Drivers of Mortality: Risk Factors and Inequality

Andrew J.G. Cairns[a], [b] Torsten Kleinow[a] and Jie Wen[a]

First version: 14 December 2019

This version: June 7, 2021

Abstract

This paper takes a detailed look at socio-economic variation in mortality across England. Local linear regression is used to analyse all-cause mortality at neighbourhood level (Lower Layer Super Output Areas) with mortality rates linked to a number of socio-economic predictive variables that determine the character of a neighbourhood. We find that income and employment deprivation are key determinants of mortality, but also that urban-rural class and the presence of care homes in a neighbourhood have an important role to play in assessing underlying mortality rates relative to national mortality. Residual spatial/regional variation in mortality is found to be much less important than socio-economic variation and much lower than the residual regional variation that results from the commonly-used Index of Multiple Deprivation. Based on these results, we propose the $LIFE$ index (Longevity Index For England) and $LIFE$ deciles as a new tool for assessing all-cause mortality.

Keywords: Mortality inequality; Lower Layer Super Output Area; $LIFE$ indices; Age and Deprivation Standardised Mortality Rate; Regional mortality variation; Local linear regression.

[a]Maxwell Institute for Mathematical Sciences, and Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK.

[b]Corresponding author: A.J.G.Cairns@hw.ac.uk

1 Introduction

1.1 Socio-economic differences in mortality

It is well known that socio-economic status and associated variables are strongly correlated with high and low mortality. The evidence base for this has been growing over a number of decades as data become available with sufficient detail to be able to investigate the dependency between particular predictive variables and mortality. Socio-economic status is not always easily available in the same form but typical information about individuals or groups includes income, income deprivation or affluence (e.g. Chetty et al., 2016 [US], Villegas and Haberman, 2014, Longevity Science Panel, 2018 [UK], Cairns et al., 2019 [Denmark], Wen et al., 2020, 2021 [UK, Canada]) and education (e.g. Mackenbach et al., 2003, 2016).

This paper seeks to delve more deeply into the the links between different measures of socio-economic status, geographic location and mortality. We do this by exploiting a large dataset for England built up from multiple datasets obtained from the UK's Office for National Statistics (ONS).

A key modelling tool in the section on all-cause mortality is *local linear regression*. As a statistical method, a version was first proposed by Cleveland (1979) and is often referred to as LOWESS (Locally Weighted Scatterplot Smoothing). As we demonstrate in this paper, it is an effective tool to handle large datasets with many covariates. Compared to e.g. Generalised Linear Models (see, for example, Macdonald et al., 2018) the model fits the data effectively without having to specify in advance the functional relationship between predictive variables (including interactions) and the outcome.

1.2 Questions addressed in this paper

This paper contains a wide ranging analysis of all-cause and cause-of-death data as part of this we seek to address the following questions:

- What are the most significant socio-economic factors that influence mortality rates?
- Does it make a difference if a neighbourhood is in an urban or rural area?
- Where are care homes located and what impact do they have on mortality?
- Can regional differences in mortality be explained entirely by differences in the socio-economic mix and other non-spatial predictive variables?
- After socio-economic and non-spatial effects have been filtered out, what remains in terms of spatial or regional variation in mortality across England?

- How much inequality is there in mortality rates at different ages?
- Have mortality inequalities been widening over the last 16 years?

In answering some of these questions, we develop a new index: the Longevity Index For England, or $LIFE$. $LIFE$ is, in fact, a group of indices based on different age groups and sex. It can be adopted as a continuous index that actuaries and other stakeholders can use as a predictive variable alongside other variables that are available at the individual level such as pension level and geodemographic profiling (e.g. Richards, 2008). Additionally, it can be used to divide the population up into deciles instead of using the Index of Multiple Deprivation or income deprivation.

1.3 Outline of the paper

In Section 2 we introduce the several datasets that we will use in our various analyses, with further details in Appendix A. Section 3 sets the scene for what is to come with some introductory analysis of the data. This also introduces the problem of regional variation in mortality. We then take a deep dive into all-cause mortality at the small-neighbourhood level in Section 4. Section 5 concludes.

2 Data

Data used in this study have been sourced from the UK's Office for National Statistics (ONS), and are available at the level of Lower Layer Super Output Areas (LSOAs; small, socially-homogeneous, geographical areas with, on average, 1600 people) and are of the following types:

- mid-year population data by LSOA;
- all-cause death counts by LSOA;
- predictive variables by LSOA.

There are 32,844 LSOAs across England.

We outline key elements of the data below, with further details to be found in Appendix A.

2.1 Population data

For each LSOA, we have mid-year population estimates from the ONS by single year of age and single calendar years from 2001-2018, $E(g, i, t, x)$, where i is the

LSOA index, g is the sex, t is the calendar year, and x is the age last birthday. As with standard practice we equate this with the central exposed to risk for age x last birthday across the whole of year t.[1]

Exposures can be aggregated into specified deciles, regions or the whole of England as required.

2.2 All cause mortality

For all cause mortality, the ONS make available death counts by sex, LSOA, single age and single year, $D(g, i, t, x)$.

Death rates at the LSOA level are then

$$m_L(g, i, t, x) = \frac{D(g, i, t, x)}{E(g, i, t, x)}$$

and at the regional level for region r

$$m_R(g, r, t, x) = \frac{\sum_{i \in R(r)} D(g, i, t, x)}{\sum_{i \in R(r)} E(g, i, t, x)}$$

where the set $R(r)$ contains all LSOA's, i, that lie within region r.

Similar expressions can be written down for decile-based death rates, $m_D(g, j, t, x)$, for decile j, and national mortality, $m(g, t, x)$.

2.3 Migration

It needs to be noted that the exposures data do not contain information about migration between LSOAs or between regions[2]. This means that caution needs to be exercised when viewing death rates across age groups as there might be some variation in the underlying population mix.

2.4 Predictive variables

The focus of this paper is to identify socio-economic and other variables that have significant predictive power in estimating mortality rates. A significant number of variables are available at the LSOA level. A key source of predictive variables is the

[1] We have not adjusted exposures at this stage as proposed by Cairns et al. (2016). However, even if systematic, cohort-based errors exist at the total population level it is unlikely that this would have an impact on our assessments of relative risk that are the focus of this paper.

[2] However, it is possible to assess net migration by comparing changes in exposures by cohort with death counts but not determine where individuals are coming from or going to.

Index of Multiple Deprivation (see ONS, 2015) and its domains and subdomains including income deprivation, employment deprivation, education deprivation, crime, barriers to housing and services, and living environment. Additionally, we considered: average number of bedrooms; level of educational attainment by age group; occupation group by age group; proportion born in the UK; proportions, by age group, of people living in various types of communal establishment; region; and urban-rural class (UR1 conurbation non-London, UR2 city or town, UR3 rural town, UR4 rural village and dispersed, and UR5 conurbation London). Further details of these can be found in Appendix A.

3 Preliminary analysis of all-cause mortality

Before we move into a more detailed analysis on the various datasets, it is worthwhile summarising the main patterns that exist in socio-economic data. Building on the recommendations of the Longevity Science Panel (LSP 2018) we will, here, use income deprivation rather than the main IMD to divide the LSOAs into deciles. Income deprivation is, of course, not a causal effect, but it was found by LSP to be more strongly *correlated* with high and low mortality than other domains of the IMD.[3]

In Figure 1 we have plotted age-specific death rates in 2018 for each of the income-deprivation deciles. For both males and females we can make the following observations:[4]

- Even though the data contain a certain amount of sampling variation, the deciles are clearly ranked from decile 1 (most deprived, highest mortality) down to decile 10 (least deprived, lowest mortality).
- There are very significant differences between the deciles at ages 40 to 60 (the mortality inequality gap) before gradually narrowing as the population gets older. This narrowing is very typical for mortality differences between socio-economic groups using different measures and in different countries (see, for example, Cairns et al., 2019, Wen et al., 2020, Redondo Lourés and Cairns, 2020).

  At ages 40 to 50, death rates in the most deprived group are around 4 times (males) and 3 times (females) the corresponding death rates in the least deprived group. Both narrow to about 1.4 times at age 89.

[3]Note that the IMD itself includes a health domain. As with LSP (2018) we do not consider the health domain. Instead, we seek to identify socio-economic covariates that are predictors of poor health and increased mortality.

[4]For a more detailed analysis of English deciles, see Wen et al. (2021).

The importance of income deprivation is investigated further in Figure 2. In this figure we consider centiles rather than deciles, and use Age Standardised Mortality Rates (ASMRs) over ages 65 to 89 rather than crude age-specific death rates to dampen sampling variation.[5]

We see that income deprivation has a clear impact across all centiles, and, indeed, steepens towards both edges.

- At the right hand end (high deprivation and high mortality), the steeper curve might reflect the possibility that prolonged ill-health drives some people into more deprived areas.
- At the left hand end for the least deprived (which we interpret as most affluent), the rationale for a steepening curve is less clear, although (speculating) it might be that higher levels of wealth might facilitate better health care in old age. This pattern at the level of centiles can also be seen in US data (Chetty et al., 2016).

[5]For an overview of ASMRs, see Appendix B.

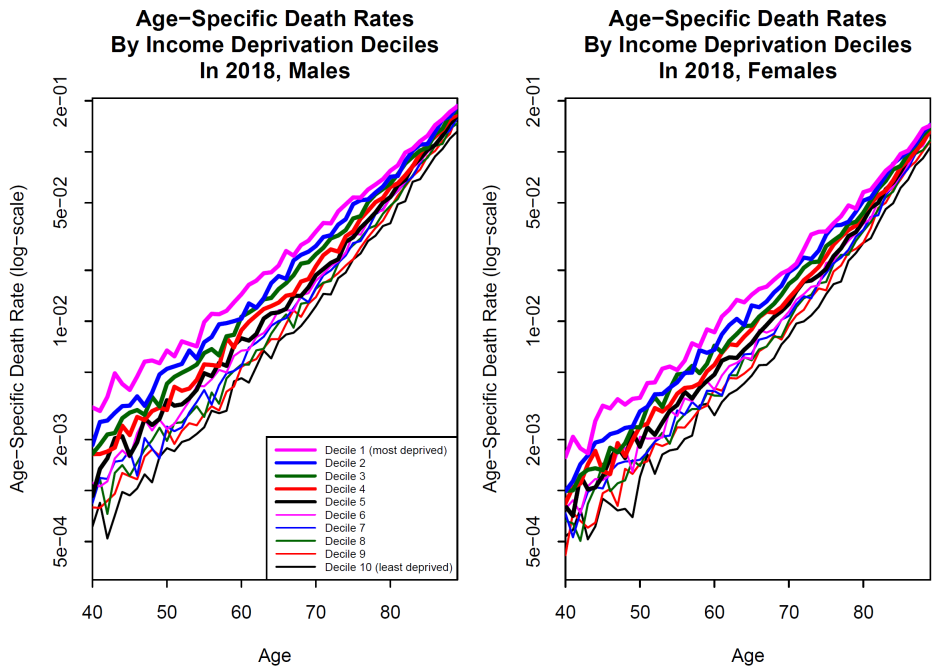

Figure 1: Age-specific death rates for English males and females in 2018, by income-deprivation deciles.

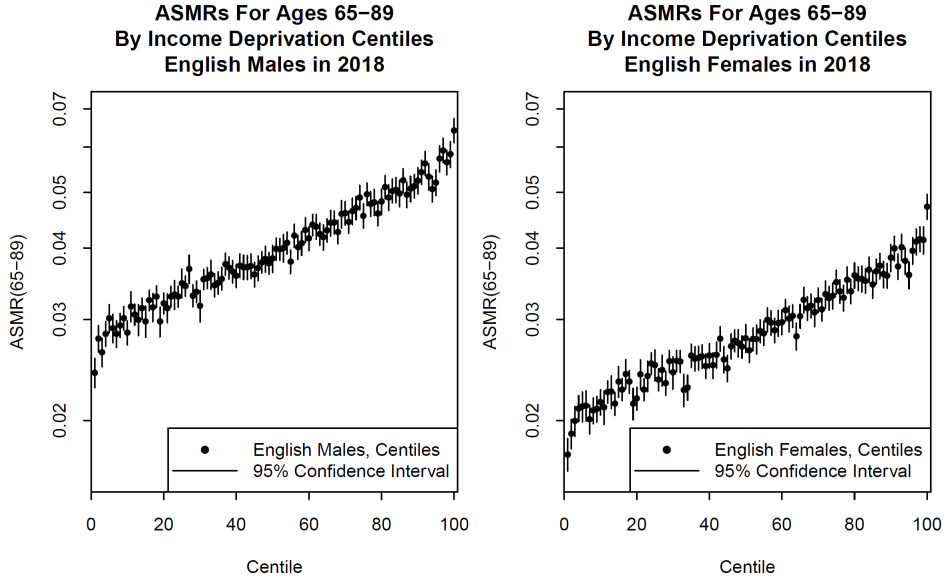

Figure 2: Estimated ASMRs in 2018 by income-deprivation centile for ages 65 to 89. Centile 1: least deprived. Centile 100: most deprived. Bars show approximate 95% confidence intervals for ASMR estimates.

In Figure 3, we illustrate how the ASMRs for the deciles have changed over time for two age groups: 40 to 64 and 65 to 89. These plots reveal the following features:

- Similar to Figure 1, the gap in ASMRs between groups 1 and 10 is wider in the upper plots for younger ages and that females' mortality is lower.
- We can see a widening gap between groups 1 and 10. This is more marked for females and for the older age group.
- Mortality improvements can be seen to have slowed down since 2010 or 2011 in all four sub-plots. However, assessing the impact on deciles 1 to 10 needs care. The slowdown is more marked in the older age group. And it is also more marked in the most deprived groups, even after taking account of the fact that they had been experiencing slower improvements since 2001.

A clear takeaway from Figure 3 concerns the setting of future mortality improvement rates in applications such as population projections and actuarial valuations. Specifically, short-term improvement rates should be different for different socio-economic groups, with higher short-term improvement rates for the least deprived groups.

Lastly, we consider how mortality rates vary from region to region across England. In Table 1 we give the ratio of actual versus expected deaths by region using English national mortality for expected deaths. Corresponding values by income-deprivation decile are given in Table 2 for comparison.

We can see significant differences between regions leading to the well-publicised North/South divide. But we can also see that differences by region are dwarfed by differences between income-deprivation deciles.

This leads to a key objective of this paper:
to what extent are regional differences explained by socio-economic differences between the regions?

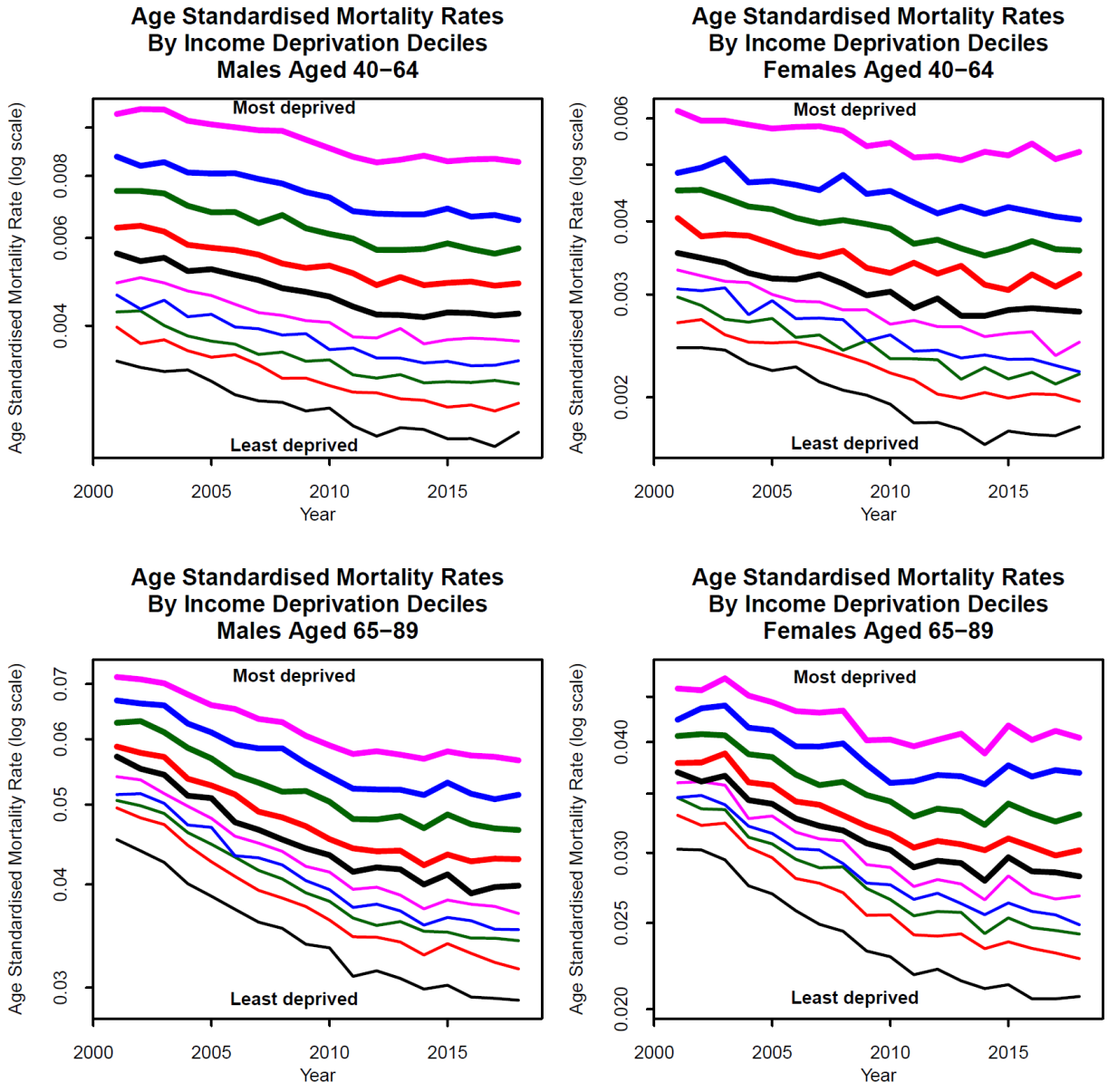

Figure 3: Estimated ASMRs by income-deprivation deciles over the period 2001 to 2018 for males and females and for age groups 40 to 64 (top) and 65 to 89 (bottom).

