



# **Modelling the dead and what they died from**

Theo Rashid

*Supervisors:* Majid Ezzati, James E Bennett, Seth Flaxman

A thesis submitted for the degree of Doctor of Philosophy  
Department of Epidemiology and Biostatistics, School of Public Health

February 2023



## Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.



## Declaration of Authorship

I, Theo Rashid, hereby declare that the work in this thesis is my own original research, and that I have appropriately cited any work within that is not my own.



*“Nobody is going to read your thesis.”*

Kyle Foreman



# Abstract

People died in England and we modelled the all cause and cause-specific death rates.  
This took longer than expected.



## Acknowledgements

Thanks be to James Bennett.



# Table of contents

<b>Copyright</b>	<b>iii</b>
<b>Declaration of Authorship</b>	<b>v</b>
<b>Abstract</b>	<b>ix</b>
<b>Acknowledgements</b>	<b>xi</b>
<b>1 Rationale, aims and objectives</b>	<b>1</b>
1.1 Rationale . . . . .	1
1.2 Objectives . . . . .	2
1.3 Structure of the thesis . . . . .	2
<b>2 Background</b>	<b>3</b>
2.1 Overview . . . . .	3
2.2 Small area health statistics (unit) . . . . .	3
2.2.1 Disease mapping at SAHSU . . . . .	5
2.3 Small area analyses of mortality . . . . .	8
2.4 Mortality by specific causes of death (counting what they died from) .	10
2.5 Health inequalities in the UK . . . . .	12
<b>3 Counting and modelling the dead</b>	<b>17</b>
3.1 Overview . . . . .	17
3.2 Counts of the dead . . . . .	17
3.2.1 Geographies of England . . . . .	19
3.2.2 Counts of the living . . . . .	21
3.2.3 Deprivation data . . . . .	21

3.2.4	Migration data . . . . .	22
3.3	Modelling the dead . . . . .	23
3.4	Inference . . . . .	23
3.5	Clean code and open source . . . . .	25
<b>4</b>	<b>Small: Life expectancy trends in England at the MSOA level</b>	<b>27</b>
4.1	Overview . . . . .	27
4.2	Methods . . . . .	27
4.2.1	Mortality and population data . . . . .	27
4.2.2	Statistical analysis . . . . .	28
4.3	Results . . . . .	31
4.3.1	Inequalities in life expectancy . . . . .	31
4.3.2	Change in life expectancy . . . . .	33
4.3.3	Life expectancy and deprivation . . . . .	37
4.3.4	Inequalities in probability of survival . . . . .	37
4.4	Discussion . . . . .	39
4.4.1	Strengths and limitations . . . . .	39
4.4.2	Comparison with previous literature . . . . .	41
4.4.3	Is England the new USA (in a bad way)? . . . . .	42
4.4.4	Summary . . . . .	43
<b>5</b>	<b>Smaller: Life expectancy inequality in London at the LSOA level</b>	<b>45</b>
5.1	Overview . . . . .	45
5.2	Methods . . . . .	45
5.2.1	Mortality and population data . . . . .	45
5.2.2	House price estimates . . . . .	46
5.2.3	Statistical analysis . . . . .	47
5.3	Results . . . . .	48
5.3.1	Inequalities in life expectancy . . . . .	48
5.3.2	Change in life expectancy and inequality . . . . .	49
5.3.3	Correlations with house price . . . . .	51
5.4	Discussion . . . . .	54
5.4.1	Strengths and limitations . . . . .	54

5.4.2 Comparison with previous literature . . . . .	56
5.4.3 Urban policy and public health implications . . . . .	57
5.5 Summary . . . . .	58
<b>6 And what they died from: cause-specific mortality in England at the district level</b>	<b>59</b>
6.1 Overview . . . . .	59
6.2 Methods . . . . .	59
6.3 Results . . . . .	62
6.3.1 All causes . . . . .	62
6.3.2 Cardiovascular diseases . . . . .	62
6.3.3 Non-communicable diseases . . . . .	64
6.3.4 Cancers . . . . .	72
6.3.5 Maternal, perinatal, nutritional and infectious causes . . . . .	72
6.3.6 Injuries . . . . .	75
6.3.7 Contribution to life expectancy change . . . . .	75
6.4 Discussion . . . . .	79
6.4.1 Comparison with previous literature . . . . .	79
6.4.2 Strengths and limitations . . . . .	81
6.5 Summary . . . . .	81
<b>7 And when they died from cancer: trends in cancer mortality at the district level in England</b>	<b>83</b>
7.1 Overview . . . . .	83
7.2 Methods . . . . .	83
7.3 Results . . . . .	84
7.4 Discussion . . . . .	84
7.5 Summary . . . . .	84
<b>8 Discussion</b>	<b>85</b>
8.1 Comparison with published literature . . . . .	85
8.2 Strengths and limitations . . . . .	85
8.3 Public health and policy implications . . . . .	85

8.4 Future work . . . . .	85
8.5 Conclusions . . . . .	86
<b>References</b>	<b>87</b>
<b>Appendices</b>	<b>96</b>
<b>A Life table methods</b>	<b>97</b>
A.1 Period life tables . . . . .	97
A.1.1 The very young ages and the very old ages . . . . .	98
A.2 Probability of dying . . . . .	99
A.3 Cause-specific decomposition of differences in life expectancy . . . . .	99
A.4 Mean age at death . . . . .	100
<b>B Model parameters and priors</b>	<b>101</b>
<b>C Hex-cartogram of life expectancy in 2019 at the MSOA level</b>	<b>107</b>
<b>D Groups of causes of death</b>	<b>109</b>

# List of Figures

3.1	Age-specific death rates for broad age groups and life expectancy in England from 2002 to 2019. . . . .	18
3.2	Age-specific death rates for broad age groups and life expectancy in England from 2002 to 2019. . . . .	19
3.3	District-level life expectancy for total deaths from 2002 to 2019. . . . .	20
4.1	Ranked MSOA life expectancies in 2002 and 2019. Each point shows the posterior median life expectancy estimate for each MSOA, forming a curve; error bars are 95% credible intervals. Arrows indicate national life expectancies in England and selected comparator countries with life expectancies within the range of English MSOAs. Hong Kong had the highest global female and male life expectancies. In the EU, Bulgaria had the lowest and Spain had the highest life expectancies for women; Latvia had the lowest and Switzerland had the highest life expectancies for men. Life expectancy for England was calculated from the data used here, and for other countries from World Bank estimates in 2019 (“Life expectancy at birth, data,” 2022). . . . .	32
4.2	Map of life expectancy and the distribution of life expectancy in 2019. The areas in white have a life expectancy equal to the national life expectancy. . . . .	33
4.3	Comparison of female and male life expectancy in 2019 and change from 2002 to 2019. . . . .	34
4.4	Geography of change in life expectancy from 2002 to 2019. . . . .	34
4.5	Map of posterior probability that the estimated change represents a true increase or decrease in life expectancy from 2002 to 2019. . . . .	35

4.6	Distribution of MSOA life expectancies in each year from 2002 to 2019. Each point shows one MSOA. The central line shows national life expectancy. . . . .	36
4.7	Maximum (highest) to minimum (lowest) and 99th to first percentile differences in life expectancy across 6791 MSOAs, 2002–19. The large difference in 2017 is due to the low life expectancy in the MSOA where the deaths in the Grenfell Tower (Kensington and Chelsea, London) fire took place. . . . .	36
4.8	Change in MSOA life expectancy in different time periods, 2002–19. Each point shows the posterior median change in one MSOA. MSOAs are coloured by their life expectancy at the beginning of each period (e.g., for 2014–19, they are coloured by life expectancy in 2014). . . . .	37
4.9	MSOA life expectancy in relation to measures of socioeconomic deprivation in the MSOA in 2002 and 2019. The socioeconomic measures are poverty, unemployment, and education. The lines show the smooth relationship fitted with locally estimated scatterplot smoothing for each year. . . . .	38
4.10	Probability of dying in specific ages in 6791 MSOAs in England in 2002 and 2019. Each point shows one MSOA. The vertical axis uses a log scale so that the large differences in survival across ages can be seen. . . . .	39
5.1	Map of life expectancy and the distribution of life expectancy in London in 2019. The areas in white have a life expectancy equal to the overall London life expectancy. . . . .	49
5.2	Geography of change in life expectancy in London from 2002 to 2019. . . . .	50
5.3	Change in LSOA life expectancy in different time periods, 2002–19. Each point shows the posterior median change in one LSOA. LSOAs are coloured by their life expectancy at the beginning of each period. . . . .	50

5.4	Life expectancy at birth and inequality in life expectancy within London's 33 local authority districts from 2002 to 2019. Each panel plots life expectancy against life expectancy inequality over time for all 33 districts, and highlights the trend line for the district which is named at the top of the panel. The districts are arranged according to their approximate location in Greater London. The large decline in life expectancy in Kensington and Chelsea in 2017 is due to the deaths caused by the fire in Grenfell Tower. . . . .	52
5.5	Distribution of estimates of Lower-layer Super Output Area (LSOA) life expectancy at birth for 2002 and 2019, and of the change from 2002 to 2019 in 33 London districts. Districts are ordered by the median life expectancy in 2002. . . . .	53
5.6	Spearman correlation between change in life expectancy and change from 2002 to 2019 in sociodemographic characteristics of the population of London Lower-layer Super Output Areas (LSOAs). LSOAs were grouped by decile of house prices in 2002 and correlations were calculated separately for each decile. Correlations close to zero imply the lack of a relationship between the ranks of the variables. . . . .	54
6.1	Total number of deaths for the 12 leading causes of death in England from 2002 to 2019, and the residual groups all other cancers, all other NCD, all other CVD, all other infections, maternal, perinatal and nutritional conditions (IMPN), and injuries. The boxes are coloured by the wider groups of CVDs; NCDs; cancers; maternal, perinatal, nutritional and infectious causes; injuries. See Table D.1 for ICD-10 codes for each category. . . . .	63
6.2	Probability of dying from ischaemic heart disease between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	65
6.3	Probability of dying from a stroke between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	66

6.4 Probability of dying from all other CVDs between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	67
6.5 Probability of dying from Alzheimer's and other dementias between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	68
6.6 Probability of dying from COPD between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	69
6.7 Probability of dying from diabetes mellitus, nephritis and nephrosis between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	70
6.8 Probability of dying for men from liver cirrhosis between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	71
6.9 Probability of dying from all other NCDs between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	73
6.10 Probability of dying from lower respiratory infections between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	74
6.11 Probability of dying from all other infections, maternal, perinatal and nutritional conditions between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. .	76
6.12 Probability of dying from injuries between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019. . . . .	77
6.13 Contributions of deaths from specific causes to life expectancy difference between the district population in 2002 and the population in 2019 in 314 local authority districts in England. The side panel shows the change in life expectancy over that time, by summing over the causes. The top panel shows the contribution for each districts for each cause.	78





# List of Tables

3.1	The number of each geographical unit of England used in thesis and their populations in 2019. . . . .	20
B.1	Specification of the Bayesian statistical model in Equation 4.3. . . . .	101
B.2	Specification of the Bayesian statistical model in Equation 5.1. . . . .	103
B.3	Specification of the Bayesian statistical model in Equation 6.1. . . . .	104
D.1	Groups of causes of death used in the district-level cause-specific analysis in Chapter 6 with ICD-10 codes. . . . .	109
D.2	Groups of causes of death used in the district-level cancer analysis in Chapter 7 with ICD-10 codes. . . . .	111



# List of Abbreviations

<b>BUGS</b>	Bayesian inference <b>U</b> sing <b>G</b> ibbs <b>S</b> ampling
<b>CrI</b>	<b>C</b> redible <b>I</b> nterval
<b>CAR</b>	<b>C</b> onditional <b>A</b> utoregressive
<b>COPD</b>	<b>C</b> hronic <b>O</b> bstructive <b>P</b> ulmonary <b>D</b> isease
<b>CVD</b>	<b>C</b> ardiovascular <b>D</b> isease
<b>GBD</b>	<b>G</b> lobal <b>B</b> urden of <b>D</b> isease
<b>GHE</b>	<b>G</b> lobal <b>H</b> ealth <b>E</b> stimates
<b>ICAR</b>	<b>I</b> ntrinsic <b>C</b> onditional <b>A</b> utoregressive
<b>ICD</b>	<b>I</b> nternational <b>C</b> lassification of <b>D</b> iseases
<b>IMD</b>	<b>I</b> ndex of <b>M</b> ultiple <b>D</b> eprivation
<b>IMPN</b>	<b>I</b> nfections, <b>M</b> aternal <b>P</b> erinatal and <b>N</b>
<b>LSOA</b>	<b>L</b> ower <b>L</b> ayer <b>S</b> uper <b>O</b> utput <b>A</b> rea
<b>MCMC</b>	<b>M</b> arkov <b>c</b> hain <b>M</b> onte <b>C</b> arlo
<b>MSOA</b>	<b>M</b> iddle <b>L</b> ayer <b>S</b> uper <b>O</b> utput <b>A</b> rea
<b>NCD</b>	<b>N</b> on- <b>cD</b> isease
<b>OA</b>	<b>O</b> utput <b>A</b> rea
<b>NUTS</b>	<b>N</b> o <b>U</b> -turn <b>S</b> ampler
<b>SAHSU</b>	<b>S</b> mall <b>A</b> rea <b>H</b> ealth <b>S</b> tatistics <b>U</b> nit



## Chapter 1

# Rationale, aims and objectives

### 1.1 Rationale

Although the UK is, by global standards, a wealthy nation with relatively high life expectancy, the country still suffers vast inequalities in mortality. While there have been studies on trends in subnational life expectancy in England, these have been limited in resolution to the district level (Bennett et al., 2015). There is vast heterogeneity in health outcomes and their determinants within districts. As a notable example of mortality from recent years, Kensington and Chelsea, which contains both the most affluent areas in the country in its south and some of the poorest in its north, was the district containing the Grenfell Tower tragedy in 2017. There is a clear need to uncover further variation in mortality below the district level.

Since around 2010, although national life expectancy has continued to increase, there is evidence that the rise in female life expectancy has reversed in the most deprived deciles of the population (Bennett et al., 2018; Marmot et al., 2020). There are likely to be important changes in the cause-specific composition of mortality that have driven this reversal.

The title of this thesis is adapted from Mathers et al. (2005): *Counting the dead and what they died from: an assessment of the global status of cause of death data*.

## 1.2 Objectives

The aim of my thesis is to estimate trends in mortality for small areas over the past 20 years in England. There are three specific objectives which will help achieve this aim:

1. To estimate trends in life expectancy for Middle-layer Super Output Areas in England from 2002 to 2019.
2. To investigate trends in life expectancy for Lower-layer Super Output Areas in estimate from 2002 to 2019.
3. To investigate trends in cause-specific mortality for districts in England from 2002 to 2019.

## 1.3 Structure of the thesis

Chapter 2 reviews the literature on spatial methods for disease mapping using the studies at the Small Area Health Statistics Unit - who hold the data for the research in this thesis - as examples. I also review studies on mortality for small subnational regions, followed by the literature of separating total mortality into different causes of death, before exploring inequalities in UK over the past few decades through to the present. Chapter 3 provides an overview of the data sources and statistical modelling choices common to all objectives of this thesis. Chapter 4 concerns the first objective of the thesis - estimating trends in life expectancy for small areas in England. Chapter 5 addresses the second objective by focussing on London at a finer scale than the previous chapter. Chapter 6 addresses objective three of this thesis, breaking down total mortality in England into specific causes of deaths at a coarser scale, and looking at potential drivers for the observed trends in life expectancy. Chapter 7 follows the methods of Chapter 6, but focussing only on deaths from cancers. Chapter 8 concludes with a discussion on the public health implications of the findings and areas for future research building on the work in this thesis.

## Chapter 2

# Background

### 2.1 Overview

The section begins with a brief history of the Small Area Health Statistics Unit (SAHSU), who manage the mortality data used in the thesis. In particular, I will focus on spatial methods, which were developed at the unit and I extend upon in the thesis, in the context of disease mapping studies of SAHSU.

There is a section on literature from the UK and beyond looking at mortality for small subnational regions, followed by a history of separating total mortality into different causes of death and the epidemiologic transition theory.

I will finish by exploring the picture of inequalities in UK over the past few decades through to the present, focussing on class, income, geography, and deprivation.

### 2.2 Small area health statistics (unit)

In 1983, a documentary on the fallout from a fire at the Sellafield nuclear site in Cumbria claimed that there was a ten-fold increase in cases of childhood leukaemia in the surrounding community. This anomaly had gone undetected by public health authorities, raising concern that routinely collected data were not able to identify local clusters of disease. The subsequent enquiry confirmed the excess, and recommended that a research unit was set up to monitor small area statistics and respond quickly

to *ad hoc* queries on local health hazards. The Small Area Health Statistics Unit (SAHSU) was established in 1987 (Elliott et al., 1992).

Beyond producing substantive research on environment and health, a core aim of SAHSU is to develop small area statistical methodology (Wakefield and Elliott, 1999) for:

- *Point source type studies.* Is there an increased risk close to an environmental hazard?
- *Geographic correlation studies.* Is there a correlation between disease risk and spatially-varying environmental variables?
- *Clustering.* Does a disease produce non-random spatial patterns of incidence? If the aetiology is unknown, this could suggest the disease is infectious.
- *Disease mapping.* Summarising the spatial variation in risk.

In a pilot study for SAHSU, Elliott et al. (1992) investigated the mortality near the Plymouth docks from mesothelioma and asbestosis – both of which are related to industrial exposure and asbestos. Death registrations with postcode information were held by SAHSU, and concentric circular bands were drawn around the Plymouth dockyards as a way to approximate the exposure from a point source of environmental pollution. There was a clear increase in risk within 3km of the docks. A similar distance-based approach was adopted to look at excess respiratory disease mortality near two factories in Barking and Havering (Aylin et al., 1999), and kidney disease mortality near chemical plants in Runcorn (Hodgson et al., 2004). In response to public concern over exposure to toxic chemicals in waste materials, SAHSU conducted the most extensive study ever into the health effects of landfill sites. Postcodes within a 2km buffer of a landfill site were classified as exposed. Compared to those living beyond 2km, SAHSU found a small unexplained excess of congenital anomalies (Elliott et al., 2001a), no increase in rates of cancer (Jarup et al., 2002b), and no excess risk of Down syndrome (Jarup et al., 2007).

Distance from source is, however, only a basic model for the exposure, which can exhibit more complex, directional spatial patterns. A number of SAHSU studies have

employed physics-informed models to create an exposure surface, and assess the geographic correlations between this surface and the health outcome, notably for a plume of mercury pollution (Hodgson et al., 2007), exposure to mobile phone base stations during pregnancy (Elliott et al., 2010), noise from aircraft near Heathrow (Hansell et al., 2013), road traffic noise in London (Halonen et al., 2015), and particulate matter from incinerators during pregnancy (Parkes et al., 2020). SAHSU published an environment and health atlas for England and Wales, showing the spatial patterns of 14 health conditions at census ward level over an aggregated 25 year period alongside five environmental exposure surfaces (Hansell, Anna L. et al., 2014).

### 2.2.1 Disease mapping at SAHSU

Many of the studies at SAHSU focus on prevalence or mortality from rare diseases at small areas. The number of cases, or number of deaths, in a region are likely to be small numbers. This sparseness issue is even more pertinent when the population is further stratified by age group. Rates calculated from observed data present apparent variability between spatial units, which is larger than the true differences in risk. There is a need for statistical smoothing techniques to obtain robust estimates of rates by sharing information between strata. Aylin et al. (1999) mapped diseases for wards in Kensington, Chelsea and Westminster using a simple model that smoothed rates towards the mean risk across the region. Thereafter, SAHSU published a plethora of studies for disease mapping models with explicit spatial dependence, which are designed to give more weight to nearby areas than those further away.

There are three main categories for modelling spatial effects. First, we can treat space as a continuous surface, as with Gaussian processes or splines. Second, we can use areal models, which make use of spatial neighbourhood structure of the units. Thirdly, we can explicitly build effects based on a nested hierarchy of geographical units, for example between state, county and census tract in the US.

In the context of disease mapping, events are usually aggregated to areas rather than assigned specific geographical coordinates. Wakefield and Elliott (1999) model aggregated counts as realisations of a Poisson process, in which the expected number of cases is calculated by integrating a continuous surface that generates the cases

over the area of the spatial unit. The surface was a function of spatially-referenced covariates. Kelsall and Wakefield (2002) describe an alternative model, where the log-transformed risk surface is modelled by a Gaussian process, whose correlation function depends on distance.

Best et al. (2005) provide a review of the use of hierarchical models with spatial dependence for disease mapping. In particular, the authors focus on Bayesian estimation, and different classes of spatial prior distributions.

The first prior proposed for spatial effects  $\mathbf{S} = S_1, \dots, S_n$  is the multivariate normal

$$\mathbf{S} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (2.1)$$

where  $\boldsymbol{\mu}$  is the mean effect vector,  $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$  and  $\boldsymbol{\Omega}$  is a symmetric, positive semi-definite matrix defining the correlation between spatial units. A common choice when specifying the structure of the correlation matrix is to assume a function that decays with the distance between the centroids of the areas, so that places nearby in space share similar disease profiles. Note, this is mathematically equivalent to the practical implementation of a Gaussian process, which uses a finite set of points. An example in Elliott et al. (2001b) chooses the exponential decay function to map cancer risk in northwest England.

A more popular prior is the conditional autoregressive (CAR) prior, also known as a Gaussian Markov random field. These form a joint distribution as in Equation 2.1, but the covariance is usually defined instead in terms of the precision matrix

$$\mathbf{P} = \boldsymbol{\Sigma}^{-1} = \tau(\mathbf{D} - \rho \mathbf{A}), \quad (2.2)$$

where  $\tau$  controls the overall precision of the effects,  $\mathbf{A}$  is the spatial adjacency matrix formed by the small areas,  $\mathbf{D}$  is a diagonal matrix with entries equal to the number of neighbours for each spatial unit, and the autocorrelation parameter  $\rho$  describes the amount of correlation. This can be seen as tuning the degree of spatial dependence, where  $\rho = 0$  implies independence between areas, and  $\rho = 1$  full dependence. The

case with  $\rho = 1$  is called the intrinsic conditional autoregressive (ICAR) model. Besag et al. (1991) proposed the model (hereafter called BYM)

$$S_i = U_i + V_i, \quad (2.3)$$

where  $U_i$  follow an ICAR distribution, and  $V_i$  are independent and identically distributed random effects. The BYM distribution was employed to model spatial variation in the relative risk of testicular (Toledano et al., 2001) and prostate (Jarup et al., 2002a) cancers for small areas in regions of England.

Further disease mapping studies at SAHSU using spatially structured effects have also extended the methodology to look at age patterns and trends over time. Asaria et al. (2012) analysed cardiovascular disease death rates by fitting a spatial model for all wards in England separately for each age group and time period. Bennett et al. (2015) designed a model to jointly forecast all-cause mortality for districts in England, age groups and years. The model used BYM spatial effects and random walk effects over age and time to capture non-linear relationships. It is also possible to borrow information across causes of death, as applied in Foreman et al. (2017) to forecast cause-specific mortality for states in the US. Random walk effects were again used for temporal non-linearities, a CAR prior was used for spatial effects, and a multivariate normal prior for causes of death whereby the covariance matrix describes the correlation structure between the 15 cause groups. The model did not, however, share information between age groups.

The relationships between different levels of a hierarchy of geographical units are often incorporated into models as a nested hierarchy of random effects. These models account for when spatial units lie within common administrative boundaries. This is often a desirable property of the model for certain geographies, like states in the US, which are administrative. Policy is decided at these geographies, so there is reason to believe these boundaries may have a greater effect on health outcomes than spatial structure. Although not used in previous SAHSU studies, Finucane et al. (2014) demonstrate how country-level blood pressure can be modelled by exploiting the hierarchy global, subregion, region and country. Note, although these models

group by geographical region, these models are not *spatial* as they do not contain any information on the relative position of the areas.

## 2.3 Small area analyses of mortality

In order to compare the health status between areas, health authorities require a measure of mortality that collapses age-specific information into a single number. Indirectly standardised measures such as the standardised mortality ratio – the ratio between total deaths and expected deaths in an area – are easy to calculate, but are not easily understood by laypeople. Directly standardised methods, in contrast, require knowledge of the full age structure of death rates rather than just the total number of deaths. Age-standardised death rates, however, suffer the same interpretability issue as the standardised mortality ratio, and are only comparable between studies if the same reference population is used. An alternative choice is *life expectancy*. Silcocks et al. (2001) explain that life expectancy is a “more intuitive and immediate measure of the mortality experience of a population, [and] is likely to have greater impact... than other measures that are incomprehensible to most people.”

The estimation of death rates requires two data sources: deaths counts and populations. Modern vital registrations systems are complete and accurate, so data on deaths are usually reliable and comprehensive. On the other hand, although usually treated as a known quantity, the population denominator is often problematic. Populations for small geographies are only recorded during a decennial census, and estimates are generated for the years in-between using limited survey data on births, deaths and migration. And although the census is considered the “gold standard”, it is subject to enumeration errors, particularly for areas with special populations such as students or armed forces (Elliott et al., 2001b).

Beyond the population issue, finer scale studies are restricted by data availability. Where data are available, there is still the need to overcome small number issues before feeding death rates through the life table. Eayres and Williams (2004) recommend a minimum population size of 5000 when using traditional life table methods, below which the calculation of life expectancy is unstable, impossible, or the error estimates

become so large that any comparison between subgroups becomes meaningless. One approach, often taken by statistical agencies, is to build larger populations by either aggregating multiple years of data (Bahk et al., 2020; “Health Expectancies at Birth for Middle Layer Super Output Areas (MSOAs), England,” 2015; “Local Health - Small Area Public Health Data,” 2021) or combining spatial units (Ezzati et al., 2008). Here, we focus on studies using Bayesian hierarchical models to generate robust estimates of age-specific death rates by recognising the correlations between spatial units and age groups, which produce more accurate estimates for small population studies of life expectancy (Congdon, 2009; Jonker et al., 2012).

Jonker et al. (2012) demonstrated the advantages of the Bayesian approach for 89 small areas in Rotterdam using a joint model for sex, space and age effects, finding a 8.2 year and 9.2 year gap in life expectancy for women and men. Stephens et al. (2013) employed the same model for 153 administrative areas in New South Wales, Australia.

Bayesian spatial models for mortality have been scaled to small areas for entire countries, and also consider trends in these regions over time. Bennett et al. (2015) forecasted life expectancy for 375 districts in England and Wales using an ensemble of spatiotemporal models trained over a 31 year period, and Dwyer-Lindgren et al. (2017a) explored mortality trends 3110 US counties from 1980 to 2014.

There have also been studies on specific cities at a finer resolution. In order to improve estimates for disability-free life expectancy, Congdon (2014) considered both ill-health and mortality in a joint likelihood with spatial effects for 625 wards in London, finding more than a two-fold variation in the percent of life spent in disability for men. Bilal et al. (2019) looked at 266 subcity units for six large cities in Latin America. As there is no contiguous boundary in this case, a random effects model for each city was used instead of a spatial model. The largest difference between the top and bottom decile of life expectancy at birth was 17.7 years for women in Santiago, Chile.

Two studies in North America have looked below the county level, at census tracts, with wide-ranging population sizes as small as 40. Dwyer-Lindgren et al. (2017b) studies trends for life expectancy and many causes of death for 397 tracts in King

County, Washington, uncovering an 18.3 year gap in life expectancy for men. Using the same model for Vancouver, Canada, Yu et al. (2021) found widening inequalities over time and a difference of 9.5 years for men.

## 2.4 Mortality by specific causes of death (counting what they died from)

In the mid-twentieth century, a team in the US Public Health Service, led by Iwao Moriyama, began looking into the cause-specific composition of mortality into all diseases and injuries for the first half of the century. Notably, Moriyama and Gover (1948) grouped vital registration data into primary causes, and found as the US saw an overall downward trend in mortality, the leading causes of death changed from communicable diseases, such as tuberculosis and diphtheria, toward non-communicable, “chronic diseases of older ages”, such as heart diseases and cancers. The success of the reduction – and in the case of typhoid fever, near-elimination – of infectious diseases was attributed to the strategy of the health officer in the early 1900s, who was focussed on improving water and sanitation, and public health interventions such immunisation and quarantines.

By comparing vital registration data over several centuries, Abdel Omran observed this shift of mortality from communicable to non-communicable diseases (NCDs) in many countries (Omran, 1977, 1971). Although the pace and determinants of the transition varied between countries, Omran formalised three common successive stages of the shift in mortality:

1. *The Age of Pestilence and Famine.* Mortality is high and largely governed by malthusian “positive checks” – epidemics, famines, and wars.
2. *The Age of Receding Pandemics.* Mortality decreases as epidemics become less frequent, but infectious diseases remain the leading causes of death.
3. *The Age of Degenerative and Man-made diseases.* Mortality declines further along with fertility, increasing the average age of population and NCDs take over as the leading causes of death.

He termed this the *Epidemiologic Transition theory*. Omran (1971) explained that England and Wales took the classic transition path followed by western societies, whereby socioeconomic factors such as improvements to living standards are crucial in causing easily preventable diseases to subside and shifting towards the third phase of the transition, and medical and other public health technology only help society much later in the final stage. Later, Olshansky and Ault (1986) would propose a fourth stage to the theory, *the Age of Delayed Degenerative Diseases*, in which the structure of causes of death is stable, but the age at which degenerative diseases kill is postponed, thus decreasing older age mortality. There are, however, questions around the universality and unidirectionality of the theory, with many examples in which age-specific death rates for population subgroups have risen over time, most notably the HIV/AIDS pandemic (Gaylin and Kates, 1997). Gersten and Wilmoth (2002) also criticise the lack of attention Omran's theory pays towards the role of infection in chronic and degenerative diseases, in particular certain cancers.

Around the same time as Omran, Samuel Preston collated cause-specific mortality data for a huge number of populations, spanning 48 nations and nearly a century (Preston, 1970; Preston and Nelson, 1974). This would enable international comparisons of groups of causes of death over different time periods, and a deeper understanding of the upward trends in life expectancy. In particular, by plotting cause-specific disease rates against overall mortality, Preston and Nelson (1974) saw that, over time, the contribution of infectious diseases to a particular *level* of mortality had become ever smaller. That is to say, as mortality declines, the contribution from infectious diseases also declines. Preston attributed this to an accelerating rate of medical progress guided by the “germ theory of disease”, which public health and science were not able to replicate for NCDs. Preston also traced the excess deaths in older males observed in western societies to cardiovascular diseases, cancer and bronchitis – a direct result of dramatic increases in cigarette smoking (Preston, 1970).

