a
- $E(i, t, a)$ = exposures in LSOA i, year t, age a
- Decile $d = 1, \ldots, 10$:

$$\hat{m}(d, t, a) = \frac{\sum_{i \in \mathcal{L}(d)} D(i, t, a)}{\sum_{i \in \mathcal{L}(d)} E(i, t, a)}$$

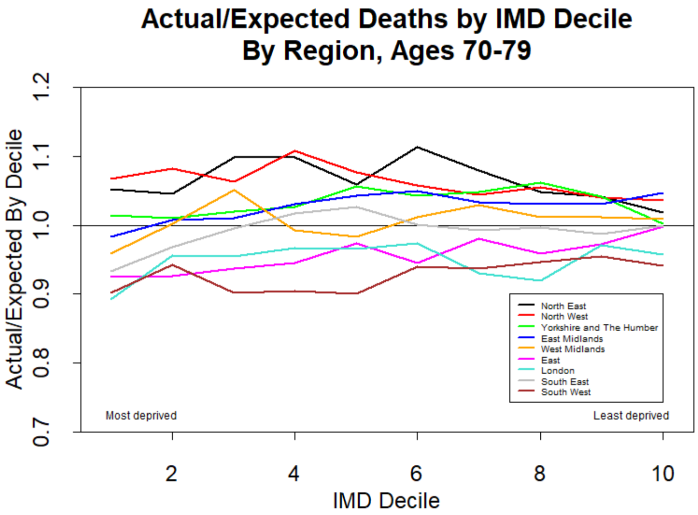
  where $\mathcal{L}(d)$ is the set of all LSOAs in IMD decile d
- Now compare actual versus expected deaths for subgroups within each decile.

IMD As a Predictive Variable: Urban-Rural Differences by IMD decile



- Compare A/E for urban-rural subgroups within each IMD decile
- E.g. **Rural** deaths in Decile 4 are **80%** of the average over all of Decile 4 **Conurbations (excl London)**: **110%**
- Significant differences between the middle deciles
- Convergence at either end
- Similar pattern at other ages

IMD As a Predictive Variable: Regional Differences



- Significant differences by region
- Smaller than urban-rural differences
- Different pattern from urban-rural

Objectives

- To publish a robust and reliable mortality index at the neighbourhood (LSOA) level that is open access
- To explain as much as possible of the variation that we observe in LSOA-level mortality using
  - publicly available data
  - socio-economic predictive variables
  - care-home population
  - urban-rural class
- To minimise unexplained urban-rural and regional differences
- To provide an open-access toolkit for actuaries
- To provide a tool and benchmark for mortality actuaries for comparison with other rating/valuation models
- To facilitate debate and action on how to tackle mortality inequality

Outline

- Introduction
- **Construction of the LIFE index**
  - Relative mortality risk - actual vs. expected deaths
  - Predictive variables
  - The Random Forest algorithm
  - Index construction
- Results
- The LIFE App

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

What we observe in each LSOA?

In each LSOA $i = 1, \ldots, N = 32,844$ we observe

- Deaths counts: $D(i, t, a)$ in calendar year t at age a
- Exposure (mid-year population estimates): $E(i, t, a)$ in calendar year t at age a
- $m = 12$ predictive variables, $X_1, \ldots, X_m$: deprivation/socio-economic data, care-home data and urban-rural class

Measuring Relative Mortality Risk - Actual vs. Expected Deaths

- First, define a baseline mortality rate $m_b(t, a)$ in year t at age a for England:

$$m_b(t,a) = \frac{\sum_{i=1}^{N} D(i,t,a)}{\sum_{i=1}^{N} E(i,t,a)} = \frac{\text{total deaths in England in year } t \text{ at age } a}{\text{total exposure in England in year } t \text{ at age } a} \qquad (1)$$

- Without any additional information the expected total number of deaths $\hat{D}^0(i)$ across a defined range of ages a (e.g. 70-79) and years t (e.g. 2001-2018) in LSOA i is then given by

$$\hat{D}^0(i) = \sum_{t,a} m_b(t,a) E(i,t,a)$$

- Define the empirical relative risk $R^0(i)$ for LSOA i:

$$R^0(i) = \frac{\sum_{t,a} D(i,t,a)}{\hat{D}^0(i)} = \frac{\text{actual deaths in LSOA } i}{\text{expected deaths in LSOA } i} \qquad (2)$$

Predictive variables used in construction of the LIFE index

Our aim is to predict the relative risk in an individual LSOA i using observed predictive variables available at the LSOA level:

x_1	old age income deprivation
x_2	employment deprivation (i.e. unemployment)
x_3	education deprivation
x_4	housing standard (number of bedrooms)
x_5	proportion of the population born outside the UK
x_6	deprivation in housing/living environment
x_7	employment/occupation: proportion in a management position
x_8	crime rate
x_9	proportion working more than 49h per week
x_{10}	proportion of population aged 60+ in a care home with nursing
x_{11}	proportion of population aged 60+ in a care home without nursing
x_{12}	urban-rural classification