Region	Regional Relative To National Mortality (%) Males	Females
North East	115.5	120.5
North West	112.9	116.1
Yorkshire and The Humber	107.6	108.3
East Midlands	101.8	102.1
West Midlands	104.2	102.6
East	91.5	92.5
London	99.5	95.2
South East	90.4	89.3
South West	89.2	87.3

Table 1: Comparison of regional death counts over the age range 70 to 79, and 2001-2018 versus expected deaths using English national mortality.

Income Deprivation Decile	Decile Relative To National Mortality (%) Males	Females
1	148.2	146.8
2	133.2	130.0
3	121.3	119.2
4	110.6	109.4
5	101.8	101.0
6	94.3	93.0
7	88.1	88.5
8	83.9	83.9
9	79.1	78.2
10	70.1	69.6

Table 2: Comparison of mortality rates for individual income-deprivation deciles over the age range 70 to 79, and 2001-2018 versus English national mortality.

4 All-cause mortality by LSOA

This section will consider *relative risk*: a scaling factor that can be applied to standard mortality rates that takes account of specified predictive variables.

For each LSOA we have, as outlined previously:

- deaths and exposures, $D(i,t,x)$ and $E(i,t,x)$;
- a vector of predictive variables, $X(i)$.

Our challenges are:

- Which predictive variables are the best at predicting high or low mortality, and at which ages?
- How many predictive variables do we need to get a reasonable model?
- Can we identify an effective approach to estimate relative risks using these predictive variables?

We will first outline the non-parametric method of *local linear regression* for assessing the relative risk for a specific LSOA. We will then use the method to discuss which covariates are most useful, and what purpose they play. And finally we will look at the results, including a look at residual regional effects.

4.1 Multivariate local linear regression

In general, we have a set of data $(X(i), Y(i))$ for $i = 1, \ldots, L$ where $X(i)$ is an $n_P \times 1$ vector of predictive variables and $Y(i)$ is the response variable with $E[Y(i)|X(i)] = f(X(i))$ for some continuous and smooth function $f : \mathbb{R}^{n_P} \to \mathbb{R}$. Observation errors around $\hat{Y}(i) = f(X(i))$ are not assumed to be homoscedastic.

Local linear regression attempts to estimate $f(x)$ for a general x as follows:

- Minimise over the scalar $a \equiv a(x)$ and vector $b \equiv b(x)$ the weighted sum of squares
$$S(a, b; x) = \sum_{j=1}^{L} w(x, j) \left(Y(j) - a - b^T X(j)\right)^2$$
where, for each x, the weights tend to zero for further $X(j)$ is from x. In other words, we fit a local regression plane through points that are in the neighbourhood of x.

- The resulting least-squares estimators, specific to each x, are $\hat{a}(x)$ and $\hat{b}(x)$ and
$$\hat{Y}(x) = \hat{a}(x) + \hat{b}(x)^T x.$$
Specifically, we then have $\hat{Y}(i) = \hat{a}(i) + \hat{b}(i)^T X(i)$ based on weights $w(i,j) \equiv w(X(i),j)$. However, the more general form of $\hat{Y}(x)$ allows us, for example, to carry out out-of-sample testing.

- We only make use of this specific point estimate and discard further information about $\hat{a}(i)$ and $\hat{b}(i)$.

4.2 Advantages of local linear regression

A key reason for using local linear regression rather than, for example, kernel smoothing is that it captures the local slope in the data (see the last bullet in Section 4.1). This is important where the response variable (e.g. the relative risk) is believed to be increasing or decreasing as the underlying predictive variables change. This is specifically useful for any data point i whose (socio-economic) neighbours are mostly to one side rather than evenly distributed round about: e.g. near the edges of the data.

Local linear regression offers an approach that can be easily implemented in multiple dimensions. This is in contrast to some alternatives such as P-splines, which can be numerically very intensive or challenging as the number of dimensions increases.

An additional advantage of local linear regression is that estimates of relative risk are not especially sensitive to non-linear transformations of the predictive variable, X (e.g. a log-transform). The exception to this might be at the edges of the data when those points with an appreciable weight, $w(i,j)$, are spread out over a wider range of values of X. This contrasts, e.g. with Generalised Linear Models (GLMs), where, by design, the relative risk is log-linear in X: so X needs to be scaled and transformed accordingly before model fitting to get a good fit.

Finally, local linear regression automatically captures interactions between different predictive variables in multiple dimensions. With GLMs, interactions must be investigated in a systematic way, and this could be very challenging computationally in multiple dimensions.

4.3 Local linear regression and socio-economic mortality

We now move from the general formulation to the mortality setting.

Define $m(t,x)$ to be the *crude* age-specific national death rate in year t, age x. We now seek to model the *underlying* LSOA-specific death rates, $m(i,t,x)$, in LSOA i

linking to a vector, $X(i)$, of predictive variables. The death rate satisfies the usual assumption that

$$E[D(i,t,x)] = m(i,t,x)E(i,t,x).$$

Our general model is then that, *over the limited age range* (x_0, x_1) *and range of years* (t_0, t_1),

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

where

- $F_1(i)$ is the relative risk due to the socio-economic (and other) characteristics of LSOA i;
- $F_2(i)$ is a relative risk that captures residual spatial effects once the $F_1(i)$ have been fitted.

Having a constant $F_1(i)$ and $F_2(i)$ over a range of ages and years then means that if we define

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

and

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

then

$$E[D(i)] = F_1(i)F_2(i)\hat{D}_0(i)$$

where

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

In taking this approach,

- $\hat{D}_0(i)$ represents the baseline expected deaths with no allowance for socio-economic or other effects.
- $\hat{D}_0(i)F_1(i)$ represents our best estimate of the expected deaths based on socio-economic drivers alone.
- $\hat{D}_0(i)F_1(i)F_2(i)$ represents our best estimate of the expected deaths based on soecio-economic and spatial drivers.

4.4 Stage 1: estimate the socio-economic relative risk, $F_1(i)$

To fit this model we, first, calculate the actual-over-expected based on the baseline model that there are no socio-economic effects:

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

We then use the $R_0(i)$ to derive estimates of the $F_1(i)$ making use of the socio-economic predictive variable $X(i)$.

Unadjusted predictive variables take different ranges: some are on the scale $(0, 1)$, some $(0, 100)$, some $(-\infty, \infty)$. To aid comparison and computation of the local-linear-regression weights, the predictive variables are (unless stated otherwise) standardised as follows:

- Suppose that the LSOAs are indexed by $i = 1, \ldots, L$, and that the predictive variables are indexed by $j = 1, \ldots, n_P$.
- Let $P(i, j)$ be the unadjusted predictive variable.
- Define $X(i, j) = (P(i, j) - \mu_j)/\sigma_j$ where μ_j and σ_j are the empirical mean and standard deviation of $P(1, j), \ldots, P(L, j)$.

Hence, the $X(i, j)$ all have mean 0 and variance 1.

Variables that we chose not to standardise include

- Urban-Rural classification and region (which are categorical rather than ordinal variables);
- The proportions in communal establishments (including care homes with and without nursing) in a particular LSOA (which remain as proportions on a $(0, 1)$ scale).[6]

We now locally regress the $R_0(i)$ against the vector X. Thus, for each LSOA, i, fit the linear (in the vector x) function

$$F(i, x) = a(i) + b(i)^T x$$

by minimising, over the scalar $a(i)$ and vector $b(i)$, the weighted sum of squares

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

[6]From a statistical perspective, and in this analysis, care homes can be considered to be nuisance parameters. Specifically, the presence of a care home in an LSOA should not impact on the socio-economic characteristics and mortality experience of the non-care-home population in the same LSOA.

where the $w_1(i,j)$ are weights that are discussed below. We then set

$$\hat{F}_1(i) \;=\; \hat{a}(i) + \hat{b}(i)^T X(i), \tag{1}$$

and update estimated deaths

$$\hat{D}_1(i) \;=\; \hat{D}_0(i) F_1(i) \tag{2}$$

and the actual-over-expected

$$R_1(i) = D(i)/\hat{D}_1(i) = D(i)/\hat{D}_0(i)\hat{F}_1(i).$$

4.4.1 Stage 1 weights

The calculations above need the weights to be specified. Broadly speaking, the weights are large when two LSOAs, i and j, share similar characteristics $X(i)$ and $X(j)$, and weights tend to zero as $X(i)$ and $X(j)$ get further apart. The idea here is that LSOAs with similar socio-economic characteristics should have similar relative risks.

A number of different ways could be used to generate the weights. Here we start with the socio-economic distance between two LSOA's

$$d_1(i,j) \;=\; ||X(i) - X(j)||_S$$

where the distance measure is

$$||X(i) - X(j)||_S \;=\; \left(\sum_{k \in S} (X(i,k) - X(j,k))^2 \right)^{0.5} \tag{3}$$

and S is a subset of $(1, \ldots, n_P)$. The reason why S might not be the full set of predictive variables $1, \ldots, n_P$, is to allow for the exclusion of "nuisance" variables such as care-home proportions that do not contribute to the underlying socio-economic characteristics of an LSOA but do influence mortality rates. We next rescale the distances

$$v_1(i,j) = d_1(i,j)/d_{1\max}(i)$$

where we choose to use an adaptive bandwidth, $d_{1\max}(i)$. Lastly, we define the weights as

$$w_1(i,j) = \begin{cases} \hat{D}_0(j)\left(1 - v_1(i,j)^3\right)^3 & \text{if } v_1(i,j) \le 1 \text{ and } u(j) = u(i) \\ 0 & \text{if } v_1(i,j) > 1 \text{ or } u(j) \neq u(i) \end{cases}$$

where $u(j)$ is the urban-rural class of LSOA j: that is, we give zero weight to LSOAs that are not in the same urban-rural class.

When we are estimating the parameters, we need to be mindful of the fact that for some LSOAs, i, the position of $X(i)$ lies in a relatively dense region while for other LSOAs $X(i)$ lies in a relatively sparse region. If we use a constant bandwidth, then, in the latter case, relatively few neighbours of $X(i)$ will contribute to the estimation of the parameters leading to a high standard error. To avoid this we use an adaptive bandwidth, $d_{1\max}(i)$, so that the weight function for LSOAs in sparse regions includes data from a sufficiently large number of neighbours. In our case, we have chosen $d_{1\max}(i)$ to equal the distance to the K_1th nearest neighbour in the same urban-rural class: that is, only the $K_1 - 1$ nearest neighbours in the same urban-rural class will have a non-zero weight, with the weights diminishing to zero the further they are away from $X(i)$. More specifically, we have chosen $K = 2500$.

There are two further elements in the weight function: the commonly used tricube kernel function; and the expected deaths, $\hat{D}_0(j)$. The purpose of scaling the weights by $\hat{D}_0(j)$ is to give greater weight to LSOAs that have larger numbers of people in the age range of interest: larger numbers means that sampling variation in the actual-over-expected variable $R_0(j)$ for a particular LSOA is lower. This contrasts with, and generalises, the standard assumption in non-linear regression that all individual observations have the same variance.[7]

4.5 Stage 2: estimate the residual spatial relative risk

In this second stage, we assess the residual spatial relative risk based on the updated actual-over-expected, $R_1(i)$, using only the longitude and latitude, $Y(i) = (Y_1(i), Y_2(i))$, of each LSOA, i, and the urban-rural class.

Unlike the socio-economic predictive variables, there is no *ex-ante* reason why we might expect any systematic trends in relative risk as we move from west to east or north to south. We will, therefore use the simpler method of kernel smoothing to generate estimates of the relative risk:

$$\hat{F}_2(i) = \frac{\sum_j w_2(i,j) R_1(j)}{\sum_j w_2(i,j)}$$

where the weights depend on the physical distance between LSOAs i and j, with adjustment for the urban-rural class of each.

[7] To justify this particular scaling, suppose we have a sequence of random variables $Y_j \sim N(\mu, \sigma_j^2)$ where the σ_j are known and μ has to be estimated. An estimator for μ is $\sum_j w_j Y_j$ with $\sum_j w_j = 1$. The estimator with the lowest variance has $w_j \propto 1/\sigma_j^2$. In our case the observations $R_0(j)$ have a variance that is proportional to $1/\hat{D}_0(j)$ (in the absence of socio-economic effects).

4.5.1 Stage 2 weights

These are based on physical distance between LSOAs adjusted for urban-rural class.

$$\begin{aligned} d_2(i,j) &= \left[\frac{(Y_1(i)-Y_1(j))^2+(Y_2(i)-Y_2(j))^2)}{\phi(u(i))\phi(u(j))}\right]^{0.5} \\ v_2(i,j) &= d_2(i,j)/d_{2\max}(i) \\ w_2(i,j) &= \begin{cases} \hat{D}_1(i)\left(1-v_2(i,j)^3\right)^3 & \text{for } u_2(i,j)\le 1 \\ 0 & \text{for } v_2(i,j)>1 \end{cases} \end{aligned}$$

As with the $w_1(i,j)$ we allocate greater weight to LSOAs with higher expected deaths. The $\phi(u)$ scaling parameters are dependent on the urban-rural class $u = 1,\ldots,5$. The rationale for this feature is that, e.g., rural LSOAs tend to be much further apart than urban LSOAs. Thus a rural neighbouring LSOA 10km away might carry more weight than a city-based LSOA 1km away. Here, we have used $\phi = (1.5, 1.5, 5, 25, 1)$ and the adaptive bandwidth, $d_{2\max}(i)$, is set equal to the distance to the K_2th nearest neighbour (based on the $d_2(i,j)$ distances), with $K_2 = 250$ giving satisfactory results.

4.6 The core model

We now present results for our core model, using the data from 2001 to 2018. We experimented with different groups of predictive variables and settled on those listed in Table 3 on the basis of (a) lower variance of randomised probability-transformed residuals (see subsection 4.13) and (b) lower variance of the residual spatial relative risk. In aiming to minimise the residual spatial relative risk *we are seeking to explain as much as possible of the observed variation in mortality using socio-economic variables.*

The zeros in the weight function, $w_1(i,j)$, mean that the model for each urban-rural (UR) class is fitted independently of other urban-rural classes. Within each UR class we have 8 predictive variables, out of which care home proportions are to be treated as nuisance parameters rather than socio-economic parameters. Hence, in equation (3), $S = \{1,2,3,4,5,6\}$.

The analysis produces 32,844 estimates each for the $F_1(i)$ and $F_2(i)$. Empirical distributions (CDFs) for these are plotted in Figure 4. The two plots show CDFs for each of the five age groups. First, consider the left hand plot: socio-economic relative risk. We can see that the wide spread for the age 40-49 age group gradually narrows with age (consistent with Figure 1). For 40-49, the plot reveals how large the mortality inequality gap is between the top and bottom groups: more than 7% of LSOAs have a death rate that is less than half of the national rate, while 6% have mortality that is more than double the national rate. The factor of 4× is approximately equivalent to an effective age difference of 14 years.

	Predictive Variable	Scaling	Short Name
1	income deprivation (older people)	standardised	IDO
2	employment deprivation	standardised	EMP
3	average number of bedrooms	standardised	BED
4	living environment deprivation	standardised	LIV
5	wider barriers deprivation	standardised	WID
6	high educated 65+	standardised	EDL7
7	care home *with* nursing, 60+	$[0, 1]$ proportion	COM1
8	care home *without* nursing, 60+	$[0, 1]$ proportion	COM0
9	urban-rural class	categories 1 to 5	UR

Table 3: Predictive variables and scalings used in the core analysis. (See Appendix A for a more detailed description of the predictive variables.)

Second, consider the right-hand plot: residual spatial relative risk. The most striking feature of these CDFs is that they are much narrower than the CDFs for socio-economic risk. Indeed, the empirical variance of $F_2(i)$ is in the range 1.4% to 3.0% of the variance of $F_1(i)$ for the same age group. This indicates that, with the right choice of socio-economic variables, socio-economic information by far outweighs geographical location as a determinant of mortality.

Figure 5 (males, ages 40-49) plots geographically the estimated values of the spatial relative risk, $\hat{F}_2(i)$ (top row), and the combined relative risk, $\hat{F}_1(i)\hat{F}_2(i))$ (bottom row), plotting values using a limited colour range.

The top left shows how $\hat{F}_2(i)$ varies across England. We can see that there are patches of oranges/reds and greens/blues (higher and lower mortality than predicted by the socio-economic model only), but there is no systematic north/south or other divide. The top right plot shows the same data but zoomed in on London. The left hand plot has an interesting arc of higher mortality to the west of London, and, in London, there is a substantial red hotspot in west London around and to the north of Heathrow airport. The reasons for these patterns is currently not clear.

The lower plots in Figure 5 show the combined relative risk, $\hat{F}_1(i)\hat{F}_2(i)$, for England and London. Across the map, there is apparently much more blue, but blues tend to be in less densely populated LSOAs, sometimes rural, and therefore much larger in area. The oranges and reds, representing high relative risk, tend to be in high-density inner-city areas in London, and northern cities.