Since its first edition in 1990, the subject of international comparisons of the cause-specific composition of mortality has been the remit of the Global Burden of Disease (GBD) studies (Murray and Lopez, 1996). The studies aim to quantify and compare the burden of diseases, injuries, and risk factors, usually through cross-sectional

methods but occasionally by examining trends and subnational populations (Dwyer-Lindgren et al., 2017a; Ezzati et al., 2008). An important innovation of the GBD study was the introduction of a hierarchical classification of groups of causes, with the broadest level divided into three groups: communicable, maternal, perinatal, and nutritional diseases (Group 1), NCDs (Group 2), and injuries (Group 3). Salomon and Murray (2002) made use of the wide-ranging dataset and grouping from the GBD to revisit the epidemiologic transition for the second half of the twentieth century, finding the majority of the change in cause structure occurred among children, with a shift from Group 1 to Groups 2 and 3, and in young adults, where the role of injuries is more dominant for men.

## 2.5 Health inequalities in the UK

While the UK is, by global standards, a wealthy nation with relatively high life expectancy, and the breadth of health inequalities are nowhere near the extremes seen in many other countries, the nation suffers still vast, preventable inequalities in mortality and morbidity. There are several ways to stratify the UK population and compare inequalities between subgroups. Here, we focus on class, income, geography, and deprivation.

The notion of class is prominent in UK society, but health outcomes between classes are difficult to separate from other risk factors such as hazards in manual labour or smoking rates. The Whitehall study of 1967 followed 17,530 men working in the civil service and recorded their mortality over a 10-year period. Marmot et al. (1984) found, by classifying the civil servants into social class according to their employment grade, there was a three-fold difference in mortality between the highest class, administrators, and men in the lowest class, mainly messengers and unskilled manual workers. They found, in general, a strong inverse association between grade and mortality – a term Marmot has coined a “social gradient”. The men were working stable, sedentary jobs in the same office building in London, so the gradient could not be fully explained by smoking or industrial exposure alone. The authors concluded there must be other factors inherent to social class (defined here by employment), which explain the mortality differences. A second cohort of Whitehall employees

from 1985 to 1988, this time including women as well as men, were screened and asked to answer questions on self-reported ill-health. Marmot et al. (1991) found the social gradient in health had persisted in the 20 years separating the studies. In 2008, Marmot was asked by the Secretary of State to conduct a review into the state of health inequalities in the UK and to use the evidence to design policy for reducing these inequalities. A key plot in the first Marmot Review, released in 2010, depicted the social gradient in mortality for regions in England by socio-economic classification of employment (Marmot et al., 2010).

Income is not a routinely collected statistic in the UK. Nevertheless, using a small survey of 7000 people on three measures of morbidity, Wilkinson (1992) showed health improved sharply from the lowest to the middle of the income range.

In 2015, the GBD Study released its first subnational estimates of mortality, starting with the UK and Japan. Steel et al. (2018) assess these data, which divided the UK into 150 regions, finding mortality from all-causes varied twofold across the country, with the highest years of life lost in Blackpool and the lowest in Wokingham. In a study on forecasting subnational life expectancy in England and Wales, Bennett et al. (2015) estimated a 8.2 year range in life expectancy for men and 7.1 year range for women in 2012 between 375 districts. The lowest life expectancies were seen in urban northern England, and the highest in the south and London's affluent districts. Within London itself, male and female life expectancy showed 5-6 years of variation.

There have been substantial efforts in the UK to measure the deprivation of an area. Since 2004, the standard deprivation indicator in England is the Index of Multiple Deprivation (IMD) – a composite indicator for each Lower-layer Super Output Area (LSOA) covering income, unemployment, health, crime and environmental data sources (“English indices of deprivation 2019,” 2019). The Marmot Report presented life expectancy and disability-free life expectancy against IMD at the Middle-layer Super Output Area, which exhibit strong social gradients (Marmot et al., 2010). The GBD study found the 15 most deprived UTLAs had consistently raised mortality, especially for all causes, lung cancer and chronic obstructive pulmonary disease. Deprived UTLAs in London, such as Tower Hamlets, Hackney, Barking and Dagenham

did better than expected for that level of deprivation (Steel et al., 2018). Bennett et al. (2018) jointly estimated death rates by age, year and deprivation decile. They found since 2011, although national life expectancy has continued to increase, the rise in female life expectancy has reversed in the two most deprived deciles. The second Marmot Review in 2020 also found female life expectancy declined in the most deprived decile between the periods 2010-12 and 2016-18 (Marmot et al., 2020). Diving further into these trends by region, the report found this trend was seen in all regions except London, the West Midlands and the North West, and that male life expectancy in the bottom decile also decreased in the North East, Yorkshire and the Humber, and the East of England.

Since the turn of the millennium, there have been two periods of contrasting health policy in the UK. The early 2000s saw the implementation of the English health inequalities strategy under New Labour, with explicit goals of reducing geographical inequalities in life expectancy. The strategy saw a large increase in public spending targeting the social determinants of health, with policies on supporting families, tackling deprivation, and preventative healthcare. Barr et al. (2017) analysed the trends in life expectancy for different quantiles of deprivation and provided evidence that the strategy achieved its aim of reducing the gap in life expectancy between the 20% most-deprived areas and the rest of the English population.

Following the change in government in 2010, the strategy came to an end. The Conservative government implemented a widespread series of cuts to public services, collectively known as austerity. The cuts were geographically unequal, with reductions to funding for local governments particularly severe to local authorities with an old industrial base (Gray and Barford, 2018). The study by Barr et al. (2017) saw that the trends in inequality reduction were reversing since 2012. These trends have been found at both ends of the life course: rising infant mortality associated with childhood poverty (Taylor-Robinson et al., 2019), and falls in female life expectancy at 65 and 85 (Hiam et al., 2018). Although it is difficult to uncover causal relationships, Alexiou et al. (2021) found strong associations between cuts to local government and the change in district-level life expectancy from 2013 to 2017. As written in the *The New York Times*, “after eight years of budget cutting, Britain is looking less like the rest of

Europe and more like the United States, with a shrinking welfare state and spreading poverty” (Goodman, 2018) – a comparison only compounded by the Brexit vote in 2016.

Policies of austerity were brought about as a response to the financial crash of 2008, which were global in scale, and many countries adopted similar fiscal strategies. However, in an international study comparing mortality trends in England and Wales to 22 industrialised countries, Leon et al. (2019) show that although there was a general slowdown in improvement of life expectancy across many nations, the slowdown in the most recent period of the study, 2011-16, was more pronounced in England and Wales. More recently, The Economist found the same evidence, comparing the long-run trend from 1980-2011 through to 2022 for 12 European countries: “longer-run slowdowns in life expectancy are observable in other European countries... but none has stalled quite as much as Britain” (“Britain has endured a decade of early deaths. Why?” 2023).

After a decade of cuts, the UK entered the 2020s facing the greatest public health challenge for a generation: the Covid-19 pandemic. Unsurprisingly, England and Wales suffered one of the highest excess deaths tolls relative to other industrialised countries (Kontis et al., 2020). At the current time of writing in 2023, the health service is yet to recover, with waiting lists for operations and waiting times for emergency care unhealthily high (Dorling, 2023). There is no suggestion that the picture of health and health inequalities in the UK should improve in the coming years.



## Chapter 3

# Counting and modelling the dead

### 3.1 Overview

This chapter presents the datasets, data cleaning, and modelling choices that are common between the proceeding chapters.

### 3.2 Counts of the dead

This thesis is primarily concerned with modelling death rates for small areas in England. This requires two data sources: counts of deaths, and populations counts. The counts of deaths come from de-identified civil registration data for all deaths in England from 2002 to 2019. In other words, every death in England from 2002 to 2019.

The data is extracted from the Office for National Statistics (ONS) database and held by SAHSU in a secure environment as individual death records are identifiable data. The data are updated every year and are mostly complete for previous years, but a handful of deaths are registered in later extracts if the ONS have been waiting on coroner's report to identify the underlying cause of death.

Each record comes with information on postcode of residence, allowing us to assign each death into a spatial unit for analysis. For each analysis, deaths were stratified into the following age groups: 0, 1–4, 5–9, 10–14, then 5-year age groups up to

80–84, and 85 years and older. There are also a series of ICD-10 (International Classification of Diseases, Tenth Revision) codes from the death certificate associated with the underlying and contributory causes leading to the death. Here, I focus only on the underlying cause of death, which has been assigned using selection algorithms to improve consistency between doctors (“User guide to mortality statistics,” 2022).

Before fitting any models, it’s good practice to explore the data visually, in this case looking at how total mortality varies over different cross sections: sex, age, space, time.

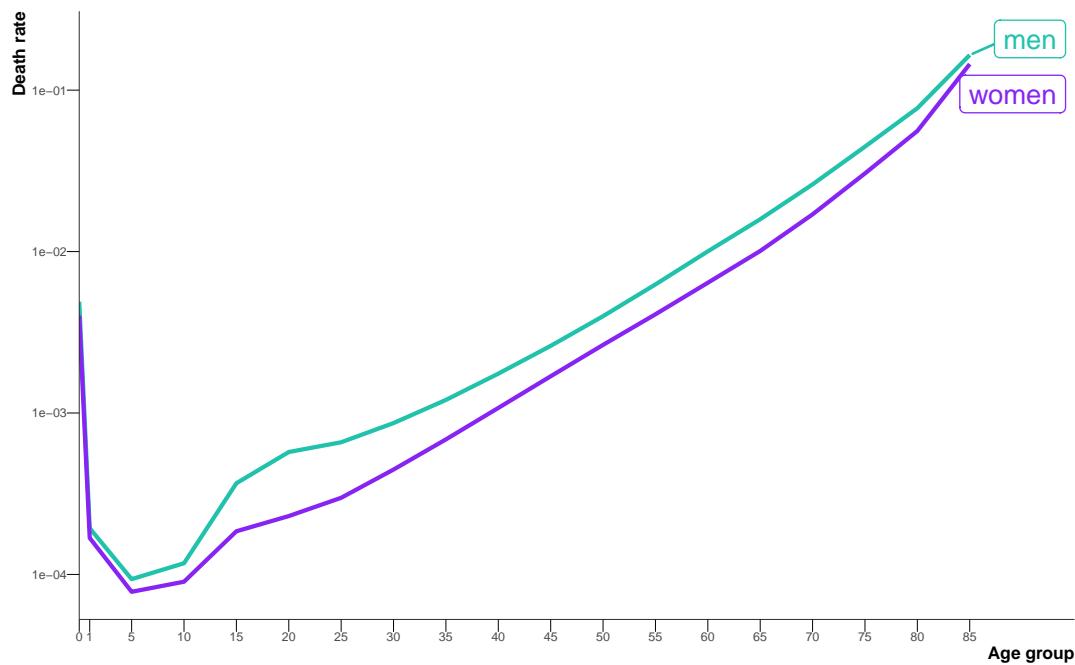


FIGURE 3.1: Age-specific death rates for broad age groups and life expectancy in England from 2002 to 2019.

The general age pattern, after aggregating all years and separating into age groups, follows a J-shape curve, with raised infant and older age mortality (Figure 3.1). Male mortality is higher at all ages, but particularly in young adulthood (15–29 years) due to injuries resulting from risky behaviour.

Figure 3.2 looks at the trends over time for wider age groups. I did not use the 5-year age groups because the number of deaths for certain age groups at single years were small, and any data presented should be non-disclosive in accordance with SAHSU’s data sharing agreement. In general, death rates have decreased from 2002 to 2019 in all age groups, but with slowing progress in young adulthood and working ages

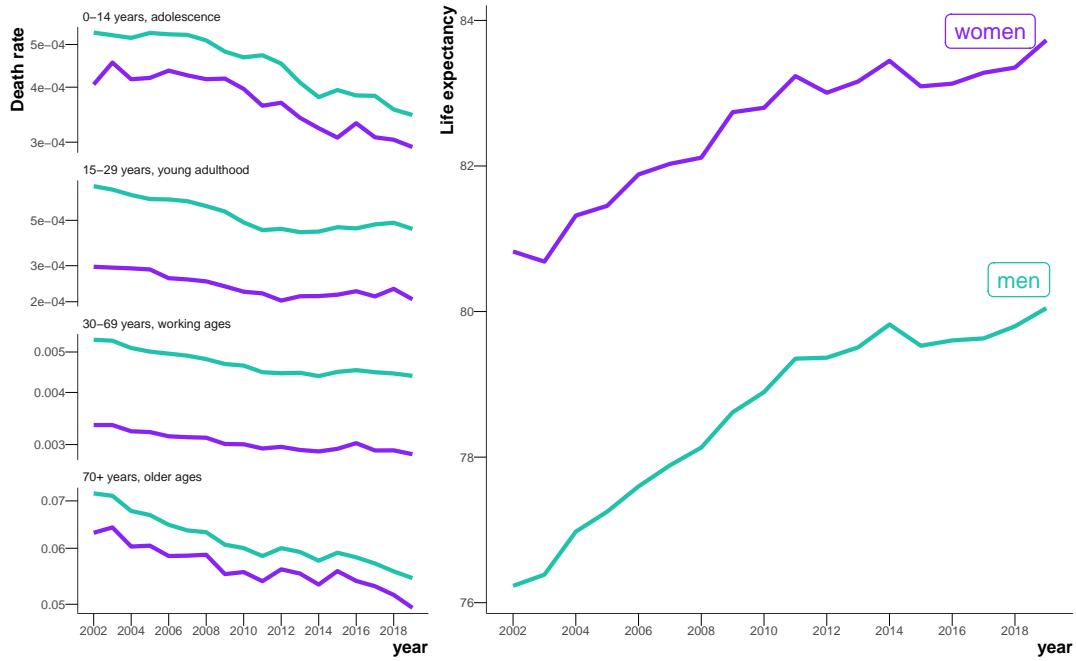


FIGURE 3.2: Age-specific death rates for broad age groups and life expectancy in England from 2002 to 2019.

(30-69 years). Likewise, life expectancy has improved throughout the study period, but has stalled since around 2010 for both sexes.

Figure 3.3 shows the geography of life expectancy after aggregating deaths over the entire study period by district. For both sexes, the picture is similar: pockets of low life expectancy in the urban northwest, northeast, west midlands, and coastal areas east of London.

Here, I have taken slices across each dimension, but the aim in the following chapters is to calculate death rates for each sex-age-space-time stratum.

### 3.2.1 Geographies of England

Having already introduced the term “district” in Figure 3.3, I’ll set out the lay of the land in terms of geographies used in this thesis. This thesis is concerned only with England, as Scotland, Wales, and Northern Ireland each have their own separate deprivation data which are not comparable. The geographies in England form a nested hierarchy of spatial units from regions to districts to Middle-layer Super Output Areas (MSOAs) to Lower-layer Super Output Areas (LSOAs). The number of units for each geography are summarised in Table 3.1.

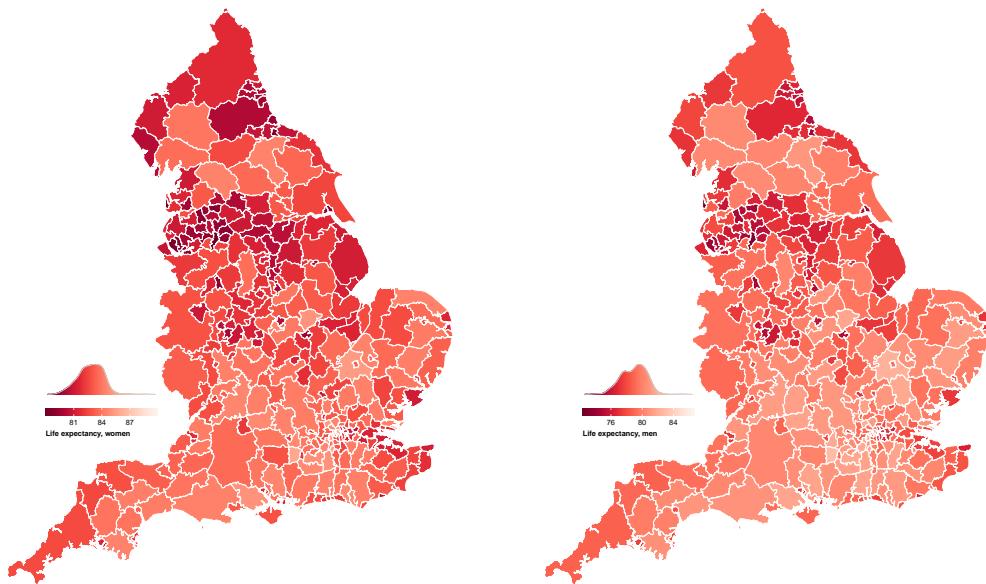


FIGURE 3.3: District-level life expectancy for total deaths from 2002 to 2019.

TABLE 3.1: The number of each geographical unit of England used in thesis and their populations in 2019.

Geography	Number of units	Median (5th - 95th percentile) population in 2019
region	9	5,934,037 (3,536,336-9,092,877)
district	314	140,271 (68,238-380,483)
MSOA	6791	7985 (5760-11,917)
LSOA	32844	1620 (1235-2468)

England is divided into 9 regions (London, North West, West Midlands, etc). Within these regions, there are 314 local authority districts. Districts are administrative geographies formed from a mixture of London boroughs, metropolitan and non-metropolitan districts, and unitary authorities. They are responsible for local policies, and are therefore subject to local government restructuring and boundary changes. For stability, I chose the district boundaries from 2020 throughout.

LSOAs are a type of census geography made up of around four or five smaller units called Output Areas (OAs). OAs are the smallest building block for spatial census statistics, with between 40 and 250 households and typically 100 to 625 people, and

are designed to have some socioeconomic homogeneity. MSOAs are then comprised of around four or five LSOAs, and these MSOAs fit within district boundaries. OAs, LSOAs, and MSOAs are all statistical units are designed by the ONS purely for analysis purposes, so researchers can use spatial units with similar, but small, population sizes. No policies are created using these boundaries (“2011 Census geographies,” 2022).

### 3.2.2 Counts of the living

This second data sources we require are populations counts. These are taken from mid-year population estimates of the usual resident population by the ONS (“Lower layer Super Output Area population estimates,” 2021; “Middle layer Super Output Area population estimates,” 2021). The ONS estimates inter-censal populations on a rolling basis, updating the previous year’s value using the change in the population in the GP patient registration data as an indicator of the true population change. The LSOA populations are fully consistent with estimates for higher levels in the nested geographical hierarchical including MSOAs, districts, regions and the national total for England (“Population estimates by output areas, electoral, health and other geographies, England and Wales,” 2021).

### 3.2.3 Deprivation data

I used data for the following measures of socioeconomic deprivation from the English Indices of Deprivation:

- Income deprivation (also referred to as *poverty*). The proportion of the geographical population claiming income-related benefits due to being out of work or having low earnings.
- Employment deprivation (also referred to as *unemployment*). The proportion of the relevant population of the geography involuntarily excluded from the labour market due to unemployment, sickness or disability, or caring responsibilities.
- Education, skills and training deprivation (also referred to as *low education*). Lack of attainment and skills, including education attainment levels, school attendance, and language proficiency indicators in the geographical population.

The above measures are the three largest contributors to the Index of Multiple Deprivation (IMD), excluding a domain on health that also uses mortality data. The data are produced at the LSOA level (“English indices of deprivation 2019,” 2019).

IMD data are not available for every year. The analysis period for the thesis is 2002 to 2019, so I used data for these measures for 2004, as data for 2002 were not available, and 2019. The 2004 data on deprivation domains were reported for LSOA boundaries from the 2001 census. I mapped these data to the 2011 census LSOA boundaries by assigning the 2001 LSOA score to all postcodes contained within it, then overlaying the 2011 LSOA boundaries, and averaging the score for all constituent postcodes of each LSOA, to obtain the corresponding score for each 2011 LSOA.

The definition of the indicators can change over time. Further, the indicator used for measuring education, skills and training deprivation (low education) is not directly interpretable because it combines multiple concepts and cannot be simply expressed as a proportion of the population. Therefore, I used ranking rather than scores so that comparisons can be made not only across spatial units in a single year, but also across the different years.

The deprivation data for geographies larger than LSOAs in Table 3.1 were created by ranking the population-weighted average of scores for all constituent LSOAs, as done previously for districts (“English indices of deprivation 2019,” 2019).

### 3.2.4 Migration data

I also used estimates of population turnover, defined as the proportion of households in each LSOA in 2019 who were different from those who had lived there in 2002, from the Consumer Data Research Centre. The Consumer Data Research Centre estimates these proportions by using the names of households members, individually and in combination, and addresses and dates of records from electoral and consumer registers and land registry sales data (van Dijk et al., 2021). Estimates of population turnover for MSOAs were created by taking the mean across all constituent LSOAs (“English indices of deprivation 2019,” 2019).

### 3.3 Modelling the dead

For each chapter, the quantity of interest is the same: mortality in each age group, spatial unit and year. Empirically, death rates can be calculated from observed data as the number of deaths divided by the population in each strata. Formally, using  $a$ ,  $s$ , and  $t$  to index age, spatial unit and time respectively, we write

$$\hat{m}_{ast} = \frac{\text{deaths}_{ast}}{\text{population}_{ast}}, \quad (3.1)$$

where  $\hat{m}_{ast}$  is the death rate. When the number of deaths becomes small, however, the empirical death rate presents an apparent variability from year to year, or from spatial unit to spatial unit, which is larger than the true differences in the risk of death. The problem is exacerbated for the young ages or rare diseases, where the number of deaths might be zero, or for smaller geographical units, where the population might be zero. Thus, I have used Bayesian hierarchical models to obtain stable estimates of death rates by sharing information across age groups, spatial units, and years. An added advantage of the Bayesian paradigm is the robust estimation of error.

This is a regression task. We simply want to smooth over the data – the models aren’t being used for prediction. I tried to design a model that captures as much of the true variation in the data as possible using epidemiological knowledge to choose plausible effects. In other words, the model is “full”, with enough parameters to capture all the true variability. The downside of this approach is that models with more parameters are harder to fit, whereas models with fewer parameters, or parsimonious models, make Bayesian inference easier but can mask some of the variance.

### 3.4 Inference

The decision was made early in the PhD to use Markov chain Monte Carlo (MCMC) sampling methods for inference, as this is the “gold standard” with guarantees that the sequence of samples will asymptotically converge to the true posterior. Furthermore, the state-of-the-art approximate inference package for spatial models, INLA, scales badly with the number of effects, and hence would struggle with the high dimensionality of the models in this thesis.

Bayesian models are specified in a probabilistic programming language. The starting point for this project was the **NIMBLE** package (de Valpine et al., 2022, 2017). **NIMBLE** uses the BUGS (“Bayesian inference Using Gibbs Sampling”) syntax for defining a hierarchical model, which my research group has a lot of experience with, as **WinBUGS**, one of the earliest software packages for Bayesian analysis, was developed largely in the department for use on SAHSU studies. **NIMBLE** has an **R** interface but compiles models to **C++** for speed and scalability. It also increases the efficiency of Gibbs sampling by automatically finding conjugate relationships between parameters in the model and marginalising over them wherever possible. The group also has a close relationship with the lead developer of **NIMBLE**.

Nevertheless, Bayesian inference is difficult to scale, and some of the models in this thesis had in excess of  $10^6$  parameters and took **NIMBLE** between 10 and 14 days to collect enough posterior samples. One of the main issues with **NIMBLE** was that the vast majority of the parameters in the model could not exploit efficient conjugate samplers, and instead used variants of basic Metropolis-Hastings samplers, which, despite numerous efforts at tuning, were inefficient. Although **NIMBLE** could execute a reasonable number of samples per second, the MCMC chains were struggling to explore the posterior efficiently so the *effective* sample size per second was low. This is a common problem in spatial and spatiotemporal models, where the parameters are correlated by design. To overcome these mixing issues, the chains had to be run for longer and thinned (i.e. take every  $n^{\text{th}}$  sample so the Markov chain samples are closer to independent).

I spent a lot of time trying different probabilistic programming languages across **R**, **python** and **Julia**, in particular packages that implemented the more efficient No U-Turn Sampler (NUTS) (Rashid, 2022). In the end, I settled on **NumPyro** (Phan et al., 2019) because it was the fastest and inference could be performed on a GPU, rather than CPUs, which is faster for large models (Lao et al., 2020). The major downside was that **NumPyro** had not been used extensively by the spatial modelling community, so I had to implement the CAR distribution from Equation 2.2 myself, which has since been contributed to the source code. Rewriting the model in **NumPyro** and sampling on a GPU cut the runtime down to around a day. **NumPyro** also has built-in methods

for approximate variational inference, such as the Laplace approximation, but these failed to converge to sensible values for these models without heavy customisation of variational function, so I stuck with sampling methods.

### 3.5 Clean code and open source

I am strong believer in open source science, and I have put a lot of attention into open sourcing code for all analyses during the PhD. With open science, not only do we facilitate the scientific method as our process and results are transparent, but we also allow future researchers to reuse and build on our models easily. It can also generate interest from researchers in different fields using similar models and from developers looking to challenge their software on complicated research questions, both of which I have seen first-hand during the course of my studies. The code is clean, version-controlled and follows best practices for scientific software engineering. As well as code contributed to open source projects along the way, the code for [statistical models](#), [plots and analysis](#), and the [thesis itself](#) can be found on GitHub.



## Chapter 4

# Small: Life expectancy trends in England at the MSOA level

This chapter is based on the peer-reviewed publication *Life expectancy and risk of death in 6791 communities in England from 2002 to 2019: high-resolution spatiotemporal analysis of civil registration data*, published in *The Lancet Public Health* (Rashid et al., 2021), for which I was first author. The figures have been reproduced for the thesis, and the text has been updated but remains much the same as the original.

### 4.1 Overview

### 4.2 Methods

#### 4.2.1 Mortality and population data

I performed a high-resolution spatiotemporal analysis of civil registration data in which I extracted de-identified data for all deaths in England from 2002 to 2019 (8,646,878 death records, extract date: 25th June 2021). Deaths were stratified by 19 age groups (0, 1–4, 5–9, ..., 80–84, 85+) and 6791 MSOAs. I did not use 129 death records (<0.001%) for which sex was not recorded. In 48 (0.001%) age-MSOA-year combinations, the number of deaths exceeded population. Most of these were in people aged 85 years and older. In these cases, the population was set equal to the number of deaths.

### 4.2.2 Statistical analysis

I used a Bayesian hierarchical model to obtain stable estimates of death rates by sharing information across age groups, MSOAs, and years. I conducted all analyses for women and men separately because mortality and trends differ by sex. In the model, the number of deaths in age group  $a (= 1, \dots, 19)$ , MSOA  $s (= 1, \dots, 6791)$  and year  $t (= 1, \dots, 18)$  follows a negative binomial distribution

$$\text{deaths}_{ast} = \text{Negative Binomial}(p_{ast}, r). \quad (4.1)$$

The parameter  $p_{ast}$  is

$$p_{ast} = \frac{r}{r + m_{ast} \cdot \text{Population}_{ast}}. \quad (4.2)$$

where  $r \geq 0$  is the overdispersion parameter, which accounts for extra variability not captured by other components in the model, and  $m_{ast}$  is the death rate. The negative binomial<sup>1</sup> likelihood can be thought of as a generalisation of the Poisson likelihood, which allows for overdispersion, with larger values of  $r$  indicating more similarity to a Poisson distribution. A Poisson distribution is a suitable approximation for rare events, and I found that mortality at the MSOA-level was sufficiently rare such that death rates never came close to 1.

Log-transformed death rates were modelled as a function of time, age group and MSOA. The model contains terms to capture the overall level and trend over time of mortality, as well as age-specific and MSOA-specific terms to allow deviations from these terms. Specifically, log-transformed death rates are modelled as

$$\log(m_{ast}) = \alpha_0 + \beta_0 t + \alpha_{1s} + \beta_{1s} t + \alpha_{2a} + \beta_{2a} t + \xi_{as} + \nu_{st} + \gamma_{at}, \quad (4.3)$$

where  $\alpha_0$  is the overall intercept across all age groups and MSOAs.  $\beta_0$  quantifies the overall trend (over time) across all age groups and MSOAs.  $\alpha_{1s}$  and  $\beta_{1s}$  measure deviation from the overall intercept and trend terms, respectively, for each MSOA.  $\alpha_{2a}$  and  $\beta_{2a}$  measure deviation from the global level and trend, respectively, for each

---

<sup>1</sup>The name “negative binomial” is quite difficult to understand in this context. A better, but less popular name, is the gamma-Poisson distribution. Here, the story is simpler: we have a mixture of Poisson distributions where the rate parameters of the Poisson distributions follow a gamma distribution (McElreath, 2020).

age group. I used first-order random walk priors on  $\alpha_{2a}$  and  $\beta_{2a}$  so that they vary smoothly over adjacent age groups, with the form  $A_a \sim \mathcal{N}(A_{a-1}, \sigma_A^2)$  for both age-specific terms  $\alpha_{2a}$  and  $\beta_{2a}$ . I constrained  $\alpha_{21} = 0$  and  $\beta_{21} = 0$  so each random walk was identifiable and centred on the corresponding overall term.

$\xi_{as}$  is an age group-MSOA interaction term, which quantifies MSOA-specific deviations from the overall age group structure given by  $\alpha_{2a}$ . This allows different MSOAs to have different age-specific mortality patterns, and each age group's death rate to have a different spatial pattern. This interaction term was modelled as  $\mathcal{N}(0, \sigma_\xi^2)$ .

$\nu_{st}$  and  $\gamma_{at}$  allow MSOA- and age group-specific non-linearity in the time trends. For each MSOA and age group, I again used first-order random walk priors with  $\nu_{s1} = \gamma_{a1} = 0$  so that the terms were identifiable.

For the main analysis, the MSOA intercepts and slopes,  $\alpha_{1s}$  and  $\beta_{1s}$ , were modelled as nested hierarchical random effects, with MSOAs nested in districts, which were, in turn, nested in regions. The regional terms are centred on zero to allow the spatial effects to be identifiable.

For comparison, I also modelled the spatial effects using a BYM model, as in Equation 2.3. The CAR component of the model requires all spatial units to have neighbours. Thus, the MSOAs containing the Isle of Wight, Hayling Island, the Isles of Scilly and Canvey Island were each joined to the nearest mainland MSOA based on road or ferry connections.

The results of the spatial model were virtually identical to the hierarchical random effects model (correlation coefficient 0.999 for female and male sexes; mean difference 0.03 years for women and 0.009 years for men; mean absolute difference 0.07 years for women and 0.09 years for men for life expectancy estimates from the two approaches). I present results from the hierarchical model for two reasons. First, it allows neighbouring MSOAs that fall in different districts to differ more than those within the same district, reflecting the relevance of district as a unit of resource allocation and policy implementation. Second, the hierarchical model was computationally less demanding with run times about 1.4 times faster than the spatial model.