Predictive variables used in construction of the LIFE index

There are five urban-rural classes, $x_{12} \in \{1, 2, 3, 4, 5\}$:

1. Urban conurbation (except London)
2. Urban city and town
3. Rural town and village
4. Rural hamlet and isolated dwellings
5. Urban conurbation (London only)

Predicting the Relative Risk using Explanatory Data - the LIFE index

- Mathematically, we model the conditional expectation of the relative mortality risk R^0 given characteristics x:

$$f(x) := \mathbb{E}[R^0|x] \text{ for any predictive vector } x \tag{3}$$

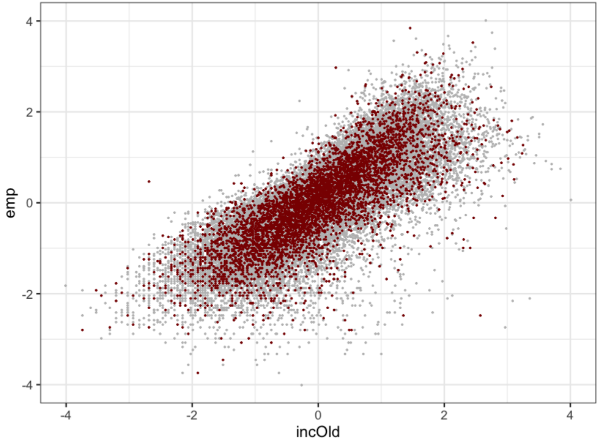
  where $x = (x_1, \dots, x_{12})$ is the vector of our twelve predictive variables.
- This is a regression problem - there are several statistical methods to estimate f.
- We estimate f with a random forest method – a supervised machine learning algorithm, and we will denote this estimator of f by $\hat{f}^{\text{RF}}$.
- Other estimation methods could be used, for example, kernel smoothers, local linear regression or fitting a parametric model to f.

Fitting the Relative Risk with a Random Forest

- We divide our data set into a training set and a validation set
- The function f is fitted to the training set only, and the fit is then evaluated using data from the validation set
- We use a Random Forest estimator to fit f.
- A Random Forest averages over a large number of random decision trees.
- For each decision tree:
  1. choose a random subset of LSOAs from the training set
  2. fit a decision tree to this subset only
  3. fitting involves a sequence of splits: each split is restricted to four randomly-selected predictive variables independent of previous splits
- Repeat steps 1 to 3 many times to get a "random forest" of individual decision trees and average over all the trees in the forest.

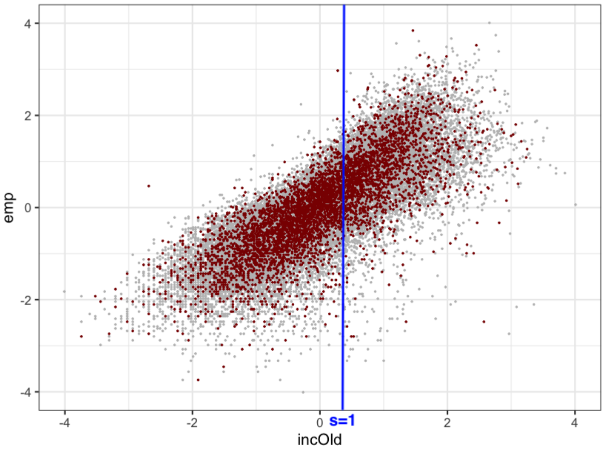
Stylised Example: Constructing an Individual Decision Tree

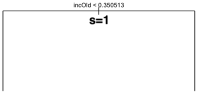
Consider only two predictive variables: old age income deprivation (incOld) and employment deprivation (emp)



- Grey dots: training set of LSOAs
- Red dots: subset of the training set used to construct tree b
- Each dot (LSOA) has its own empirical relative risk $R^0(i)$

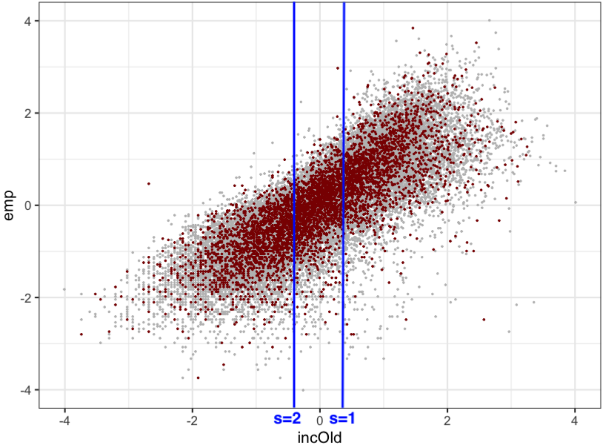
Constructing an Individual Decision Tree: First Split