Equivalent plots for the higher age groups and for females can be seen in Appendix C. In terms of residual spatial relative risk, there is is a gradual shift in the geographical distribution as well as a narrowing of the range of values and the emergence of a split in the form of south coast and east (low) versus the rest of the country, and persistent higher mortality in west London. For the combined relative risk (lower plots) the

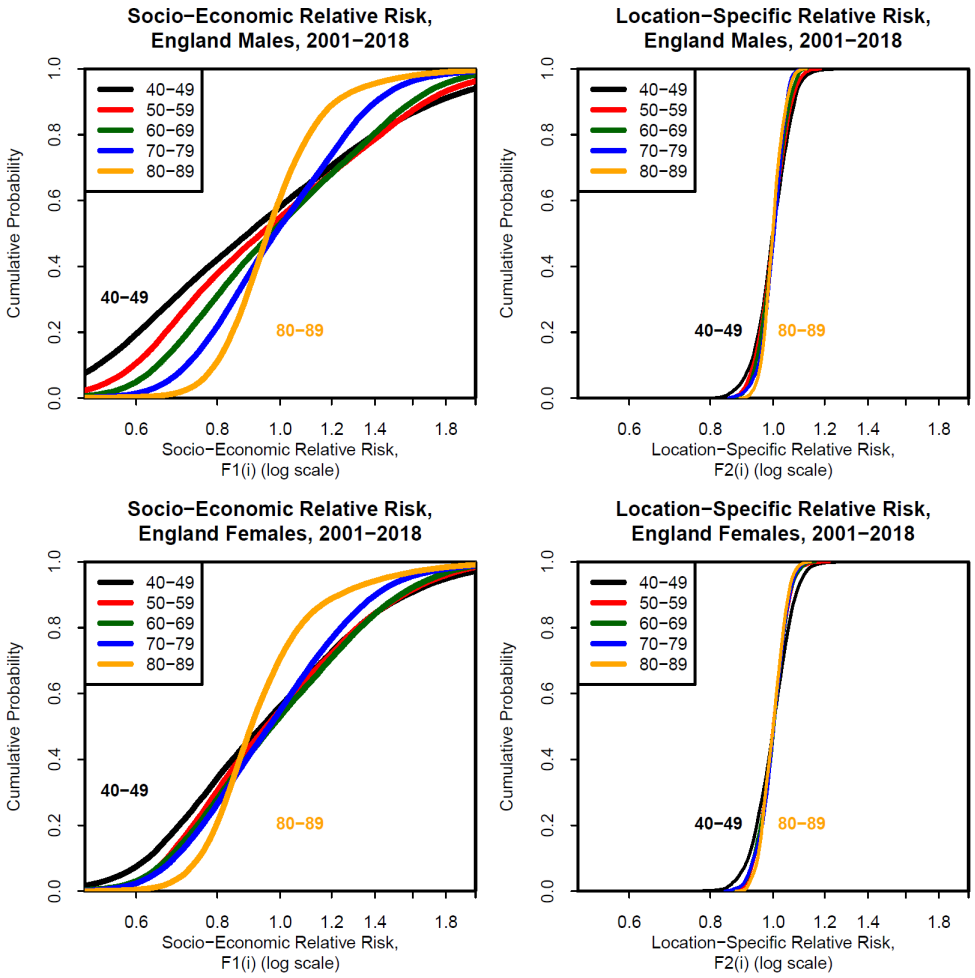

Figure 4: Left: Empirical cumulative distribution functions for the estimated socio-economic relative risk $F_1(i)$ for age groups 40-49, ..., 80-89, each covering years 2001-2018. Right: Empirical cumulative distribution functions for the estimated residual spatial relative risk $F_2(i)$ for age groups 40-49, ..., 80-89.

pictures are more consistent with Figure 5 but with a gradually diminishing range.

4.7 Care homes

The presence of a care home in an LSOA can clearly have an impact on observed mortality. For example, an affluent LSOA that we would expect to have low mortality might have higher than expected deaths if there is a large care home in the neighbourhood. As remarked above, therefore, care homes with and without nursing have been included as predictive variables, but, as they are additionally nuisance variables, they do not influence the weight function, $w_1(i, j)$.

In the data, there are 23,464 LSOAs with no care home, 6,023 have care homes without nursing only, 2,058 have care homes with nursing only, and 1,299 have care homes of both types. In 2011, there were 177,530 persons above age 60 in a care home without nursing[8] and 131,158 persons in a care home with nursing.[9]

Care home data are available for each LSOA (proportions in the 60+ age group in care homes with and without nursing). It is therefore of interest to consider the following questions: are care homes concentrated more in either less or more deprived areas; can we estimate mortality rates for the healthier non-care-home population; what do mortality rates look like if each LSOA has an average sized care home population?

The first question is addressed in Figure 6. On the left we plot the rank of the estimated relative risk, $\hat{F}_1(i)\hat{F}_2(i)$ for the 80-89 age group versus the proportion in the 60+ age group in a care home in the LSOA, $X_7(i) + X_8(i)$. As the relative risk includes the effect on mortality of having a care home, this scatterplot is skewed to the right. Indeed it is not surprising that the 5% of LSOAs with the highest relative risk all have significant proportions of elderly in care homes. On the right, we plot the adjusted relative risk $\hat{F}_1^B(i)F_2(i)$ that removes the impact on mortality of excess care home mortality. Specifically, in Equation 1 we have

$$\hat{F}_1(i) = \hat{a}(i) + \sum_{j=1}^{8} \hat{b}(i, j)X(i, j).$$

Below, we refer to this as $\hat{F}_1^A(i)$ (case A below). The adjusted relative risk, $\hat{F}_1^B(i)$ (see case B below), estimates what the relative risk would be if there were no care homes in the LSOA: that is, as if $X(i, 7) = X(i, 8) = 0$. The right-hand scatterplot is much more evenly distributed from left to right. This is an indication that the location of care homes is not significantly influenced by the socio-economic mix of the area.

In 7 we compare the base case A with two hypothetical cases:

[8] Average 18 persons in those LSOAs with a care home without nursing.

[9] Average 53 persons in those LSOAs with a care home with nursing.

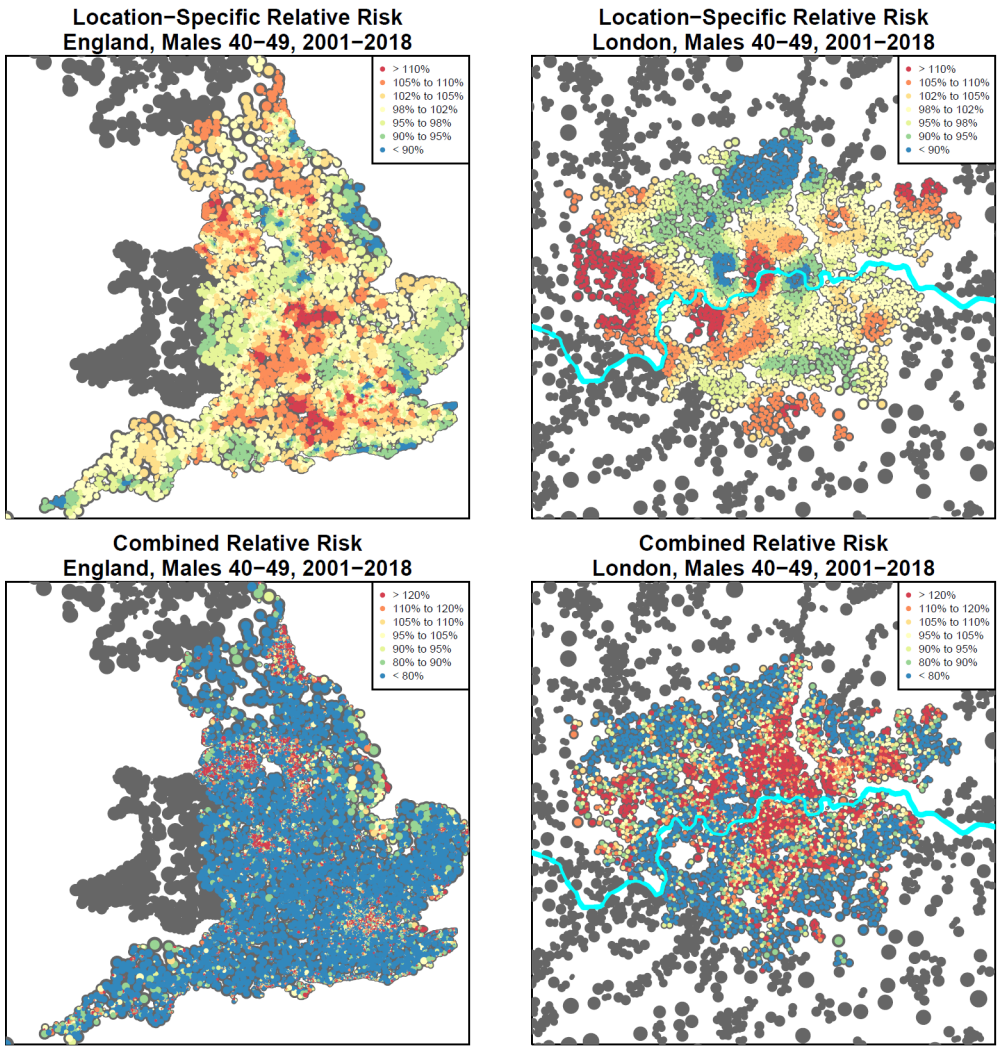

Figure 5: Top row: estimated spatial relative risk (coloured dots), $F_2(i)$, by LSOA for England (left) and London (right) for males, ages 40-49. Bottom row: combined relative risk, $F_1(i)F_2(i)$. Dot sizes reflect the physical size of each LSOA. Light blue: River Thames (for geographical reference).

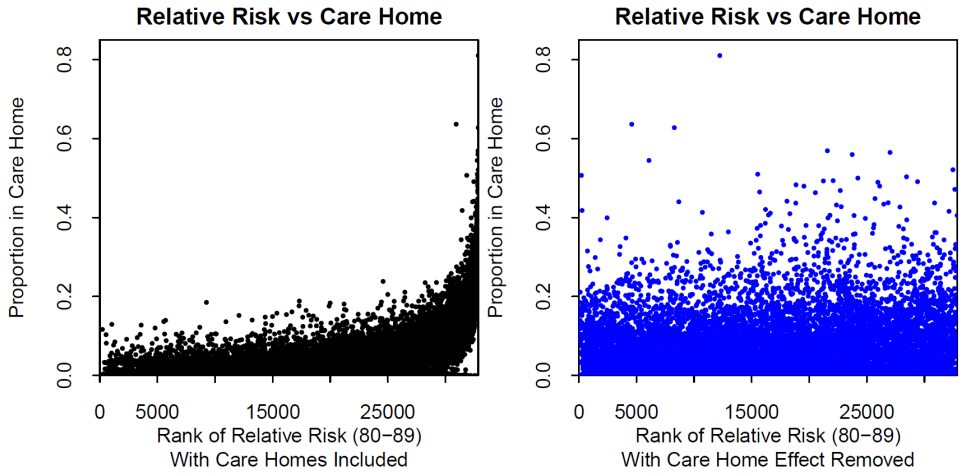

Figure 6: Left: scatterplot of the rank of $\hat{F}_1^A(i)\hat{F}_2(i)$ for the 80-89 age group versus $X_7(i)+X_8(i)$. Right: scatterplot of the rank of the relative risk of the non-care-home population relative risk $\hat{F}_1^B(i)\hat{F}_2(i)$ for the 80-89 age group versus $X_7(i)+X_8(i)$.

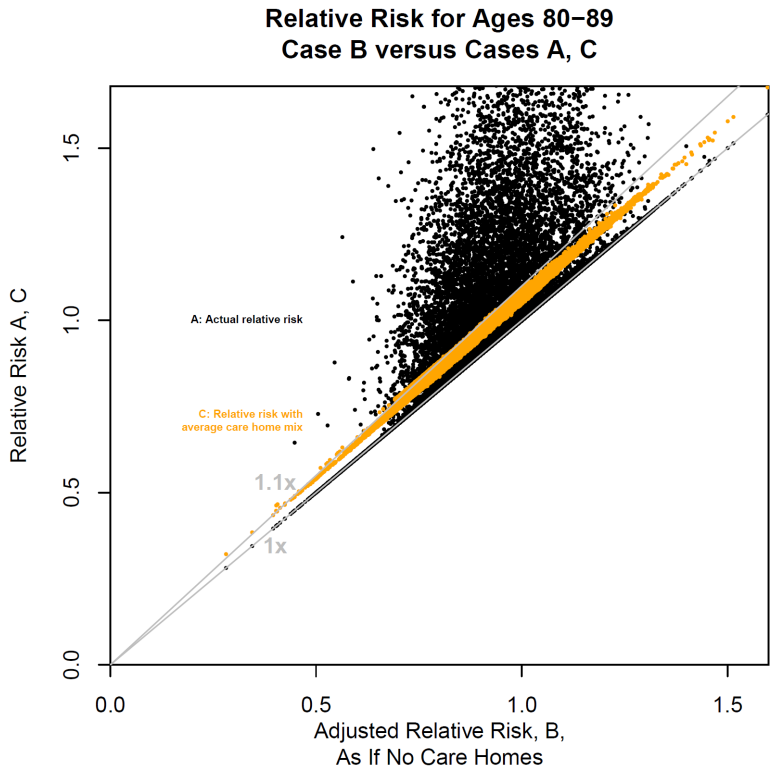

Figure 7: Scatterplot of the adjusted relative risk $\hat{F}_1^B(i)\hat{F}_2(i)$ (case B) versus cases A and C. A (black dots): actual relative risk $\hat{F}_1^A(i)\hat{F}_2(i)$. C (orange dots): relative risk assuming average proportion in care homes with and without nursing $\hat{F}_1^C(i)\hat{F}_2(i)$.

- case A: this is our base case using Equation (1).
- case B: we assume that none of the population is resident in a care home with or without nursing:

$$\hat{F}_1^B(i) \;=\; \hat{a}(i) + \sum_{j=1}^{6} \hat{b}(i,j)X(i,j). \tag{4}$$

- case C: we assume that the average proportions $\bar{X}_7$ and $\bar{X}_8$ of the 80-89 population reside in a care home with and without nursing,

$$\hat{F}_1^C(i) = \hat{a}(i) + \sum_{j=1}^{6} \hat{b}(i,j)X(i,j) + \hat{b}(i,7)\bar{X}_7 + \hat{b}(i,7)\bar{X}_8.$$

The second question also uses the adjusted relative risk, $\hat{F}_1^B(i)$. This is illustrated in Figure 7 where, for ages 80-89, we compare relative risks with and without inclusion of the care home effect (Case A versus Case B in the figure, black dots). All LSOAs with no care homes lie on the 1x diagonal. Those with a care home lie above the 1x diagonal. The ratio with to without is affected by three factors: the number of people in a care home with or without nursing; and the estimated "impact" on individuals of being in a care home (the magnitude of the $\hat{b}(i,7)$ and $\hat{b}(i,8)$ parameters). As we can see, in some LSOAs the presence of care homes more than doubles the estimated mortality rate within an LSOA.

In Figure 7 we contrast the two cases A and C versus the adjusted case with no care homes. In case C, the impact of having the average care home population is closer to a parallel shift than a proportional adjustment to the adjusted case with no care homes. In aggregate, case C (as with case A) should correspond approximately to national mortality, whereas case B will have lower aggregate mortality than the national population as it excludes the effect of care homes.

4.8 Impact of urban-rural class

It is interesting to compare the baseline results with two cases:

- case D: urban-rural class plays no role in the local linear regression. This produces relative risks $\hat{F}_1^D(i)$ as well as $\hat{F}_2^D(i)$.
- case E: all LSOAs benchmarked against urban-rural class 2 (cities and towns).[10] This produces relative risks $\hat{F}_1^E(i)$ as well as $\hat{F}_2^E(i)$.

In case D, we simply treat each LSOA as though they are all in the same urban-rural class. In Figure 8 we show a scatterplot of $\hat{F}_1^D(i)$ versus $\hat{F}_1^A(i)$ coloured by urban-rural class for ages 50-59. If there was no urban-rural effect then all points would lie on the $x = y$ diagonal. In reality, for example, the blue dots (UR class 4: very rural) are less steep indicating that this class is less sensitive to changes in the socio-economic predictive variables than other urban-rural classes. Rural towns (green) and London (orange) can also be seen to be slightly less steep, with notable differences between the with and without cases towards the top right.

When we consider the combined relative risk, $\hat{F}_1(i)\hat{F}_2(i)$, we find that cases A and D produce similar results in terms of goodness of fit in each age group. However, the influence of $\hat{F}_2(i)$ is quite different. Specifically, the variance of $\hat{F}_2^D(i)$ is significantly higher than $\hat{F}_2^A(i)$. This goes against our objective to minimise variation in the residual spatial relative risk in order to explain as much of the observed variation as possible using socio-economic and non-geographical predictive variables.