All standard deviation parameters of the random effects had  $\sigma \sim \mathcal{U}(0, 2)$  priors. For the global intercept and slope, we used the diffuse prior  $\mathcal{N}(0, \sigma^2 = 10^5)$ . The overdispersion parameter  $r$  had the prior  $\mathcal{U}(0, 50)$ .

Table B.1 shows all model parameters, their priors and dimensions.

Inference was performed using MCMC in NIMBLE package (de Valpine et al., 2022, 2017). I monitored convergence using trace plots and the R-hat diagnostic (Vehtari et al., 2021) and thinned post burn-in samples to reduce memory and storage use. For women, I ran four chains for 150,000 iterations, discarding the first 50,000 and thinning the remainder by 400 to obtain 1,000 post-burn-in draws from the posterior distribution of model parameters. For men, due to slower mixing, I ran eight chains for 150,000 iterations, discarding the first 100,000 and thinning the remainder by 400.

In 2017, the MSOA in Kensington and Chelsea, London where Grenfell Tower is located had 119 deaths, compared with 48 in 2016 and 51 in 2018. The additional deaths were caused by a fire in a highrise residential building. This outlier year led to unstable estimates of the longterm trend in life expectancy in this MSOA, and also slightly changed estimates in other MSOAs in the district. To avoid this instability, when applying the statistical model, I replaced the number of deaths for this year with the mean of those in 2016 and 2018 for each age and sex group. When making estimates for 2017, the difference between actual and interpolated deaths was added back to the posterior estimates so that these deaths were counted in the corresponding year. The Grenfell Tower fire incident was, to the best of our knowledge, the only spatially-specific, long-tailed mortality event in England over the study period.

I calculated life expectancy at birth, and probability of dying at specific ages by sex and MSOA using life table methods <sup>2</sup>.

The reported 95% credible intervals (CrIs) represent the 2.5th to 97.5th percentiles of the posterior distribution of estimated life expectancies. I also report the posterior probability that the estimated change over time in an MSOA represents an increase versus a decrease in life expectancy. Posterior probability represents the inherent

---

<sup>2</sup>See Appendix A for full details on demographic methods used in this thesis.

uncertainty in life expectancy trends. In an MSOA in which the entire posterior distribution of life expectancy in 2019 is greater than in 2002, there is around a 100% posterior probability of an increase, and hence around a 0% probability of a decrease, and vice versa. Posterior probabilities more distant from 50% indicate more certainty.

## 4.3 Results

### 4.3.1 Inequalities in life expectancy

In 2019, there was a 20.6 year (95% CrI 17.5–24.2) gap for women between the MSOA with the highest life expectancy (an MSOA in Camden, London; 95.4 years (92.4–98.7)) and the MSOA with the lowest life expectancy (an MSOA in Leeds; 74.7 years (73.4–76.2)). The gap was 27.0 years (23.4–31.1) for men, between an MSOA in Kensington and Chelsea, London (95.3 years (92.1–99.3)) and an MSOA in Blackpool (68.3 years (66.9–69.6)). The difference between the first and 99th percentiles of life expectancy in 2019 was 14.2 years (13.9–14.5) for women and 13.6 years (13.4–13.9) for men.

When all MSOAs were ranked on the basis of their life expectancy, the difference between successively ranked MSOAs was particularly large for the approximately 5% of MSOAs with the lowest and highest life expectancy (seen as the sharper decline or rise at the two ends of the ranked life expectancy curve in Figure 4.1), indicating distinct groups at extreme advantage and disadvantage.

The 124 (1.8%) of 6791 MSOAs with the lowest female life expectancy and 262 (3.9%) MSOAs with the lowest male life expectancy in 2019 were located in urban areas, particularly in the north, including Blackpool, Leeds, Liverpool, Manchester, and Newcastle. Many of the MSOAs with the highest life expectancy, especially for men, were in London and its neighbouring districts (Figure 4.2).

---

<sup>3</sup>For an interactive version of this figure, see the [visualisation](#).

<sup>4</sup>For a hex-cartogram of life expectancy in 2019, see Figure C.1.

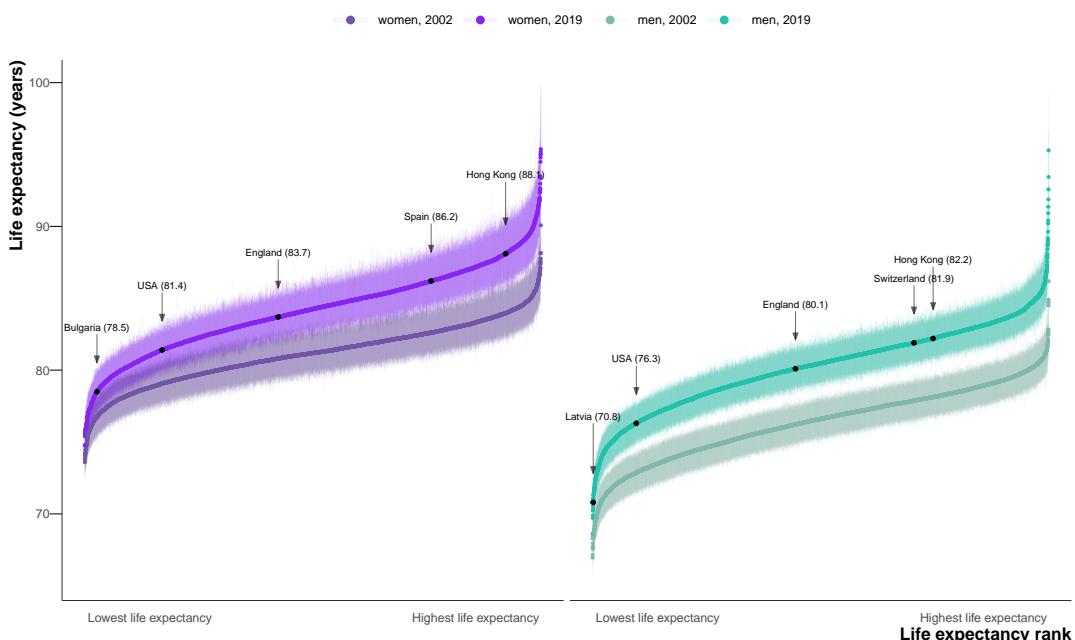


FIGURE 4.1: Ranked MSOA life expectancies in 2002 and 2019. Each point shows the posterior median life expectancy estimate for each MSOA, forming a curve; error bars are 95% credible intervals. Arrows indicate national life expectancies in England and selected comparator countries with life expectancies within the range of English MSOAs. Hong Kong had the highest global female and male life expectancies. In the EU, Bulgaria had the lowest and Spain had the highest life expectancies for women; Latvia had the lowest and Switzerland had the highest life expectancies for men. Life expectancy for England was calculated from the data used here, and for other countries from World Bank estimates in 2019 (“Life expectancy at birth, data,” 2022).

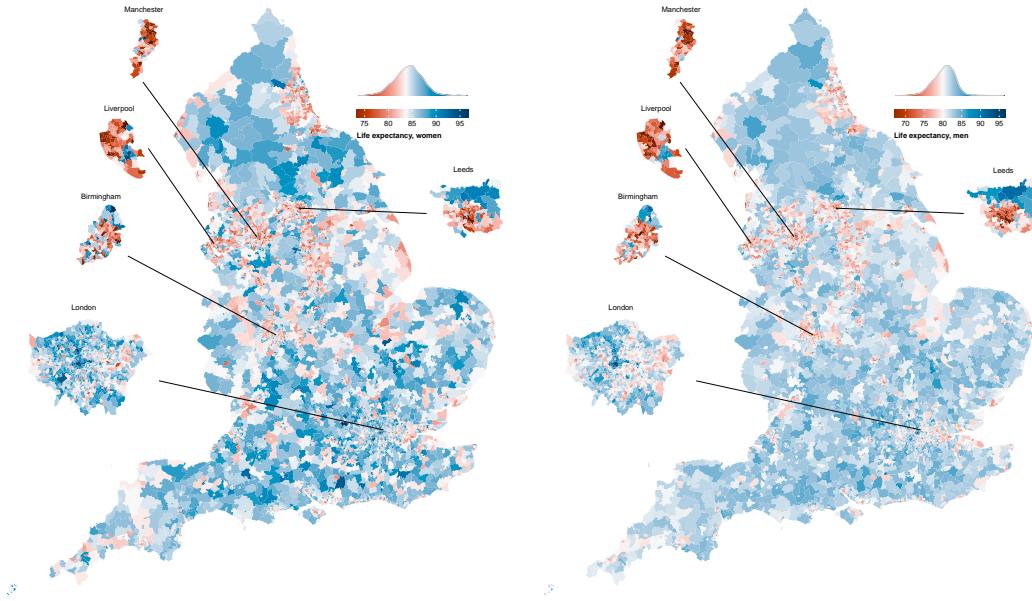


FIGURE 4.2: Map of life expectancy and the distribution of life expectancy in 2019. The areas in white have a life expectancy equal to the national life expectancy. <sup>3</sup> <sup>4</sup>

### 4.3.2 Change in life expectancy

Female and male life expectancy were correlated across MSOAs with a correlation coefficient of 0.87 (Figure 4.3). Female life expectancy was higher than male life expectancy in all but 15 MSOAs. The female advantage was more than 5 years in 1498 (22.1%) of 6791 MSOAs and 1–5 years in another 5187 (76.4%). From 2002 to 2019, a decline in life expectancy was more probable than an increase in 124 mostly urban MSOAs of 6791 (1.8% of all MSOAs) for women, with posterior probabilities of greater than 80% in 34 of these. The largest estimated decline of 3.0 years (0.9–5.3; posterior probability of the estimated decline being a true decline 99.6%) occurred in an MSOA in Leeds (Figure 4.4, Figure 4.5).

Elsewhere, median posterior change was positive, ranging from less than 1 year in 408 MSOAs to more than 7 years in 63 MSOAs. Posterior probability of an increase in male life expectancy was more probable than a decrease in all but one MSOA in Blackpool, in which life expectancy changed by –0.4 years (–2.3 to 1.6; posterior probability of being a true decline 64%). For the other MSOAs, the increase ranged from less than 1 year in 31 MSOAs to more than 7 years in 114 MSOAs. The largest

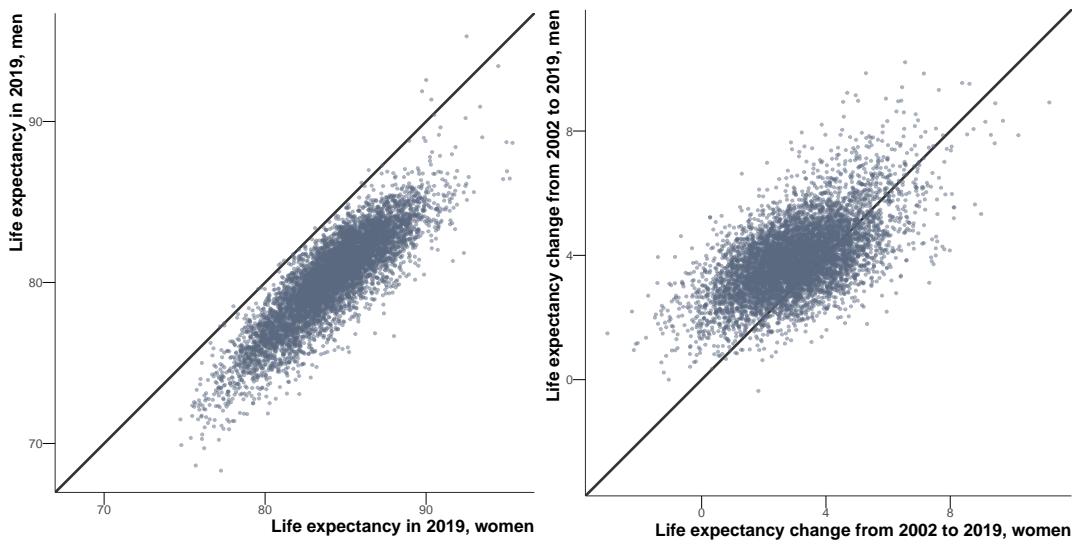


FIGURE 4.3: Comparison of female and male life expectancy in 2019 and change from 2002 to 2019.

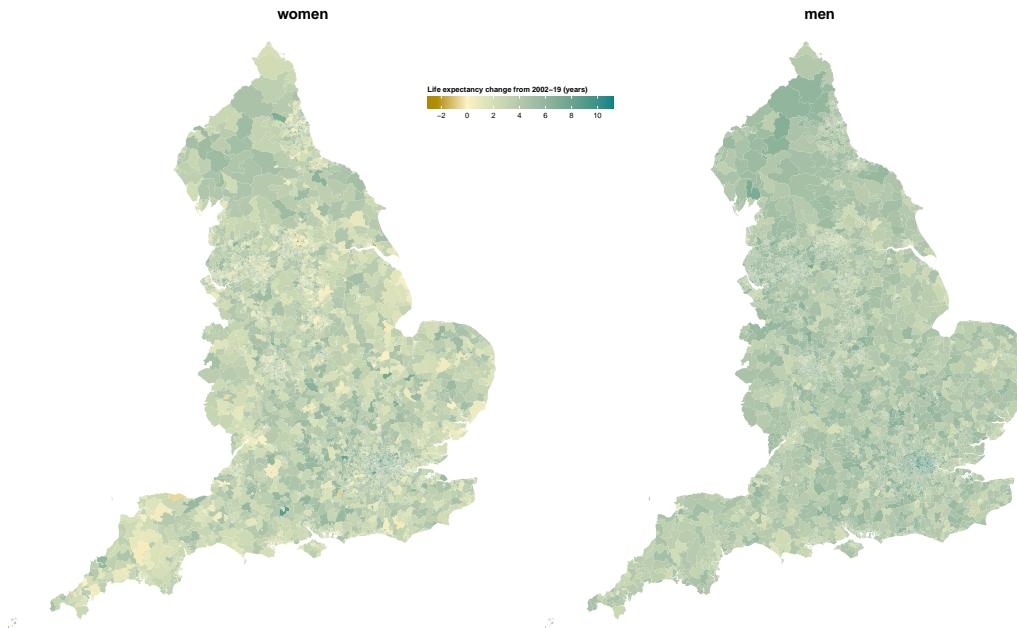


FIGURE 4.4: Geography of change in life expectancy from 2002 to 2019.

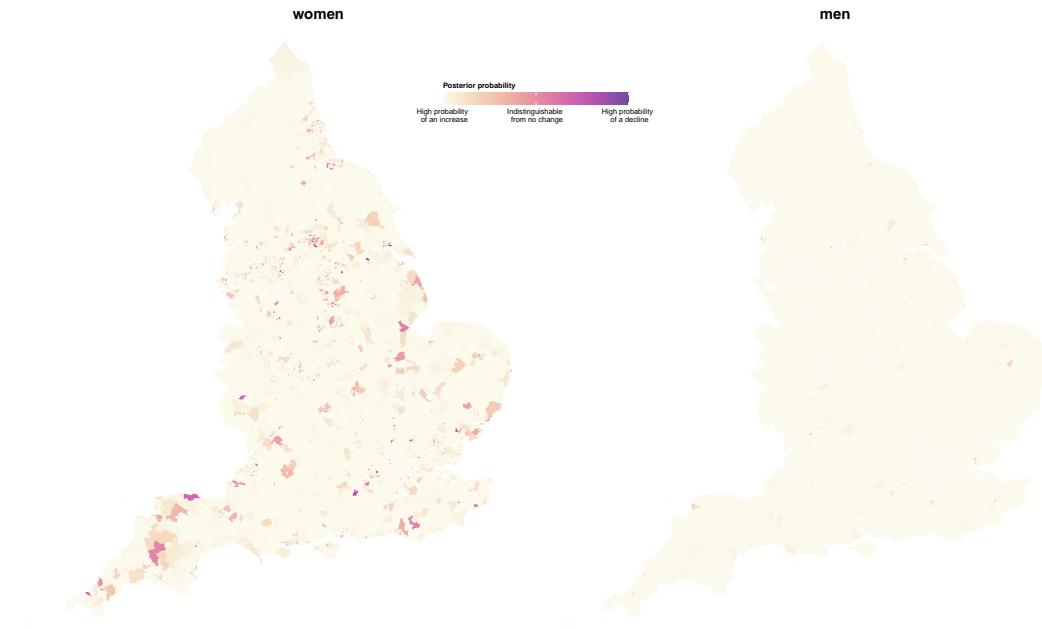


FIGURE 4.5: Map of posterior probability that the estimated change represents a true increase or decrease in life expectancy from 2002 to 2019.

increases in female and male life expectancies were seen in some MSOAs in and around London (e.g., in the London Borough of Camden). In 5,133 (75.6%) MSOAs, male life expectancy increased more than female life expectancy (Figure 4.3), leading to a closing of the life expectancy gap between female and male sexes.

The life expectancy increase from 2002 to 2019 was smaller in MSOAs where life expectancy had been lower in 2002, and vice versa, especially for women, which led to a larger life expectancy inequality across MSOAs in 2019 than in 2002 (Figure 4.6, Figure 4.7). Specifically, the aforementioned 20.6 year (17.5–24.2) gap for women and 27.0 year (23.4–31.1) gap for men between the lowest and highest MSOA life expectancies in 2019 were larger than those in 2002 by 4.3 years (−1.3 to 9.3) for women and 7.7 years (4.0 to 11.7) for men. Similarly, the gap between the first and 99th percentiles of MSOA life expectancy for women increased from 10.7 years (10.4–10.9) in 2002 to reach 14.2 years (13.9–14.5) in 2019, and for men increased from 11.5 years (11.3–11.7) in 2002 to 13.6 years (13.4–13.9) in 2019.

When broken down by time period, the vast majority of MSOAs saw a life expectancy increase in 2002–06 and 2006–10 (Figure 4.8). By contrast, women in 351 (5.2%)

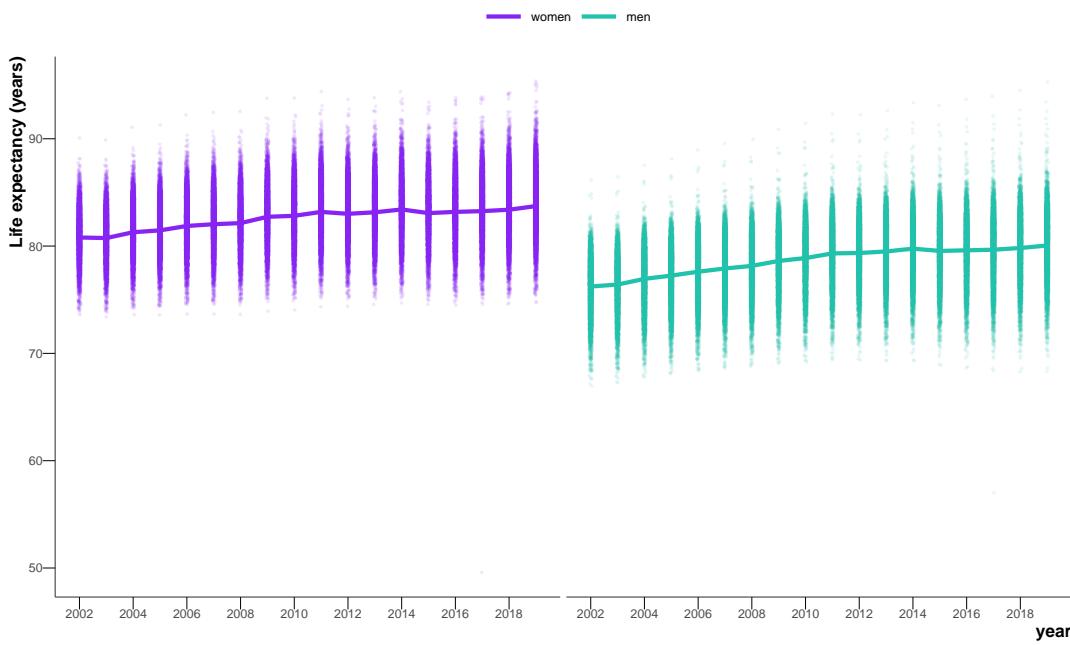


FIGURE 4.6: Distribution of MSOA life expectancies in each year from 2002 to 2019. Each point shows one MSOA. The central line shows national life expectancy.

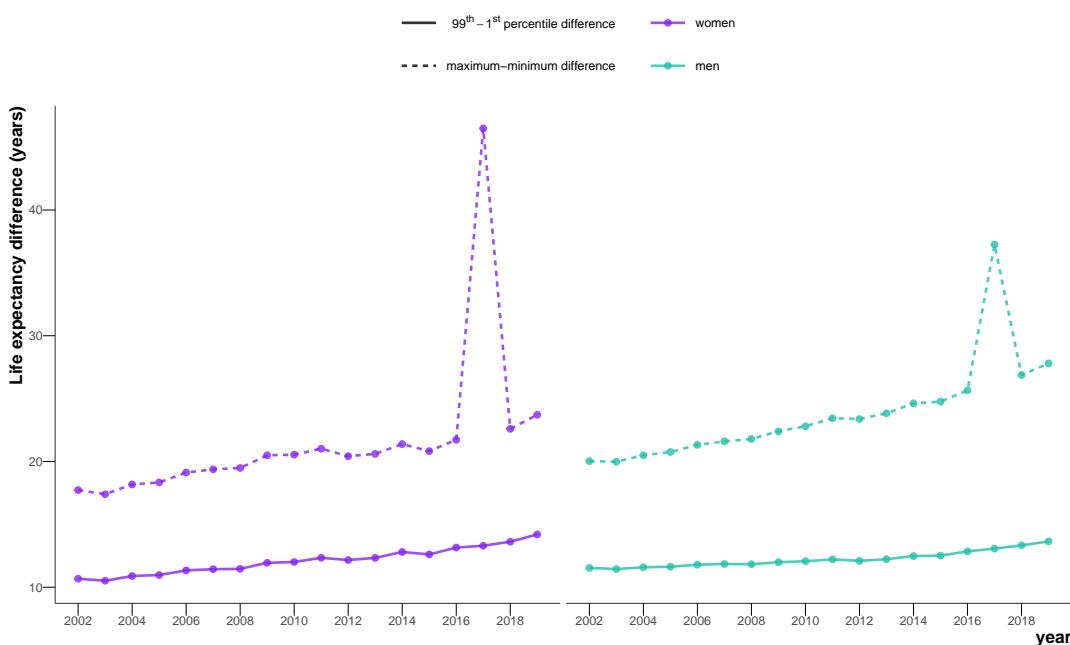


FIGURE 4.7: Maximum (highest) to minimum (lowest) and 99th to first percentile differences in life expectancy across 6791 MSOAs, 2002–19. The large difference in 2017 is due to the low life expectancy in the MSOA where the deaths in the Grenfell Tower (Kensington and Chelsea, London) fire took place.

MSOAs had a median posterior change in life expectancy in 2010–14 that was negative. By 2014–19, the number of MSOAs with a negative median posterior change had risen to 1270 (18.7%) for women, with men in 784 (11.5%) MSOAs also showing a decline. These MSOAs tended to be places in which life expectancy was already low.

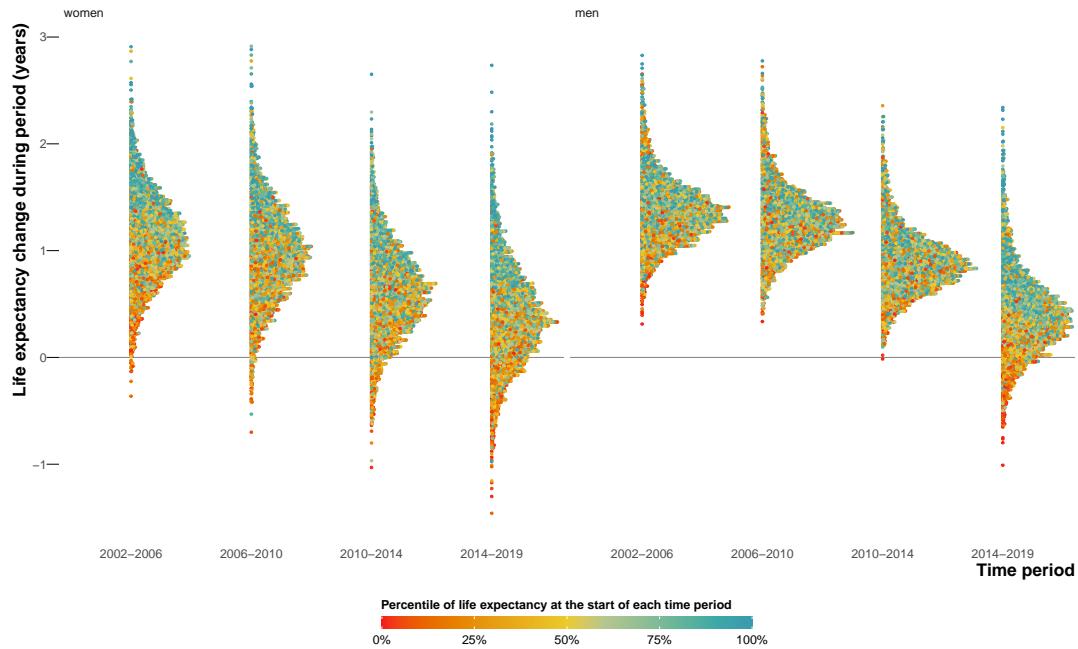


FIGURE 4.8: Change in MSOA life expectancy in different time periods, 2002–19. Each point shows the posterior median change in one MSOA. MSOAs are coloured by their life expectancy at the beginning of each period (e.g., for 2014–19, they are coloured by life expectancy in 2014).

### 4.3.3 Life expectancy and deprivation

Life expectancy at birth was inversely associated with the extent of unemployment, poverty, and low education in MSOA in 2002 and 2019 (Figure 4.9). There was substantial variation in life expectancy across MSOAs at any level of poverty or unemployment seen in the vertical spread of points in figure 6. From 2002 to 2019, there were, on average, smaller gains in life expectancy in the MSOAs with the highest levels of unemployment, poverty, and low education than in those in the lowest levels, especially for women.

### 4.3.4 Inequalities in probability of survival

Similar to life expectancy, there were large inequalities in the probability of surviving from birth to 80 years, which ranged from 42% to 87% in women and 27% to 85%

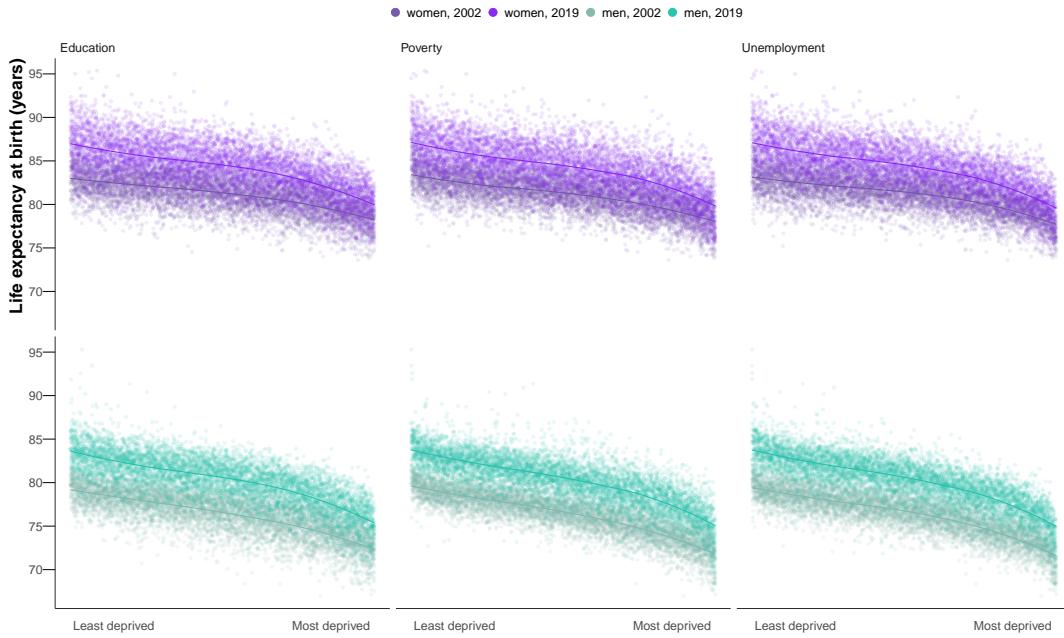


FIGURE 4.9: MSOA life expectancy in relation to measures of socioeconomic deprivation in the MSOA in 2002 and 2019. The socioeconomic measures are poverty, unemployment, and education. The lines show the smooth relationship fitted with locally estimated scatterplot smoothing for each year.

in men across MSOAs in 2019. These large survival inequalities were present at every stage of the lifecourse including childhood and early adolescence (0–14 years), young adulthood (15–29 years), working ages (30–69 years), and older ages (70–79 years) (Figure 4.10). Specifically, the probability of dying at different stages of the lifecourse in the 99th percentile of MSOAs was between 2.6 and 3.1 times that of the first percentile for female and male sexes in 2019. From 2002 to 2019, the relative inequality across MSOAs (ie, ratio of the 99th to the first percentile) in the probabilities of dying increased at every stage of the life course; the absolute inequality (ie, difference between the 99th and first percentiles) decreased slightly in all combinations except for working age women (30–69 years). Within childhood and adolescence, there were particularly large inequalities in infant mortality (0 to <12 months), with a ratio of the 99th to the first percentile of MSOAs being 3.2 for female and male sexes in 2019. Infant mortality increased from 2014 to 2019 in 1378 (20.3%) MSOAs for girls and 888 (13.0%) for boys, many of which experienced a decline in life expectancy.