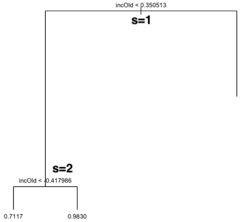




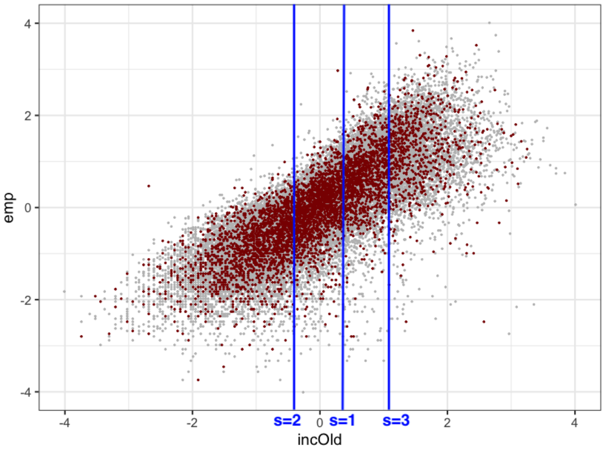
There are now two nodes: to the left and right of the $s = 1$ split

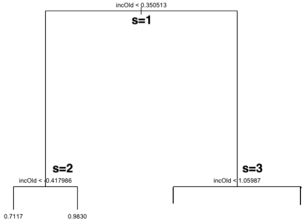
Constructing an Individual Decision Tree: Second Split

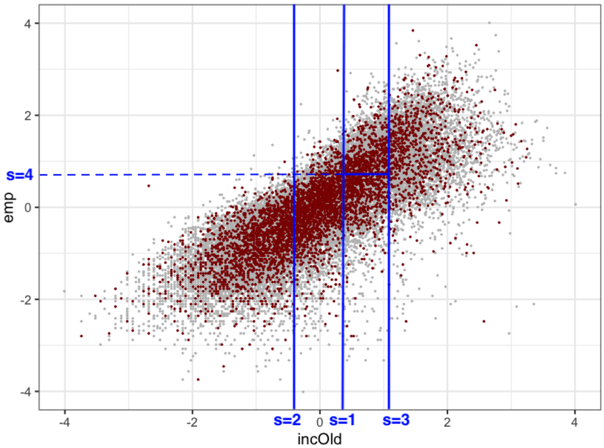


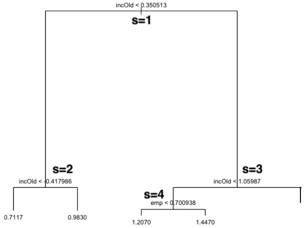


Constructing an Individual Decision Tree: Third Split



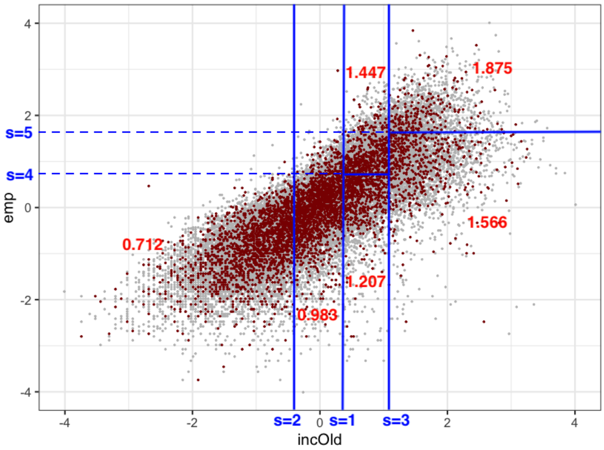


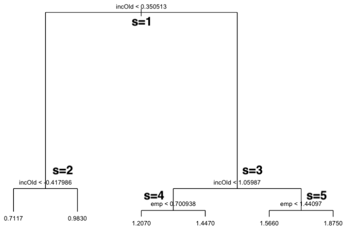




- Each split divides a single node into two, either horizontally ($s = 4$) or vertically
- Other nodes stay intact

Constructing an Individual Decision Tree: Fifth Split





We denote with $\hat{f}^{(b)}$ the piecewise constant function that represents one tree.
E.g. $\hat{f}^{(b)}(x) = 0.712$ for any x to the left of split $s = 2$.

- Our final random forest estimator $\hat{f}^{\text{RF}}$ for the regression function $f(x) = \mathbb{E}[R^0|x]$ is obtained by taking the average over all individual decision trees $\hat{f}^{(b)}$, that is,

$$\hat{f}^{\text{RF}}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{(b)}(x) \tag{4}$$

for any vector $x = (x_1, \dots, x_{12})$ of predictive variables.