In case E, we benchmark all LSOA's against their nearest neighbours socio-economically, in urban-rural class 2. Thus we use weights derived from the modified distances:

$$d_1(i,j) = \begin{cases} ||X(i) - X(j)||_S & \text{for } j \neq i,\ u(j) = 2 \\ \infty & \text{for } j = i \\ \infty & \text{for } u(j) \neq 2 \end{cases}$$

that is, we only assign a non-zero weight to LSOAs in urban-rural class 2. Results are plotted in Figure 9. Here, we plot $\hat{F}_1^E(i)$ versus $\hat{F}_1^A(i)$. Red dots for class 2 lie on a straight line as there is no change in how class 2 is fitted. Urban-rural classes 4 (rural, blue dots) and 5 (London, orange dots) stand out as being well below the the main diagonal. This suggests that, on a like for like basis (i.e. similar socio-economic predictive variables) LSOAs in class 2 (cities and towns) have significantly higher mortality than very rural areas and in London, particularly at the more deprived end of the spectrum. Systematic differences by urban-rural class highlighted here diminish with age.

[10] Class 2 is the largest of the five classes and has the widest range of predictive variables, allowing LSOAs in other classes to be sensibly matched with LSOAs with similar predictive variables in class 2.

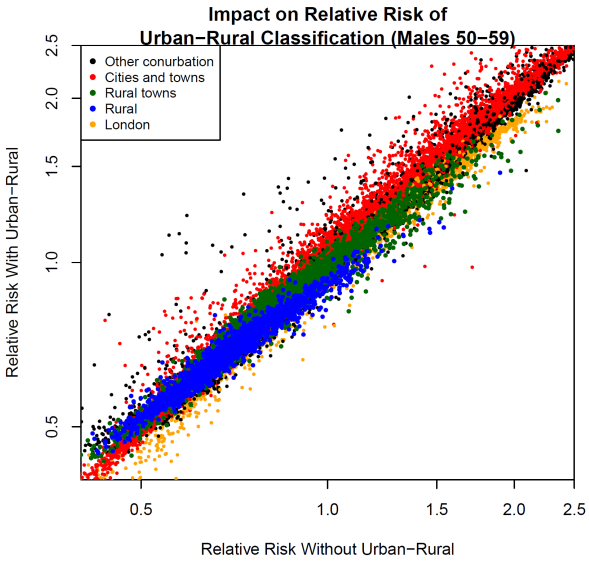

Figure 8: Comparison of socio-economic relative risk, $\hat{F}_1(i)$, in the two cases D (where urban-rural class is not used as a predictive variable) versus A (where each LSOA is benchmarked against its own urban-rural class). Scatterplot of $(\hat{F}_1^D(i), \hat{F}_1^A(i))$.

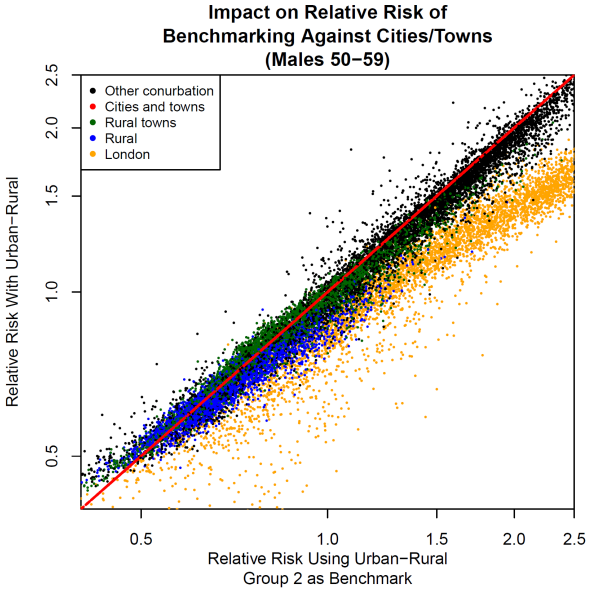

Figure 9: Comparison of socio-economic relative risk, $\hat{F}_1(i)$, in the two cases E (urban-rural class 2 is used as a benchmark) versus A (where each LSOA is benchmarked against its own urban-rural class). Scatterplot of $(\hat{F}_1^E(i), \hat{F}_1^A(i))$.

4.9 Summary of the role of predictive variables

The nine predictive variables (including urban-rural class) play different roles.

- Income deprivation (elderly) and employment deprivation are the principal *socio-economic* drivers. Employment deprivation is the main driver for mortality in younger age groups. Income deprivation for the elderly is the main driver for older age groups.
- Urban-rural class is also an important driver, particularly in minimising residual spatial relative risk.
- The proportion of an LSOA in care homes is also very important in terms of its influence over an LSOAs relative risk. However, this is an effect that we wish to remove when we seek to assess the underlying mortality characteristics of each LSOA.
- Average bedrooms, living environment deprivation, wider barriers deprivation, and high education in the 65+ age group are all statistically significant but play a lesser role. Additionally, the importance of each depends on the age group and the urban-rural class.

By way of example, Figure 10 illustrates how the relative risk with care homes removed, $F_1^B(i)$, depends on given pairs of predictive variables. Predictive variables are expressed as ranks to spread the points out. Individual points (LSOAs) are coloured according to which decile $F_1^B(i)$ falls in.[11] In the upper plot (males 70-79), the coloured bands are close to vertical indicating the dominance of income-deprivation amongst the old (IDO) over high education amongst the elderly (EDL7). However, the slope of the bands indicates that higher levels of education do result in lower mortality rates. Additionally, the bands in the lower right are a bit steeper than the bands towards the upper left indicating a non-linear relationship between the two predictive variables in terms of their impact on mortality in the males 70-79 group. In the lower plot (males 50-59) we look at the relationship between employment deprivation, wider barriers and relative risk. Employment is the dominant effect, but we can see that increased values for the wider barriers variable also increases relative risk.

[11] The extent to which the coloured bands are fuzzy or lack crisp divisions is an indication that the two predictive variables do not, on their own, provide a complete picture: there is additional information in the other variables.

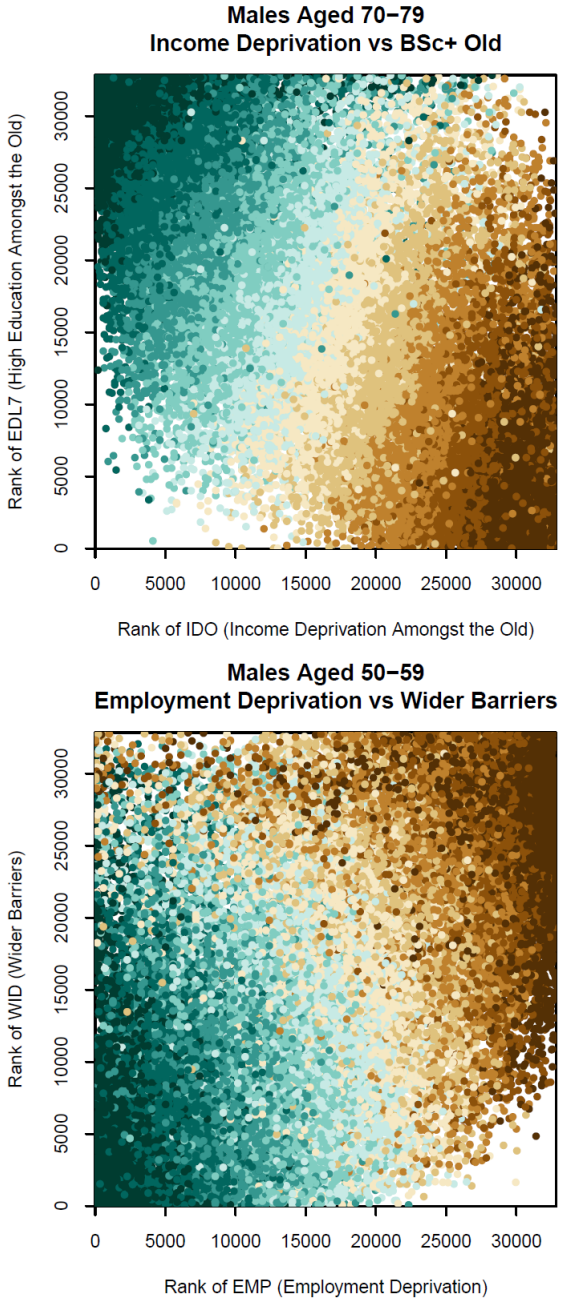

Figure 10: Scatterplots showing the relationship between pairs of predictive variables and socio-economic relative risk, $F_1^B(i)$, with the care home effect removed. Top: males aged 70-79, income-old deprivation (rank) versus proportion amongst the old with a high education (rank). Bottom: males aged 50-59, employment deprivation versus wider barriers. Dots: individual LSOAs. Colours: dark blue/green – 10% lowest relative risk $F_1^B(i)$; dark brown 10% highest relative risk, $F_1^B(i)$.

4.10 The $LIFE$ indices

We are now in a position to consider how the results of local linear regression can be used as an alternative to IMD or income-deprivation deciles or centiles.

In choosing an index, it is appropriate to strip out the impact of care homes to focus on the underlying mortality of each LSOA. To this effect, we propose the use of the socio-economic relative risk, $\hat{F}_1^B(i)$ (equation 4). Further, we have the choice of which age group to use 40-49, ..., 80-89, with the five options being labelled, respectively, as the *Longevity Indices For England*: $LIFE^4$, $LIFE^5$, $LIFE^6$, $LIFE^7$ and $LIFE^8$.

In what follows, we will use the $LIFE^7$ index as this gives the best results for ASMRs over the age range 40-89. However, the other $LIFE$ indices give similar results. The $LIFE^7$ deciles are calculated as follows:

- Decile 1 is the group of LSOAs with the 10% highest values of the $LIFE^7$ index, $\hat{F}_1^B(i)$, for age group 70-79.
- Decile 2 is the group of LSOAs with the next 10% highest values of the $LIFE^7$ index.
- ...
- Decile 10 is the group of LSOAs with the 10% lowest values of the $LIFE^7$ index.

Decile 1 will have the highest mortality and decile 10 the lowest.

The use of the relative risk $\hat{F}_1^B(i)$ with care home bias removed, in combination with the results illustrated in Figure 6, means that each of the deciles will have roughly similar proportions of people in care homes.

The $LIFE$ indices can also be used as additional covariates when analysing life insurance and pension scheme data. That is, it does not need to be used directly as a relative risk: simply that it might be used as a predictor of higher or lower mortality alongside other covariates such as pension amount and geodemographic profiling to enhance model fit. In particular, the $LIFE$ index for a particular LSOA is indicative of the relative risk that would apply to an *average* person in that LSOA (i.e. if a male was chosen at random from the 50-59 age group). On the other hand, if we have further information about a specific individual within a given LSOA then this might mark them out as being likely to have higher or lower mortality. For example, it might be known that they are a smoker, or that they are an active member of an occupational pension scheme.

Region	Regional Relative To National Mortality			
	Males		Females	
	Unadjusted	Adjusted	Unadjusted	Adjusted
North East	115.5	101.5	120.5	103.1
North West	112.9	102.7	116.1	104.2
Yorkshire and The Humber	107.6	101.3	108.3	101.4
East Midlands	101.8	100.7	102.1	99.6
West Midlands	104.2	99.7	102.6	98.2
East	91.5	97.1	92.5	98.0
London	99.5	100.5	95.2	100.5
South East	90.4	100.1	89.3	99.5
South West	89.2	95.8	87.3	94.3

Table 4: Comparison of regional death counts over the age range 70 to 79, and 2001-2018 versus expected deaths using English national mortality, without (unadjusted) and with (adjusted) allowance for socio-economic relative risk, $F_1^A(i)$.

4.11 Regional and decile mortality

In Table 1 earlier, we noted how much variation there was in regional mortality versus national mortality. We now compare expected deaths at the regional level before and after adjusting for the socio-economic relative risk $\hat{F}_1(i)$. These are reported in Table 4. As can be seen, much of the regional variation that we observed previously (Table 4, males/females "unadjusted") can be explained by socio-economic effects (males/females "adjusted"). However, some differences still remain with the south-west and north-west standing out.

In Figure 11 we plot Age Standardised Mortality Rates (ASMRs) for males for the age range 40-89 based on two sets of deciles:

- deciles based on income deprivation (ID);
- deciles based on the $LIFE^7$ index, $\hat{F}_1^B(i)$, using the 70-79 age group.

We can see that the two plots have broadly similar patterns: improving mortality and a widening gap. However, the $LIFE^7$ deciles produce a slightly wider spread, consistent with greater predictive power.

In Figure 12 we plot Age and Deprivation Standardised Mortality Rates (ADSMRs) where we standardise using either income deprivation deciles or $LIFE^7$ deciles.[12]

[12]See Appendix B.3 for definitions and a description of ADSMRs, and how ASDMRs mitigate the differences between ASMRs that are simply caused by the fact that some regions are more deprived than others.

We can make the following observations:

- Use of the $LIFE^7$ deciles results in a much narrower range of ADSMRs than the income deprivation deciles: again an indication that the new approach in this paper explains much more of the differences between regions compared to income deprivation as a sole predictive variable. If London is excluded, then the gap between regions is almost halved.
- London is a clear outlier amongst the nine regions with much stronger improvements over the 18 year period than all other regions.
- All regions have seen a slowdown in mortality improvements since 2010/11. For London, this is a slowdown relative to its previously faster rate of improvement.
- For females (not plotted) the patterns are quite similar, but rates are lower, although regional differences are a bit bigger.

Explaining the London effect is beyond the scope of this paper, but it clearly needs to be better understood. Possible reasons are: changing demographics/gentrification[13]; a widening gap in NHS spending; more effective use of NHS and public-health funding. Note, also, that the predictive variables are single values covering the whole of the period 2001-2018 rather than time dependent, so it is possible that any drift in the predictive variables (e.g. income or employment deprivation) over time that is significantly different in London from other regions might explain the London effect.

[13]E.g. faster growth in London than elsewhere of GDP, or higher levels of education, or patterns of migration within England and from outside its borders might have benefitted London mortality (e.g. the *healthy immigrant effect*; see Vang et al., 2017, and Wen et al., 2020).

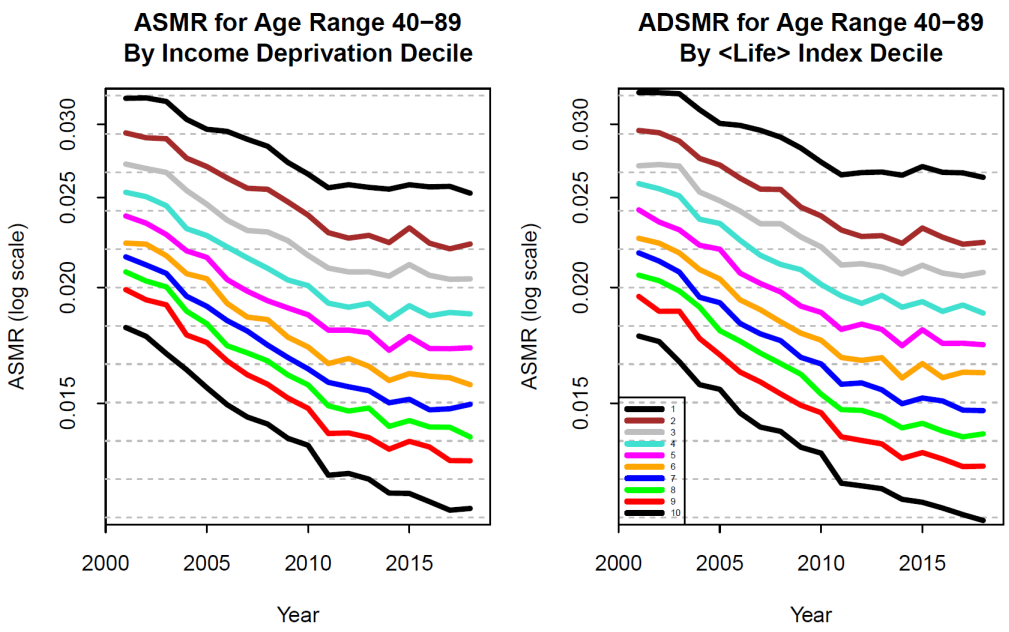

Figure 11: ASMRs for males for the age range 40-89 for each of 10 deciles. Left: deciles based on income deprivation deciles. Right: deciles based on the $LIFE^7$ index.

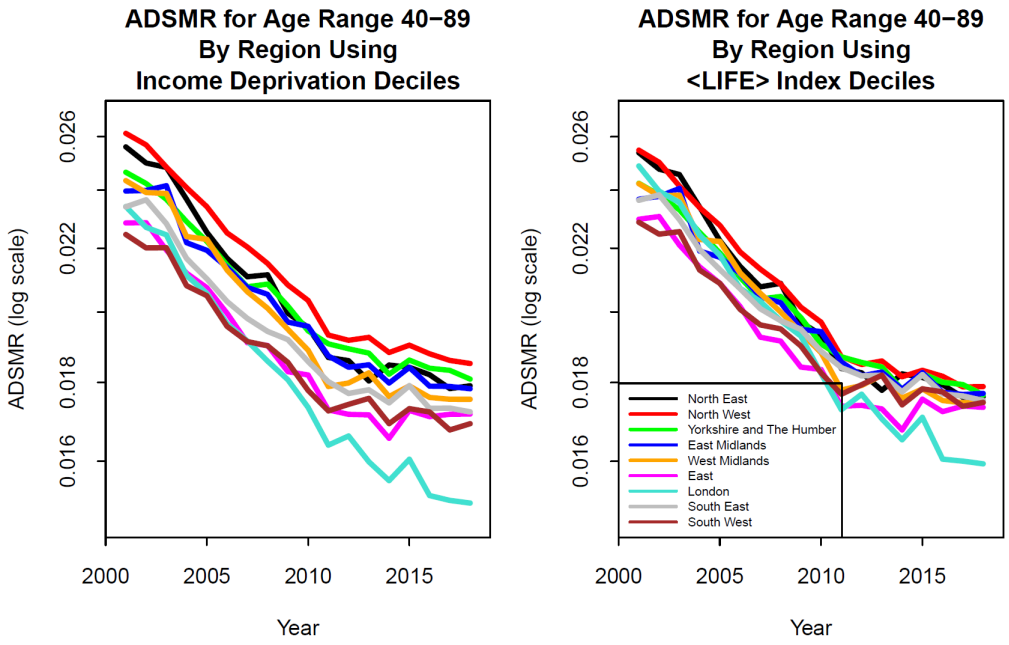

Figure 12: ADSMRs for males for the age range 40-89 for each region. Left: deprivation standardisation using income deprivation. Right: deprivation standardisation using the $LIFE^7$ index.