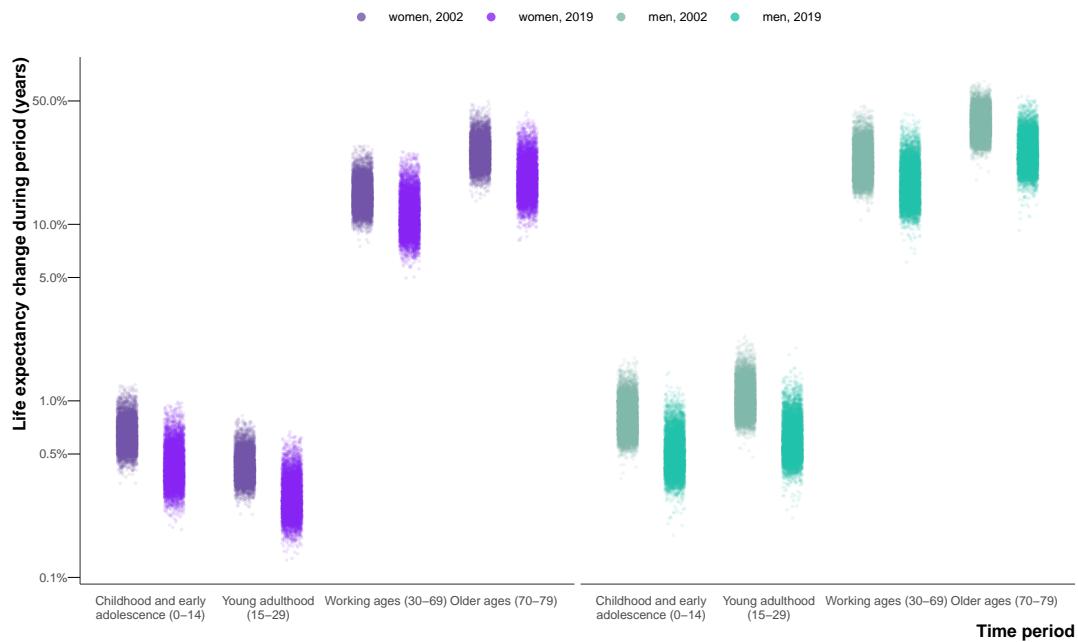


FIGURE 4.10: Probability of dying in specific ages in 6791 MSOAs in England in 2002 and 2019. Each point shows one MSOA. The vertical axis uses a log scale so that the large differences in survival across ages can be seen.

## 4.4 Discussion

The high-resolution analysis over space and time shows that life expectancy has not only ceased to increase, but has declined in many communities in England since 2010. The decline has accelerated since 2014, affecting the female population of 18.7% of MSOAs and the male population of 11.5% of MSOAs. In 1.8% of MSOAs, women have had a long-term decline in life expectancy over two decades. MSOAs that have gained the least in longevity since 2002 were those that started with the lowest life expectancy, located in northern urban areas with high levels of poverty and unemployment, and with relatively low education. Conversely, those MSOAs with higher life expectancies in 2002 had some of the largest gains. As a result, England has seen widening inequalities in longevity, with the life expectancy gap surpassing 20 years for women and 27 years for men.

### 4.4.1 Strengths and limitations

The main strength of the analysis is the presentation of high-resolution data for mortality and longevity across England over a period of substantial change in economic,

health, and social care policy. By applying a hierarchical model based on patterns of mortality over age, space, and time, I obtained robust yearly estimates of mortality and life expectancy, together with the uncertainty in these estimates, for small areas. By contrast, studies that had not used a coherent model produced unstable (i.e., very large uncertainty) or implausible life expectancy estimates in some MSOAs, despite having aggregated deaths over 5 years, nor could they analyse trends at the MSOA level (“Health Expectancies at Birth for Middle Layer Super Output Areas (MSOAs), England,” 2015; “Local Health - Small Area Public Health Data,” 2021). Comparison of estimates at MSOA and district level shows that the estimated MSOA life expectancy range was about 1.8 times the district-level range for women and 2.0 times the district-level range for men in 2019.

A limitation of the work in this chapter is that I did not break down age beyond 85 years, which might mask some differences in old-age mortality and survival patterns. Although MSOAs have small populations and are designed to have some socioeconomic homogeneity, there are inevitable variations in socioeconomic status and health within them. To understand life expectancy inequalities in relation to individual socioeconomic characteristics requires linking health and other data such as census records, education, and taxes, as done in countries like New Zealand and Sweden.

The people who live in each MSOA can change due to both within-country and international migration. Regression of the change in life expectancy from 2002 to 2019 in each MSOA against population turnover, the proportion of households in each MSOA in 2019 who were different from those who had lived there in 2002 (van Dijk et al., 2021), was not able to explain the variation in life expectancy change for women ( $R^2 < 0.001$ ) or for men ( $R^2 = 0.01$ ) at the national level. Studies in both the UK (Connolly et al., 2007) and USA (Ezzati et al., 2008) have also shown that migration is not sufficient to explain the trends in health and health inequalities, and that these trends are largely due to real changes in population health. Even if rising inequalities are partly due to health-selective migration, this phenomenon has social and economic origins that should be addressed through employment opportunities, affordable housing, high-quality education, and health care.

Population and mortality statistics in the UK are generated independently from one another. As a result, we encountered a situation of having more deaths than population in a small percentage (0.001%) of age-MSOA year combinations, a phenomenon that was more common in those aged 85 years and older. This finding might be due to errors in population estimates in years between censuses or because some people (e.g., those living in longterm care facilities such as care homes), are counted in one MSOA for the population statistic but have their death registered in another. Furthermore, care home residents might have relocated from other MSOAs, with different socioeconomic characteristics from that in which the care home is located. This factor could attenuate the association between socioeconomic variables and life expectancy. The extent of this underestimation is modest; however, because a large part of life expectancy variation is due to deaths at earlier ages, when people are less likely to live and die in care homes (Bennett et al., 2018).

#### 4.4.2 Comparison with previous literature

The life expectancy estimates in specific years are similar to the snapshots presented by the ONS and Public Health England (“Health Expectancies at Birth for Middle Layer Super Output Areas (MSOAs), England,” 2015; “Local Health - Small Area Public Health Data,” 2021), with correlation coefficients of 0.92-0.95 and mean differences of -0.004 to 0.19 years. However, these reports could not analyse trends because data were aggregated over 5 years (2009–13, 2013–17, or 2015–19). In terms of trends, studies that grouped small area units into deciles of deprivation have detected a decline in female life expectancy in the one or two most deprived deciles (Bennett et al., 2018; Marmot et al., 2020). By analysing trends at the MSOA level, we could identify the communities in which longevity is declining and show that the decline, which began around 2010 in women in some MSOAs, has spread and accelerated since 2014.

Congdon (2019) also used spatial models to smooth over MSOA-level data of mortality, but specifically drug-related deaths and suicides between 2012-16. The author singled out the district of Blackpool as containing many of the MSOAs with the most extreme relative risks of death from these causes, including the MSOA I found had the lowest

life expectancy for men in 2019.

#### 4.4.3 Is England the new USA (in a bad way)?

Over the period of this analysis, from 2002 to 2019, national life expectancy increased in high-income countries in Australasia, Europe, and North America. Female life expectancy has stagnated or declined in various intervals since 2010 in the UK (84% of the UK population in 2019 lived in England) and in some other high-income countries including France, Germany, Italy, and the USA; the UK and USA have had some of the poorest performances in terms of the duration or extent of slowdown or reversal in longevity gain. The comparative performance of high-income countries' longevity trends has been attributed to differences in risk factors such as smoking, health care, and social inequalities (Leon et al., 2019).

To our knowledge, nationwide trend data for small-area life expectancy are available only in the USA. The declining life expectancy in numerous English MSOAs since 2014, especially those that already had a low life expectancy, resembles a trend spanning nearly three decades in the USA in two ways (Chetty et al., 2016; Dwyer-Lindgren et al., 2017a; Ezzati et al., 2008). First, in both countries, there is substantial variation in life expectancy at any level of poverty, which might be due to geographical variations in health behaviours, the public health programmes that influence these behaviours or otherwise prevent disease, and health services (Burn-Murdoch, 2023; Chetty et al., 2016). The second similarity in small area life expectancy trends is that the decline in life expectancy was more widespread in women than in men (Ezzati et al., 2008). Historically, women and men had similar life expectancies in high-income nations before a rise in traffic injuries and diseases associated with specific occupations and health behaviours such as smoking and alcohol use created a male mortality disadvantage in the 20th century (Beltrán-Sánchez et al., 2015). The closing of female and male life expectancy in the late 20th century and early 20st century in many highincome nations (Kontis et al., 2017) is partly due to the dynamics of smoking, which peaked later in women than in men, and affects causes of death such as respiratory diseases and lung cancer that have stagnated or even increased in women in deprived communities (Bennett et al., 2018; Leon et al., 2019). However, it

is rare for the convergence of female and male life expectancies to occur in the form of female life expectancy decline (Ezzati et al., 2008), which might be due to a combination of the worsening economic, psychosocial (eg, poverty, stress, and domestic violence), and behavioural (smoking and alcohol use) determinants of mortality in English women.

In both countries, the decline in life expectancy was associated with the economic trends of unemployment and insecure and low-wage employment following late 20th century deindustrialisation. In England, these economic trends led to a larger loss of jobs in the north than in London and the southeast (Blundell et al., 2020; Davenport and Zarenko, 2020). These long-term changes were followed by a reduction in social support and welfare payments and in funding to the local governments during the austerity period, which increased poverty, including in-work poverty (Alexiou et al., 2021; Blundell et al., 2020), and also had larger effects in the north than in London and southern parts of the country and worsened the effects of loss of secure employment (Alexiou et al., 2021; Davenport and Zarenko, 2020; Gray and Barford, 2018). Poverty and reduced funding to services increase mortality through health behaviours such as smoking and alcohol use, poor nutrition and living environment, psychosocial pathways, and lower provision or use of preventive and curative health care.

#### 4.4.4 Summary

I performed a high-resolution spatiotemporal analysis of all deaths in England from 2002 to 2019, using a Bayesian hierarchical model to obtain estimates of age-specific death rates by age, sex, and MSOA. I used life table methods to calculate life expectancy at birth and probabilities of death in different ages by sex and MSOA.

In 2002–06 and 2006–10, all but a few (0–1%) MSOAs had a life expectancy increase for women and men. In 2010–14, female life expectancy decreased in 351 (5.2%) of 6791 MSOAs. And by 2014–19, the number of MSOAs with declining life expectancy was 1270 (18.7%) for women and 784 (11.5%) for men. The life expectancy increase from 2002 to 2019 was smaller in MSOAs where life expectancy had been lower in 2002 (mostly northern urban MSOAs), and larger in MSOAs where life expectancy had been higher in 2002 (mostly MSOAs in and around London).

These results show that numerous communities in England had begun to have a decline in longevity, mirroring an earlier trend in the USA. That these trends happened in the decade before the Covid-19 pandemic – so-called “normal times” – is worrying, and signals ongoing policy failures to tackle the social determinants of health.

## Chapter 5

# Smaller: Life expectancy inequality in London at the LSOA level

This chapter is based on the peer-reviewed publication *Changes in life expectancy and house prices in London from 2002 to 2019: hyper-resolution spatiotemporal analysis of death registration and real estate data*, published in *The Lancet Regional Health Europe* (Bennett et al., 2023), for which I was joint first author. The paper combined two main datasets: estimates of life expectancy in London at LSOA level and estimates of house prices at the OA level. Unlike the original text, I will largely focus on the life expectancy estimates, for which I was responsible, rather than the house price estimates produced by James Bennett.

### 5.1 Overview

### 5.2 Methods

#### 5.2.1 Mortality and population data

I performed a high-resolution spatiotemporal analysis of civil registration data in which I extracted de-identified data for all deaths in London from 2002 to 2019 (909,097 death records, extract date: 25th June 2021). Deaths were stratified by 19 age groups (0, 1–4, 5–9, ..., 80–84, 85+) and 4835 LSOAs. These LSOAs fit inside

33 districts: London's 32 boroughs and the City of London. I did not use 16 death records (0.002%) for which sex was not recorded. In 3267 (0.099%) age-LSOA-year combinations, the number of deaths exceeded the population. In 88% of these combinations, the deaths exceeded population by only one or two, and 96% of these combinations were in people aged 80 years and older. In these cases, the population was set equal to the number of deaths.

### **5.2.2 House price estimates**

We scraped data from real estate website Rightmove for all London postcodes, obtaining 2.1 million sales records for years 2002–2019. We used postcode to map homes to 24,131 (96.4% of 25,031) OAs in London which belonged to 4830 LSOAs (>99% of 4835). The 900 (3.6%) OAs with no sales over this period tended to be places with higher proportions of rented housing, especially rented social housing.

We then used a mixed-effects model to estimate the house price at OA level. The model was formulated to take into account how prices vary in relation to location, time and characteristics of the property. Specifically, the model included effects for the following: OA and district of the house, quarter of year in which the transaction took place, type of house (flat, terraced, semi-detached, detached house), status of land ownership (freehold (owns the property and the land upon which the property is built), leasehold (owns the property but not the land, which is leased)), whether the property is an existing property or newly constructed, and number of bedrooms, as well as interactions between the effects. Further details can be found in the paper (Bennett et al., 2023). We used the median of all constituent OAs for each LSOA's price. Here, I present the price estimates for a two-bedroom, leasehold, existing (not newly constructed) flat sold in the spring season. Two-bedroom leasehold flats were the most common type of sale during the study period with 19.7% of sales.

### 5.2.3 Statistical analysis

The model was largely the same as Chapter 4, with a few changes: First, the negative binomial likelihood from Equation 4.1 was replaced with a beta-binomial likelihood,

$$\text{deaths}_{ast} = \text{Beta-Binomial}(m_{ast}\rho, (1 - m_{ast})\rho, \text{Population}_{ast}). \quad (5.1)$$

where  $m_{ast}$  is the death rate and  $\rho \geq 0$  is the overdispersion parameter. I found that the variability of the LSOA-level mortality data was such that death rates did near 1, violating the assumption for Equation 4.1 that deaths are rare events. In fact, when I tested a negative binomial or Poisson likelihood, I found the death rates for some age-LSOA-year combinations exceeded 1, which of course is impossible in the real world. The beta-binomial likelihood is a generalisation of the binomial distribution that allows for overdispersion. The death rates were logit-transformed and the model structure is otherwise the same as Equation 4.3, but using nested hierarchical random effects with LSOAs nested in MSOAs, which were nested in districts. A  $\rho \sim \exp(0.1)$  hyperprior was used for the overdispersion term, and  $\sigma \sim \mathcal{N}^+(1)$  hyperpriors for the standard deviation parameters, as recommended by the Stan development team (“Prior Choice Recommendations,” 2020).

Table B.2 shows all model parameters, their priors and dimensions.

Again, I used NIMBLE for MCMC (de Valpine et al., 2022, 2017), with four chains of 80,000 iterations, discarding the first 30,000, thinning the remainder by 200 and checking for convergence using trace plots and R-hat (Vehtari et al., 2021).

The local effect of the Grenfell Tower fire was even more pronounced at the LSOA level. In 2017, the LSOA in Kensington and Chelsea where Grenfell Tower is located had 75 deaths, compared to less than 15 in 2016 and 2018. The additional deaths were caused by a fire in a highrise residential building. I used the same treatment as in Chapter 4 to remove the outlier from the modelling stage and add the excess mortality in at a later stage before calculating life expectancies. I defined life expectancy inequality as the difference between 2.5th and 97.5th percentiles of LSOA life expectancies.

To investigate the extent that life expectancy change is accompanied with changes in

the characteristics of LSOA populations, I report the Spearman's (rank) correlations of change in life expectancy with the following characteristics, separately for each decile of house price in 2002:

- Change in house price.
- Change in households living in the LSOA during the period 2002–2019 as measured by population turnover.
- Change in share of the LSOA population in the age groups 0–14, 15–29, 30–69 and 70+ years.
- Change in rank of LSOA in terms of the poverty, unemployment and low education measures of deprivation.

Scores of deprivation variables are not comparable between years, so I only consider the change in ranking. In order to use a consistent measure of correlation, and account for potential non-linear associations, I used Spearman's rank correlation for all variables.

## 5.3 Results

### 5.3.1 Inequalities in life expectancy

Life expectancy at birth for London increased from 80.9 (95% CrI 80.8–81.0) years in 2002 to 85.4 (85.3–85.5) years in 2019 for women, and from 76.1 (76.0–76.2) years to 81.6 (81.5–81.7) years for men. Life expectancy inequality (difference between 2.5th and 97.5th percentiles) was 19.1 (18.4–19.7) years for women and 17.2 (16.7–17.8) years for men, in 2019 which is a substantial increase from 11.1 (10.7–11.5) years for women and 11.6 (11.3–12.0) years for men in 2002. The corresponding estimates for life expectancy inequality in 2019 calculated using MSOA level analysis in Chapter 4 were 10.8 (10.3–11.3) years for women and 11.0 (10.5–11.6) years for men.

Life expectancy in 2019 was highest in LSOAs in central London districts of Kensington and Chelsea, Westminster, City of London and Camden, in the southwest (Richmond upon Thames and Kingston upon Thames) and parts of the northwest (e.g., parts of Harrow and Barnet), with life expectancy in many LSOAs surpassing

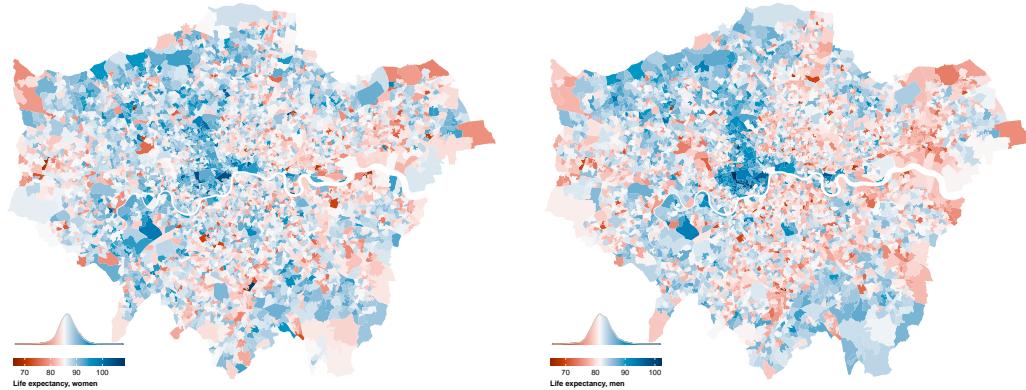


FIGURE 5.1: Map of life expectancy and the distribution of life expectancy in London in 2019. The areas in white have a life expectancy equal to the overall London life expectancy.

90 years (Figure 5.1). Low life expectancy was spread in LSOAs throughout the city but was more common in outer east and southeast London.

### 5.3.2 Change in life expectancy and inequality

There was large variation across London in how much life expectancy increased from 2002 to 2019, ranging from <2 years in 537 (11.1%) LSOAs for women and 214 (4.4%) for men to >10 years in 220 (4.6%) for women and 211 (4.4%) for men (Figure 5.2). In 134 (2.8%) LSOAs for women and 32 (0.7%) for men, life expectancy may have declined, with a posterior probability of a decline greater than 80% in 41 (0.8%) and 14 (0.3%) LSOAs for women and men, respectively. Life expectancy increased more in LSOAs in central and inner east and south London than in outer areas.

The rate of increase in life expectancy was smaller after 2010 compared to the earlier years for both sexes (Figure 5.3), and the number of LSOAs in which life expectancy may have fallen was larger, especially from 2010 to 2014. The slow-down or reversal of LSOA life expectancy after 2010 is consistent with the pattern seen throughout England's MSOAs in Chapter 4, which may be partly due to increase in poverty due to low-wage employment and cuts in services in the austerity era. However, the

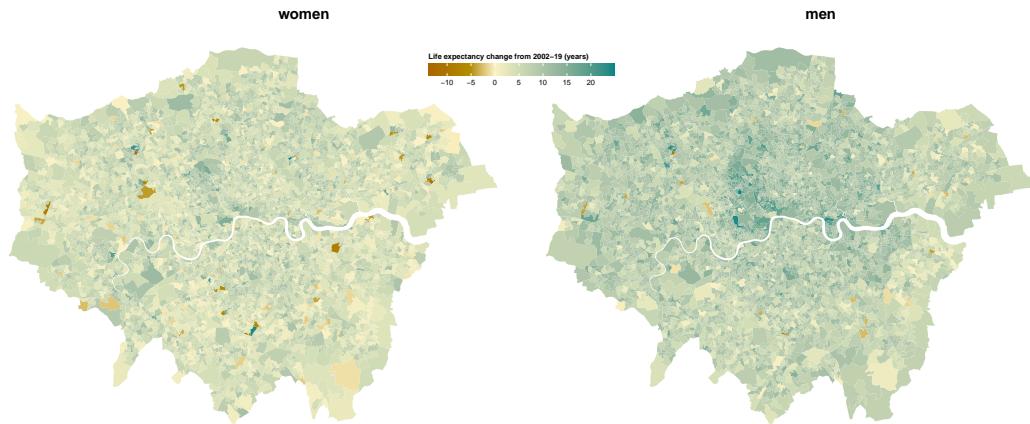


FIGURE 5.2: Geography of change in life expectancy in London from 2002 to 2019.

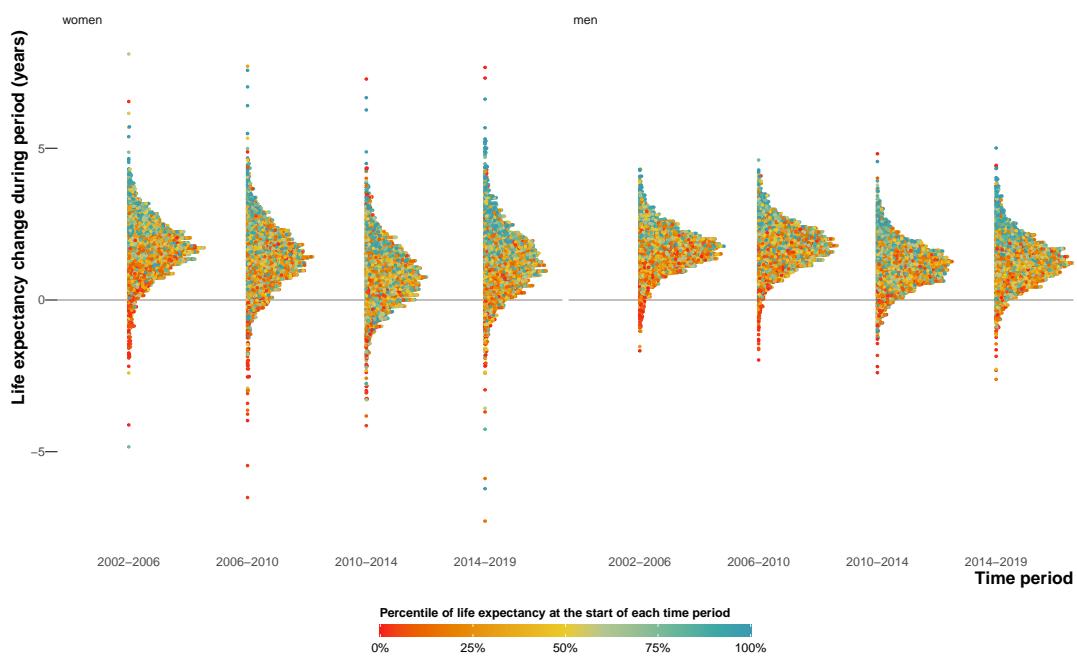


FIGURE 5.3: Change in LSOA life expectancy in different time periods, 2002–19. Each point shows the posterior median change in one LSOA. LSOAs are coloured by their life expectancy at the beginning of each period.

progressive worsening of life expectancy gain and loss from 2010–2014 to 2014–2019 which was seen in other parts of England did not happen in London (Figure 5.3).

There was substantial variation in the size of life expectancy increase over short distances. As a result of this spatial heterogeneity, life expectancy inequality increased not only in London as a whole, but also in every district in London alongside increasing average life expectancy (Figure 5.4 and Figure 5.5).

By 2019, female and male life expectancies had a 2.5th–97.5th percentile range that was >12 years in every district, and >20 years in three districts for men and two for women. In 2002, within district inequality had been >12 years in only two districts for men and none for women. At the same time, the difference between districts with the highest and lowest life expectancy increased from 5.9 (5.2–7.7) years in 2002 to 8.5 (5.9–11.9) years in 2019 for females, and from 6.8 (5.9–8.7) years to 10.9 (8.2–14.2) years for males indicating that the rise in life expectancy inequality took place both within and between districts.

### 5.3.3 Correlations with house price

Life expectancy increase was associated with changes in the LSOA resident population and their socioeconomic characteristics (Figure 5.6). The association was weaker in women and in LSOAs that were already expensive in 2002. There was some association between increase in life expectancy and increase in house price for the LSOAs that were least expensive in 2002. Although there was no obvious pattern between life expectancy increases and the share of households who had not lived there in 2002, male life expectancy was also associated with a larger rise in the share of population in working ages (30–69 years), and a larger decline in the share of population in young adulthood (15–29 years). The LSOAs that improved their rankings in terms of poverty and employment experienced the largest increases in life expectancy.

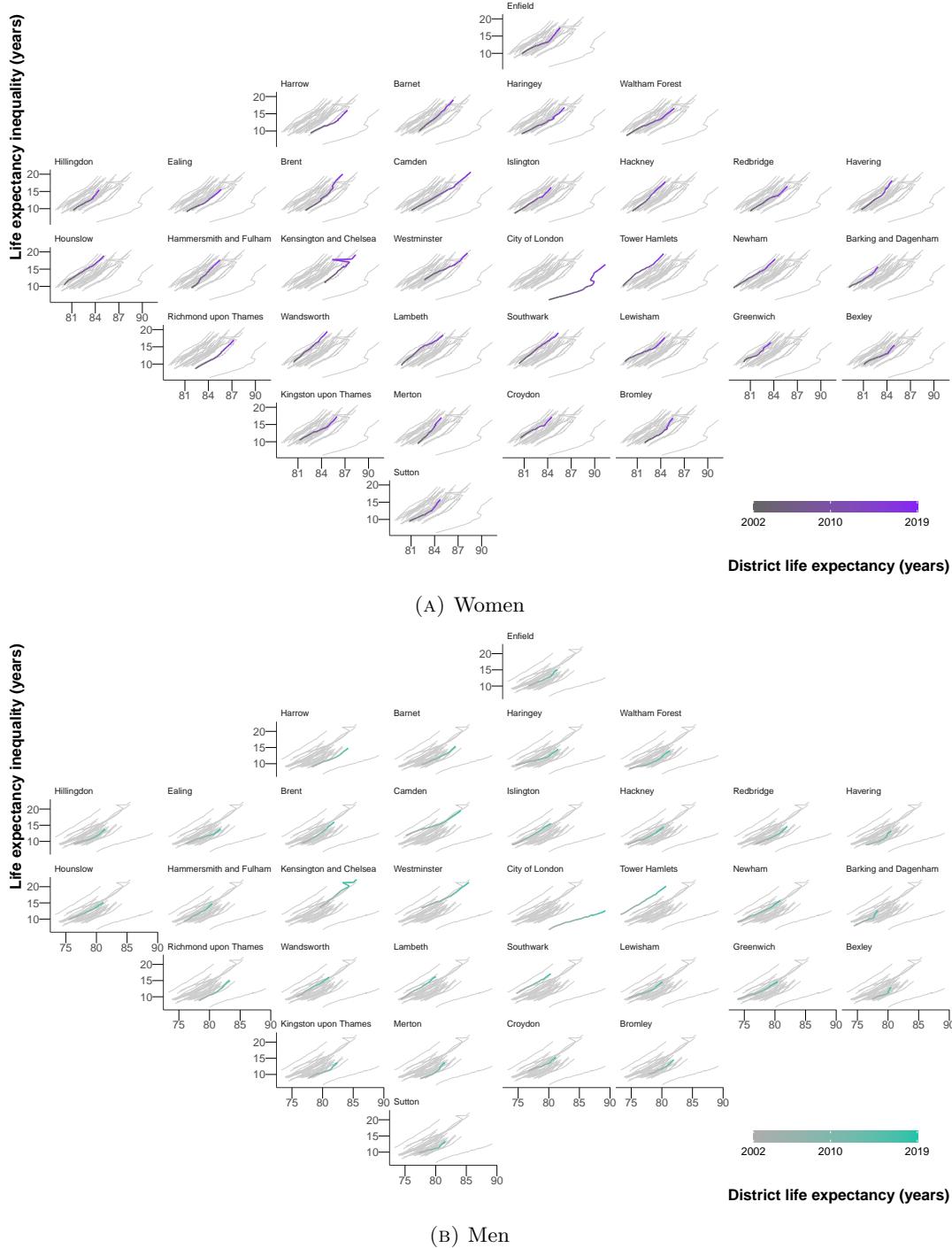


FIGURE 5.4: Life expectancy at birth and inequality in life expectancy within London's 33 local authority districts from 2002 to 2019. Each panel plots life expectancy against life expectancy inequality over time for all 33 districts, and highlights the trend line for the district which is named at the top of the panel. The districts are arranged according to their approximate location in Greater London. The large decline in life expectancy in Kensington and Chelsea in 2017 is due to the deaths caused by the fire in Grenfell Tower.



FIGURE 5.5: Distribution of estimates of Lower-layer Super Output Area (LSOA) life expectancy at birth for 2002 and 2019, and of the change from 2002 to 2019 in 33 London districts. Districts are ordered by the median life expectancy in 2002.

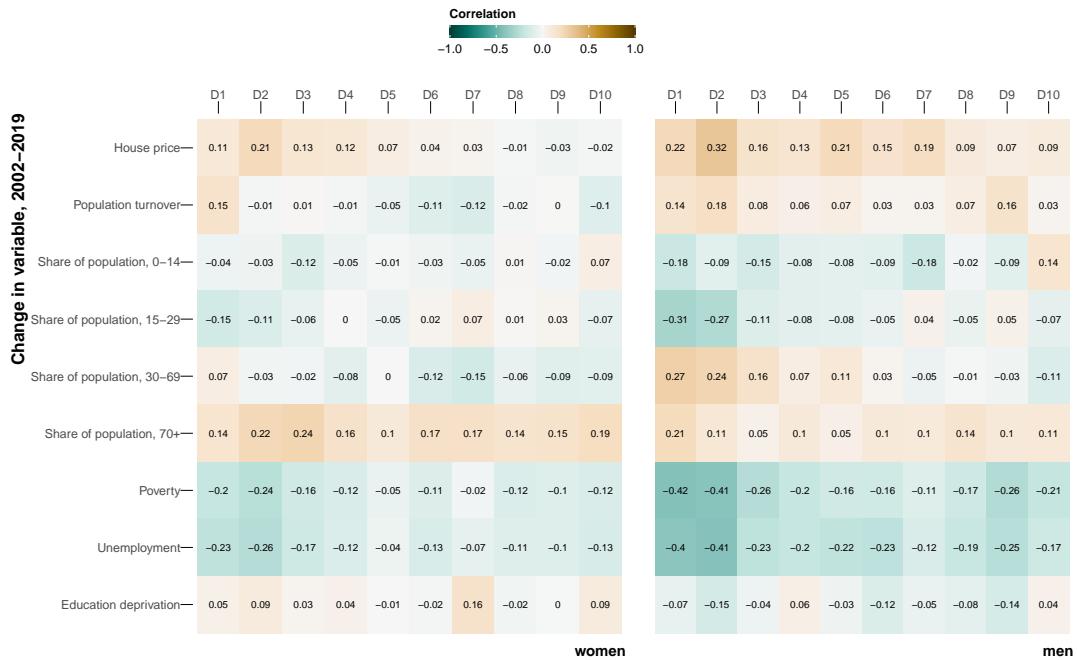


FIGURE 5.6: Spearman correlation between change in life expectancy and change from 2002 to 2019 in sociodemographic characteristics of the population of London Lower-layer Super Output Areas (LSOAs). LSOAs were grouped by decile of house prices in 2002 and correlations were calculated separately for each decile. Correlations close to zero imply the lack of a relationship between the ranks of the variables.