- Note that $\hat{f}^{\text{RF}}$ is piecewise constant as it is an average over a finite number of piecewise constant regression tree functions $\hat{f}^{(b)}$. However, $\hat{f}^{\text{RF}}$ can take many more values compared to any individual tree $\hat{f}^{(b)}$.
- The random-forest estimator is more reliable and robust than individual decision trees

- We define the Longevity Index For England (LIFE) as the value of f for a specific LSOA using the socio-economic characteristics of this neighbourhood ...
- ... but replacing the proportion of people living in care homes with the average for the whole of England.
- More precisely, the LIFE index for LSOA i is

$$R(i) = f\left(\tilde{X}(i)\right) \text{ with } \tilde{X}(i) = (X_{i,1}, \ldots, X_{i,9}, \bar{X}_{10}, \bar{X}_{11}, X_{i,12}) \quad (5)$$

  where $\bar{X}_k$ = mean of all values of x_k.
- Use of $\bar{X}_{10}$ and $\bar{X}_{11}$ (carehomes) $\Rightarrow$
  two, otherwise identical, LSOA's that have different X_{10}, X_{11} will have the same LIFE index value.

For each sex, and a chosen range of calendar years (t_0, t_1) and ages (a_0, a_1):

1. Define the empirical relative risk $R^0(i)$ for LSOA i:

$$R^0(i) = \frac{\sum_{t,a} D(i,t,a)}{\hat{D}^0(i)} = \frac{\text{actual deaths in LSOA } i}{\text{expected deaths in LSOA } i} \qquad (6)$$

2. Construct an estimator $\hat{f}^{\text{RF}}(x)$ for the regression function

$$f(x) := \mathbb{E}[R^0|x] \text{ for any predictive vector } x$$

3. Estimate the index value as

$$\color{red}{R(i, \text{male/female}, \text{age})} \;\equiv\; R(i) \;=\; \hat{f}^{\text{RF}}\left(\tilde{X}(i)\right)$$

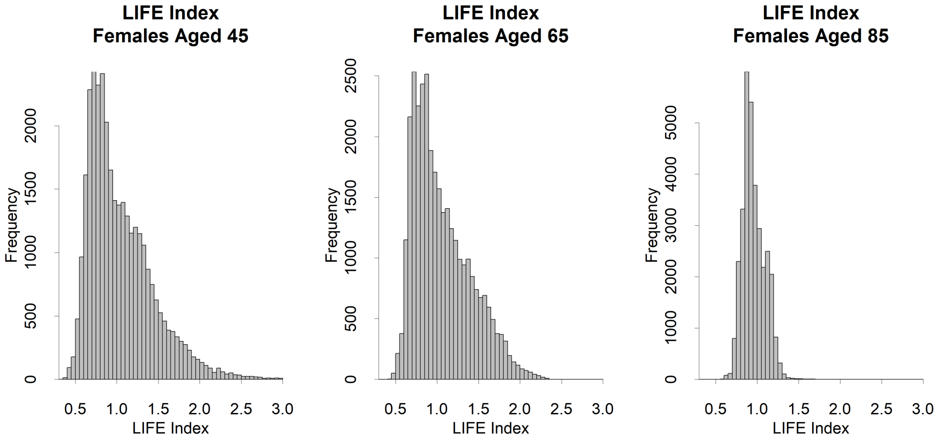
   where $\tilde{X}(i)$ are the socio-economic characteristics of LSOA i with the populations in care homes replaced by the national average

In our research the range of years is 2001 to 2018, and the age range is ten years centred around the age of interest

Outline

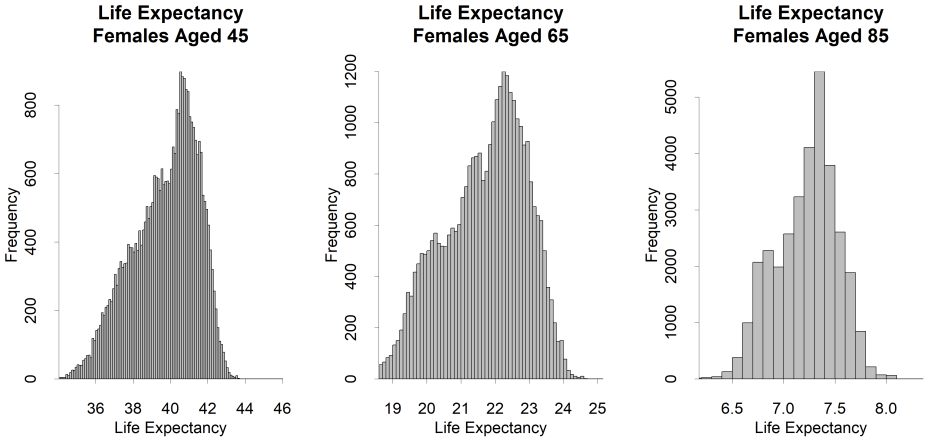
- Introduction
- Construction of the LIFE index
- **Results**
  - Distribution of index values
  - Regional variation
  - IMD vs. LIFE
- The LIFE App