4.12 Time dependency

It is of interest to investigate how much the relative risks vary through time. Here, we do this by fitting the model again to the non-overlapping time periods 2001-2009 and 2010-2018.

Figure 13 shows how the socio-economic relative risk changes between the two periods. As expected, there is a very high correlation between estimates for the two periods. The scatterplots also highlight differences between the different urban-rural classes.

- For the 40-49 age group, most points lie close to the main diagonal. This is an indication that levels of inequality have remained roughly similar. Much of the "fuzziness" around the main diagonal will be due to sampling variation in the death counts. However, for London (orange dots) the dots clearly lie on a flatter slope, and this is an indication that levels of inequality in this age group have *fallen* between the two periods in London.
- For the 80-89 age group, the scatterplot is clearly steeper than the main diagonal: a clear indication that levels of inequality have risen sharply in this age group in recent years.

In Figure 14 we mimic the earlier Figure 7 where we illustrate the impact of different levels of care home proportions. We can note the following points:

- Case A (black dots): with current levels of care home provision we see that the relative risks are a bit higher in 2010-2018, indicating that those in care homes have worse prospects now compared to the earlier period.
- Case C (orange dots): similar but slightly higher in the 2010-2018 period also pointing to worsening prospects.

Possible reasons (amongst others) for this worsening might be:

- Provision of care home places is not keeping up with demand for places, and so people are even less healthy at the point of entry than they were in the past, pushing up death rates.
- Our data only tell us about the total population above age 60 in care homes. They do not tell us about the age profile of the care home populations. A changing age profile could have an impact on case C.[14]

[14]For example, if there is a shift towards a greater proportion of the care home population being in their 80's, then this could result in a rise in the black cluster A in the scatterplot from 2001-2009 to 2010-2018.

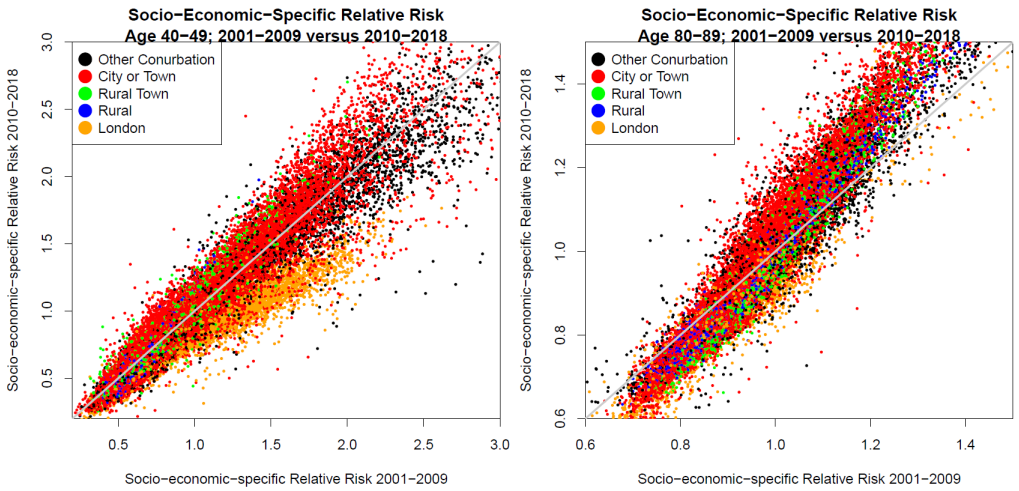

Figure 13: Socio-economic relative risk for males, ages 40-49 and 80-89. 2001-2009 estimates versus 2010-2018 estimates.

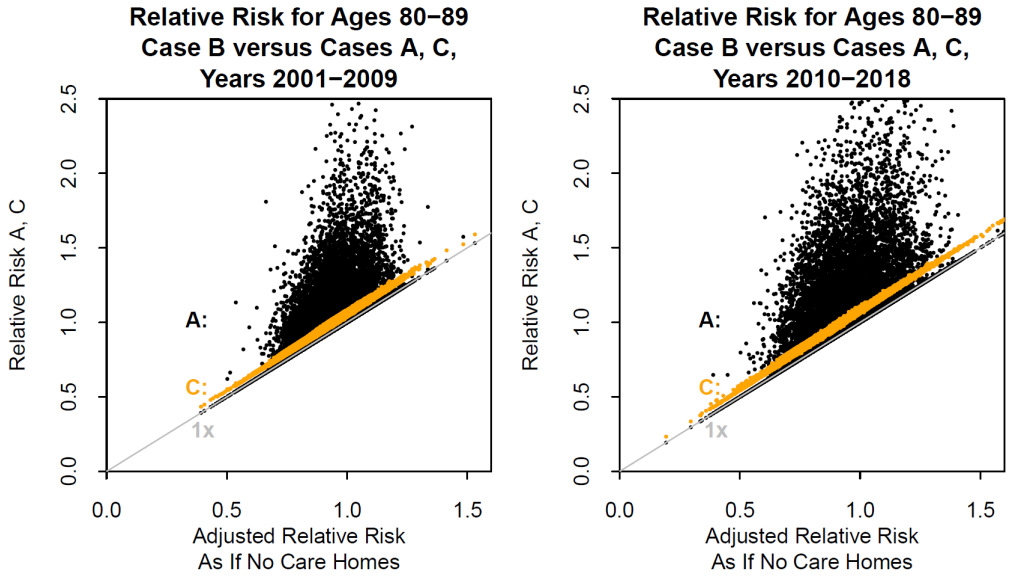

Figure 14: Impact of care homes by time period for case B versus cases A, C and D (see subsection 4.7).

4.13 Analysis of residuals

Local linear regression itself does not require any assumption about the distribution of the deaths, simply that $E[D(i)] = \hat{D}_0(i)F_1(i)F_2(i)$. In our analysis of the residuals we refine this further by investigating if the death counts have a Poisson distribution.

We propose here the use of *randomised probability-transformed residuals.*[15] These are calculated as follows:

- Suppose we have observations $y_1, \ldots, y_n$ of the random variables $Y_1, \ldots, Y_n$ with the null hypothesis that $Y_i \sim \text{Poisson}(\lambda_i)$ and that the Y_i are independent of one another.
- For each $i = 1, \ldots, n$, let

$$\begin{aligned} q_{0i} &= Pr(Y_i < y_i|\lambda_i) \\ q_{1i} &= Pr(Y_i \leq y_i|\lambda_i) \end{aligned}$$

  where the probabilities are calculated under the Poisson assumption with mean λ_i. If $y_i = 0$ then $q_{0i} = 0$.
- Simulate the *randomised probability-transformed residual* $U_i \sim U(q_{0i}, q_{1i})$.[16]

If the Poisson null hypothesis is true then the U_i will be independent and identically distributed uniform$(0, 1)$ random variables. In our mortality setting, each U_i is derived from the observed deaths, $D(i)$, and the estimated mean, $\hat{D}_0(i)F_1(i)F_2(i)$. This provides us with the potential for a number of graphical diagnostics. These could include, for example, QQ plots and histograms of the U_i. However, here we consider two types of scatterplots.

- Plot the (V_i, U_i) where V_i is the rank of the expected deaths $\hat{D}_2(i) = \hat{D}_0(i)\hat{F}_1(i)\hat{F}_2(i)$ out of $\hat{D}_2(1), \ldots, \hat{D}_2(L)$. This is illustrated in Figure 15 for three of the age ranges.
- Plot the (W_i, U_i) where W_i is the rank of the combined relative risk $\hat{F}(i) = \hat{F}_1(i)\hat{F}_2(i)$ out of $\hat{F}(1), \ldots, \hat{F}(L)$. This is illustrated in Figure 16.

In both cases, if the Poisson hypothesis is true then the scatterplots should look uniform and random with no clustering[17].

[15] Randomised probability-transformed residuals are a generalised version of probability-transformed residuals for continuous random variables. If Z is a continuous random variable with cumulative distribution function $F(z)$, then the probability-transformed residual $U = F(Z)$ has a uniform$(0, 1)$ distribution. Here randomisation is required because the random variables of interest are discrete valued rather than continuous.

[16] That is, uniformly distributed on the interval (q_{0i}, q_{1i}).

[17] Note that the V_i and W_i are evenly distributed on the integers $1, \ldots, L$ rather than randomly distributed.

In both Figures 15 and 16, the scatterplots do look reasonably uniform and random for age groups 40-49 and 60-69. But in both cases the scatterplots are less uniform for age group 80-89 indicating that there are further effects that have not been captured in the model.

In Figure 15 we can see some clustering in the top left and bottom right for ages 80-89. In Figure 16 the clustering is more evenly distributed along the top and bottom for ages 80-89: an indication of overdisperion of some sort. The differences in clustering in the two Figures are consistent with the fact that exposures are *estimated* exposures and, therefore, subject to estimation error. The pattern in Figure 15 is simply telling us that if exposures are underestimated then expected deaths are low (so towards the left of the scatterplot) and deaths tends to be higher than estimated (so a high value of U_i). Similarly, if the exposures are overestimated then expected deaths are high (so towards the right of the scatterplot) and deaths tends to be lower than estimated (so a low value of U_i). This illustrates just one impact of the challenge that the ONS faces when estimating the population between censuses in each LSOA. Figures 15 and 16 suggest that the problem is only significant at higher ages.[18]

[18]Note, the the presence of overdispersion has no direct impact on the estimation methodology, which only relies on the expected number of deaths rather than the distribution around that mean. Methods such as Generalised Linear Models would, on the other hand, would need to specify how to handle overdispersion: e.g. by replacing the Poisson assumption above with the negative binomial distribution.

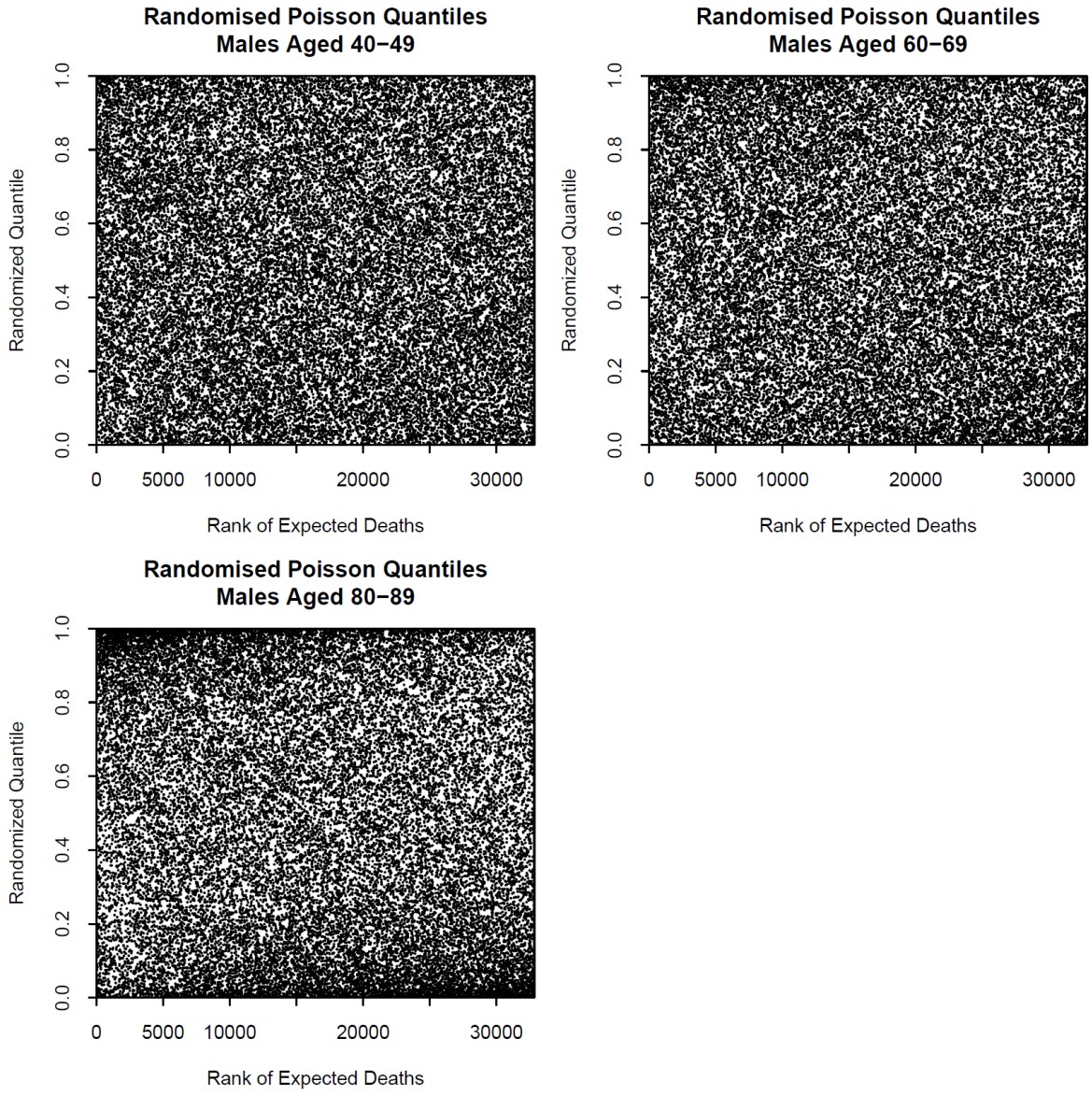

Figure 15: Scatterplots of the rank of the expected deaths, $\hat{D}_2(i)$ versus the randomised probability-transformed residual, U_i, for three age groups.

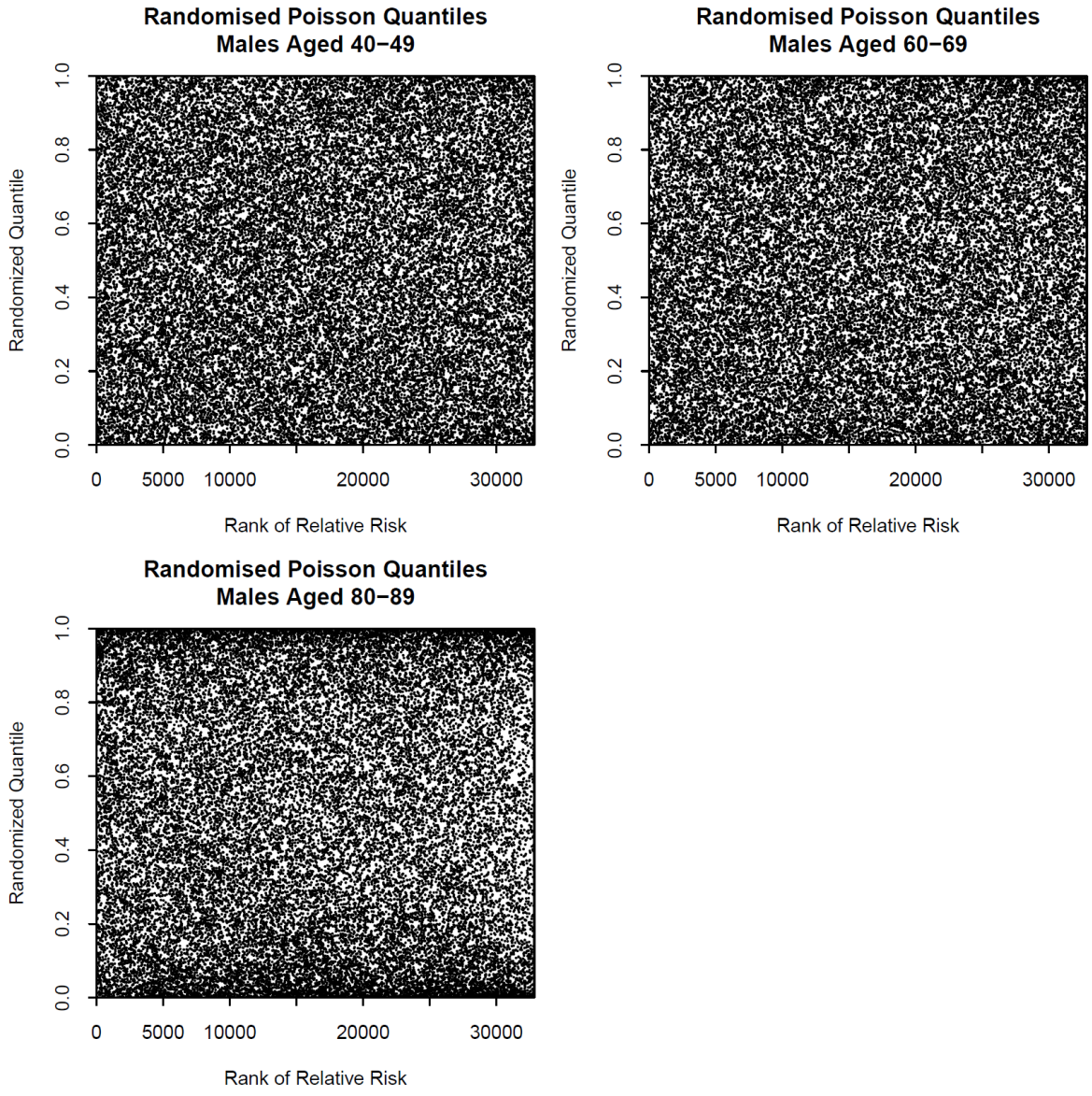

Figure 16: Scatterplots of the rank of the relative risk, $\hat{F}_1(i)\hat{F}_2(i)$ versus the randomised probability-transformed residual, U_i, for three age groups.