## 5.4 Discussion

This high-resolution analysis revealed that in London, a major global metropolis, not only has life expectancy increased substantially, but this increase has happened both between districts and within every district.

Life expectancy in London has previously only been reported at MSOA level, as discussed in Chapter 4. When aggregated to MSOA level, the LSOA life expectancies in this work were correlated with MSOA estimates in Chapter 4 (correlation coefficient  $> 0.99$  and mean difference  $\leq 0.01$  years for both sexes).

### 5.4.1 Strengths and limitations

The main strength of the study is the presentation of very high-resolution data on life expectancy and house prices in consistent spatial and temporal units in a major global city over a period of substantial policy interest and investment in urban renewal. Carrying out the study at the LSOA level uncovered inequalities to a fuller extent.

I defined life expectancy inequality as the difference between 2.5th and 97.5th percentiles of LSOA life expectancies rather than the difference between the maximum and minimum as in Chapter 4. This was because the LSOAs with extremely low life expectancies tended contain age-LSOA-year combinations in which the number of deaths exceeded the population. Small-area population, which is the denominator of age-specific death rates, is estimated by the ONS for intercensal years, and may be subject to error. This is especially the case in older ages when some people live and die in a long-term care facility, and may be counted towards population (denominator of death rates) in their original LSOA of residence and towards deaths (numerator of death rates) in the LSOA where the care facility is located. There were a higher proportion (0.099% compared to 0.001% in Chapter 4) of spatial units with this issue at the LSOA level. In 2019, 11 LSOAs for women and one LSOA for men had life expectancies in which the median estimate was greater than 100 years. Given the MSOAs with the highest life expectancies for each sex in Chapter 4 were both in London and exceeded 95 years, these estimates may be plausible for some LSOAs, which contain around a fifth of the population of MSOAs. However, upon inspection, four out of the five LSOAs with the largest increases in life expectancy, which were also LSOAs with life expectancies above 100, were adjacent to LSOAs with very large decreases. And in four out of five LSOAs with the largest estimated decreases in life expectancy, estimated population counts were lower than the recorded death counts in the older age groups in several years of the data, implying that populations or deaths may have been allocated incorrectly. Subsequent inspection showed that postcodes coinciding with, or adjacent to, LSOAs with the largest increases and decreases contained care homes, which can lead to the place of residence or death being inconsistently recorded for some of their residents.

Mortality and house prices were estimated using two separate models because a joint model would be extremely complex and would require additional assumptions about how house prices are associated with age- and LSOA-specific mortality. It was only possible to indirectly evaluate, through LSOA-level population turnover and population characteristics, whether population change is a potential mechanism for the observed change in life expectancy because routine death registration in England

only records place of residence at the time of death. Further, there are currently limited time-series data on quality of housing, access to jobs, services and amenities, and other home and neighbourhood characteristics that affect health. To understand whether changes in the life expectancy of communities arises from changes in the health of the population, itself due to changes in their economic status and/or local environment and amenities, versus a change in the resident population requires linked datasets which are able to track over time environmental characteristics of areas together with individuals' place of residence, socioeconomic status and mortality records.

#### 5.4.2 Comparison with previous literature

The inequalities of 19.1 (18.4–19.7) years for women in 2019 and 17.2 (16.7–17.8) years for men in London in 2019 are some of the largest for any city in the literature and the largest for women. Further, some of the inequality in the tails of the life expectancy distribution is masked by our choice of inequality measure, the 2.5th–97.5th percentile difference, rather than the full range across all LSOAs.

In a study of the 397 census tracts in King County, Washington, a county ranking in the top 5% of the US in terms of both income and life expectancy, Dwyer-Lindgren et al. (2017b) found an 18.3 year gap in life expectancy for men between the top and bottom tracts in 2014, and a 14.8 gap for women. For Vancouver, Canada, Yu et al. (2021) found a difference in 2016 of 9.5 years for men and 8.3 years for women between the 90th and 10th percentiles of 368 census tracts. Using the same measure of inequality as the Vancouver study, Jonker et al. (2012) uncovered a 8.2 year and 9.2 year gap in life expectancy for women and men for 89 small areas in Rotterdam. Across the six large cities in Latin America, Bilal et al. (2019) observed the largest difference between the top and bottom decile of life expectancy at birth was 17.7 years for women in Santiago, Chile.

To some extent, as this is one of the highest-resolution studies in the literature, it is natural that we would find large inequalities. As we go to finer and finer resolutions, we will always uncover more inequality, until we reach the physical limit of the individual and the inequality is between the earliest neonatal death and the oldest person

to die in the city.

In terms of work specific to London, beyond that at the MSOA level in Chapter 4, Congdon (2014) considered both ill-health and mortality for 625 wards in London, finding more than a two-fold variation in the percent of life spent in disability for men, but did not mention the absolute inequality of life expectancy in the city. Some studies have documented changes in spatial patterns of poverty in London,(Bosetti, 2015; Travers et al., 2016) and have found that poverty rates have fallen in inner East London but may have risen in some outer London areas, with corresponding changes in the share of the population working in higher professional occupations. No previous study has reported how life expectancy change in small-area units coincided with change in house prices.

#### 5.4.3 Urban policy and public health implications

Since the turn of the millennium, London's population and economy have grown substantially. The economic growth has been highly polarised with high-pay and high-skilled employment alongside low-pay and insecure jobs (Overman and Xu, 2022). As a result, despite city-wide growth in income, nearly one half of London's population fall in the bottom two quintiles of national income deprivation ("English indices of deprivation 2019," 2019). Together with an uncontrolled property market, this has created house prices that are unfavourable to low-income families, displacing entire subgroups of the population to cheaper parts of the city, with fewer amenities and worse access to jobs, quality education, healthcare and other services and amenities (Bosetti, 2015). Many are unable to purchase (versus rent) and hence spend an increasing share of income on housing, and/or live in lower quality or smaller accommodation in the more desirable districts (Bosetti, 2015; Travers et al., 2016). These trends may have contributed to health inequalities both between nearby areas and across the entire city alongside other trends such as differences in the extent of neighbourhood improvement or provisions of health and social care services as council budgets were reduced as a result of austerity policies.

The evolution of London, and major cities in other high-income nations and emerging economies, into places where only the well-off can afford to own properties, where the

balance of the city is driven by the cost of property and wealth dominates access, poses a gloomy, non-cohesive future for these cities.

## 5.5 Summary

I performed a hyper-resolution spatiotemporal analysis of all deaths in London from 2002 to 2019, using a Bayesian hierarchical model to obtain estimates of age-specific death rates by age, sex, and LSOA. I used life table methods to calculate life expectancy at birth by sex and LSOA, and explored correlations between change in life expectancy and the change in sociodemographic characteristics of the resident population of LSOAs and population turnover.

In 134 (2.8%) of London's LSOAs for women and 32 (0.7%) for men, life expectancy may have declined from 2002 to 2019, with a posterior probability of a decline >80% in 41 (0.8%, women) and 14 (0.3%, men) LSOAs. The life expectancy increase in other LSOAs ranged from <2 years in 537 (11.1%) LSOAs for women and 214 (4.4%) for men to >10 years in 220 (4.6%) for women and 211 (4.4%) for men. The 2.5th-97.5th percentile life expectancy difference across LSOAs increased from 11.1 (10.7–11.5) years in 2002 to 19.1 (18.4–19.7) years for women in 2019, and from 11.6 (11.3–12.0) years to 17.2 (16.7–17.8) years for men.

These was some association between changes in male life expectancy and a rise in the share of population in working ages (30–69 years), and a larger decline in the share of population in young adulthood (15–29 years). The LSOAs that improved their rankings in terms of poverty and employment experienced the largest increases in life expectancy.

## Chapter 6

# And what they died from: cause-specific mortality in England at the district level

### 6.1 Overview

### 6.2 Methods

I performed a high-resolution spatiotemporal analysis of civil registration data in which I extracted de-identified data for all deaths in England from 2002 to 2019 (8,648,191 death records, extract date: 17th November 2022). England was divided into 314 districts according to the boundaries in 2020. Deaths were assigned into the following age groups: 0, 1–4, 5–9, ..., 80–84, 85+ years. I did not use 130 death records (<0.001%) for which sex was not recorded. There were no age-district-year combinations in which the number of deaths from all causes exceeded population.

Each death record in the post-neonatal period was assigned an ICD-10 code corresponding to the underlying cause of death. For neonates, which are not assigned an underlying cause of death, I used the ICD-10 code in the first position on the death record. I used ICD-10 codes to assign each death to 136 cause groups of the WHO Global Health Estimates (GHE) study (“Global Health Estimates,” 2020); these groups encompass causes of death with related aetiology and clinical and public

health relevance. I also grouped diabetes mellitus and nephritis and nephrosis (hereafter referred to as *diabetes*) as these deaths might have a similar history. I used the top 12 causes of death according to the total number of deaths from 2002 to 2019 for cause-specific analysis, as well as a residual groups comprising deaths from all other cancers, all other NCD, all other CVD, all other infections, maternal, perinatal and nutritional conditions (IMPN), and injuries (external causes). Together, these form a mutually exclusive, collectively exhaustive list of causes of death.

The full list of ICD-10 codes for each cause group can be found in Table D.1.

Although the size of the spatial unit of analysis has been increased to the district-level, by stratifying by cause of death, the number of death in each age-district-year-cause stratum is still very small. As in Chapter 4 and Chapter 5, I used a Bayesian hierarchical model to obtain stable estimates of death rates. I conducted all analyses separately by sex and cause group.

I used a binomial likelihood:

$$\text{deaths}_{ast} = \text{Binomial}(m_{ast}, \text{Population}_{ast}). \quad (6.1)$$

A beta-binomial was trialled but did not improve model fit. The model is as follows:

$$\text{logit}(m_{ast}) = \alpha_0 + \beta_0 t + \alpha_{1s} + \beta_{1s} t + \alpha_{2a} + \beta_{2a} t + \xi_{as} + \gamma_{at}, \quad (6.2)$$

where the terms have the same meanings as Equation 4.3. This is the same as model Chapter 5, but without a random walk for each spatial unit, which was tested but did not improve the fit despite adding a lot of complexity.

Spatial effects were modelled using a ICAR effect rather than the nested hierarchy of random effects seen in Chapter 4 and Chapter 5. The geographies here only allow a two-tier hierarchy with only nine regions mapping to 314 districts, which was less preferable than a spatial model. The ICAR model requires all spatial units to have neighbours. Thus, the districts containing the Isle of Wight and the Isles of Scilly were each joined to the nearest mainland district based on road or ferry connections.

All model parameters, their priors and dimensions are in Table [B.3](#).

I fitted the model in the probabilistic programming language NumPyro (Phan et al., 2019) and obtained 1,000 draws from the posterior distribution of model parameters that were used to calculate age-specific death rates. Where possible, non-centred parametrisations were used in the model to improve the efficiency of the NUTS samplers. I ran four chains with a warmup of 2000 iterations and then 10,000 further iterations and thinned the remainder by 40 to obtain 1000 post-burn-in draws from the posterior distribution of model parameters.

I ran a model for a combined group of all deaths to calculate life expectancy in each district. Death rates for this total mortality group, and for injuries, were corrected in 2017 in Kensington and Chelsea for the Grenfell fire by imputing using the mean deaths from injuries between 2016 and 2018 before running the model.

The primary reporting outcome is the unconditional probability of dying between birth and 80 years of age. Unconditional probability is the probability of death in the absence of competing causes of death, and is calculated using life table methods from age-specific death rates from the cause of interest. I limited the age range to 80 years because the probability of death in the absence of competing causes equals 100% when the entire life course is considered. The reported 95% credible intervals represent the 2.5th and 97.5th percentiles of the posterior distribution of estimated probability of death.

I calculated the contributions of deaths from each cause of death, in each age group, to the life expectancy change for each district between different time periods using Arriaga's method, which is widely used to decompose life expectancy differences between populations or population subgroups (Arriaga, 1984). Arriaga's method calculates how much each age group contributes to the life expectancy difference by summing how much death rate differences at that age change the years of life lived both at that age and in subsequent ages through changing the number of survivors. It then partitions the age-specific contributions to the life expectancy gap by cause of death in proportion to the difference in cause-specific death rates between the subgroups. The cause-specific death rates were scaled such that the sum over all causes

was equal to the estimate for total mortality. For this analysis, I used the sample mean death rate in each age-district-year-cause combination.

Details on the calculations for both the probability of dying and Arriaga's method can be found in Appendix A.

## 6.3 Results

### 6.3.1 All causes

There were 4,465,948 (48.4%) deaths in women and 4,182,108 (51.6%) in men in England from 2002 to 2019. 3,972,286 (45.9%) of these deaths occurred in people aged younger than 80 years of age. Figure 6.1 shows the leading causes of death for each sex.

In the following sections, I will write in turn on each of the wider cause groups (CVDs; NCDs; cancers; maternal, perinatal, nutritional and infectious causes; injuries) in order of largest proportion of total deaths.

### 6.3.2 Cardiovascular diseases

Although mortality from CVD has more than halved since the 1980s in England, it remained the wider cause group with the largest proportion of deaths at 2,712,895 (31.4%). The highest probability of dying in 2019 for any sex-cause-district combination was for ischaemic heart disease in Manchester at 0.13 (0.12-0.13) (Figure 6.2).

Asaria et al. (2012) found CVD mortality followed a persistent downward trend in nearly all wards in England from 1982 to 2006. Both ischaemic heart disease (Figure 6.2) and stroke mortality (Figure 6.3) have continued to follow this trend through to 2019 at the district level. These reflect improvements in reducing and controlling risk factors such as high blood pressure and high blood cholesterol, organisational changes to the NHS such that acute CVD episodes are treated in specific centres, improvements in the treatment of CVD including coronary angiographies and stent

---

<sup>1</sup>For brevity, the group "diabetes mellitus, nephritis and nephrosis" has been shortened to "diabetes".

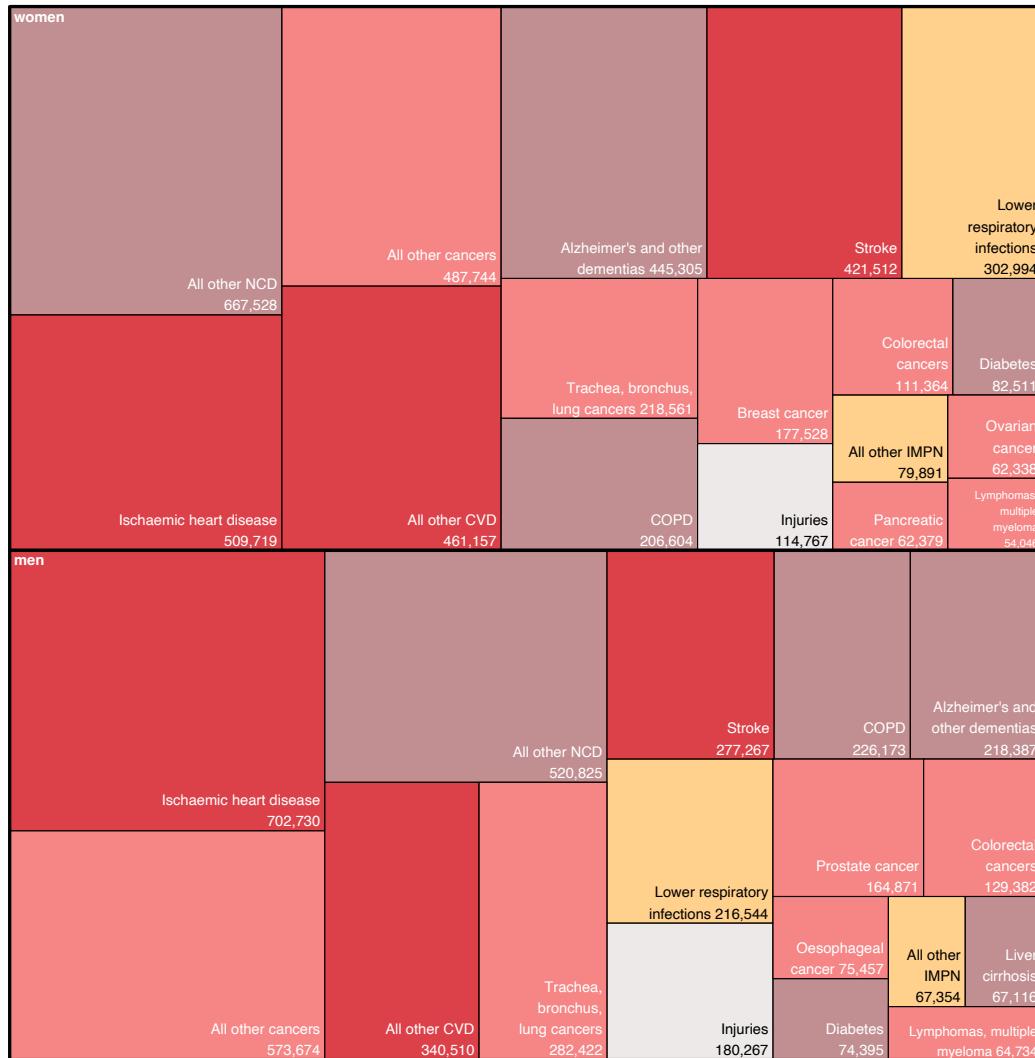


FIGURE 6.1: Total number of deaths for the 12 leading causes of death in England from 2002 to 2019, and the residual groups all other cancers, all other NCD, all other CVD, all other infections, maternal, perinatal and nutritional conditions (IMPN), and injuries. The boxes are coloured by the wider groups of CVDs; NCDs; cancers; maternal, perinatal, nutritional and infectious causes; injuries. See Table D.1 for ICD-10 codes for each category. <sup>1</sup>

insertion, and public health campaigns such as FAST (Face drooping, Arm weakness, Speech difficulties, Time) so the general public know when to seek emergency help.

All other CVD also shows a downward trend, with the exceptions of Hastings and, more noticeably, the Isle of Wight. Looking at the raw data for these districts, there is a step increase the use of ICD-10 code I51.5 – a garbage CVD code (Murray and Lopez, 1996) – since around 2012, suggesting these trends result from poor coding practice.

### 6.3.3 Non-communicable diseases

Non-communicable diseases were responsible for 2,508,844 (29.0%) deaths from 2002 to 2019 (Figure 6.1).

Alzheimer's and other dementias saw the largest increases of any cause, with all districts observing an increase for both sexes (Figure 6.5). It was also one of the most unequal causes of death, with a 4.1- and 3.6-fold difference for women and men between the highest and lowest districts in 2019. This reflects that, beyond age and family history, the main risk factors for Alzheimer's and other dementias are the same as for CVD (smoking, obesity, diabetes, high blood pressure, high cholesterol).

Some part of this trend may be due to increased diagnosis and coding of deaths as dementias, with doctors increasingly assigning mental and neurological conditions as the underlying cause of death rather than simply “dying of old age”. It is worth noting that given only 96,109 (14.5% of 663,692) of deaths from Alzheimer's and other dementias occur in those under 80 years, the probability of dying between birth and 80 is probably not the best metric for summarising this particular disease.

Chronic obstructive pulmonary disease (COPD) was had the largest geographical variation of any disease for women, with a 6.0-fold variation (4.4-fold for men). The probability of dying declined in a larger proportion of districts for men (305 (97.1%)) than for women (215 (68.5%)). This is a result of the main risk factor, tobacco smoking, peaking later in women than in men. The largest increases for both sexes were seen in Kingston upon Hull for (45.9% (24.8% to 67.9%) for women and 20.6% (2.9% to 40.7%) for men).

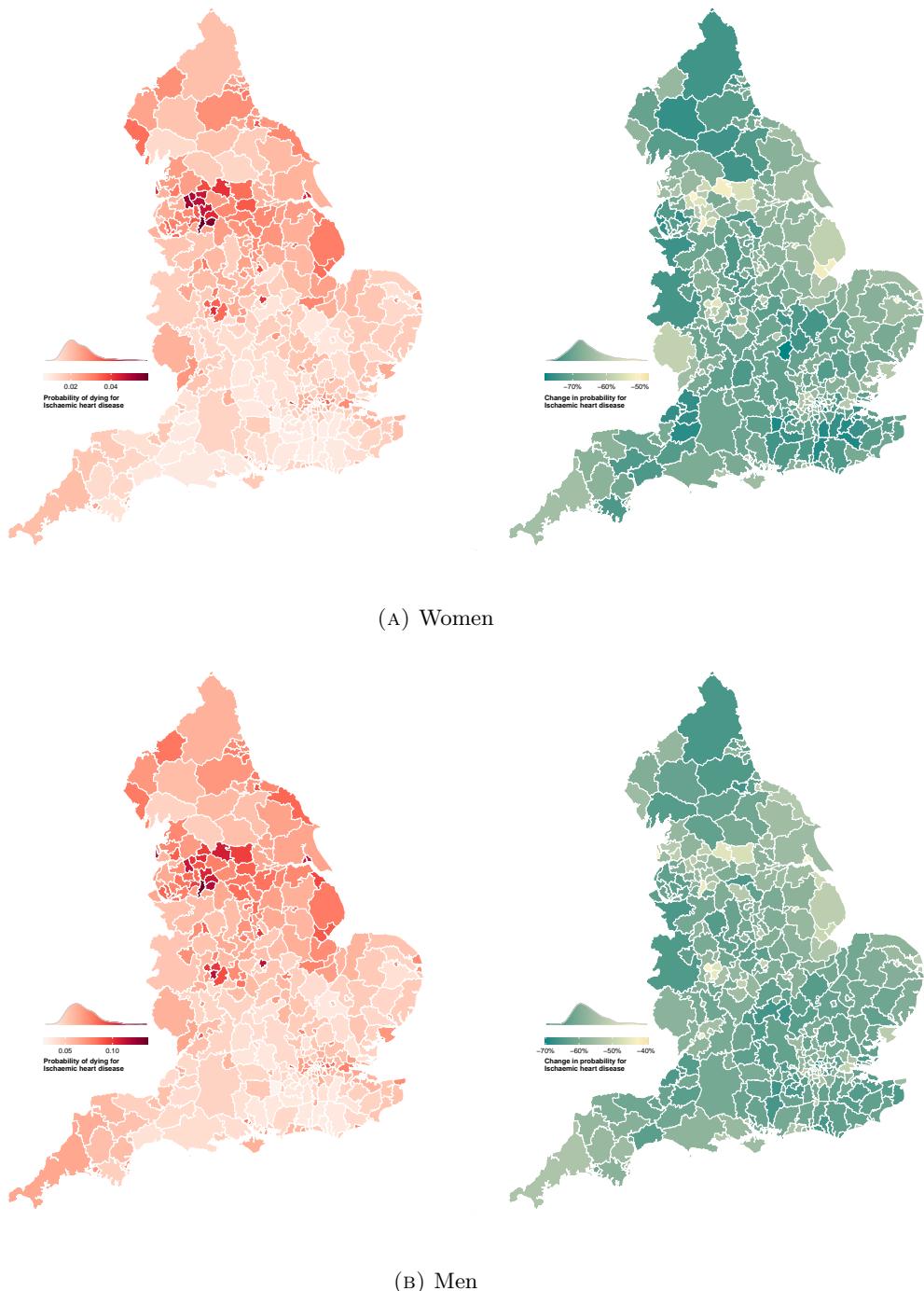


FIGURE 6.2: Probability of dying from ischaemic heart disease between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

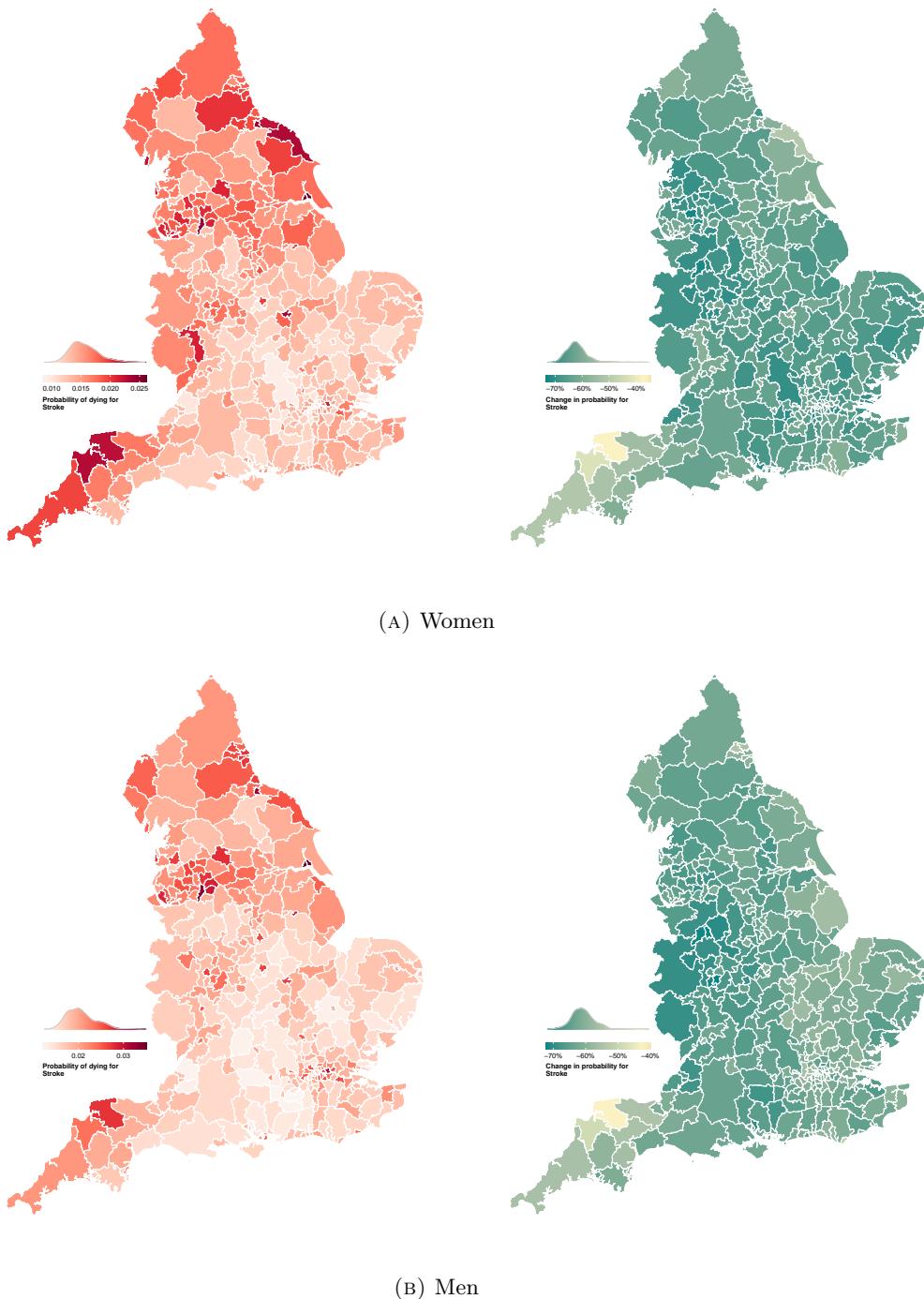


FIGURE 6.3: Probability of dying from a stroke between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

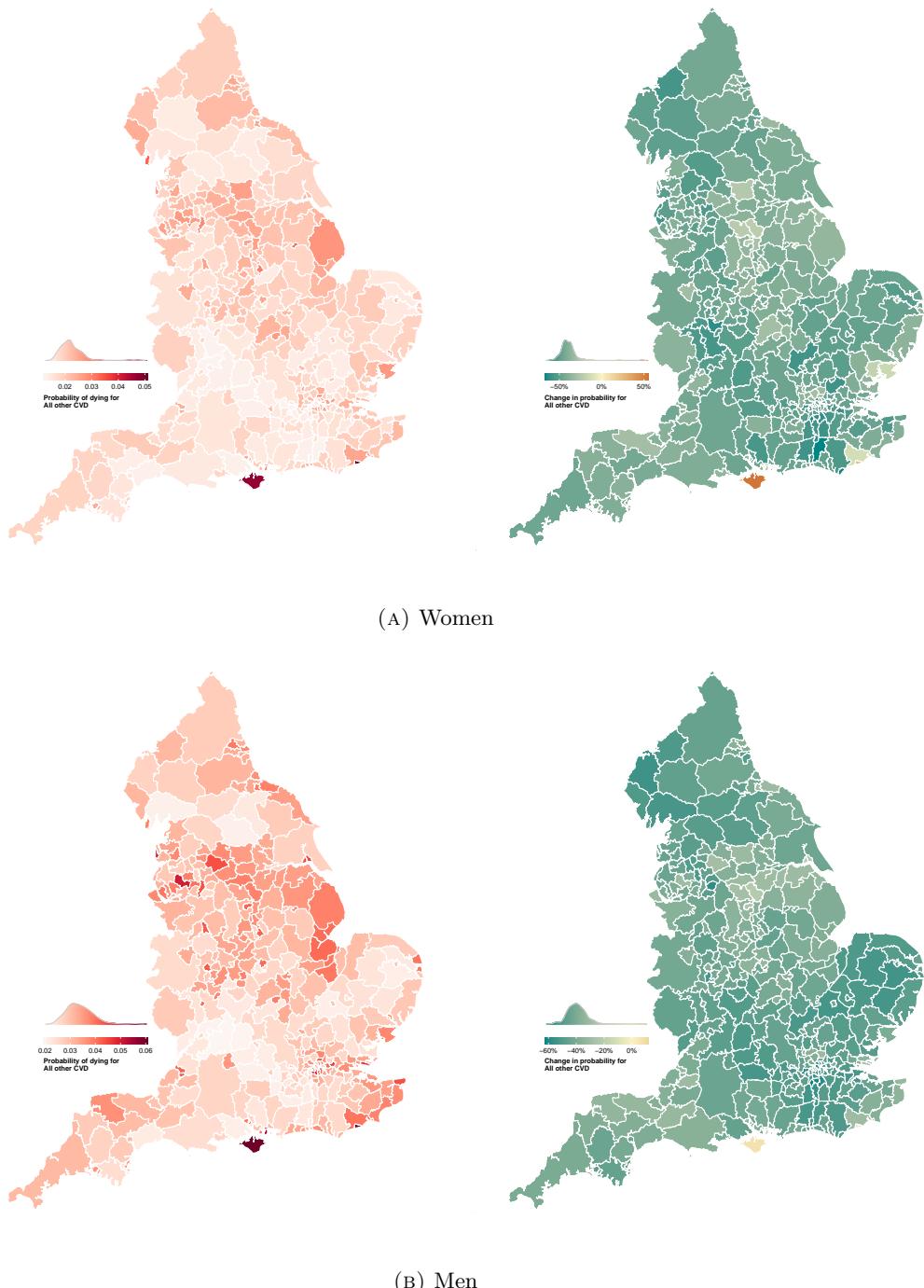


FIGURE 6.4: Probability of dying from all other CVDs between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