How much inequality does the LIFE index reveal? Females



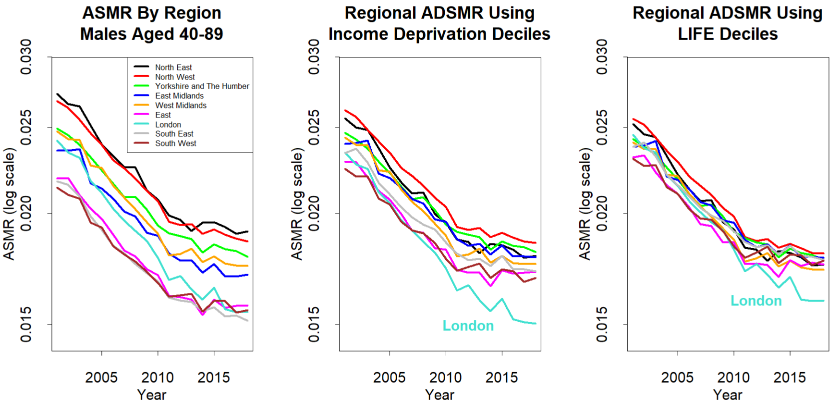
- Histogram: LIFE index values for 32,844 LSOAs, females aged 45, 65 and 85
- Very wide spread in the 40's; gradually narrowing with age

Inequality in period life expectancies: easier to interpret



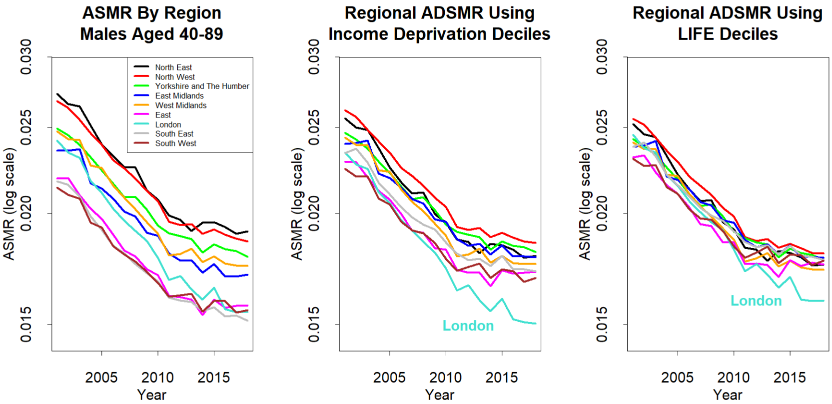
- Histogram: life expectancy values for 32,844 LSOAs, females aged 45, 65 and 85
- x-axis range is median $\pm 15\%$

The LIFE index explains much more of the regional variation



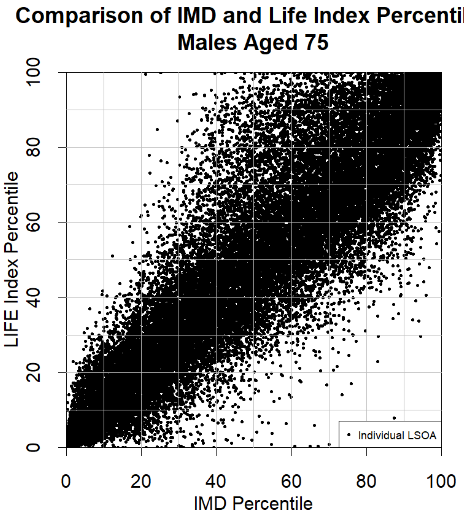
- Left: unadjusted regional ASMRs reflect underlying socio-economic variation
- Middle, right: Age *and Deprivation* Standardised Mortality Rate (ADSMR) adjusts for regional variation in Income Deprivation (middle) or LIFE (right)

The LIFE index explains much more of the regional variation



- **London** mortality has fallen at a faster rate than other regions
- Excluding London, using the LIFE index based on socio-economic and urban-rural predictive variables, a North/South divide is no longer obvious

How much of a difference between IMD and LIFE?

