Decisions in Economics and Finance
https://doi.org/10.1007/s10203-025-00511-2

Modelling mortality by cause of death and socio-economic stratification: an analysis of mortality differentials in England

Andrés M. Villegas[1] · **Madhavi Bajekal**[2] · **Steven Haberman**[3]

Received: 30 November 2023 / Accepted: 11 January 2025

Abstract

It is well known that more advantaged socio-economic groups – whether defined by educational attainment, occupation, income or area deprivation – have lower mortality rates and longer lives than less advantaged socio-economic groups. In many cases, affluent subpopulations also experience faster rates of improvement in mortality. Socio-economic differentials in mortality and longevity pose important challenges when designing public policies for tackling social inequalities; and for managing longevity risk in pension funds and annuity portfolios. The successful addressing of these social and financial challenges requires the best possible understanding of the drivers of socio-economic mortality differentials. A key step in achieving this understanding is to investigate how mortality trends for leading causes of death differ between socio-economic groups. Accordingly, the main purpose of this paper is to propose modelling techniques that enable the modelling and projection of mortality trends by cause of death and socio-economic stratification. We first extend the Lee-Carter model to allow for the consideration of coding changes in cause-specific mortality data. We then embed this model into a multiple population setting to allow for the quantification of socio-economic differences in cause-specific mortality. Using England mortality data for socio-economic subpopulations defined using the Index of Multiple Deprivation (IMD), we show that this modelling approach can be satisfactorily employed both in the assessment of the magnitude of historical mortality

✉ Andrés M. Villegas
a.villegas@unsw.edu.au

Madhavi Bajekal
m.bajekal@ucl.ac.uk

Steven Haberman
s.haberman@city.ac.uk

1 School of Risk and Actuarial Studies and ARC Centre of Excellence in Population Ageing Research (CEPAR), UNSW Sydney, Kensington, Australia

2 Department of Applied Health Research, University College London, London, UK

3 Faculty of Actuarial Science and Insurance, Bayes Business School (formerly Cass), City St George's, University of London, London, UK

Published online: 28 January 2025

differentials for the main causes of death and in the projection of the possible future evolution of these differentials.

Keywords Mortality modelling · Multipopulation models · Socio-economic circumstances · Causes of death · Lee-Carter model

1 Introduction

It is well known that more advantaged socio-economic groups – whether defined by educational attainment, occupation, income or area deprivation – have lower mortality rates and longer lives than less advantaged socio-economic groups. In many cases, affluent subpopulations also experience faster rates of improvement in mortality. These socio-economic differences in mortality are unacceptable in a fair society and policy makers need to take action to reduce the social gradient in health (The Marmot Review 2010). Furthermore, actuaries need to recognise mortality heterogeneity as not doing so could result, for instance, in an inadequate funding of annuity and pension obligations (Meyricke and Sherris 2013; Villegas and Haberman 2014). The design of effective policies for tackling social inequalities in health and the successful management of the financial implications of longevity risk therefore require the best possible understanding of the drivers of socio-economic mortality differentials. A key step in achieving this understanding is to investigate how mortality trends for leading causes of death differ between socio-economic groups.

Moreover, the utility and importance of looking at the twin issues of mortality by cause of death and by socio-economic stratification has been widely recognised by the actuarial profession. For instance, in a survey on mortality by cause of death and socio-economic stratification prepared for the International Actuarial Association, Ridsdale and Gallop (2010) highlight that these two subjects would be of interest to actuaries:

- aiming to understand mortality trends and its drivers with the objective of pricing the mortality elements of insurance, annuity and pensions products, or of reserving effectively for liabilities;
- involved in the underwriting and in the setting of premium rates for "substandard lives"; or
- looking to inform or test the credibility of their assumptions for mortality projections.

Richards (2009) also acknowledges the close links between socio-economic circumstances and causes of death and hence argues that to avoid misleading results *"cause-of-death statistics need to be broken down by socio-economic group or deprivation index before being used for forecasting purposes for actuarial work"*.

However, in spite of its many benefits and uses, the analysis of cause-specific mortality trends faces several theoretical and practical challenges; much of them summarised in GAD (2001), Continuous Mortality Investigation (2004) and Richards (2009). The main theoretical problem when considering mortality by cause of death is how to capture the correlations between causes: causes are not always independent and may be linked by poorly understood mechanisms (or causal chains). For instance, the same

risk factors can affect several causes as is the case with smoking and numerous forms of cancer and heart diseases. In addition, even if two causes of death do not share risk factors, reductions in the relative importance of one cause can lead to further mortality improvements on the other one as medical research efforts shift from one cause to the other.

In addition, the validity of the analysis and projection of cause-specific mortality trends can be affected by changes in the diagnosis and in the classification of causes of death. The World Health Organisation (WHO) has reviewed the International Classification of Diseases (ICD) regularly in the 20^{th} century to reflect new diseases, changes in medical terminology and the development of medical knowledge (Moriyama et al. 2011). Further, national statistical offices often make adjustments when adopting locally the ICD coding system (see e.g. Rooney and Devis (1996) for the case of England and Wales). These data production changes can cause substantial discontinuities in the mortality trends of some causes of death. This is for instance the case of respiratory diseases in England and Wales which, as illustrated in Fig. 2, have seen major trend-breaks due to the numerous changes in the cause of death classification system.

As summarised by Richards (2009) *"when doing projections by cause of death we must not only take great care with socio-economic differentials, and also worry about projecting correlated time series, but we must also take note of uncertainty surrounding the classification of cause of death itself"*.