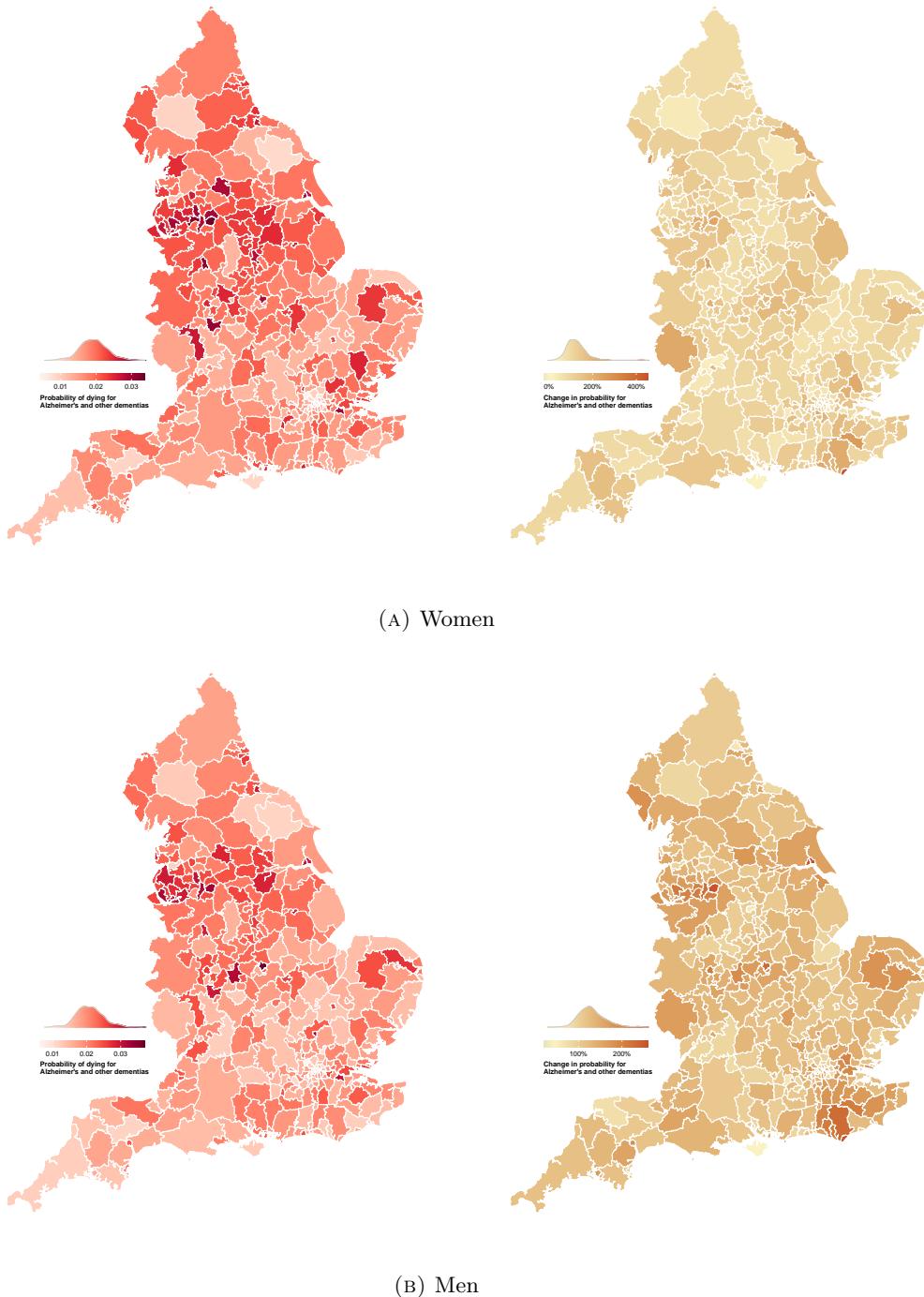


FIGURE 6.5: Probability of dying from Alzheimer's and other dementias between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

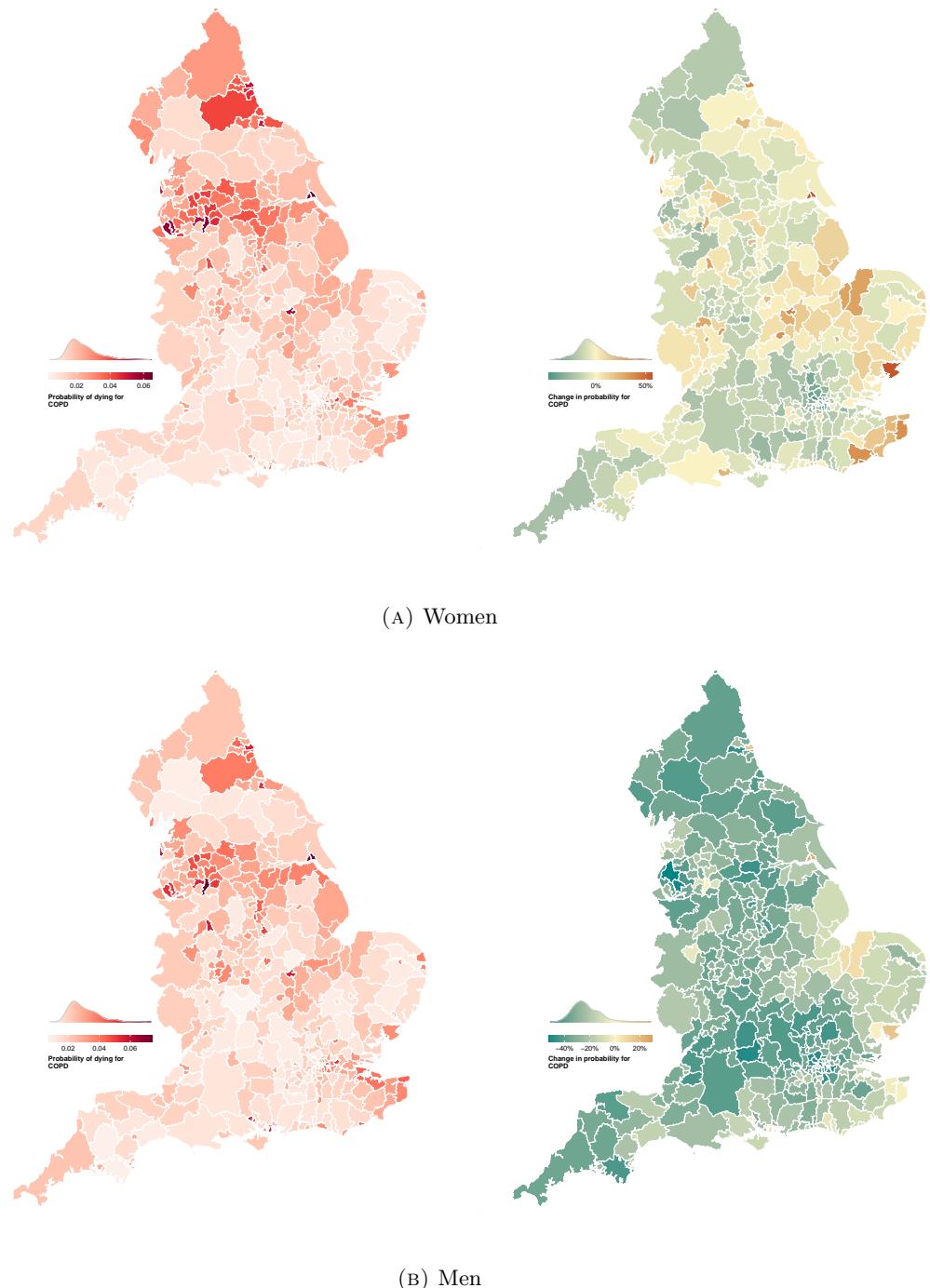


FIGURE 6.6: Probability of dying from COPD between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

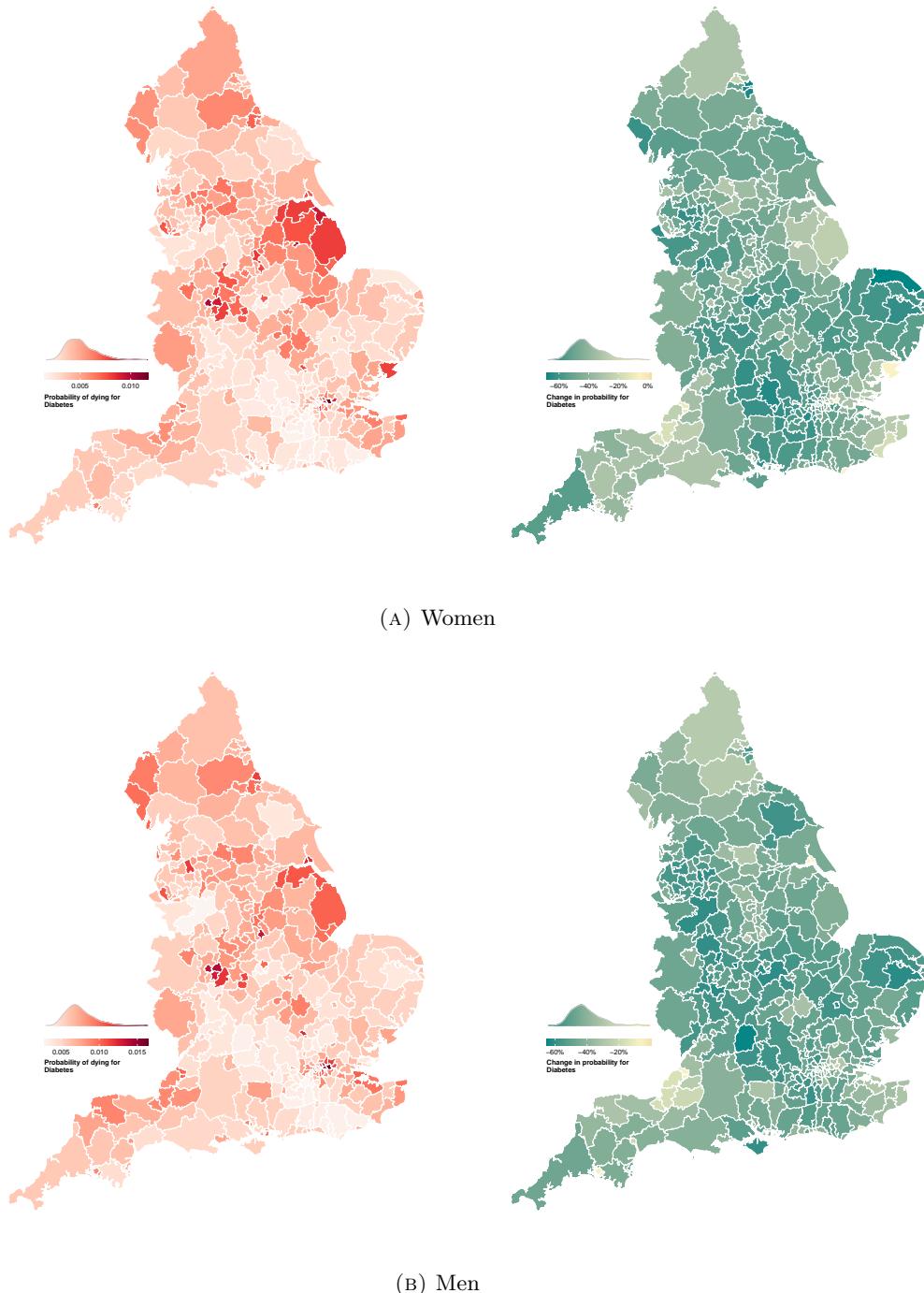


FIGURE 6.7: Probability of dying from diabetes mellitus, nephritis and nephrosis between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

Mortality from diabetes decreased in every district for both sexes (Figure 6.7), matching trends at the national level in the literature (Pearson-Stuttard et al., 2021). Those who are African-Caribbean, Black African, Chinese or South Asian are more at risk, and the highest probabilities of dying for each sex were in east London – one of the most ethnically diverse areas in the country. Specifically, the borough of Newham had the highest mortality for both sexes (0.01 (0.01-0.01) for women and 0.02 (0.01-0.02) for men).

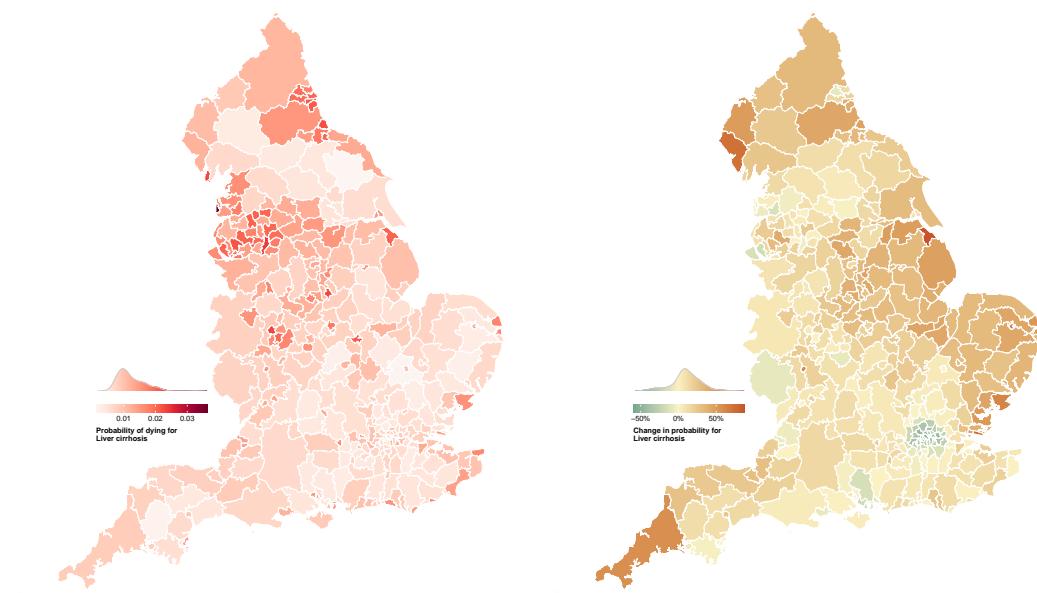


FIGURE 6.8: Probability of dying for men from liver cirrhosis between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

Liver cirrhosis, an advanced stage of liver damage, was only of the top 12 leading causes of death for men (67,116 (1.6%)) but not for women. It has perhaps the most interesting geography for change of all causes, with contrast between huge decreases in London and increases in the rest of the country, with the largest in the coastal southwest and east. The geography is perhaps indicative of the contrasting dynamics of the two main risk factors for liver cirrhosis: alcohol misuse and hepatitis B/C infection. Alcohol is the main cause of liver disease, and has driven a large proportion of increases in liver cirrhosis throughout Europe (Blachier et al., 2013). Alcohol is generally consumed less in the capital, with London having the lowest percentage of adults who abstain from drinking alcohol (23.6%) and the lowest percentage of adults

who drink over 14 units per week (20.1%) (“Local Alcohol Profiles for England,” 2021). On the other hand, prevalence of hepatitis B/C has decreased over the past decades. There has been reduced incidence and vaccination for hepatitis B. Since a large number of people acquired hepatitis C in the 1970s and 1980s, the virus was identified and transmission has reduced (Blachier et al., 2013).

All other NCDs had the highest mortality for any disease for women, with the highest probability of dying of 0.09 (0.08-0.09) in Blackpool (Figure 6.9).

#### 6.3.4 Cancers

Cancers accounted for 28.5% of all deaths, nearly the same total mortality burden as NCDs. With over 150 different types of cancer, each with their own anatomical and molecular subtypes, cancer is extremely complex, with a specialist workforce dedicated to each unique cancer. Compared with CVDs and NCDs, which are dominated by their major causes (ischaemic heart disease and stroke, and Alzheimer’s, COPD and diabetes, respectively), cancers have the largest number of individual causes in the top 12. I felt I should go deeper into the cancer story, and pay further attention to a wider variety of site-specific cancers. This is the focus of Chapter 7.

#### 6.3.5 Maternal, perinatal, nutritional and infectious causes

As is the case in most high income countries, which are in the third stage of the epidemiologic transition, there are far fewer deaths (666,783 (7.7%)) from maternal, perinatal, nutritional and infectious causes compared to the NCDs of group 2 (including CVD and cancers). Although notably, these results consider the period before the Covid-19 pandemic, so the cause-composition will have changed to include more deaths from infections over the most recent years.

Mortality from lower respiratory infections, which include influenza and pneumonia, can vary greatly between years depending on the severity of the virus and whether the flu vaccination prevents the prevalent strain. It also depends on the strength of the immune system of the infected. The ability of the lungs to recover from infection can be degraded by years of industrial exposure or smoking, which perhaps influences the high mortality in the urban northwest (Figure 6.10).

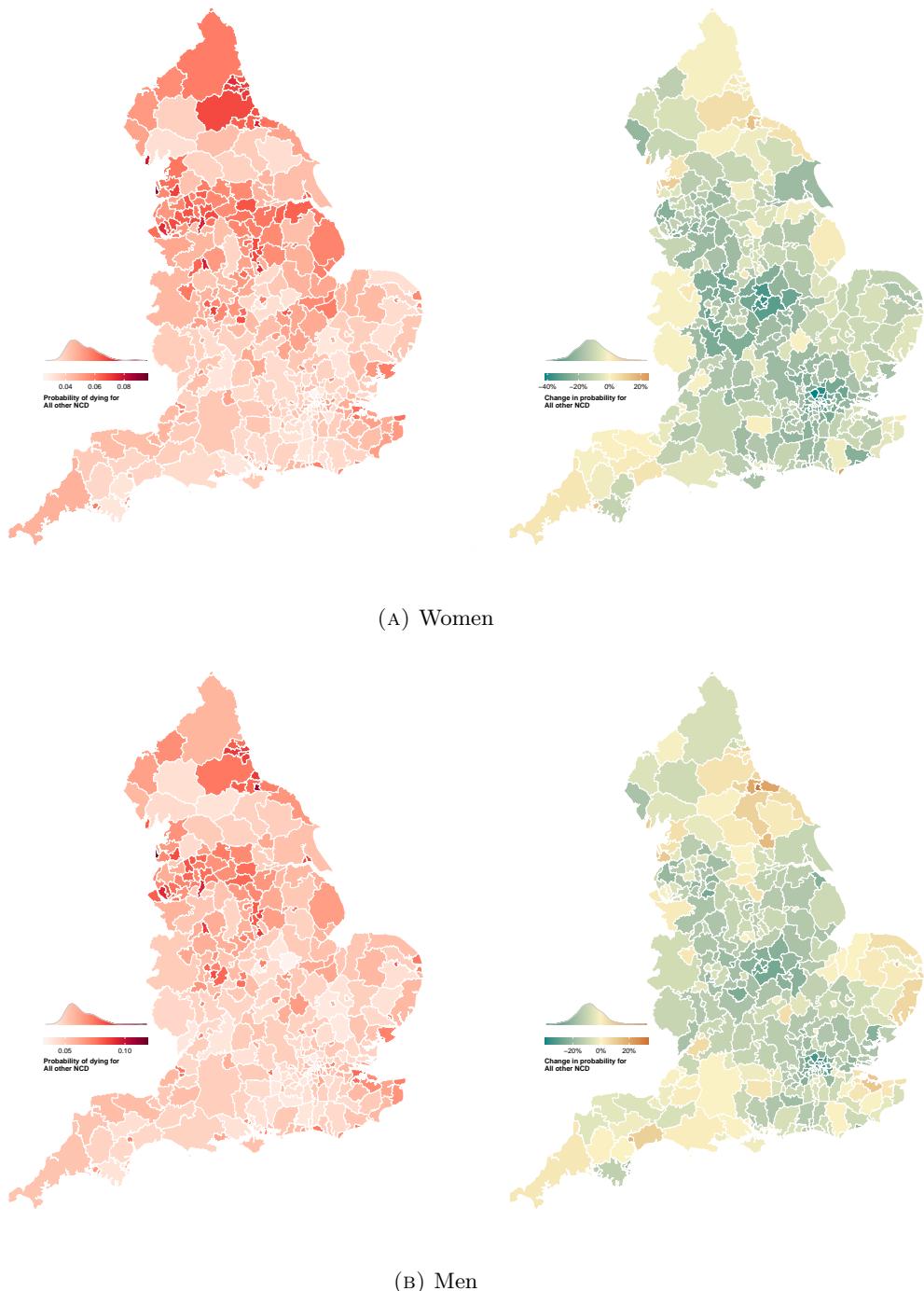


FIGURE 6.9: Probability of dying from all other NCDs between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

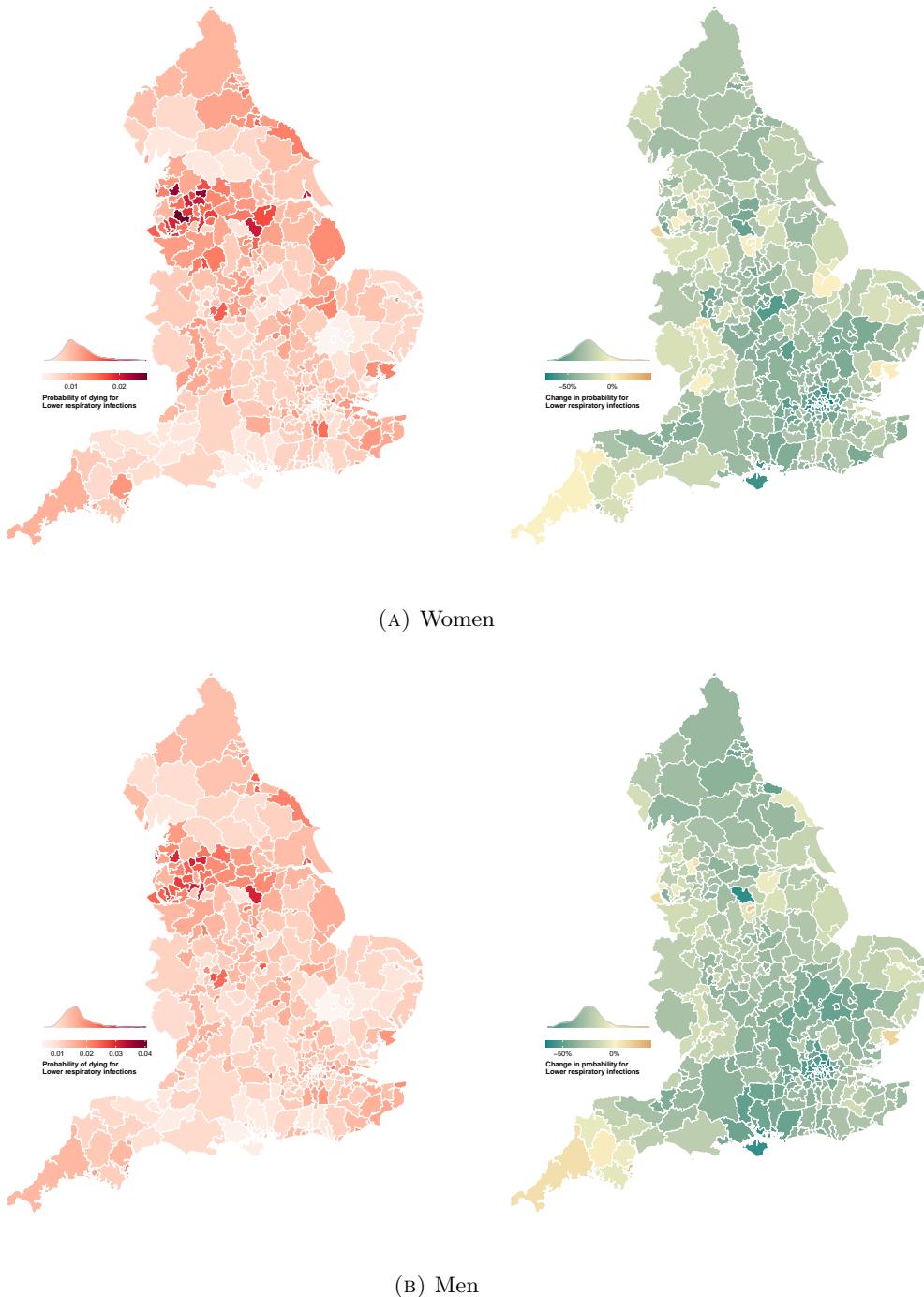


FIGURE 6.10: Probability of dying from lower respiratory infections between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

The similar profiles for men and women for the level of mortality from all other IMPN in 2019 (Figure 6.11) reflects that the three main causes by number of deaths are other infectious diseases, diarrhoeal diseases, low birth weight rather than maternal conditions. However, the change in probability of dying was greatly different between the sexes, with 71 (22.6%) districts seeing an increase for women, the largest in Barrow-in-Furness (55.6% (-0.8% to 135.2%)), compared to only 5 (0.02%) districts for men. Without further disaggregating this group, it is difficult to suggest drivers of these trends.

### 6.3.6 Injuries

Accounting for only 295,034 (3.4%) of total deaths from 2002 to 2019, injuries contribute the smallest proportion of the wider groups of causes of death. The highest mortality in 2019 were in coastal areas, where there are obvious natural hazards, are in the urban northwest (Figure 6.12). Blackpool had the highest probability of dying for men at 0.04 (0.04-0.05), which corresponds with the literature on rates of mortality from drugs and suicide (Congdon, 2019).

### 6.3.7 Contribution to life expectancy change

From 2002 to 2019, every district increased in life expectancy for both women and men (Figure 6.13). This was driven largely by decreases in ischaemic heart disease, which contributed gains of between 0.7 and 1.9 years for women between 1.4 and 4.0 years for men. In contrast, Alzheimer's and other dementias acted in the other direction, forcing life expectancy gains downwards by between 0.2 and 1.1 years for women and between 0.4 and 1.0 years for men.

In Figure 6.14, I break the overall trend down into two time periods around the year 2010. This both follows the analysis for Figure 4.8, but also it is the year there was a change in government, and a change in fiscal policy towards austerity.