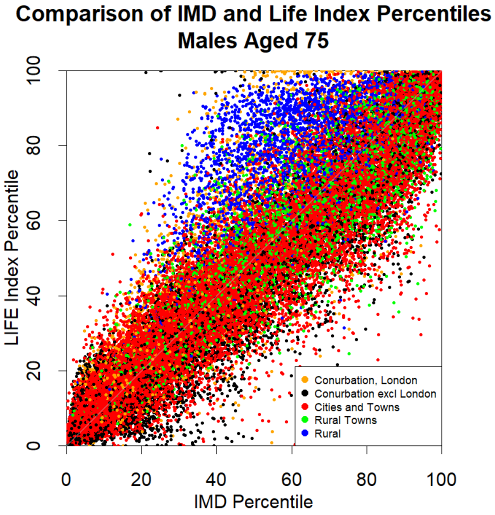


High percentile ⇒ low mortality

Difference in Deciles
E.g. LSOA in IMD-2 ⟶ LIFE-4
⇒ Difference = 2

Difference	Number of LSOAs
0	12,195
1	13,431
2	4,906
3	1,599
4	529
≥ 5	184

Urban-rural is a significant driver of differences



- **Rural LSOAs** have the biggest difference:
  significantly lower mortality than predicted by the IMD
- But significant differences in other urban-rural classes as well
- E.g. high **London** percentiles (low mortality) correspond mainly to Kensington & Chelsea

Adjusting for regional variation

There are two versions of the LIFE index available in the App. For each LSOA, i:

- **Core, version 1:**

$$m_1(i,a) = R_1(i,a)m_{\text{base}}(a)$$

  where $R_1(i,a)$ uses the Random Forest methodology
  and $m_{\text{base}}(a)$ is the base table, average mortality for England at age a
- **Version 2:** combined socio-economic and spatial relative risk
  Adjustments are made at the level of 106 Clinical Commissioning Groups (CCGs)
  (CCG's are responsible for the planning and commissioning of health care services for their local area)

$$m_2(i,a) = R_1(i,a)R_2(i,a)m_{\text{base}}(a)$$

  where $R_2(i,a) \equiv R_2(CCG(i),a)$ is an adjustment that reflects the actual deaths in each CCG relative to expected deaths based on socio-economic factors only ($R_1(i,a)$).
- *Variation in $R_2(CCG,a)$ is* much smaller *than variation in $R_1(i,a)$.*
- $R_1(i,a) \Rightarrow$ debate at the national level on mortalty inequality
- $R_2(CCG,a) \Rightarrow$ how do individual CCGs compare on a socio/urban, like-for-like basis?

LSOAs with the lowest and highest LIFE indices: $R_1(i, a)$

Top/bottom 20 LSOAs for both males and females include:

	Age 55	Age 75
Lowest	St Albans 005B	Kensington & Chelsea 012E
	Richmond upon Thames 012A	Westminster 019F
Highest	Blackpool 007C	Cannock Chase 010C
	N.E. Lincolnshire 002A	Bolton 025C
	Salford 024D	Welwyn Hatfield 010E

Recall: north/south split is the result of underlying socio-economic and urban-rural differences

Clinical Commissioning Groups: lowest/highest $R_2(CCG, a)$

Actual/Expected after applying the LIFE index:

	Age 55	Age 75
Low	Norfolk and Waveney	Thurrock
	Dorset	Barnsley
	Sheffield	Sheffield
	Ipswich & E. Suffolk	Southend
High	Portsmouth	Bury
	Warrington	Calderdale
	Southport & Formby	E. Staffordshire
	E. Staffordshire	Warrington

E.g. Norfolk and Waveney has lighter mortality than other CCGs with a similar socio-economic/urban-rural mix
E.g. Sheffield males have 8-9% lower mortality than the LIFE index predicts
Portsmouth males (50's, 60's) have 8-10% higher mortality than predicted
Most CCG's are within $\pm 5\%$.

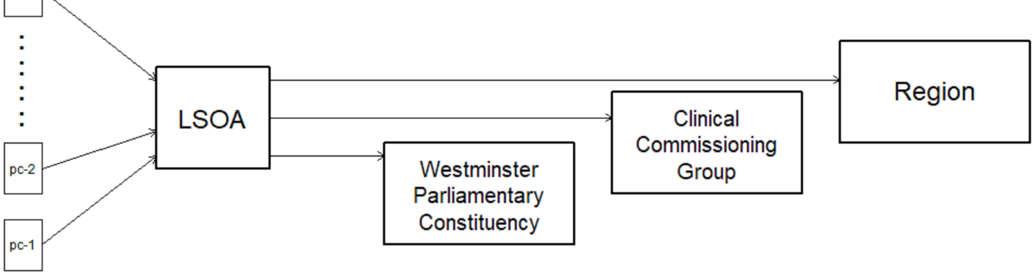
Outline

- Introduction
- Construction of the LIFE index
- Results
- **The LIFE App**
  - Overview
  - Examples and live demonstration

