Version 1.1

Datasets to Accompany the ARC Sessional Research Paper:

Drivers of Mortality: Risk Factors and Inequality

by Andrew J.G. Cairns, Torsten Kleinow and Jie Wen

The paper, supplementary plots and accompanying data are available online at

`www.macs.hw.ac.uk/~andrewc/ARCresources/Sessional2020/`

1   Introduction

Data for England are available at the level of small geographical areas known as Lower Super Output Areas (LSOAs). Each area has typically between 1000 and 3000 persons, with an average of 1500 to 1600, across a range of ages.

There are 32844 LSOAs at present. New LSOAs are created from time to time in response to growth in housing.

For each LSOA we have:

- data from 2001-2016:

  - death counts, $D(i, t, x)$, where i is the LSOA, t is the year, and x is the age last birthday;

  - central exposed to risk, $E(i, t, x)$, which are derived from the mid-year population estimates for 2001-2016;

  - Note that the exposures, necessarily are mid-year *estimates*. As a result, in a small number of cases there are zero exposures but one or more deaths. Even if mid-year population estimates are accurate, the death count could still be greater than zero if the deceased individual(s) died early in the year. Modelling, therefore, needs to take account of this, either by modelling errors in the exposures, or by aggregating data over years and ages.

- One off (i.e. not observed through time) *predictive variables* for each LSOA that might be associated with higher or lower than average rates of mortality.

2   Datasets available in R

The webpage `www.macs.hw.ac.uk/~andrewc/ARCresources/Sessional2020/` has the single R workspace `SessionalJanuary2020Data.Rdata` available for download.

The R workspace contains the following data.

1. Gender [R workspace: Gender]

   - Character vector of length 2: ("Male","Female") linked to the first dimension in the deaths and exposures

2. Year [Year]

   - observation years in the deaths and exposures data
   - 2001 to 2016

3. Age [Age]

   - ages covered in the deaths and exposures data
   - 40 to 89

4. LSOA index [lsoa]

   - This is a vector of length 32844
   - The values range from 1 to 33768
   - The values correspond to the official code for each LSOA
   - The Official code is 9 characters long.
   - The first four characters are always "E010".
   - The last five characters are the value for the LSOA that are recorded in my vector.
   - Thus if the value is 1, then the LSOA code is "E01000001" and if the value is 33768, then the LSOA code is "E01033768".
   - Each code is only used once at most.
   - Some codes are missing. These are codes that would have been used previously. However, if an LSOA has grown substantially, then it would be split, the old LSOA code deleted, and the two new LSOAs given new codes not yet used.

5. Death counts: [Deaths]

   - 4-dimensional array for males and females
   - gender x lsoa x year x age
   - 2 x 32844 x 16 x 50
   - each of the 32844 LSOAs corresponds to the LSOA index vector above

6. Central exposed to risk (derived from mid-year population estimates) [Exposures]

   - 4-dimensional array for males and females
   - gender x lsoa x year x age
   - 2 x 32844 x 16 x 50

7. Income deprivation [IncomeDeprivation]

   - this measures the proportion of the population in each LSOA who are receiving benefits from the state because they are on a low income;
   - the data are in a vector of length 32844: one entry for each LSOA.

8. Income deprivation affecting older people [IncomeDeprivationOld]

   - this measures the proportion of the population above age 60 in each LSOA who are receiving benefits from the state because they are on a low income;
   - the data are in a vector of length 32844: one entry for each LSOA.

9. Employment deprivation [EmploymentDeprivation]

   - this measures the proportion of the *working* population in each LSOA who are unemployed;
   - the data are in a vector of length 32844: one entry for each LSOA corresponding to the vector of 5-digit LSOA codes above.

10. Living environment deprivation [LivingDeprivation]

    - this measures the quality of the living environment (indoors and outdoors);
    - indoors: (poor) quality of housing;
    - outdoors: e.g. (poor) air quality and traffic accidents;
    - the data are in a vector of length 32844: one entry for each LSOA.