5 Conclusions

The increasing availability of large mortality-related datasets opens up the possibility for more detailed analyses of the key drivers of death rates.

In this paper, we have conducted a detailed analysis of mortality inequalities in England, using all-cause mortality data at the level of LSOAs. We have used the non-parametric method of local linear regression to quantify more accurately the very significant mortality inequalities that exist across England, particularly at younger ages. The method is very well suited to our large dataset and can handle, in a straightforward way, the inclusion of several predictive variables, and not particularly sensitive to non-linear transformations of the predictive variables. In particular, the method automatically captures any potential interactions between predictive variables.

Amongst all of the available predictive variables, income and employment deprivation were found to have the strongest predictive power. But we also found that urban-rural class and the presence of care homes within a neighbourhood were also important predictors. Once socio-economic effects have been filtered out, residual spatial relative risk was found to be quite small in comparison, countering the headline differences between English regions. Perhaps this is not surprising, but it emphasizes that, on a like for like basis, there is no intrinsic disadvantage in terms of life expectancy to living in the north rather than the south.

The data also confirm that inequality between different socio-economic groups has been rising including at high ages where, in the past, there has been a generally narrower inequality gap compared to younger ages.

The methodology leads us to the proposal of the $LIFE$ indices (Longevity Index for England). These can potentially be used in three ways. First, in their "raw" form directly as relative risks. Second, the $LIFE$ indices can be used to group LSOAs into deciles (with a clear improvement over income deprivation as a predictor of high or low mortality). Lastly, the $LIFE$ indices can be used as predictive variables in their own right (on a continuous scale) in the assessment of the mortality of life insurance and pensions portfolios alongside other predictive variables such as pension amount and geodemographic grouping.

The $LIFE$ indices might also be of use in developing strategies to tackle mortality inequality by identifying more accurately the worst affected groups. The companion spatial relative risks offer additional insights for policymakers into which areas have excess mortality even after taking account of socio-economic variation, potentially a result of poorer health behaviours in certain regions.

Acknowledgements

The authors gratefully acknowledge funding from the Actuarial Research Centre of the Institute and Faculty of Actuaries, the Society of Actuaries and the Canadian Institute of Actuaries through the "Modelling Measurement and Management of Longevity and Morbidity Risk" research programme (see `www.actuaries.org.uk/arc`).

References

Cairns, A.J.G., Kallestrup-Lamb, M., Rosenskjold, C.P.T., Blake, D., and Dowd, K., (2019) Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index. *ASTIN Bulletin*, 49: 555-590.

Cairns, A.J.G., Blake, D., Dowd, K., and Kessler, A.R., (2016) Phantoms Never Die: Living with Unreliable Population Data. *Journal of the Royal Statistical Society, Series A*, 179: 975-1005.

Case, A., and Deaton, A. (2015) Rising morbidity and mortality in midlife among white non-hispanic americans in the 21st century. *Proceedings of the National Academy of Sciences*, 112: 15078-15083.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A, and Cutler, D. (2016) The association between income and life expectancy in the United States, 2001-2014. *Journal of the American Medical Association,* 315: 1750-1766.

Cleveland, W.S. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 74: 829–836.

Eurostat (2013) *Revision of the European standard population. Report of Eurostat's task force. 2013 edition.* Luxembourg: Publications Office of the European Union.

Longevity Science Panel (2018) Life expectancy: Is the socio-economic gap narrowing?
`www.longevitypanel.co.uk/viewpoint/life-expectancy-is-the-socio-economic-gap-narrowing/`
(Accessed 10/12/2019)

Macdonald, A.S., Richards, S.J., and Currie, I.D. (2018) *Modelling mortality with actuarial applications.* Cambridge University Press, Cambridge.

Mackenbach, J.P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemström, Ö., Valkonen, T., and Kunst, A.E. (2003). Widening socioeconomic inequalities in mortality in six Western European countries. *International Journal of Epidemiology,* 32: 830-837.

Mackenbach, J.P., et al. (2016) Trends in inequalities in premature mortality: a

study of 3.2 million deaths in 13 European countries. *Journal of Epidemiology and Community Health,* 69: 207–217.

Office for National Statistics (2015) English indices of deprivation 2015. `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015` (Accessed 3/11/2019)

Office for National Statistics (2018) Population estimates by output areas, electoral, health and other geographies, England and Wales: mid 2017. *Statistical Bulletin*, 25 October 2018.

Redondo Lourés, C., and Cairns, A.J.G. (2020) Mortality In The US By Education Level. *Annals of Actuarial Science*, 14: 384-419

Richards, S. J. (2008) Applying Survival Models to Pensioner Mortality Data. *British Actuarial Journal*, 14: 257–303.

Vang, Z.M., Sigouin, J., Flenon, A., and Gagnon, A. (2017) Are immigrants healthier than native-born Canadians? A systematic review of the healthy immigrant effect in Canada. *Ethnicity and Health*, 22: 209-241.

Villegas, A.M., and Haberman, S. (2014) On the Modeling and Forecasting of Socioeconomic Mortality Differentials: An Application to Deprivation and Mortality in England. *North American Actuarial Journal*, 18: 168-193.

Wen, J., Cairns, A.J.G., and Kleinow, T., (2021) Fitting Multi-Population Mortality Models to Socio-Economic Groups *Annals of Actuarial Science*, 15: 144-172.

Wen, J., Kleinow, T., and Cairns, A.J.G. (2020) Trends in Canadian Mortality By Pension Level: Evidence From the CPP and QPP. *North American Actuarial Journal*, 24: 533-561

A Datasets

Data for England are available at the level of small geographical areas known as Lower Layer Super Output Areas (LSOAs). Each area has typically between 1,000 and 3,000 persons, with an average of about 1,600, across all ages.

There are 32,844 LSOAs at present. New LSOAs are created from time to time in response to growth in housing. Data relating to the LSOAs can be found on the ONS (Office for National Statistics) website www.ons.gov.uk.

A.1 Deaths and exposures

For each LSOA we have:

- data from 2001-2018:
  - death counts, $D(g, i, t, x)$, where g is the sex, i is the LSOA, t is the year, and x is the age last birthday;[19] [20]
  - central exposed to risk, $E(g, i, t, x)$, equated to the mid-year population estimates for 2001-2018 available from the ONS;[21]
- 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.

A.2 Potential predictive variables and related data

1. LSOA index.
   - LSOA codes are of the form "E010xxxxx" where the LSOA index *xxxxx* ranges from 00001 to 33768.
   - Only 32,844 indexes are currently in use and, therefore, 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. And some LSOAs have shrunk and will have been merged and allocated a new index.
2. The Index of Multiple Deprivation (IMD)[22]

   This is the official composite measure of relative deprivation in England, with a single value for each LSOA. A higher value indicates a higher level of deprivation. The IMD has seven domains:
   - income deprivation;
   - employment deprivation;

[19] User-requested deaths data used in this study can be found at `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/007807deathsbylowersuperoutputareaageandsexenglandandwales2001to2016`

[20] Death counts are *registrations* in calendar year t rather than *occurrences*. The advantage to the ONS and users of using registrations is that the tables can be produced in a much more timely manner. Death counts by year of occurrence can be delayed by the very small number of deaths that require an inquest.

[21] Mid-year population estimates can be found at `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates`. Note that data for 2012-2016 have been revised slightly since our original download by the ONS to account for revisions made to local authority population estimates (ONS, 2018). Unrevised files for 2015 and 2016 are available on the same web page. The authors have verified that use of the revised data make very little difference to estimates of relative risk, $F_1(i)$ and $F_2(i)$.

[22] For further details, see Office for National Statistics (2015).

- education, skills and training;
- health deprivation and disability;
- crime;
- barriers to housing and services;
- living environment.

Some of these have further sub-domains (which we discuss below) that we consider to be useful to refine predictions of mortality.

3. Income deprivation (a domain of the Index of Multiple Deprivation (IMD)):
    - 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 32,844: one entry for each LSOA;
    - sub-domains include *income deprivation affecting older people*, which measures income deprivation amongst people aged 60 and older.
4. Employment deprivation (a domain of the IMD)
    - this measures the proportion of the *working* population in each LSOA who are unemployed;
    - the data are in a vector of length 32,844: one entry for each LSOA corresponding to the vector of 5-digit LSOA codes above.
5. Living environment deprivation (a domain of the IMD)
    - 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 32,844: one entry for each LSOA.
6. Barriers to housing and services (a domain of the IMD)
    - 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;
    - geographical barriers measures distance to key services;

- 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 32,844: one entry for each LSOA;
- and data are available separately for wider barriers and geographical barriers.

7. Average number of bedrooms

   - this measures the average number of bedrooms per household in the LSOA
   - the published 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 32,844: one entry for each LSOA.

8. Highest level of qualification:

   - this gives the proportion of a particular group within the LSOA who have attained a particular level of education
   - data are in the form of a 3-dimensional array for males and females combined

     lsoa x age-group x education level $(32,844 \times 6 \times 8)$
   - 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;

9. Occupation group proportions

   - gives the proportion of a particular group within the LSOA who have a particular type of occupation
   - data are in the form of a 4-dimensional array
     gender x lsoa x age-group x occupation group $(2 \times 32,844 \times 14 \times 9)$
   - 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;

10. Urban-Rural Classification

    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 32,844: one entry for each LSOA.

11. 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 32,844: one entry for each LSOA.

12. Communal establishments

    - This element of the data (a user-requested dataset from the ONS) record the number of persons in each LSOA in a communal establishment at the time of the 2011 census.
    - The data count the number of persons, $C(i, g, y, \tau)$ where
      - i is the LSOA index;
      - g is sex;
      - y is the age group 0-59, and 60+;
      - τ is the type of communal establishment:
        1 Care home: Private or local authority, with nursing;
        2 Care home: Private or local authority, without nursing;
        3 Remainder of medical and care establishments;
        4 Other communal establishments.

B Age Standardised Mortality Rate (ASMR)

B.1 Basic definition

The purpose of the ASMR is to facilitate comparison of mortality rates in different populations. In particular, if the age profiles of different populations are different then some measures (e.g. deaths per 100,000 population) might simply reflect differences in the age profile even if death rates at individual ages are identical. The ASMR avoids this by using a standard population rather than the actual age profile.

The basic definition, ignoring other indices, over the age range (x_0, x_1) is

$$ASMR(x_0, x_1) = \frac{\sum_{x=x_0}^{x_1} m(x)ES(x)}{\sum_{x=x_0}^{x_1} ES(x)}$$

where $ES(x)$ is the standard population at age x[23], and $m(x)$ is the death rate at age x.

B.2 Further development

In this paper we make use of a number of variants of the ASMR. Various age ranges are considered: e.g. 40-49, 40-64, 65-89, etc. We also calculate ASMR's by region and by deprivation or other deciles.

Suppressing indices for the age range, (x_0, x_1), calendar year, t, and sex, g, we start with the following death rates:

- $m(x)$ = national death rate at age x,
- $m_I(i, x)$ = income-deprivation decile i death rate,
- $m_R(r, x)$ = region r death rate,
- $m_{RI}(r, i, x)$ = death rate at age x in region r, deprivation decile i.

[23]Here we use the European Standard Population, 2013 (Eurostat, 2013). Comparative results in this paper for different populations are unlikely to be sensitive to the choice of standard population.

Each of these has corresponding ASMRs:

$$\begin{aligned} ASMR \equiv ASMR(t) &= \frac{\sum_{x=x_0}^{x_1} m(x)ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \\ ASMR_I(i) &= \frac{\sum_{x=x_0}^{x_1} m_I(i,x)ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \\ ASMR_R(r) &= \frac{\sum_{x=x_0}^{x_1} m_R(r,x)ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \\ ASMR_{RI}(r,i) &= \frac{\sum_{x=x_0}^{x_1} m_{RI}(r,i,x)ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \end{aligned}$$

For all-cause mortality ($c = 0$) we will be summing over single ages. For cause-of-death mortality ($c > 0$), we will be summing over 5-year age groups: for example if the stated age range is 40 to 49, then we are summing over two age groups: 40-44 and 45-49.

B.3 The Age and Deprivation Standardised Mortality Rate (ADSMR)

Now, we can develop the formula for $ASMR_R(r)$ as follows:

$$ASMR_R(r) = \frac{\sum_{x=x_0}^{x_1} ES(x) \sum_{i=1}^{10} m_{RI}(r,i,x) w_{RI}(r,i,x)}{\sum_{x=x_0}^{x_1} ES(x)} \tag{5}$$

where $w_{RI}(r,i,x) = E_{RI}(r,i,x) / \sum_j E_{RI}(r,j,x)$ represents the weight carried by decile i amongst deciles 1 to 10, in region r at age x (so, for each r and x, $\sum_i w_{RI}(r,i,x) = 1$).

We then note than some regions have greater proportions of more deprived areas than other regions. The resulting differences in weights then means that some regions will have naturally higher ASMRs even if there are no differences in death rates at the level of income deprivation between regions (i.e. $m_{RI}(r,i,x) = m_I(i,x)$ for all $r = 1, \ldots, 9$).

To remedy this, we propose the ADSMR as an alternative to the regional $ASMR_R$s. Specifically we replace the weights $w_{RI}(r,i,x)$ in 5 by $\tilde{w}_{RI}(r,i,x) = 0.1$. Hence

$$\begin{aligned} ADSMR(r) &= \frac{\sum_{x=x_0}^{x_1} \sum_{i=1}^{10} m_{RI}(r,i,x) \tilde{w}_{RI}(r,i,x) ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \\ &= \frac{\frac{1}{10} \sum_{i=1}^{10} \sum_{x=x_0}^{x_1} m_{RI}(r,i,x) ES(x)}{\sum_{x=x_0}^{x_1} ES(x)} \\ &= \frac{1}{10} \sum_{i=1}^{10} ASMR_{RI}(r,i). \end{aligned}$$

The use of the ADSMR allows us to filter out the impact of differences in deprivation levels. Any differences that remain need further investigation.

C Supplementary plots

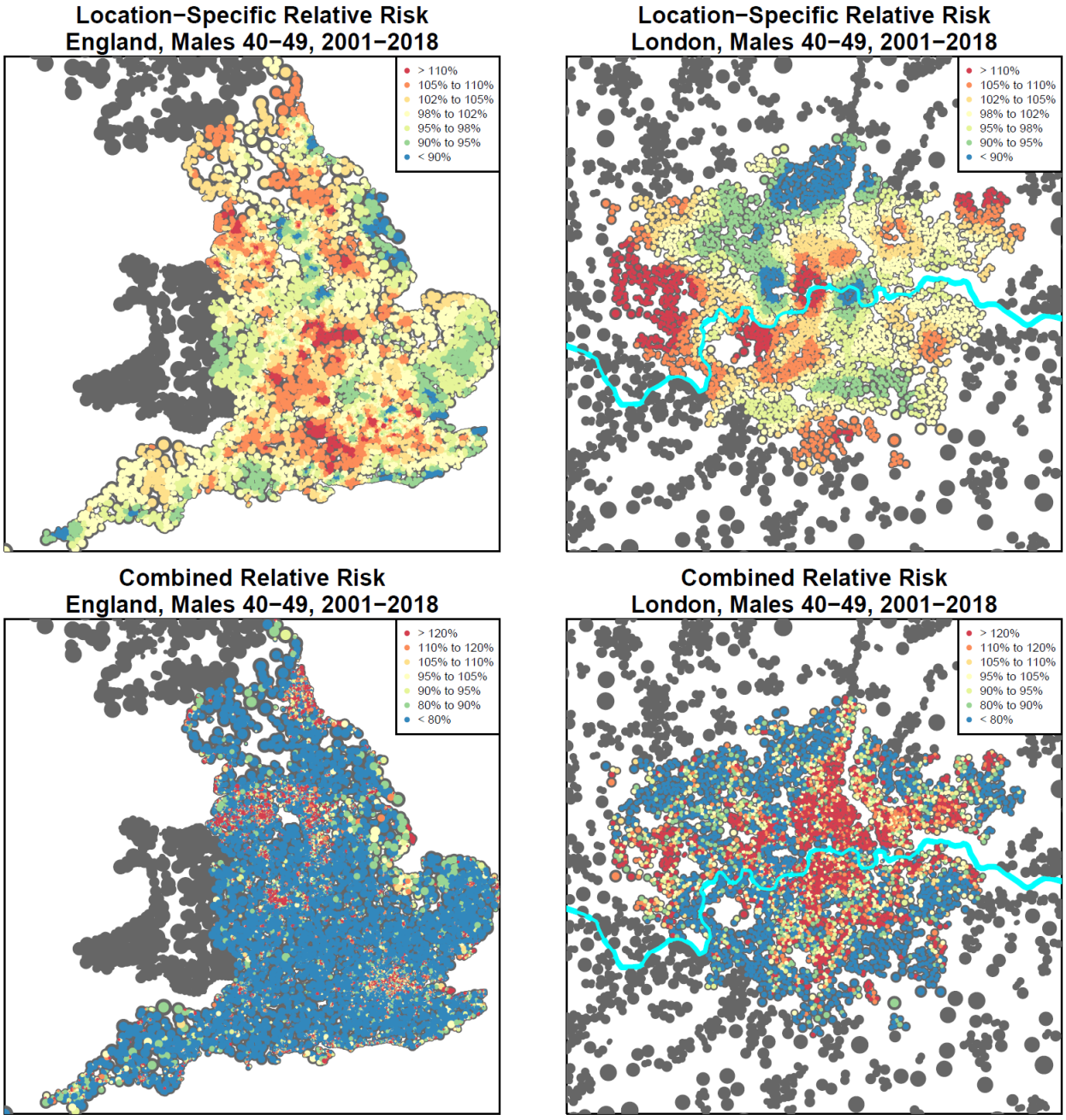

Figure 17: Top row: estimated spatial relative risk, $F_2(i)$, by LSOA for England (left) and London (right) for males, ages 40-49. Bottom row: combined relative risk, $F_1(i)F_2(i)$. Dot sizes reflect the physical size of each LSOA.