- The App has been developed to allow non-expert users to explore the index and discover mortality inequalities between different areas

- Information by LSOA: *residential* postcode → LSOA; or lookup table
- Maps by region, Clinical Commissioning Group (CCG), or Westminster parliamentary constituency
- User inputs and choices:
  - Males/Females
  - Age
  - Postcode → LSOA; region; CCG; constituency
  - Socio-economic-only index ($R_1(i,a)$), or with an additional adjustment ($R_2(CCG(i),a)$) for the local CCG

- Outputs include (the average for the LSOA)
  - LIFE index value → percentile and decile
  - Period remaining life expectancy (based on 2019 English mortality)



WPC, CCG and Regions all derive their LIFE index values from the LSOA values

The LIFE App

Managing your expectations:

- The LIFE App is currently a "Beta" version
  so we are very happy to get your feedback
  - what do you like about the app?
  - what might be improved?
- The app is hosted on a server with limited capacity
  lots of users might slow things down a bit
- We are not professional software developers!

Demo: ....

LIFE App: Index values at the postcode/LSOA level

Longevity Index for England – LIFE

Introduction and Disclaimer | Index by Postcode | Interactive Map

Show user guide

Show number of active users connected

Search by Full Postcode

M28 2GF

OR search by city/town/borough name and select LSOA

Age

55

Display index for:

- (•) Males
- ( ) Females

Choose index type:

- (•) Socio-economic relative risk only
- ( ) Socio-economic relative risk with regional adjustment

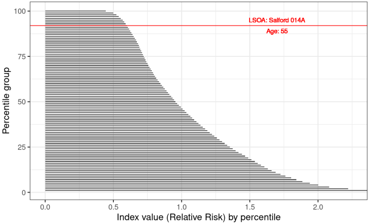
Show index

Check index value and life expectancy in the selected postcode area or LSOA, and relative position to other LSOAs in England.

Index and Life Expectancy of Selected LSOA

LSOA name	LSOA code	Region	Age	Index value	Index decile (*)	Index percentile (**)	Remaining life expectancy from age selected (in 2019) (***)
Salford 014A	E01005695	North West	55	0.6	10 /10	92 /100	29 years and 9 months

Relative Position of Selected LSOA to Others by Index Value



* **Decile 1:** 10% highest index value, i.e. highest relative mortality risk; Decile 10: 10% lowest relative mortality risk.
** **Percentile 1:** 1% highest index value, i.e. highest relative mortality risk; Percentile 10: 1% lowest relative mortality risk.
*** **Life expectancy:** Calculated as period life expectancy in 2019 using indexed mortality data up to age 104.

High decile or high percentile ⇒ low mortality/high life expectancy

LIFE App: Interactive Map – Region

Longevity Index for England – LIFE

HERIOT WATT UNIVERSITY | Actuarial Research Centre, Institute and Faculty of Actuaries | Institute and Faculty of Actuaries

Introduction and Disclaimer | Index by Postcode | Interactive Map

Show user guide

Show number of active users connected

View by:

- (•) Region
- ( ) Clinical Commissioning Group (CCG)
- ( ) Parliamentary Constituency (PC)

Region to display

North West

Age

65

Display index for:

- ( ) Males in England
- (•) Females in England

Choose index type:

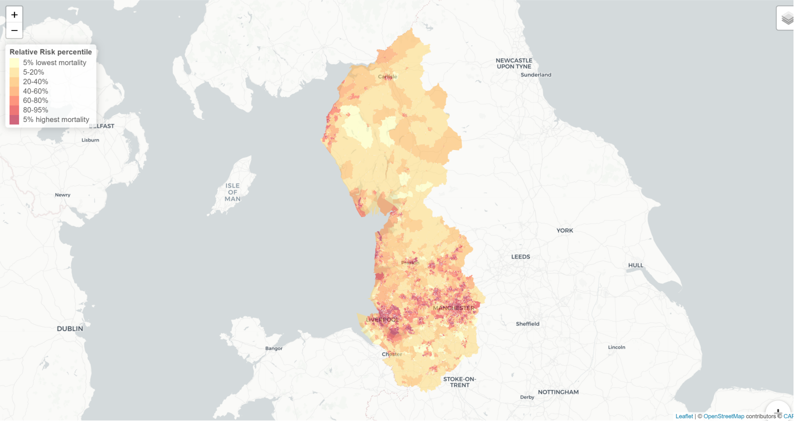
- (•) Socio-economic relative risk only
- ( ) Socio-economic relative risk with regional adjustment