In the period 2010-19, life expectancy progress slowed in England, as discussed at the MSOA level in Chapter 4, with three districts experiencing a decline in life expectancy. For women, lung cancer switched from generally contributing negatively to the change in district life expectancy (Figure 6.14a) to positively (Figure 6.14b) The contribution

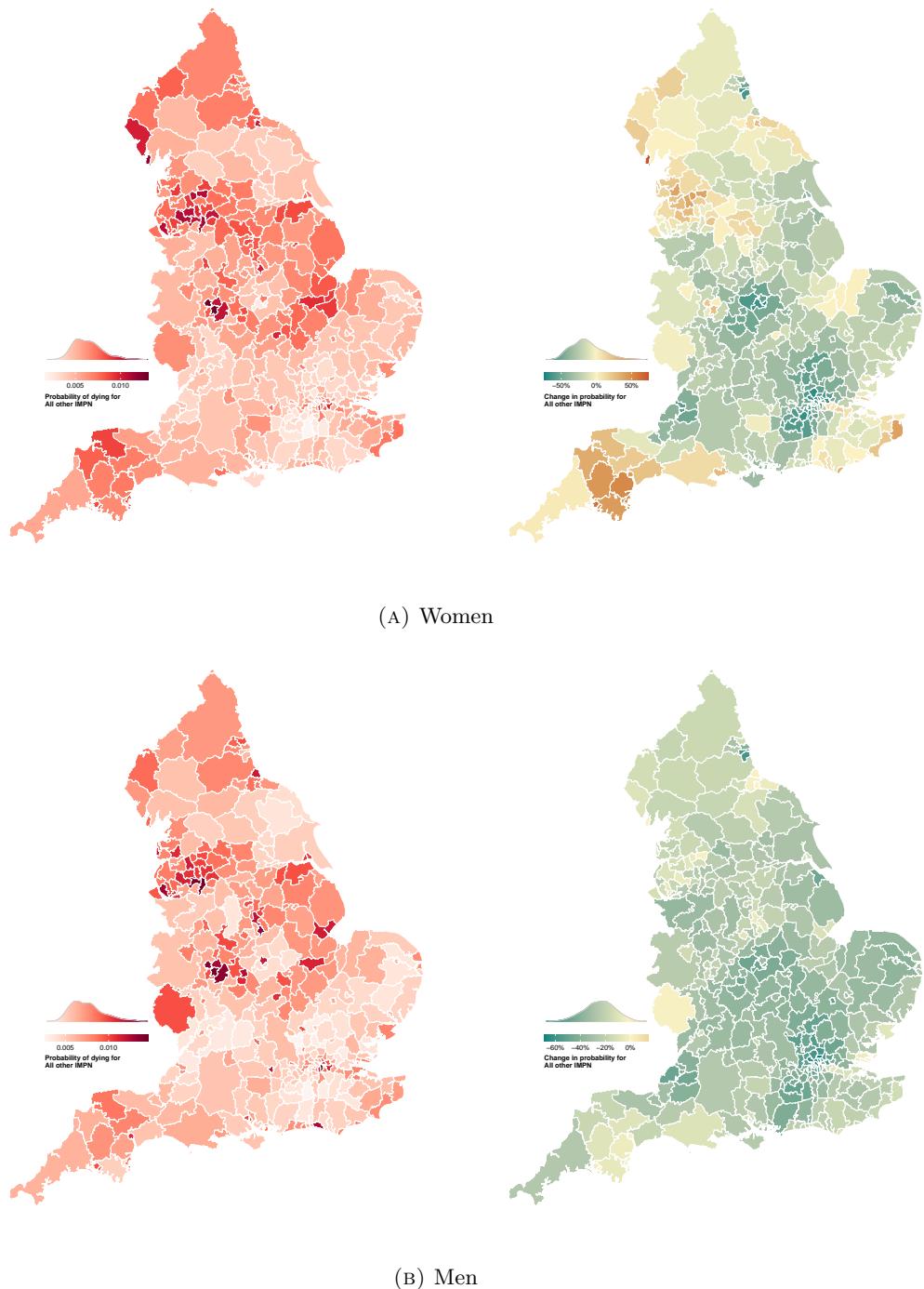


FIGURE 6.11: Probability of dying from all other infections, maternal, perinatal and nutritional conditions between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

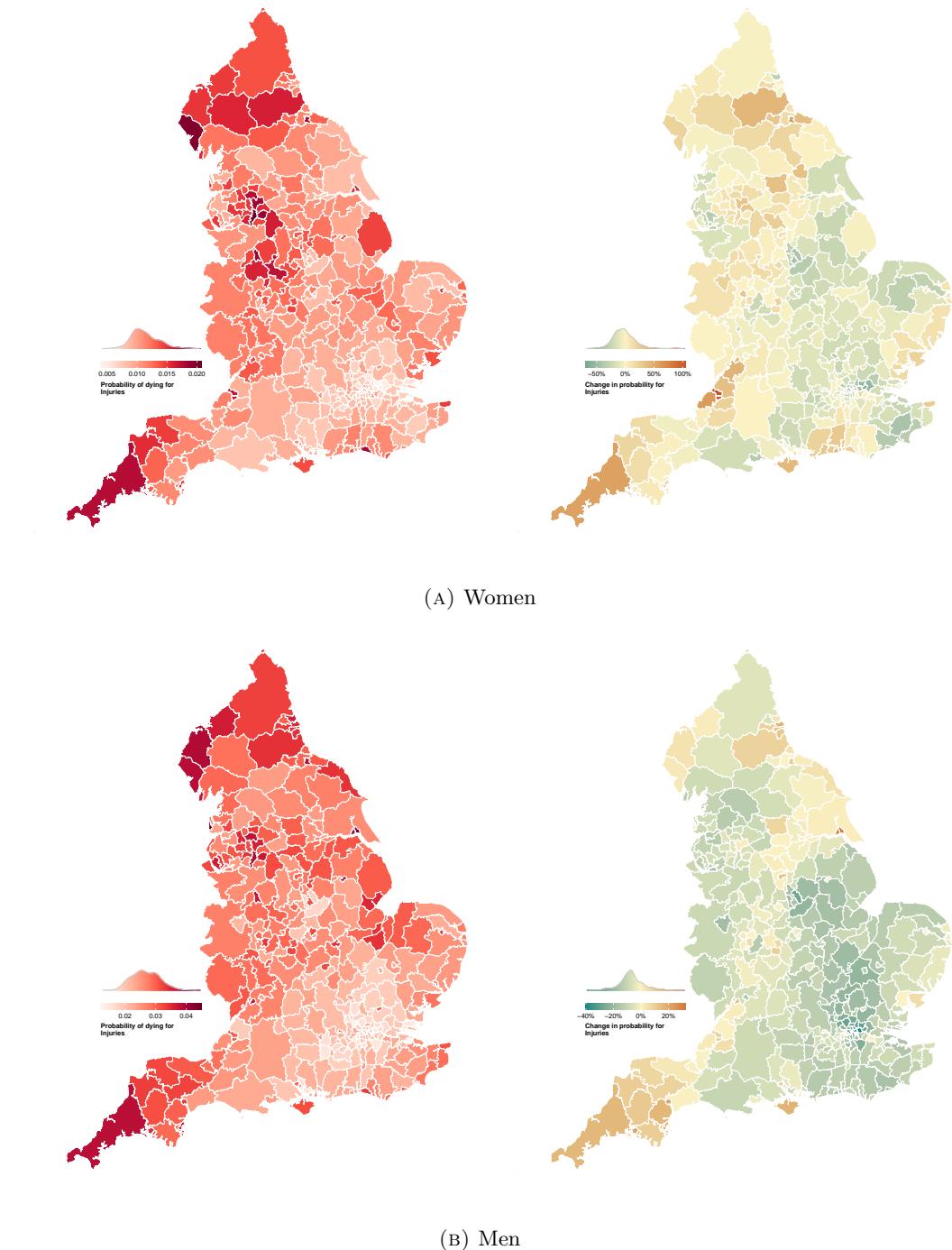


FIGURE 6.12: Probability of dying from injuries between birth and 80 years of age in 314 local authority districts in England in 2019 and change from 2002 to 2019.

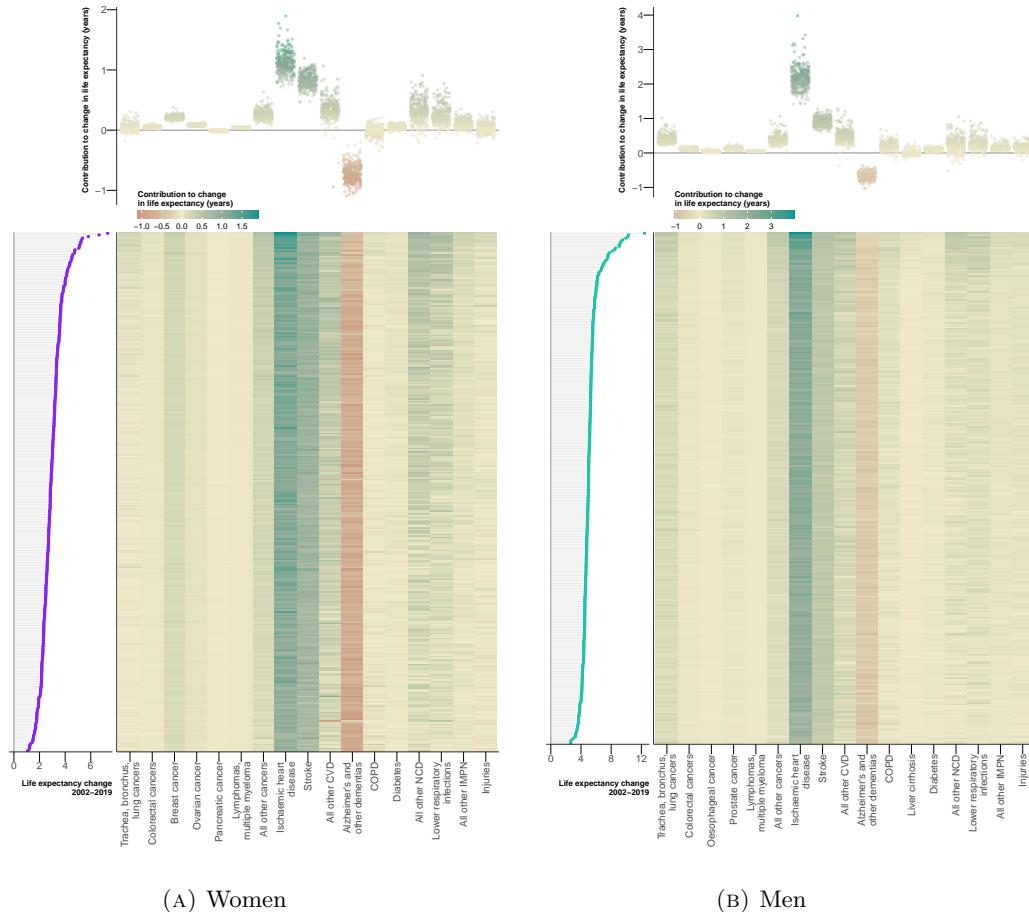


FIGURE 6.13: Contributions of deaths from specific causes to life expectancy difference between the district population in 2002 and the population in 2019 in 314 local authority districts in England. The side panel shows the change in life expectancy over that time, by summing over the causes. The top panel shows the contribution for each cause.

for the other cancer groups was similar for the two time periods. There were a larger number of districts with positive contributions for all other NCDs in the second period. The positive contribution from all the CVD groups (ischaemic heart disease, stroke, all other CVD) was weaker for the second time period as the negative forcing effect of Alzheimer's and other dementias strengthened massively. In the years 2010 to 2019, the *inequality* in progress between districts was driven largely by Alzheimer's and all other NCDs, both with a difference in contribution between the 10th and 90th percentile districts of 3.0 years.

For men, the contribution for all cancers groups, including lung cancer, maintained the same picture in each time period. As for women, the slowdown in progress was driven by reductions in the contribution from CVDs in the second time period and the strong negative contributions from Alzheimer's and other dementias. Reductions in CVD slowing on average – not uniform. Still increasing in lowest districts Injuries, which contributed positively to life expectancy increases in 2002-10 for all districts, saw a reversal with all but 19 districts contributing negatively. In the latter period, ischaemic heart disease, Alzheimer's and other dementias, and all other NCDs were driving the inequality in progress between districts, with differences in contribution between the 10th and 90th percentile districts of 2.8, 2.9 and 3.0 years respectively.

## 6.4 Discussion

Huge inequalities in some causes in 2019 Worrying increases in others Found slowing progress driven by these causes

### 6.4.1 Comparison with previous literature

There are two key studies in the literature which break the population of England down into subgroups and look at cause-specific Comparison to Bennett 2018. Deciles, largely similar groups of death Comparison to GBD 2015. 150 upper tier local authorities, rather than districts. Broader in space, but using 264 GBD causes of death, more sharing and reliance on covariates in the model What is causing the epi transition backward that we saw at the MSOA level Did not look at gap in inequality in progress, instead focussed on the gap in inequality

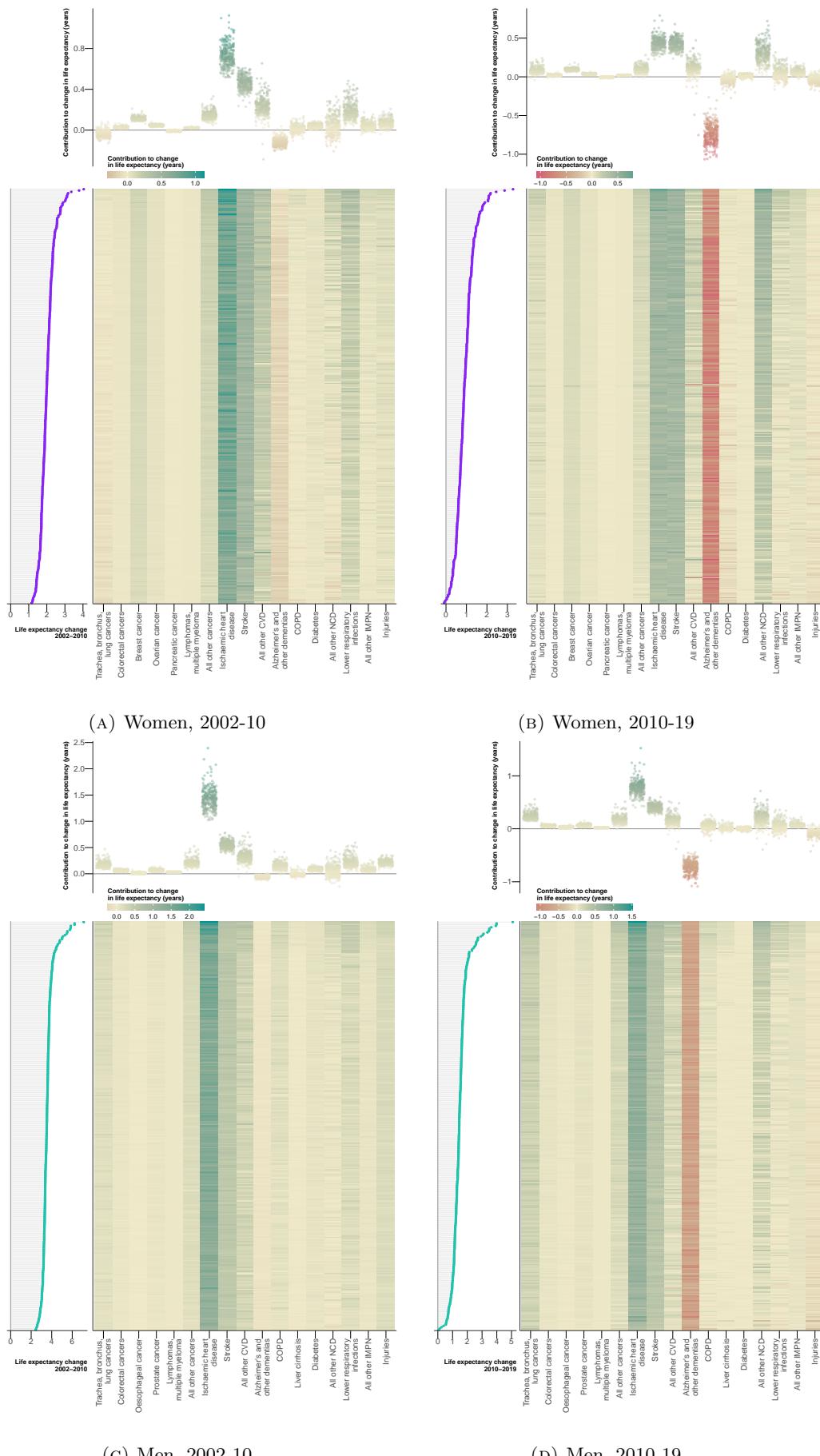


FIGURE 6.14: Contributions of deaths from specific causes to life expectancy difference between the district populations in 2002 and 2010, and in 2010 and 2019, in 314 districts in England. The side panel shows the change in life expectancy over that time, by summing over the causes. The top panel shows the contribution for each districts for each cause.

### 6.4.2 Strengths and limitations

District level rather than MSOA level Numbers get too small

Collapsed over age groups, unlike Bennett 2018 Skew towards older ages, due to how we chose the groups on number of deaths, which itself is skewed towards older ages

Split into further causes. Maternal, perinatal All other NCD Injuries – intentional or unintentional. Much attention on on deaths of despair Colin Angus, Angus Deaton etc. Risk factor that has remained constant since 2000s Groups already quite small and smaller numbers

Migration not as much of an issue Today it remains the case that the majority of moves occur over short distances: the 2011 Census records that 57.1% of the individuals aged 16 and over that changed address within the preceding 12 months moved within the same Local Authority District (LAD).

## 6.5 Summary



## Chapter 7

# And when they died from cancer: trends in cancer mortality at the district level in England

### 7.1 Overview

### 7.2 Methods

The methodology for this section if the same as Chapter 6, but I have stratified cancer groups further. I used the top ten leading cancer causes of death according to the total number of deaths from 2002 to 2019 for cause-specific analysis, as well as a residual group comprising all other cancer deaths. As a result, the residual group in this chapter is smaller than in Chapter 6. The full list of ICD-10 codes for each cause group can be found in Table D.2.

I ran a model for a combined group of all cancer deaths, which resulted in near-identical median estimates of probability of dying to those created by summing the age-district-year-cause specific death rates over all the cancer groups (correlation coefficient across all years 0.99 for both sexes). I present results from the model as the credible intervals are directly estimated.

The primary reporting outcome is, again, the unconditional probability of dying between birth and 80 years of age. I also report the mean age at death among those who died of that cause. Details on the calculations for both the probability of dying and the mean age at death can be found in Appendix A.

### **7.3 Results**

### **7.4 Discussion**

### **7.5 Summary**

## Chapter 8

# Discussion

### 8.1 Comparison with published literature

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

### 8.2 Strengths and limitations

Bayesian methods Highest resolution in a joint model, with error estimates

### 8.3 Public health and policy implications

### 8.4 Future work

Substantive and methodological

Only care about LHS of equation. Flexible models, such as Gaussian processes, which use inductive biases and knowledge, or wilder like Very familiar with research, was on a paper But ultimately, the working environment was not conducive to methods development Cluster offline, only a command line This is health data – Understand

why people still use Scottish lip cancer and reference data A lot of time spent scaling the models Future research can incorporate new models Joint likelihood for causes (e.g Best et al.), using multiple surface, or something more flexible like Kronecker GP Packages don't exist yet, but should be there in near future

Since 2019, Covid Produce a forecast of mortality Run the counterfactual, as we did in Kontis at the weekly national level, but for district level and annual See how the cause-composition changed Obviously increase in infectious disease, but were injuries reduced? Were cancer outcomes worse due to missed surgeries? Is there a longer term effect due to strain on emergency services?

## **8.5 Conclusions**

# References

- 2011 Census geographies. 2022. *Office for National Statistics*.
- Alexiou A, Fahy K, Mason K, Bennett D, Brown H, Bambra C, Taylor-Robinson D, Barr B. 2021. Local government funding and life expectancy in England: A longitudinal ecological study. *The Lancet Public Health* **6**:e641–e647. doi:[10.1016/S2468-2667\(21\)00110-9](https://doi.org/10.1016/S2468-2667(21)00110-9)
- Arriaga EE. 1984. Measuring and Explaining the Change in Life Expectancies. *Demography* **21**:83–96. doi:[10.2307/2061029](https://doi.org/10.2307/2061029)
- Asaria P, Fortunato L, Fecht D, Tzoulaki I, Abellan JJ, Hambly P, de Hoogh K, Ezzati M, Elliott P. 2012. Trends and inequalities in cardiovascular disease mortality across 7932 English electoral wards, 1982–2006: Bayesian spatial analysis. *International Journal of Epidemiology* **41**:1737–1749. doi:[10.1093/ije/dys151](https://doi.org/10.1093/ije/dys151)
- Aylin P, Maheswaran R, Wakefield J, Cockings S, Jarup L, Arnold R, Wheeler G, Elliott P. 1999. A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: The UK Small Area Health Statistics Unit. *Journal of Public Health* **21**:289–298. doi:[10.1093/pubmed/21.3.289](https://doi.org/10.1093/pubmed/21.3.289)
- Bahk J, Kang H-Y, Khang Y-H. 2020. Life expectancy and inequalities therein by income from 2016 to 2018 across the 253 electoral constituencies of the National Assembly of the Republic of Korea. *Journal of Preventive Medicine and Public Health* **53**:143–148. doi:[10.3961/jpmph.20.050](https://doi.org/10.3961/jpmph.20.050)
- Barr B, Higgerson J, Whitehead M. 2017. Investigating the impact of the English health inequalities strategy: Time trend analysis. *BMJ* **358**:j3310. doi:[10.1136/bmj.j3310](https://doi.org/10.1136/bmj.j3310)
- Beltrán-Sánchez H, Finch CE, Crimmins EM. 2015. Twentieth century surge of excess

- adult male mortality. *Proceedings of the National Academy of Sciences* **112**:8993–8998. doi:[10.1073/pnas.1421942112](https://doi.org/10.1073/pnas.1421942112)
- Bennett JE, Li G, Foreman K, Best N, Kontis V, Pearson C, Hambly P, Ezzati M. 2015. The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *The Lancet* **386**:163–170. doi:[10.1016/S0140-6736\(15\)60296-3](https://doi.org/10.1016/S0140-6736(15)60296-3)
- Bennett JE, Pearson-Stuttard J, Kontis V, Capewell S, Wolfe I, Ezzati M. 2018. Contributions of diseases and injuries to widening life expectancy inequalities in England from 2001 to 2016: A population-based analysis of vital registration data. *The Lancet Public Health* **3**:e586–e597. doi:[10.1016/S2468-2667\(18\)30214-7](https://doi.org/10.1016/S2468-2667(18)30214-7)
- Bennett JE, Rashid T, Zolfaghari A, Doyle Y, Suel E, Pearson-Stuttard J, Davies B, Fecht D, Muller ES, Nathvani RS, Sportiche N, Daby HI, Johnson E, Li G, Flaxman S, Toledano MB, Asaria M, Ezzati M. 2023. Changes in life expectancy and house prices in London from 2002 to 2019: Hyper-resolution spatiotemporal analysis of death registration and real estate data. *The Lancet Regional Health – Europe* **27**. doi:[10.1016/j.lanepe.2022.100580](https://doi.org/10.1016/j.lanepe.2022.100580)
- Besag J, York J, Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**:1–20. doi:[10.1007/BF00116466](https://doi.org/10.1007/BF00116466)
- Best N, Richardson S, Thomson A. 2005. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* **14**:35–59. doi:[10.1191/0962280205sm388oa](https://doi.org/10.1191/0962280205sm388oa)
- Bilal U, Alazraqui M, Caiaffa WT, Lopez-Olmedo N, Martinez-Folgar K, Miranda JJ, Rodriguez DA, Vives A, Diez-Roux AV. 2019. Inequalities in life expectancy in six large Latin American cities from the SALURBAL study: An ecological analysis. *The Lancet Planetary Health* **3**:e503–e510. doi:[10.1016/S2542-5196\(19\)30235-9](https://doi.org/10.1016/S2542-5196(19)30235-9)
- Blachier M, Leleu H, Peck-Radosavljevic M, Valla D-C, Roudot-Thoraval F. 2013. The burden of liver disease in Europe: A review of available epidemiological data. *Journal of Hepatology* **58**:593–608. doi:[10.1016/j.jhep.2012.12.005](https://doi.org/10.1016/j.jhep.2012.12.005)
- Blundell R, Costa Dias M, Joyce R, Xu X. 2020. COVID-19 and Inequalities. *Fiscal Studies* **41**:291–319. doi:[10.1111/1475-5890.12232](https://doi.org/10.1111/1475-5890.12232)

- Bosetti N. 2015. Inside out: The new geography of wealth and poverty in London. Centre for London.
- Britain has endured a decade of early deaths. Why? 2023. *The Economist* 19–21.
- Burn-Murdoch J. 2023. Why are Americans dying so young? *Financial Times*.
- Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D. 2016. The Association Between Income and Life Expectancy in the United States, 2001–2014. *JAMA* **315**:1750–1766. doi:[10.1001/jama.2016.4226](https://doi.org/10.1001/jama.2016.4226)
- Coale AJ, Demeny P, Vaughan B. 1983. Regional Model Life Tables and Stable Populations, Second. ed. Academic Press.
- Congdon P. 2019. Geographical Patterns in Drug-Related Mortality and Suicide: Investigating Commonalities in English Small Areas. *International Journal of Environmental Research and Public Health* **16**:1831. doi:[10.3390/ijerph16101831](https://doi.org/10.3390/ijerph16101831)
- Congdon P. 2014. Modelling changes in small area disability free life expectancy: Trends in London wards between 2001 and 2011. *Statistics in Medicine* **33**:5138–5150. doi:[10.1002/sim.6298](https://doi.org/10.1002/sim.6298)
- Congdon P. 2009. Life Expectancies for Small Areas: A Bayesian Random Effects Methodology. *International Statistical Review* **77**:222–240. doi:[10.1111/j.1751-5823.2009.00080.x](https://doi.org/10.1111/j.1751-5823.2009.00080.x)
- Connolly S, O'Reilly D, Rosato M. 2007. Increasing inequalities in health: Is it an artefact caused by the selective movement of people? *Social Science & Medicine* **64**:2008–2015. doi:[10.1016/j.socscimed.2007.02.021](https://doi.org/10.1016/j.socscimed.2007.02.021)
- Davenport A, Zarenko B. 2020. Levelling up: Where and how? IFS Green Budget 2020. Institute for Fiscal Studies.
- de Valpine P, Paciorek CJ, Turek D, Michaud N, Anderson-Bergman C, Obermayer F, Wehrhahn Cortes C, Rodríguez A, Lang DT, Paganin S, Hug J. 2022. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling.
- de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. 2017. Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *Journal of Computational and Graphical Statistics* **26**:403–413. doi:[10.1080/10618600.2016.1172487](https://doi.org/10.1080/10618600.2016.1172487)
- Dorling D. 2023. How austerity caused the NHS crisis. *openDemocracy*.
- Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Mackenbach JP, van

- Lenthe FJ, Mokdad AH, Murray CJL. 2017a. Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA Internal Medicine* **177**:1003–1011. doi:[10.1001/jamainternmed.2017.0918](https://doi.org/10.1001/jamainternmed.2017.0918)
- Dwyer-Lindgren L, Stubbs RW, Bertozzi-Villa A, Morozoff C, Callender C, Finegold SB, Shirude S, Flaxman AD, Laurent A, Kern E, Duchin JS, Fleming D, Mokdad AH, Murray CJL. 2017b. Variation in life expectancy and mortality by cause among neighbourhoods in King County, WA, USA, 1990–2014: A census tract-level analysis for the Global Burden of Disease Study 2015. *The Lancet Public Health* **2**:e400–e410. doi:[10.1016/S2468-2667\(17\)30165-2](https://doi.org/10.1016/S2468-2667(17)30165-2)
- Eayres D, Williams ES. 2004. Evaluation of methodologies for small area life expectancy estimation. *Journal of Epidemiology & Community Health* **58**:243–249. doi:[10.1136/jech.2003.009654](https://doi.org/10.1136/jech.2003.009654)
- Elliott P, Briggs D, Morris S, Hoogh C de, Hurt C, Jensen TK, Maitland I, Richardson S, Wakefield J, Jarup L. 2001a. Risk of adverse birth outcomes in populations living near landfill sites. *BMJ* **323**:363–368. doi:[10.1136/bmj.323.7309.363](https://doi.org/10.1136/bmj.323.7309.363)
- Elliott P, Toledano MB, Bennett JE, Beale L, Hoogh C de, Best N, Briggs D. 2010. Mobile phone base stations and early childhood cancers: Case-control study. *BMJ* **340**:c3077. doi:[10.1136/bmj.c3077](https://doi.org/10.1136/bmj.c3077)
- Elliott P, Wakefield J, Best N, Briggs D. 2001b. Spatial epidemiology: Methods and applications. Oxford University Press.
- Elliott P, Westlake AJ, Hills M, Kleinschmidt I, Rodrigues L, McGale P, Marshall K, Rose G. 1992. The Small Area Health Statistics Unit: A national facility for investigating health around point sources of environmental pollution in the United Kingdom. *Journal of Epidemiology & Community Health* **46**:345–349. doi:[10.1136/jech.46.4.345](https://doi.org/10.1136/jech.46.4.345)
- English indices of deprivation 2019. 2019. *Ministry of Housing, Communities & Local Government*.
- Ezzati M, Friedman AB, Kulkarni SC, Murray CJL. 2008. The Reversal of Fortunes: Trends in County Mortality and Cross-County Mortality Disparities in the United States. *PLOS Medicine* **5**:e66. doi:[10.1371/journal.pmed.0050066](https://doi.org/10.1371/journal.pmed.0050066)
- Finucane MM, Paciorek CJ, Danaei G, Ezzati M. 2014. Bayesian Estimation of

- Population-Level Trends in Measures of Health Status. *Statistical Science* **29**:18–25. doi:[10.1214/13-STS427](https://doi.org/10.1214/13-STS427)
- Foreman KJ, Li G, Best N, Ezzati M. 2017. Small area forecasts of cause-specific mortality: Application of a Bayesian hierarchical model to US vital registration data. *Journal of the Royal Statistical Society Series C (Applied Statistics)* **66**:121–139.
- Gaylin DS, Kates J. 1997. Refocusing the lens: Epidemiologic transition theory, mortality differentials, and the AIDS pandemic. *Social Science & Medicine* **44**:609–621. doi:[10.1016/S0277-9536\(96\)00212-2](https://doi.org/10.1016/S0277-9536(96)00212-2)
- Gersten O, Wilmoth JR. 2002. The Cancer Transition in Japan since 1951. *Demographic Research* **7**:271–306. doi:[10.4054/DemRes.2002.7.5](https://doi.org/10.4054/DemRes.2002.7.5)
- Goodman PS. 2018. In Britain, Austerity Is Changing Everything. *The New York Times*.
- Gray M, Barford A. 2018. The depths of the cuts: The uneven geography of local government austerity. *Cambridge Journal of Regions, Economy and Society* **11**:541–563. doi:[10.1093/cjres/rsy019](https://doi.org/10.1093/cjres/rsy019)
- Halonen JI, Hansell AL, Gulliver J, Morley D, Blangiardo M, Fecht D, Toledano MB, Beevers SD, Anderson HR, Kelly FJ, Tonne C. 2015. Road traffic noise is associated with increased cardiovascular morbidity and mortality and all-cause mortality in London. *European Heart Journal* **36**:2653–2661. doi:[10.1093/eurheartj/ehv216](https://doi.org/10.1093/eurheartj/ehv216)
- Hansell AL, Blangiardo M, Fortunato L, Floud S, Hoogh K de, Fecht D, Ghosh RE, Laszlo HE, Pearson C, Beale L, Beevers S, Gulliver J, Best N, Richardson S, Elliott P. 2013. Aircraft noise and cardiovascular disease near Heathrow airport in London: Small area study. *BMJ* **347**:f5432. doi:[10.1136/bmj.f5432](https://doi.org/10.1136/bmj.f5432)
- Hansell, Anna L., Beale, Linda, Ghosh, Rebecca E., Fortunato, Lea, Fecht, Daniela, Jarup, Lars, Elliott, Paul. 2014. The Environment and Health Atlas for England and Wales. Oxford University Press.
- Health Expectancies at Birth for Middle Layer Super Output Areas (MSOAs), England. 2015. *Office for National Statistics*.
- Hiam L, Harrison D, McKee M, Dorling D. 2018. Why is life expectancy in England and Wales “stalling”? *J Epidemiol Community Health* **72**:404–408. doi:[10.1136/jech-2017-210401](https://doi.org/10.1136/jech-2017-210401)

- Hodgson S, Nieuwenhuijsen MJ, Colvile R, Jarup L. 2007. Assessment of exposure to mercury from industrial emissions: Comparing “distance as a proxy” and dispersion modelling approaches. *Occupational and Environmental Medicine* **64**:380–388. doi:[10.1136/oem.2006.026781](https://doi.org/10.1136/oem.2006.026781)
- Hodgson S, Nieuwenhuijsen MJ, Hansell A, Shepperd S, Flute T, Staples B, Elliott P, Jarup L. 2004. Excess risk of kidney disease in a population living near industrial plants. *Occupational and Environmental Medicine* **61**:717–719. doi:[10.1136/oem.2003.010629](https://doi.org/10.1136/oem.2003.010629)
- Jarup L, Best N, Toledano MB, Wakefield J, Elliott P. 2002a. Geographical epidemiology of prostate cancer in Great Britain. *International Journal of Cancer* **97**:695–699. doi:[10.1002/ijc.10113](https://doi.org/10.1002/ijc.10113)
- Jarup L, Briggs D, de Hoogh C, Morris S, Hurt C, Lewin A, Maitland I, Richardson S, Wakefield J, Elliott P. 2002b. Cancer risks in populations living near landfill sites in Great Britain. *British Journal of Cancer* **86**:1732–1736. doi:[10.1038/sj.bjc.6600311](https://doi.org/10.1038/sj.bjc.6600311)
- Jarup L, Morris S, Richardson S, Briggs D, Cobley N, de Hoogh C, Gorog K, Elliott P. 2007. Down syndrome in births near landfill sites. *Prenatal Diagnosis* **27**:1191–1196. doi:[10.1002/pd.1873](https://doi.org/10.1002/pd.1873)
- Jonker MF, van Lenthe FJ, Congdon PD, Donkers B, Burdorf A, Mackenbach JP. 2012. Comparison of Bayesian Random-Effects and Traditional Life Expectancy Estimations in Small-Area Applications. *American Journal of Epidemiology* **176**:929–937. doi:[10.1093/aje/kws152](https://doi.org/10.1093/aje/kws152)
- Kelsall J, Wakefield J. 2002. Modeling Spatial Variation in Disease Risk. *Journal of the American Statistical Association* **97**:692–701. doi:[10.1198/016214502388618438](https://doi.org/10.1198/016214502388618438)
- Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. 2017. Future life expectancy in 35 industrialised countries: Projections with a Bayesian model ensemble. *The Lancet* **389**:1323–1335. doi:[10.1016/S0140-6736\(16\)32381-9](https://doi.org/10.1016/S0140-6736(16)32381-9)
- Kontis V, Bennett JE, Rashid T, Parks RM, Pearson-Stuttard J, Guillot M, Asaria P, Zhou B, Battaglini M, Corsetti G, McKee M, Di Cesare M, Mathers CD, Ezzati M. 2020. Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature Medicine* **26**:1919–1928. doi:[10.1038/s41591-020-1112-0](https://doi.org/10.1038/s41591-020-1112-0)