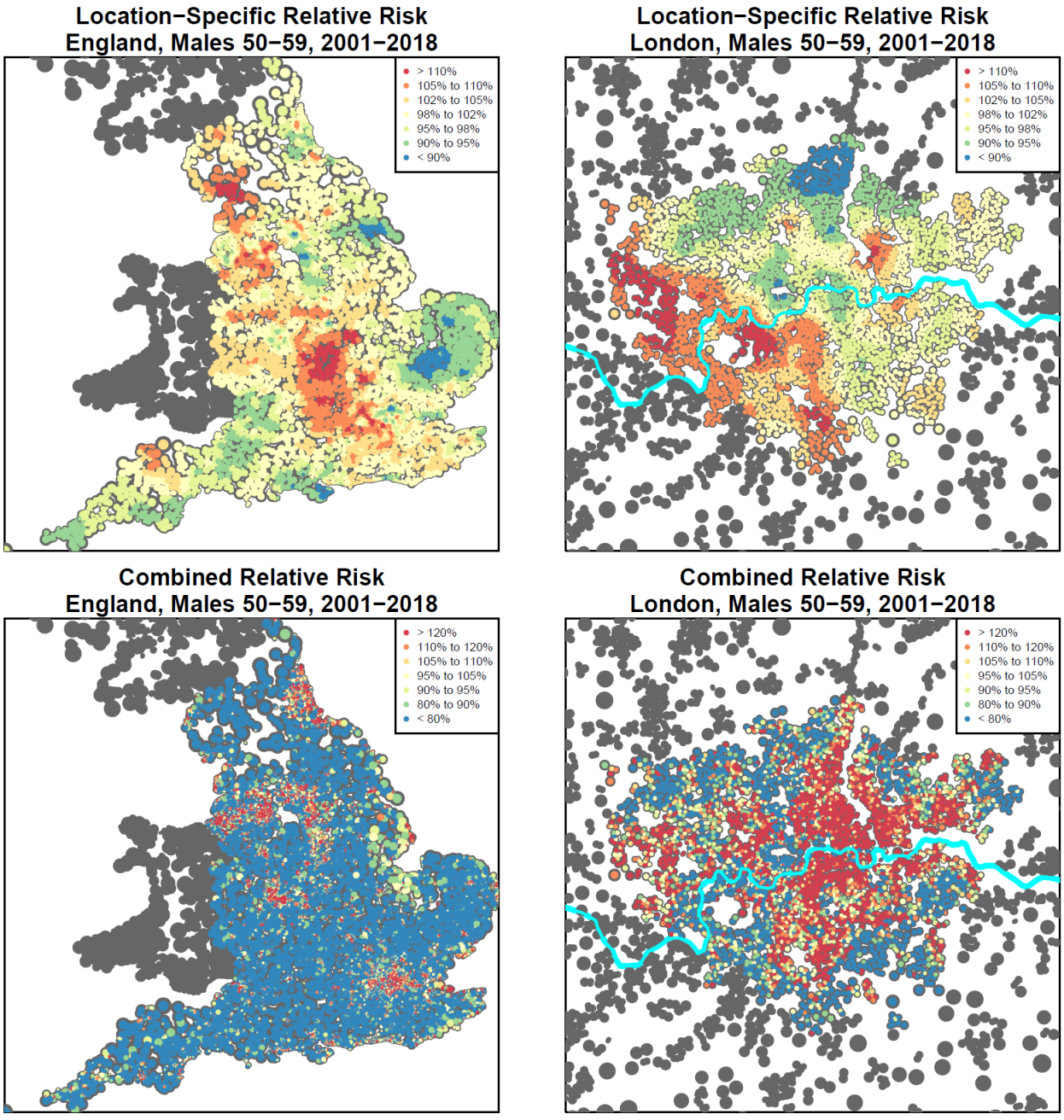

Figure 18: As Figure 17 but for ages 50-59.

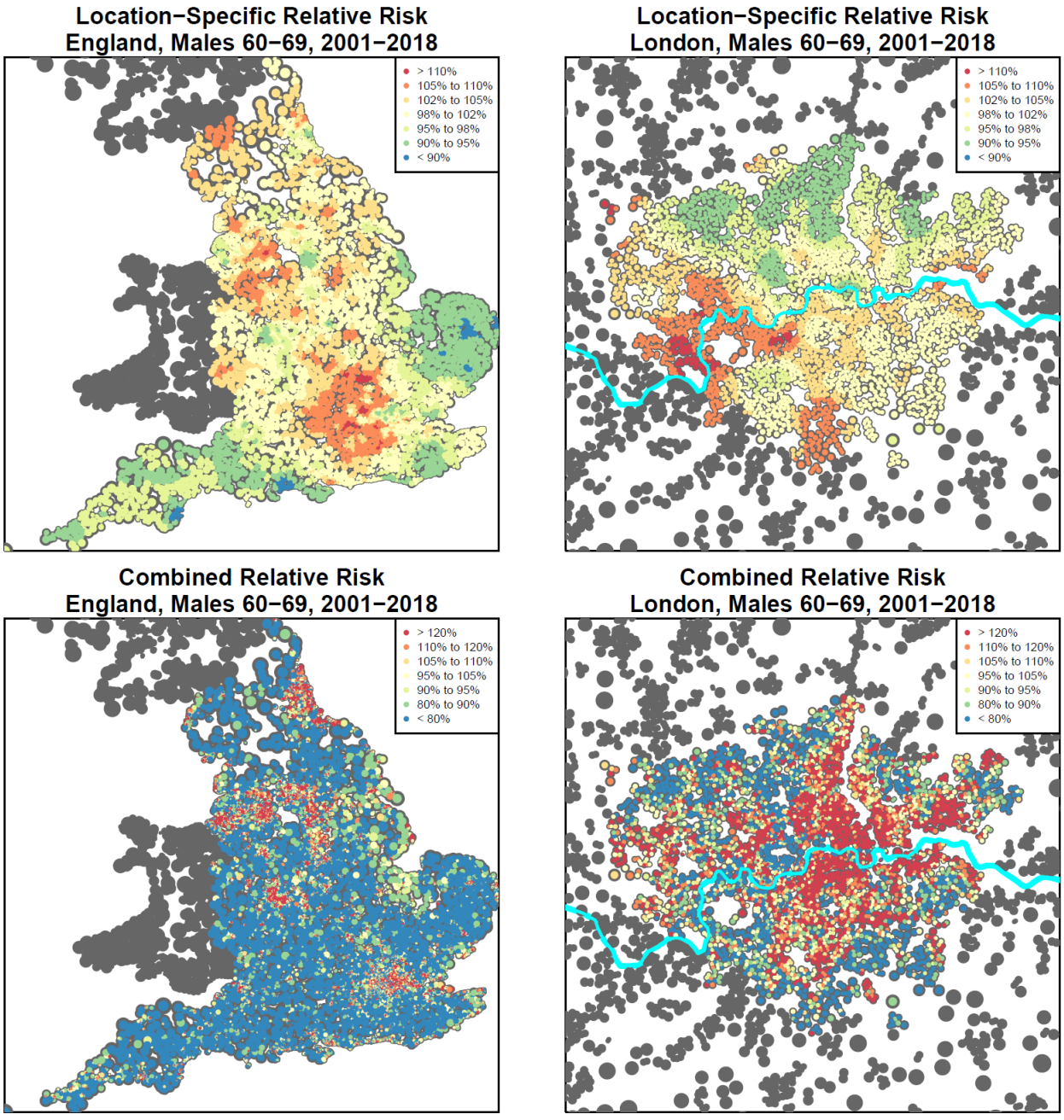

Figure 19: As Figure 17 but for ages 60-69.

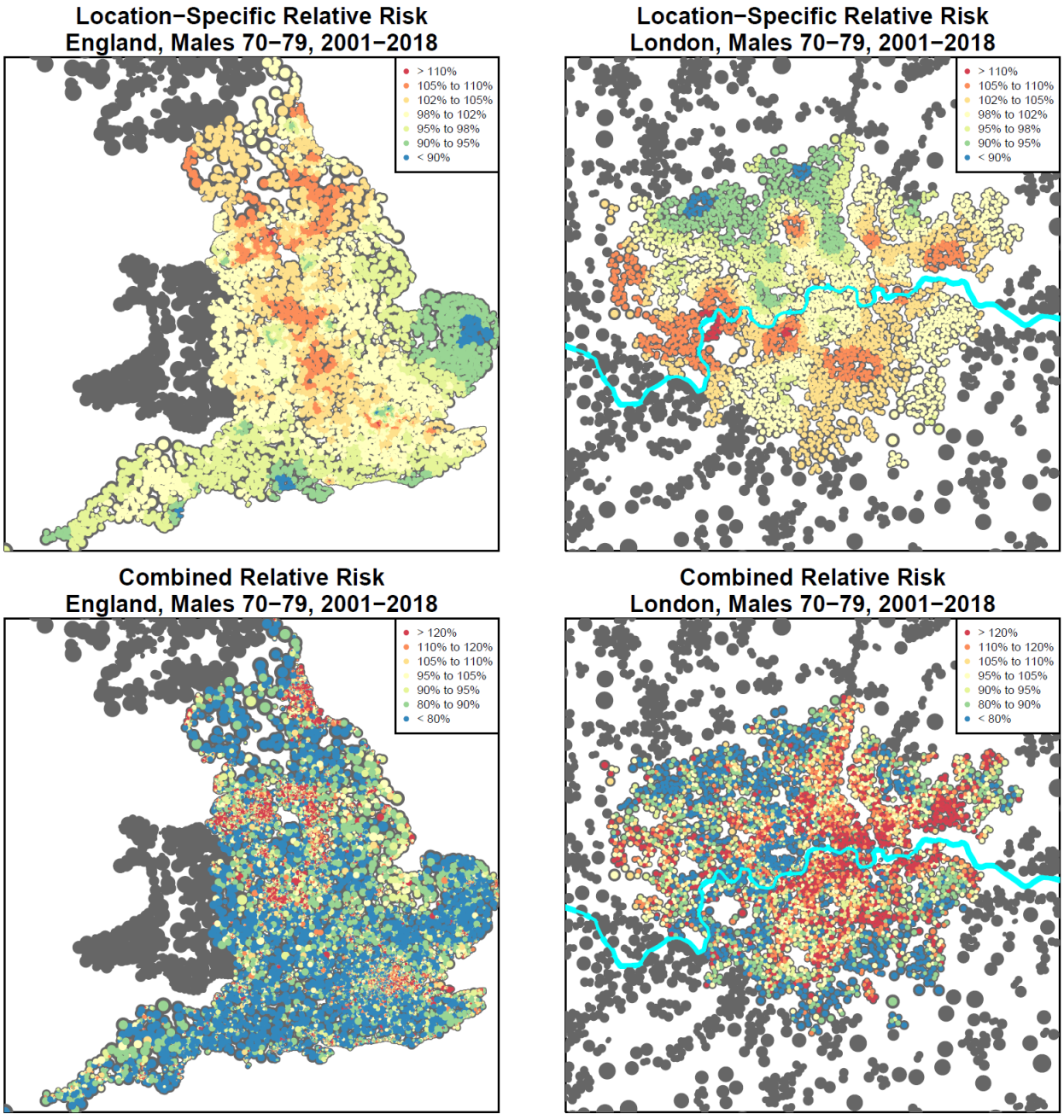

Figure 20: As Figure 17 but for ages 70-79.

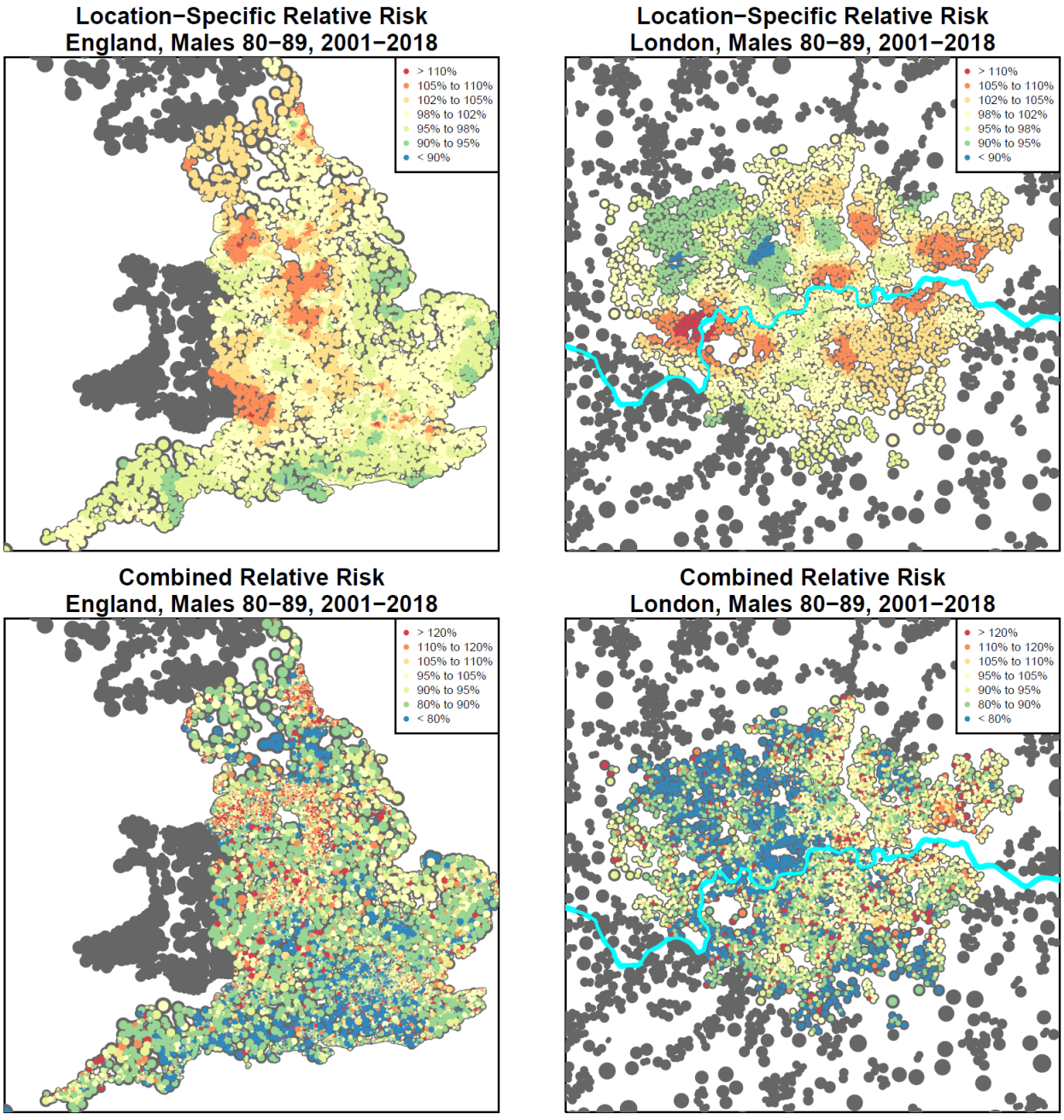

Figure 21: As Figure 17 but for ages 80-89.