Update map and metrics

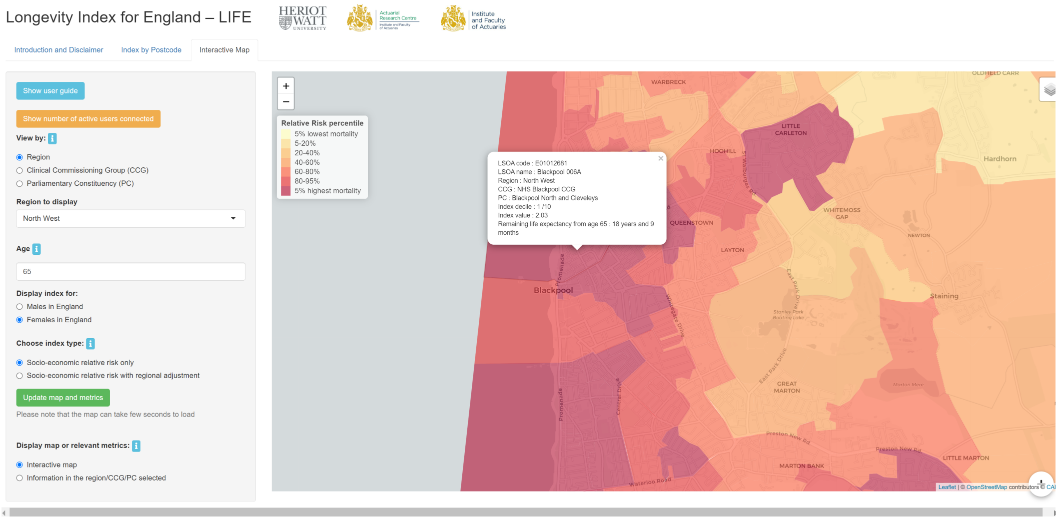
Please note that the map can take few seconds to load

Display map or relevant metrics:

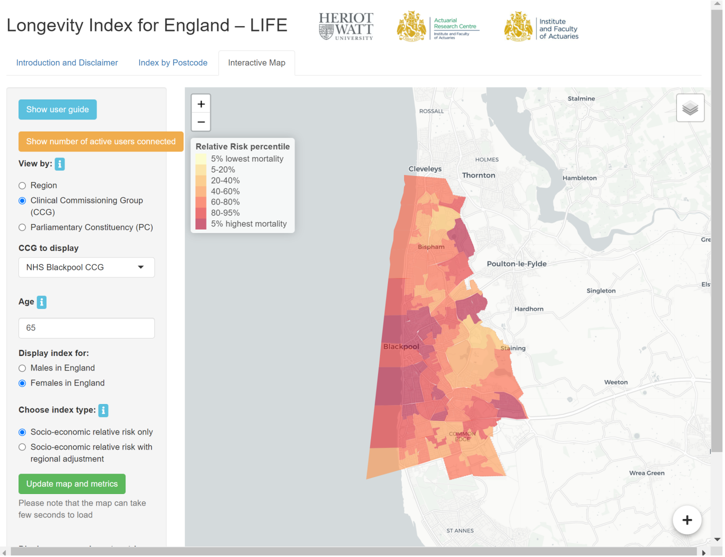
- (•) Interactive map
- ( ) Information in the region/CCG/PC selected



LIFE App: Interactive Map – zoom in and click on an LSOA



LIFE App: Interactive Map – CCG



LIFE App: Interactive Map – Information on the region/CCG/PC

Longevity Index for England – LIFE

Introduction and Disclaimer | Index by Postcode | Interactive Map

Show user guide

Show number of active users connected

View by:

- Region
- Clinical Commissioning Group (CCG)
- Parliamentary Constituency (PC)

CCG to display

NHS Blackpool CCG

Age

65

Display index for:

- Males in England
- Females in England

Choose index type:

- Socio-economic relative risk only
- Socio-economic relative risk with regional adjustment

Update map and metrics

Please note that the map can take few seconds to load

Display map or relevant metrics:

- Interactive map
- Information in the region/CCG/PC selected

Summary statistics of the selected region/CCG/PC

CCG name	NHS Blackpool CCG
CCG code	E38000015
Age selected	65
Number of LSOAs (*)	94
Average index value (**)	1.38
Rank of average index value in England	5 out of 106 CCGs
Average remaining life expectancy from selected age in 2019 (***)	20 years and 6 months

* Total number of LSOAs in the region/CCG/PC selected.
** Average over index values in all LSOAs within the region/CCG/PC selected, weighted by the LSOA-level average population size over 2001-2018 and of the age selected.
*** Average of LSOA-level period life expectancy within the region/CCG/PC selected, weighted by the LSOA-level average population size over 2001-2018 and of the age selected.

Thank You!

Questions?

Constructive comments and feedback on the LIFE index and the App are very welcome:

E: A.J.G.Cairns@hw.ac.uk
E: T.Kleinow@hw.ac.uk
W: tinyurl.com/LIFEindex
W: www.macs.hw.ac.uk/∼andrewc/ARCresources