- Lao J, Suter C, Langmore I, Chimisov C, Saxena A, Sountsov P, Moore D, Saurous RA, Hoffman MD, Dillon JV. 2020. Tfp.mcmc: Modern Markov Chain Monte Carlo Tools Built for Modern Hardware. doi:[10.48550/arXiv.2002.01184](https://doi.org/10.48550/arXiv.2002.01184)
- Leon DA, Jdanov DA, Shkolnikov VM. 2019. Trends in life expectancy and age-specific mortality in England and Wales, 1970–2016, in comparison with a set of 22 high-income countries: An analysis of vital statistics data. *The Lancet Public Health* **4**:e575–e582. doi:[10.1016/S2468-2667\(19\)30177-X](https://doi.org/10.1016/S2468-2667(19)30177-X)
- Life expectancy at birth, data. 2022. *World Bank*.
- Local Alcohol Profiles for England. 2021. *Public Health England*.
- Local Health - Small Area Public Health Data. 2021. *Public Health England*.
- Lower layer Super Output Area population estimates. 2021. *Office for National Statistics*.
- Marmot MG, Allen J, Boyce T, Goldblatt P, Morrison J. 2020. Marmot Review: 10 years on. Institute of Health Equity.
- Marmot MG, Davey Smith G, Stansfeld S, Patel C, North F, Head J, White I, Brunner E, Feeney A. 1991. Health inequalities among British civil servants: The Whitehall II study. *The Lancet* **337**:1387–1393. doi:[10.1016/0140-6736\(91\)93068-K](https://doi.org/10.1016/0140-6736(91)93068-K)
- Marmot MG, Goldblatt P, Allen J, Boyce T, McNeish D, Grady M, Strelitz J, Geddes I, Friel S, Porritt F, Reinertsen E, Bell R, Allen M. 2010. The Marmot Review: Fair society, healthy lives. Institute of Health Equity.
- Marmot MG, Shipley MJ, Rose G. 1984. Inequalities in death - specific explanations of a general pattern? *The Lancet*, Originally published as Volume 1, Issue 8384 **323**:1003–1006. doi:[10.1016/S0140-6736\(84\)92337-7](https://doi.org/10.1016/S0140-6736(84)92337-7)
- Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD. 2005. Counting the dead and what they died from: An assessment of the global status of cause of death data. *Bulletin of the World Health Organization* **83**:171–177.
- McElreath R. 2020. Statistical Rethinking, Second. ed. CRC Press.
- Middle layer Super Output Area population estimates. 2021. *Office for National Statistics*.
- Moriyama IM, Gover M. 1948. Statistical Studies of Heart Diseases: I. Heart Diseases and Allied Causes of Death in Relation to Age Changes in the Population. *Public Health Reports (1896-1970)* **63**:537–545. doi:[10.2307/4586527](https://doi.org/10.2307/4586527)

- Murray CJL, Lopez AD. 1996. The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected in 2020. World Health Organization.
- Olshansky SJ, Ault AB. 1986. The Fourth Stage of the Epidemiologic Transition: The Age of Delayed Degenerative Diseases. *The Milbank Quarterly* **64**:355–391.  
doi:[10.2307/3350025](https://doi.org/10.2307/3350025)
- Omran AR. 1977. A century of epidemiologic transition in the United States. *Preventive Medicine* **6**:30–51. doi:[10.1016/0091-7435\(77\)90003-2](https://doi.org/10.1016/0091-7435(77)90003-2)
- Omran AR. 1971. The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *The Milbank Memorial Fund Quarterly* **49**:509–538.  
doi:[10.2307/3349375](https://doi.org/10.2307/3349375)
- Overman HG, Xu X. 2022. Spatial disparities across labour markets. Institute for Fiscal Studies.
- Parkes B, Hansell AL, Ghosh RE, Douglas P, Fecht D, Wellesley D, Kurinczuk JJ, Rankin J, de Hoogh K, Fuller GW, Elliott P, Toledano MB. 2020. Risk of congenital anomalies near municipal waste incinerators in England and Scotland: Retrospective population-based cohort study. *Environment International* **134**:104845.  
doi:[10.1016/j.envint.2019.05.039](https://doi.org/10.1016/j.envint.2019.05.039)
- Pearson-Stuttard J, Bennett JE, Cheng YJ, Vamos EP, Cross AJ, Ezzati M, Gregg EW. 2021. Trends in predominant causes of death in individuals with and without diabetes in England from 2001 to 2018: An epidemiological analysis of linked primary care records. *The Lancet Diabetes & Endocrinology* **9**:165–173.  
doi:[10.1016/S2213-8587\(20\)30431-9](https://doi.org/10.1016/S2213-8587(20)30431-9)
- Phan D, Pradhan N, Jankowiak M. 2019. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. doi:[10.48550/arXiv.1912.11554](https://arxiv.org/abs/1912.11554)
- Population estimates by output areas, electoral, health and other geographies, England and Wales. 2021. *Office for National Statistics*.
- Preston SH. 1970. An International Comparison of Excessive Adult Mortality. *Population Studies* **24**:5–20. doi:[10.2307/2173259](https://doi.org/10.2307/2173259)
- Preston SH, Heuveline P, Guillot M. 2001. Demography: Measuring and Modeling Population Processes. Blackwell Publishing.

- Preston SH, Nelson VE. 1974. Structure and change in causes of death: An international summary. *Population Studies* **28**:19–51. doi:[10.1080/00324728.1974.10404577](https://doi.org/10.1080/00324728.1974.10404577)
- Prior Choice Recommendations. 2020. *GitHub*.
- Rashid T. 2022. Probabilistic-programming-packages.
- Rashid T, Bennett JE, Paciorek CJ, Doyle Y, Pearson-Stuttard J, Flaxman S, Fecht D, Toledano MB, Li G, Daby HI, Johnson E, Davies B, Ezzati M. 2021. Life expectancy and risk of death in 6791 communities in England from 2002 to 2019: High-resolution spatiotemporal analysis of civil registration data. *The Lancet Public Health* **6**:e805–e816. doi:[10.1016/S2468-2667\(21\)00205-X](https://doi.org/10.1016/S2468-2667(21)00205-X)
- Salomon JA, Murray CJL. 2002. [The Epidemiologic Transition Revisited: Compositional Models for Causes of Death by Age and Sex](#). *Population and Development Review* **28**:205–228.
- Silcock PBS, Jenner DA, Reza R. 2001. Life expectancy as a summary of mortality in a population: Statistical considerations and suitability for use by health authorities. *Journal of Epidemiology & Community Health* **55**:38–43. doi:[10.1136/jech.55.1.38](https://doi.org/10.1136/jech.55.1.38)
- Steel N, Ford JA, Newton JN, Davis ACJ, Vos T, Naghavi M, Glenn S, Hughes A, Dalton AM, Stockton D, Humphreys C, Dallat M, Schmidt J, Flowers J, Fox S, Abubakar I, Aldridge RW, Baker A, Brayne C, Brugha T, Capewell S, Car J, Cooper C, Ezzati M, Fitzpatrick J, Greaves F, Hay R, Hay S, Kee F, Larson HJ, Lyons RA, Majeed A, McKee M, Rawaf S, Rutter H, Saxena S, Sheikh A, Smeeth L, Viner RM, Vollset SE, Williams HC, Wolfe C, Woolf A, Murray CJL. 2018. Changes in health in the countries of the UK and 150 English Local Authority areas 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **392**:1647–1661. doi:[10.1016/S0140-6736\(18\)32207-4](https://doi.org/10.1016/S0140-6736(18)32207-4)
- Stephens AS, Purdie S, Yang B, Moore H. 2013. Life expectancy estimation in small administrative areas with non-uniform population sizes: Application to Australian New South Wales local government areas. *BMJ Open* **3**:e003710. doi:[10.1136/bmjopen-2013-003710](https://doi.org/10.1136/bmjopen-2013-003710)
- Taylor-Robinson D, Lai ETC, Wickham S, Rose T, Norman P, Bambra C, Whitehead M, Barr B. 2019. Assessing the impact of rising child poverty on the unprecedented rise in infant mortality in England, 2000–2017: Time trend analysis. *BMJ Open*

- 9:e029424. doi:[10.1136/bmjopen-2019-029424](https://doi.org/10.1136/bmjopen-2019-029424)
- Thatcher AR, Kannisto V, Andreev KF. 2002. The Survivor Ratio Method for Estimating Numbers at High Ages. *Demographic Research* **6**:1–18. doi:[10.4054/DemRes.2002.6.1](https://doi.org/10.4054/DemRes.2002.6.1)
- Toledano MB, Jarup L, Best N, Wakefield J, Elliott P. 2001. Spatial variation and temporal trends of testicular cancer in Great Britain. *British Journal of Cancer* **84**:1482–1487. doi:[10.1054/bjoc.2001.1739](https://doi.org/10.1054/bjoc.2001.1739)
- Travers T, Sims S, Bosetti N. 2016. Housing and inequality in London. Centre for London.
- User guide to mortality statistics. 2022. *Office for National Statistics*.
- van Dijk JT, Lansley G, Longley PA. 2021. Using linked consumer registers to estimate residential moves in the United Kingdom. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **184**:1452–1474. doi:[10.1111/rssa.12713](https://doi.org/10.1111/rssa.12713)
- Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. 2021. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. *Bayesian Analysis* **16**. doi:[10.1214/20-BA1221](https://doi.org/10.1214/20-BA1221)
- Wakefield J, Elliott P. 1999. Issues in the statistical analysis of small area health data. *Statistics in Medicine* **18**:2377–2399. doi:[10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18<2377::AID-SIM263>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2377::AID-SIM263>3.0.CO;2-G)
- WHO methods and data sources for country-level causes of death 2000–2019. 2020.
- Wilkinson RG. 1992. Income distribution and life expectancy. *British Medical Journal* **304**:165–168. doi:[10.1136/bmj.304.6820.165](https://doi.org/10.1136/bmj.304.6820.165)
- Yu J, Dwyer-Lindgren L, Bennett J, Ezzati M, Gustafson P, Tran M, Brauer M. 2021. A spatiotemporal analysis of inequalities in life expectancy and 20 causes of mortality in sub-neighbourhoods of Metro Vancouver, British Columbia, Canada, 1990–2016. *Health & Place* **72**:102692. doi:[10.1016/j.healthplace.2021.102692](https://doi.org/10.1016/j.healthplace.2021.102692)

## Appendix A

# Life table methods

Most of the content in this appendix is taken directly from Preston et al. (2001), but I have reproduced it here for reference and for completeness.

### A.1 Period life tables

Calculating life expectancy for a cohort is possible, but you have to wait until every member of the cohort has died. Instead, demographers use period (or “current”) life tables, which consider what would happen to a hypothetical cohort that are subjected to the death rates in each age group at an exact period in time. Life tables can be constructed using discrete age bands starting at age  $x$  and ending at age  $x + n$ . We supply the age-specific death rates,  ${}_n m_x$ , and the average person-years lived by those dying in the interval,  ${}_n a_x$ , and the life table calculates the mean age at death – the life expectancy,  $e_x$ .

We start with a hypothetical cohort of size  $l_0 = 100,000$  and sequentially apply the probability of dying in each age group, calculated as

$${}_n q_x = \frac{n \cdot {}_n m_x}{1 + (n - {}_n a_x) {}_n m_x}. \quad (\text{A.1})$$

The open interval  ${}_\infty q_x = 1$ , as nobody is immortal. Using the probability of surviving in each age group,  ${}_n p_x = 1 - {}_n q_x$ , the number of survivors is given by

$$l_{x+n} = l_x \cdot {}_n p_x. \quad (\text{A.2})$$

The number of person-years lived is the sum of the number of survivors weighted by the band width and number of people who died weighted by  ${}_n a_x$

$${}_n L_x = n l_x \cdot + {}_n a_x l_{x n} q_x \quad {}_\infty L_x = \frac{l_x}{{}_\infty m_x}, \quad (\text{A.3})$$

and the total number of person-years lived above  $x$  is

$$T_x = \sum_{x=a}^{\infty} {}_n L_x. \quad (\text{A.4})$$

Then, life expectancy is given by diving the number of person-years lived by the number of people who will live them

$$e_x = \frac{T_x}{l_x}. \quad (\text{A.5})$$

Throughout the thesis, I only consider life expectancy at birth.

### A.1.1 The very young ages and the very old ages

On average, it's a good approximation is to assume deaths occur halfway through the age interval:  ${}_n a_x = n/2$ . But for younger ages, particularly at lower levels of mortality, the majority of infant deaths lie further towards the earliest stages of infancy. Coale and Demeny used regression on a series of international datasets to recommend suitable values for  ${}_1 a_0$  and  ${}_4 a_1$  instead of the midpoint (Coale et al., 1983).

The start of the open age group can be many years away from some of the ages at death, particularly in ageing populations. In order to produce reliable estimates of death rates at high ages, I used the Kannisto-Thatcher method to expand the terminal age group ( $\geq 85$  years) of the life table and adjust  ${}_n a_x$  above 70 years (Thatcher et al., 2002).

## A.2 Probability of dying

The probability of dying from a specific cause of death,  $i$ , is calculated as in Equation A.1. Equally, we can subtract the probability of surviving to that age group,  $1 - \prod_x n p_x^i$ . Note, even for the smallest death rates,  $\infty q_x^i = 1$  – if you live to infinity, you’ll die of it eventually.

## A.3 Cause-specific decomposition of differences in life expectancy

Arriaga (1984) proposed a method to calculate the age-specific contributions to the difference in life expectancy between two populations as

$${}_n\Delta_x = \frac{l_x^1}{l_0^1} \left( \frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^1} \right) + \frac{T_{x+n}^2}{l_0^1} \left( \frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2} \right). \quad (\text{A.6})$$

We then assume the age- and cause-specific contributions are proportional to the difference in cause-specific death rates:

$${}_n\Delta_x^i = {}_n\Delta_x \cdot \frac{{}_nm_x^i(2) - {}_nm_x^i(1)}{{}_nm_x(2) - {}_nm_x(1)} \quad (\text{A.7})$$

Arriaga showed the sum of the age- and cause-specific contributions are equal to the difference in life expectancy,

$$e_0(2) - e_0(1) = \sum_x {}_n\Delta_x = \sum_x \sum_i {}_n\Delta_x^i. \quad (\text{A.8})$$

So, we can collapse over age groups to get the cause-specific contributions to life expectancy as  $\sum_x {}_n\Delta_x^i$ .

#### A.4 Mean age at death

The mean age at death among those who died from a specific cause of death was calculated as

$$\text{mean age at death} = \frac{\sum_x {}_n d_x \cdot {}_n a_x}{\sum_x {}_n d_x}, \quad (\text{A.9})$$

where  ${}_n d_x$  is the number of deaths in an age band, calculated as the product of the death rate and the population.

## Appendix B

# Model parameters and priors

This appendix contains full information on the parameters of each model.

TABLE B.1: Specification of the Bayesian statistical model in Equation 4.3.

Parameter name	Symbol	Prior	Dimension
Overall intercept	$\alpha_0$	$\mathcal{N}(0, \sigma^2 = 10^5)$	1
Overall slope	$\beta_0$	$\mathcal{N}(0, \sigma^2 = 10^5)$	1
Regional intercept	$\alpha_{1r}$	$\mathcal{N}(0, \sigma_{\alpha_{1r}}^2)$	9
Regional intercept standard deviation	$\sigma_{\alpha_{1r}}$	$\mathcal{U}(0, 2)$	1
District intercept	$\alpha_{1d}$	$\mathcal{N}(0, \sigma_{\alpha_{1d}}^2)$	314
District intercept standard deviation	$\sigma_{\alpha_{1d}}$	$\mathcal{U}(0, 2)$	1
MSOA intercept	$\alpha_{1s}$	$\mathcal{N}(0, \sigma_{\alpha_{1s}}^2)$	6791
MSOA intercept standard deviation	$\sigma_{\alpha_{1s}}$	$\mathcal{U}(0, 2)$	1
Regional slope	$\beta_{1r}$	$\mathcal{N}(0, \sigma_{\beta_{1r}}^2)$	9
Regional slope standard deviation	$\sigma_{\beta_{1r}}$	$\mathcal{U}(0, 2)$	1
District slope	$\beta_{1d}$	$\mathcal{N}(0, \sigma_{\beta_{1d}}^2)$	314
District slope standard deviation	$\sigma_{\beta_{1d}}$	$\mathcal{U}(0, 2)$	1

Parameter name	Symbol	Prior	Dimension
MSOA slope	$\beta_{1s}$	$\mathcal{N}(0, \sigma_{\beta_{1s}}^2)$	6791
MSOA slope standard deviation	$\sigma_{\beta_{1s}}$	$\mathcal{U}(0, 2)$	1
Age group intercept	$\alpha_{2a}$	$\mathcal{N}(\alpha_{2,a-1}, \sigma_{\alpha_{2a}}^2)$	18
Age group intercept standard deviation	$\sigma_{\alpha_{2a}}$	$\mathcal{U}(0, 2)$	1
Age group slope	$\beta_{2a}$	$\mathcal{N}(\beta_{2,a-1}, \sigma_{\beta_{2a}}^2)$	18
Age group slope standard deviation	$\sigma_{\beta_{2a}}$	$\mathcal{U}(0, 2)$	1
Age group MSOA interaction	$\xi_{as}$	$\mathcal{N}(0, \sigma_\xi^2)$	19 x 6791
Age group MSOA interaction standard deviation	$\sigma_\xi$	$\mathcal{U}(0, 2)$	1
MSOA random walk over time	$\nu_{st}$	$\mathcal{N}(\nu_{s,t-1}, \sigma_\nu^2)$	6791 x 17
MSOA random walk over time standard deviation	$\sigma_\nu$	$\mathcal{U}(0, 2)$	1
Age group random walk over time	$\gamma_{at}$	$\mathcal{N}(\gamma_{a,t-1}, \sigma_\gamma^2)$	19 x 17
Age group random walk over time standard deviation	$\sigma_\gamma$	$\mathcal{U}(0, 2)$	1
Overdispersion parameter	$r$	$\mathcal{U}(0, 50)$	1

TABLE B.2: Specification of the Bayesian statistical model in Equation 5.1.

Parameter name	Symbol	Prior	Dimension
Overall intercept	$\alpha_0$	$\mathcal{N}(0, \sigma^2 = 10^5)$	1
Overall slope	$\beta_0$	$\mathcal{N}(0, \sigma^2 = 10^5)$	1
District intercept	$\alpha_{1d}$	$\mathcal{N}(0, \sigma_{\alpha_{1d}}^2)$	33
District intercept	$\sigma_{\alpha_{1d}}$	$\mathcal{N}^+(1)$	1
standard deviation			
MSOA intercept	$\alpha_{1m}$	$\mathcal{N}(0, \sigma_{\alpha_{1m}}^2)$	983
MSOA intercept	$\sigma_{\alpha_{1m}}$	$\mathcal{N}^+(1)$	1
standard deviation			
LSOA intercept	$\alpha_{1s}$	$\mathcal{N}(0, \sigma_{\alpha_{1s}}^2)$	4835
LSOA intercept	$\sigma_{\alpha_{1s}}$	$\mathcal{N}^+(1)$	1
standard deviation			
District slope	$\beta_{1d}$	$\mathcal{N}(0, \sigma_{\beta_{1d}}^2)$	33
District slope	$\sigma_{\beta_{1d}}$	$\mathcal{N}^+(1)$	1
standard deviation			
MSOA slope	$\beta_{1m}$	$\mathcal{N}(0, \sigma_{\beta_{1m}}^2)$	983
MSOA slope standard deviation	$\sigma_{\beta_{1m}}$	$\mathcal{N}^+(1)$	1
LSOA slope	$\beta_{1s}$	$\mathcal{N}(0, \sigma_{\beta_{1s}}^2)$	4835
LSOA slope standard deviation	$\sigma_{\beta_{1s}}$	$\mathcal{N}^+(1)$	1
Age group intercept	$\alpha_{2a}$	$\mathcal{N}(\alpha_{2,a-1}, \sigma_{\alpha_{2a}}^2)$	18
Age group intercept	$\sigma_{\alpha_{2a}}$	$\mathcal{N}^+(1)$	1
standard deviation			
Age group slope	$\beta_{2a}$	$\mathcal{N}(\beta_{2,a-1}, \sigma_{\beta_{2a}}^2)$	18
Age group slope	$\sigma_{\beta_{2a}}$	$\mathcal{N}^+(1)$	1
standard deviation			
Age group LSOA interaction	$\xi_{as}$	$\mathcal{N}(0, \sigma_\xi^2)$	19 x 4835
interaction			

Parameter name	Symbol	Prior	Dimension
Age group LSOA interaction standard deviation	$\sigma_\xi$	$\mathcal{N}^+(1)$	1
LSOA random walk over time	$\nu_{st}$	$\mathcal{N}(\nu_{s,t-1}, \sigma_\nu^2)$	4835 x 17
LSOA random walk over time standard deviation	$\sigma_\nu$	$\mathcal{N}^+(1)$	1
Age group random walk over time	$\gamma_{at}$	$\mathcal{N}(\gamma_{a,t-1}, \sigma_\gamma^2)$	19 x 17
Age group random walk over time standard deviation	$\sigma_\gamma$	$\mathcal{N}^+(1)$	1
Overdispersion parameter	$\rho$	$\exp(0.1)$	1

TABLE B.3: Specification of the Bayesian statistical model in Equation 6.1.

Parameter name	Symbol	Prior	Dimension
Overall intercept	$\alpha_0$	$\mathcal{N}(0, 10)$	1
Overall slope	$\beta_0$	$\mathcal{N}(0, 10)$	1
District intercept	$\alpha_{1s}$	ICAR	314
District intercept standard deviation	$\sigma_{\alpha_{1s}}$	$\mathcal{N}^+(1)$	1
District slope	$\beta_{1s}$	ICAR	314
District slope standard deviation	$\sigma_{\beta_{1s}}$	$\mathcal{N}^+(1)$	1
Age group intercept	$\alpha_{2a}$	$\mathcal{N}(\alpha_{2,a-1}, \sigma_{\alpha_{2a}}^2)$	18
Age group intercept standard deviation	$\sigma_{\alpha_{2a}}$	$\mathcal{N}^+(1)$	1

---

Parameter name	Symbol	Prior	Dimension
Age group slope	$\beta_{2a}$	$\mathcal{N}(\beta_{2,a-1}, \sigma_{\beta_{2a}}^2)$	18
Age group slope standard deviation	$\sigma_{\beta_{2a}}$	$\mathcal{N}^+(1)$	1
Age group district interaction	$\xi_{as}$	$\mathcal{N}(0, \sigma_\xi^2)$	19 x 314
Age group district interaction standard deviation	$\sigma_\xi$	$\mathcal{N}^+(1)$	1
Age group random walk over time	$\gamma_{at}$	$\mathcal{N}(\gamma_{a,t-1}, \sigma_\gamma^2)$	19 x 17
Age group random walk over time standard deviation	$\sigma_\gamma$	$\mathcal{N}^+(1)$	1

---



## Appendix C

# Hex-cartogram of life expectancy in 2019 at the MSOA level

This is an alternative version of Figure 4.2, using a hexagon-based cartogram where every MSOA has equal size. Hence, London, which has a lot of small MSOAs, takes up a larger proportion of the map. For an interactive version of the original figure, see the [visualisation](#).

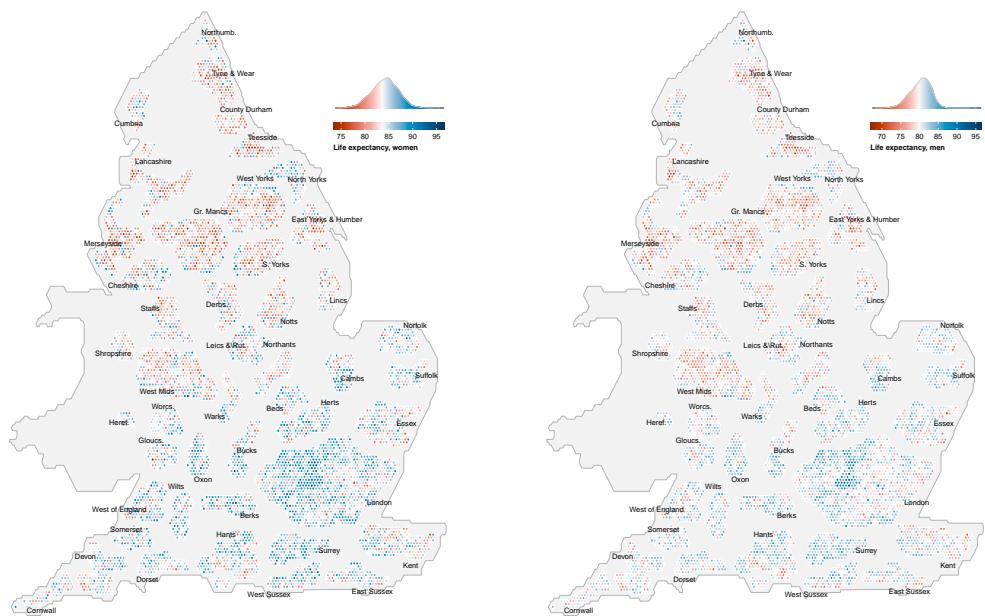


FIGURE C.1: Cartogram of life expectancy and the distribution of life expectancy in 2019. The areas in white have a life expectancy equal to the national life expectancy.

## Appendix D

# Groups of causes of death

TABLE D.1: Groups of causes of death used in the district-level cause-specific analysis in Chapter 6 with ICD-10 codes.

Cause	ICD-10 codes
Trachea, bronchus and lung cancers	C33-C34
Breast cancer	C50
Prostate cancer	C61
Colorectal cancers	C18-C21
Pancreatic cancer	C25
Ovarian cancer	C56
Lymphomas, multiple myeloma	C81-C90, C96
Oesophageal cancer	C15
All other cancers	C00-C14, C16-C17, C22-24, C26-C32, C37-C41, C43-C49, C51-C54, C57-C67, C48-C81, C91-C95, C97, D00-D48
Ischaemic heart disease	I20-I25
Stroke	I60-I69
All other CVD	I00-I19, I26-I59, I70-I99
Alzheimer's and other dementias	F00-F03, G30
COPD	J40-44

Cause	ICD-10 codes
Diabetes mellitus, nephritis and nephrosis	E10-14, N00-19
Liver cirrhosis	K70, K74
All other NCD	D55-D648, D65-D89, E03-E07, E15-E16, E20-E34, E65-E88, F01-F99, G06-G13, G15-G98, H00-H61, H68-H93, J30-J39, J45-J98, K00-K14, K20-K69, K71-K73, K75-K92, L00-L98, M00-M99, N20-64, N75-N99, Q00-Q99, R95, X41-X42, X45
Lower respiratory infections	J09-18, J20-J22
All other IMPN	A00-99, B00-99, D50-53, D649, E00-E02, E40-46, E50-54, G00, G03-G04, G14, H65-H66, N70-N73, J00-J06, O00-O99, P00-P96, Z353
Injuries	U00-U01, U509, V00-V99, W00-W99, X00-X40, X43-X44, X46-X99, Y00-Y01, Y10-Y36, Y381, Y40-Y86, Y870-Y872, Y88-Y89

Ovarian cancer (women) and prostate cancer (men) are sex specific. Liver cirrhosis was not in the top 12 leading causes of death for women, so it was included within all other NCD.

The residual groups (all other cancers, all other CVD, all other NCD, all other IMPN) also contained deaths from the “ill-defined diseases” GHE group (R00-R94, R96-R99, U07, U99). There were 196,055 deaths from ill-defined diseases. These were proportionately assigned between the residual groups.

Causes of death with the ICD-10 code S00-S99 or T00-T99 are not valid underlying causes of death. These were all neonatal deaths, which are not assigned an underlying cause of death, and were imputing using the code in the first position on the death record. These deaths were classified as injuries.

TABLE D.2: Groups of causes of death used in the district-level cancer analysis in Chapter 7 with ICD-10 codes.

Cancer	ICD-10 codes
Trachea, bronchus and lung cancers	C33-C34
Breast cancer	C50
Prostate cancer	C61
Colorectal cancers	C18-C21
Pancreatic cancer	C25
Ovarian cancer	C56
Lymphomas, multiple myeloma	C81-C90, C96
Oesophageal cancer	C15
Bladder cancer	C67
Leukaemia	C91-C95
Corpus uteri cancer	C54
Stomach cancer	C16
Liver cancer	C22
All other cancers	C00-C14, C17, C23-24, C26-C32, C37-C41, C43-C49, C51-C53, C57-C66, C48-C81, C97 D00-D48

The residual group also contained deaths from the “ill-defined diseases” GHE group (R00-R94, R96-R99, U07, U99), which were proportionately assigned between the residual groups for cancers, non-communicable diseases, cardiovascular diseases and infections as part of a wider study on causes of death in England. The residual group also includes breast cancer for men. Ovary cancer, corpus uteri cancer (women) and prostate cancer (men) are sex specific. Bladder cancer and liver cancer were not leading cancers for women, so they are included in the residual group.

The next leading cancers in the residual group of all other cancers were bladder cancer, brain and nervous system cancers, and liver cancer for women, and brain and nervous

system cancers, mesothelioma, and melanoma and other skin cancers for men.