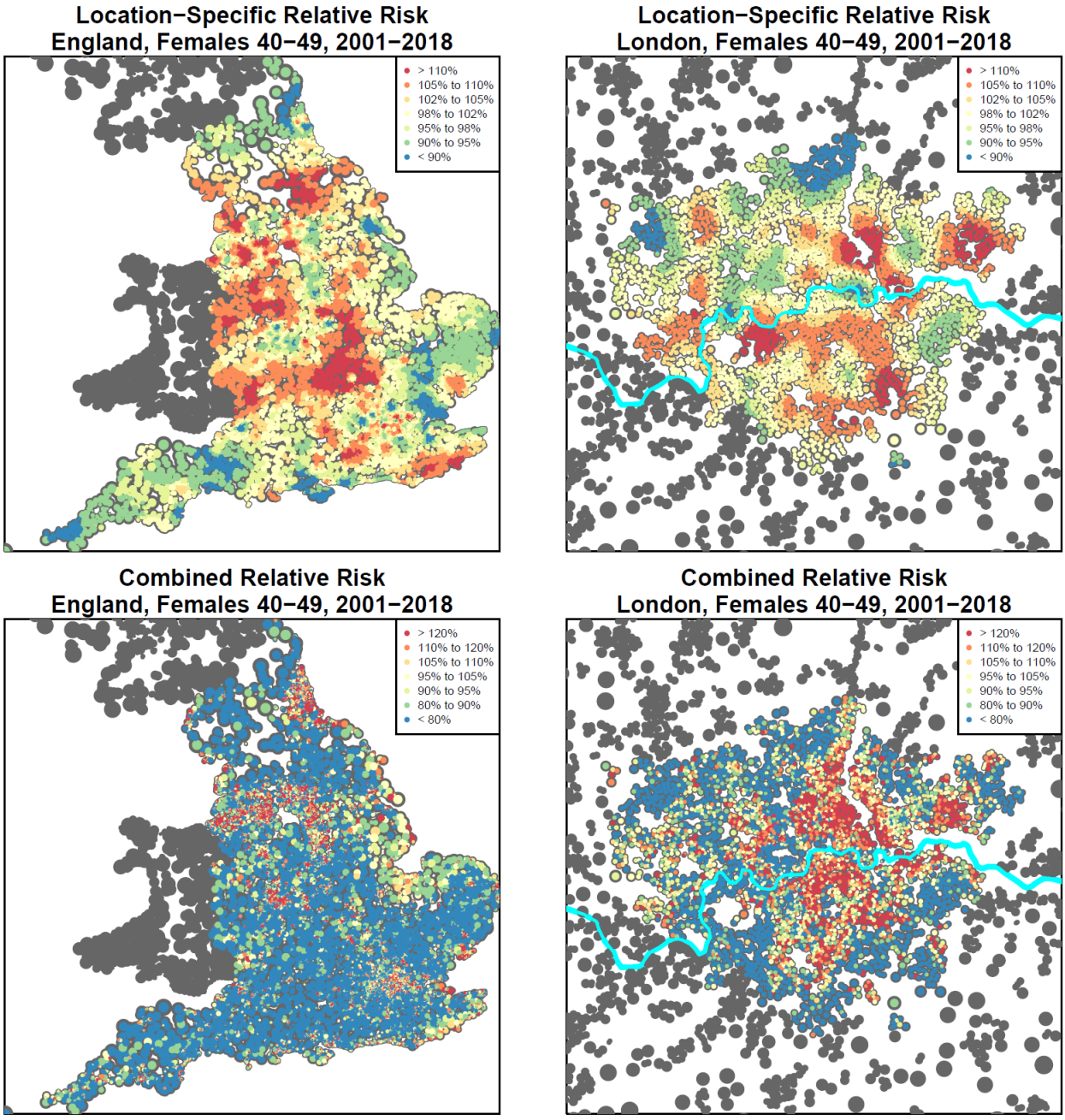

Figure 22: Top row: estimated spatial relative risk, $F_2(i)$, by LSOA for England (left) and London (right) for females, ages 40-49. Bottom row: combined relative risk, $F_1(i)F_2(i)$. Dot sizes reflect the physical size of each LSOA.

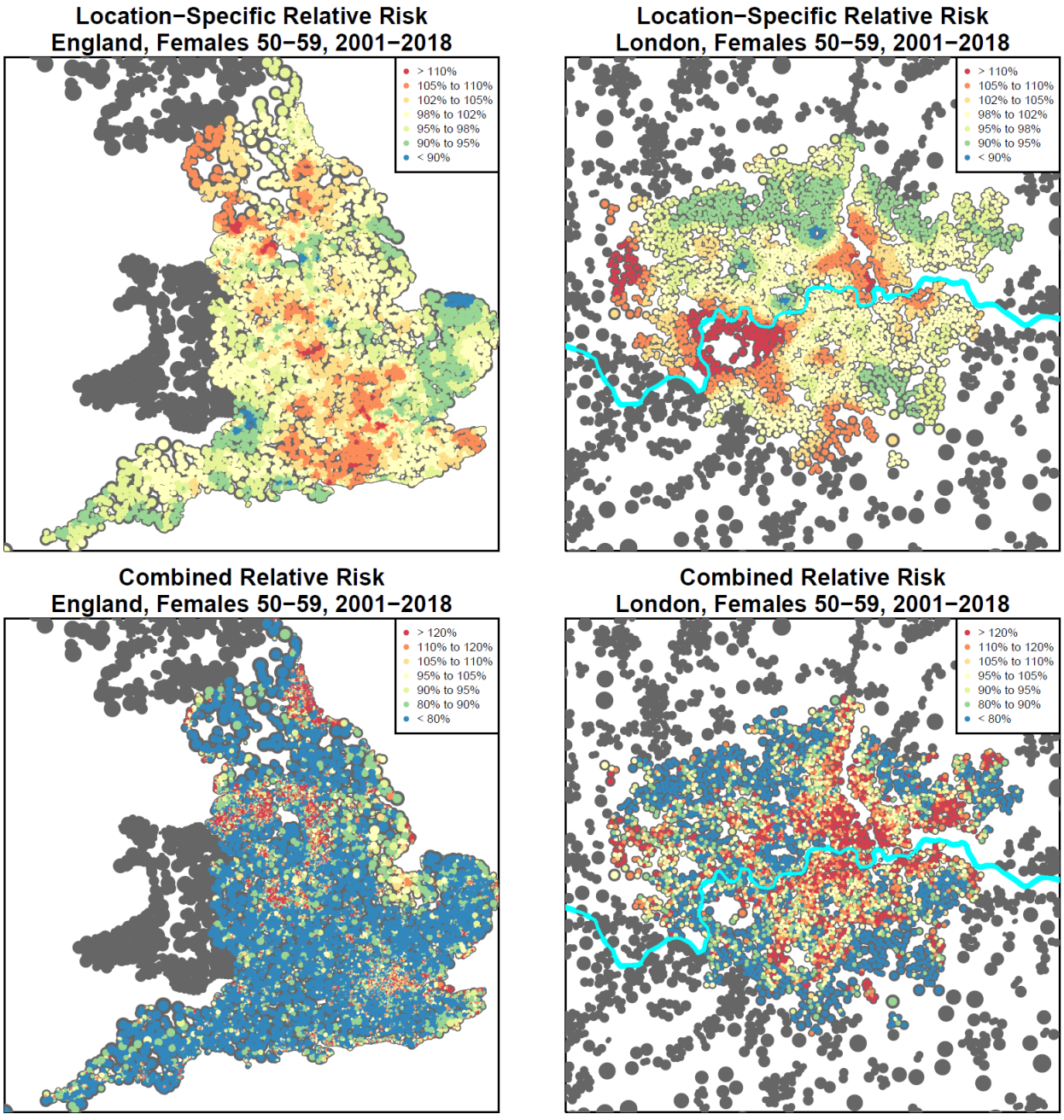

Figure 23: As Figure 22 but for ages 50-59.

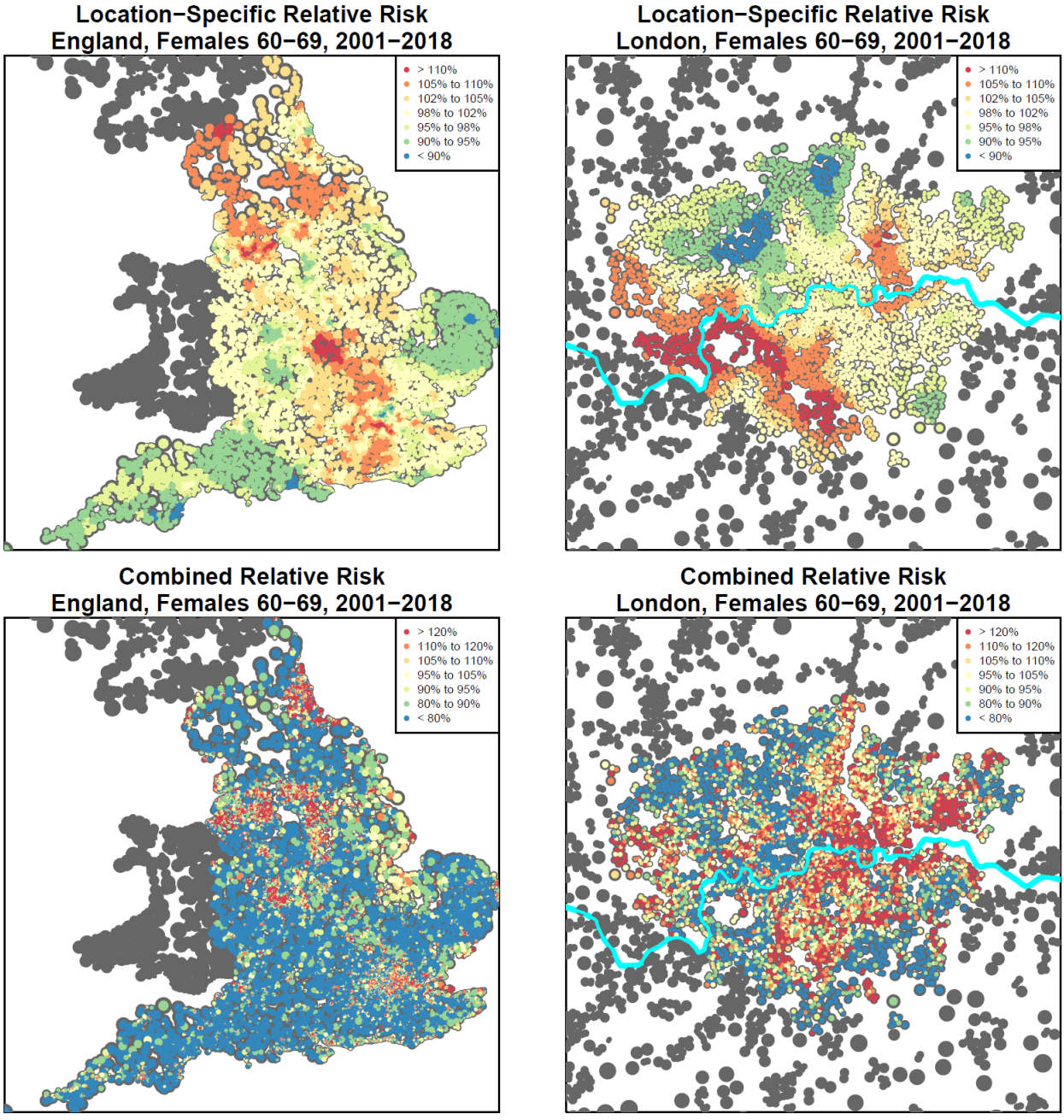

Figure 24: As Figure 22 but for ages 60-69.

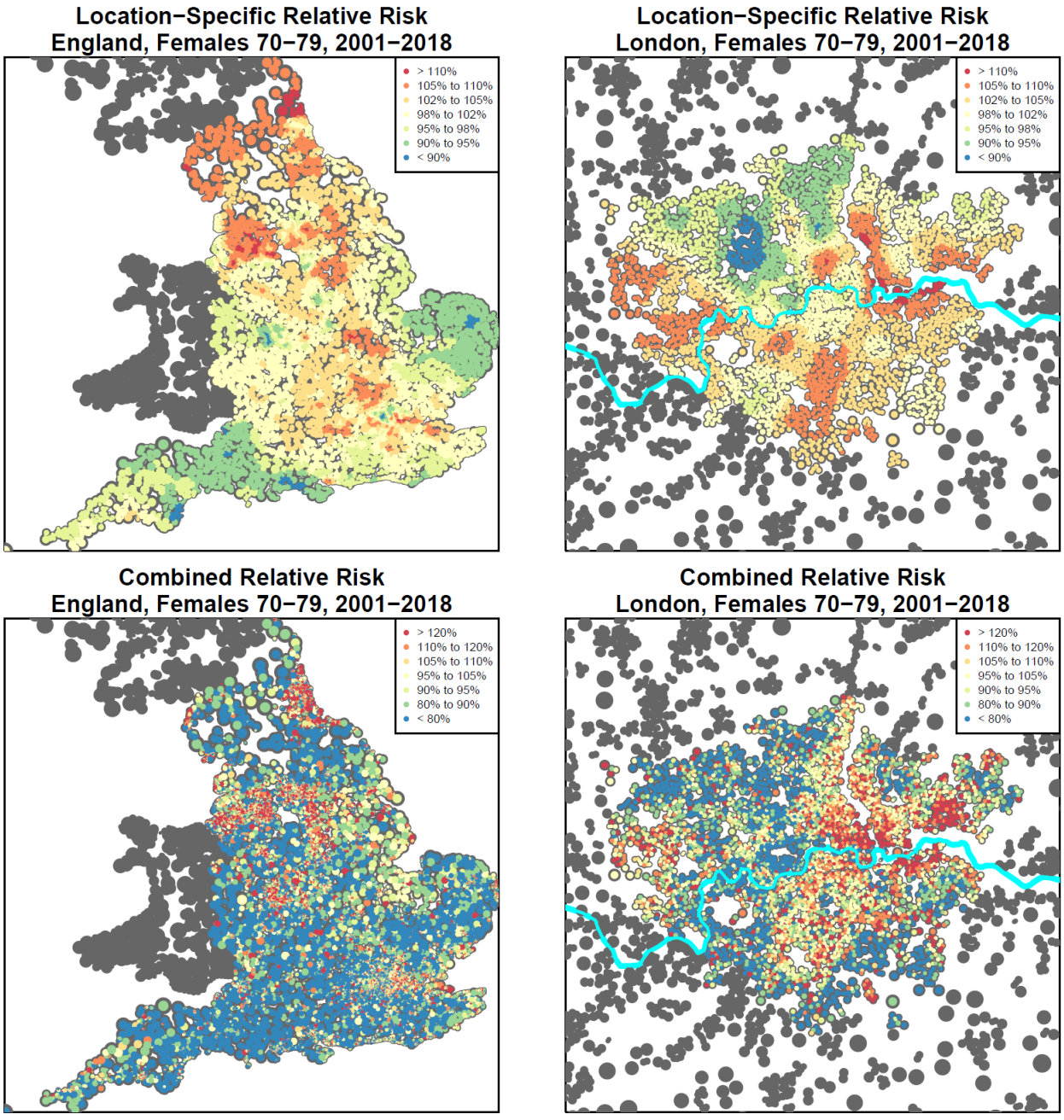

Figure 25: As Figure 22 but for ages 70-79.

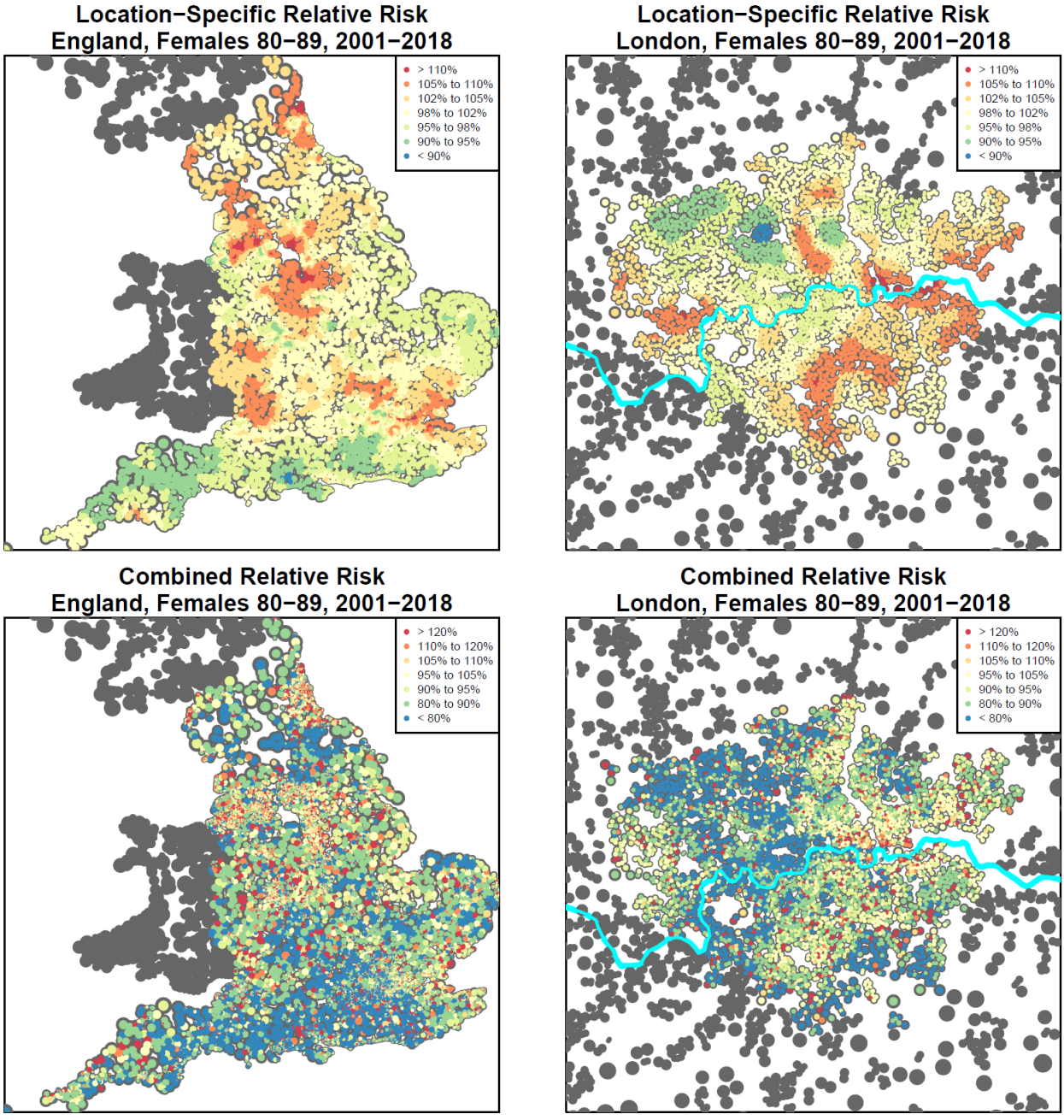

Figure 26: As Figure 22 but for ages 80-89.