11. Barriers to housing and services (deprivation)

    - like living environment deprivation, this measures a number of different things;
    - this measures 'wider barriers' and 'geographical barriers';
    - wider barriers includes overcrowding in households and homelessness; [WiderBarriers]
    - geographical barriers measures distance to key services; [GeoBarriers]

- although a higher value for geographical barriers implies more 'deprived', it can also be associated with lower mortality; for example, greater distances to services might indicate that the LSOA is more affluent or rural with housing more spaced out; in fact, the geographical barriers variable is negatively correlated with income deprivation;

- the data are in a vector of length 32844: one entry for each LSOA.

12. Average number of bedrooms [AverageBed]

- this measures the average number of bedrooms per household in the LSOA

- the data vector has been standardised to a $N(0, 1)$ distribution;

- in contrast to the deprivation indices, a high value (more bedrooms) is likely to be associated with lower mortality;

- the data are in a vector of length 32844: one entry for each LSOA.

13. Highest level of qualification [EducationLevelProportions]

- gives the proportion of a particular group within the LSOA who have attained a particular level of education

- 3-dimensional array for males and females combined

- lsoa x age-group x education level

- 32844 x 6 x 8

- 32844 LSOAs

- 6 age groups: All; 16 to 24; 25 to 34; 35 to 49; 50 to 64; 65 plus;

- 8 education groups:

  (a) All categories: Highest level of qualification
  (b) No qualifications
  (c) Level 1 qualifications
  (d) Level 2 qualifications
  (e) Apprenticeship
  (f) Level 3 qualifications
  (g) Level 4 qualifications and above
  (h) Other qualifications

- see
  www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels;

- you can use the education data to construct a vector of predictive variables: e.g.

- the proportion of people in the LSOA aged 50-64, who have no qualification or level 1 only;
- an average level of educational attainment in a particular age group;

14. Occupation group proportions [OccupationGroupProportions]

- gives the proportion of a particular group within the LSOA who have a particular type of occupation
- 4-dimensional array
- gender x lsoa x age-group x occupation group
- 2 x 32844 x 14 x 9
- 2 genders (1=males, 2=females)
- 32844 LSOAs
- 14 age groups: All; 16-19; 20-24; 25-29; 30-34; 35-39; 40-44; 45-49; 50-54; 55-59; 60-64; 65-69; 70-74; 35-64
- most age groups are small, so there will be a lot of sampling variation, weakening their predictive ability. This is less of a problem for the 35-64 age group;
- 9 occupation groups

  (a) Higher managerial, administrative and professional occupations
  (b) Lower managerial administrative and professional occupations
  (c) Intermediate occupations
  (d) Small employers and own account workers
  (e) Lower supervisory and technical occupations
  (f) Semi-routine occupations
  (g) Routine occupations
  (h) Never worked, long-term unemployed and full-time students
  (i) Total: NS-SeC

- you can use the occupation data to construct a vector of predictive variables: e.g.

  - the proportion of people in the LSOA aged 35-64, who fall into the "higher managerial" group;

15. Communal establishments [COMM]

- gives the *proportion* of the population aged 60+ within the LSOA who live in a particular type of communal establishment
- 2-dimensional array
- lsoa x type of communal establishment

- 32844 x 4
- 32844 LSOAs
- 4 types:
    - Care home without nursing (public and private)
    - Care home with nursing (public and private)
    - Other medical establishment
    - Other communal establishment

16. Urban Rural Classification [UrbanRural]

    1 Conurbation: non London

    2 City or town

    3 Rural town

    4 Rural village and dispersed

    5 Conurbation: London

    - the data are in a vector of length 32844: one entry for each LSOA.

17. Region [Region]

    1 North East

    2 North West

    3 Yorkshire and Humber

    4 East Midlands

    5 West Midlands

    6 East

    7 London

    8 South East

    9 South West

    - the data are in a vector of length 32844: one entry for each LSOA.