Accordingly, the main purpose of this paper is to propose modelling techniques that facilitate the modelling and projection of mortality trends by cause of death and socio-economic stratification, with a particular focus on addressing potential issues arising from changes in the cause of death coding system. To achieve this, we integrate well-established modelling tools from actuarial science and demography: specifically, the Three-way Lee-Carter model by Russolillo et al. (2011) for multiple population mortality modelling, and the statistical correction method for cause of death coding changes proposed by Rey et al. (2011). While these two independent modelling tools are not new, the combination of these techniques into a framework for cause-of-death and subpopulation modelling is novel. This constitutes our main methodological contribution.

A second purpose is to provide an analysis of the socio-economic differentials for the leading causes of death in England. This study is complementary to the work of Villegas and Haberman (2014) where the modelling of socio-economic differences in all-cause mortality was addressed and which provided an investigation of all-cause mortality differentials in England across deprivation subgroups.

The paper is organised as follows. In Sect. 2 we describe the proposed modelling approach for the quantification and projection of socio-economic mortality differentials by cause of death. Specifically, in Sect. 2.1 we introduce an extension of the widely-use model of Lee and Carter (1992) to control for coding changes in cause-specific mortality data. Then, in Sect. 2.2, we embed this model in to a multiple population setting to allow for the consideration of socio-economic differences in mortality. In Sect. 3, we apply the proposed modelling approach in the examination of the relationship between deprivation and mortality for the leading causes of death

in England. Finally, in Sect. 4, we discuss our main findings and discuss possible limitations and extensions of our work.

2 Modelling and forecasting mortality by cause of death and socio-economic stratification

In this section we describe a modelling approach to quantify cause-specific mortality differentials in population subgroups induced by an additional covariate. Although our main interest is on population subgroups defined by socio-economic factors (e.g. occupation, level of education, social class), these subgroups could also be associated to geographical factors (e.g. regions within a country) or other factors such as marital status, gender, ethnicity and so on. The proposed approach is based on an extension of the Lee-Carter model (Lee and Carter 1992) to deal with data production changes in cause of death data, combined with the multipopulation Three-way Lee-Carter model proposed by Russolillo et al. (2011).

2.1 Lee-Carter model with cause of death coding adjustments

Assume that the death counts from a given cause of death are independent Poisson responses

$$D_{xt} \sim Poisson(e_{xt}\mu_{xt})$$

so that the force of mortality from the given cause at exact age x in time t, μ_{xt}, is given by $\mu_{xt} = \mathbb{E}(D_{xt})/e_{xt}$, with e_{xt} denoting the central exposed to risk at age x and time t. Also assume that mortality data are available for consecutive years $t_0, t_1, \ldots, t_n$ and let $\mathcal{S} = \{s_1, s_2, \ldots, s_h\}$ be the times at which the coding changes occur with the convention that $s_0 = t_0$. In order to account for data production changes we extend the widely-used model of Lee and Carter (1992) to assume that the force of mortality from certain cause is modelled by:

$$\log \mu_{xt} = \alpha_x + \beta_x \kappa_t + \sum_{i=1}^{h} \delta_x^{(i)} f^{(i)}(t) \tag{1}$$

where, as in the standard Lee-Carter model, α_x captures the general age-specific mortality pattern, κ_t represents the overall level of mortality at time t and β_x measures the age-specific responses to changes in the general level κ_t. In addition, for $i = 1, \ldots, h$, $f^{(i)}(t) = \mathcal{I}_{s_{i-1} \le t < s_i}$ denotes the indicator function taking value 1 if $t \in [s_{i-1}, s_i)$ and $\delta_x^{(i)}$ are new parameters which measure the magnitude of coding changes at age x due to the change in coding system occurred at time s_i. This model controls for data production changes by assuming that there are different age-patterns $\alpha_x + \delta_x^{(i)}$ for each subperiod $[s_{i-1}, s_i)$, $i = 1, \ldots, h$, where a particular cause of death coding system is in force. It is worth noting that the model in Equation (1) can only correct for data production changes that result in sustained jumps in mortality rates and cannot remove gradual changes in coding that result in changes in mortality trends.

In a similar manner to the standard Lee-Carter model, the model defined by Equation (1) is only identifiable up to a set of transformations. Specifically, given constants b_1, $b_2 \neq 0$ and a_i, $i = 1, \ldots, h$, we can transform the parameters in Equation (1) in the following ways

$$\{\tilde{\alpha}_x, \tilde{\kappa}_t\} = \{\alpha_x + b_1\beta_x, \kappa_t - b_1\} \tag{2}$$

$$\left\{\tilde{\beta}_x, \tilde{\kappa}_t\right\} = \left\{\frac{1}{b_2}\beta_x, b_2\kappa_t\right\} \tag{3}$$

$$\left\{\tilde{\delta}_x^{(i)}, \tilde{\kappa}_t\right\} = \left\{\delta_x^{(i)} + a_i\beta_x, \kappa_t - a_i f^{(i)}(t)\right\}, \quad i = 1, \ldots, h, \tag{4}$$

leaving the fitted mortality rates in Equation (1) unchanged. Transformations (2) and (3) correspond to the transformations that can be applied to the standard Lee-Carter model, whilst the new family of transformations defined by (4) is induced by the new parameters $\delta_x^{(i)}$. This latter set of transformations mean that we are free to shift mortality levels from $\delta_x^{(i)}$ to κ_t.

In order to ensure the identifiability of the model, some constraints need to be imposed. To deal with transformations (2) and (3) we impose the constraints

$$\begin{aligned} \sum_x \beta_x &= 1 \\ \kappa_{t_n} &= 0 \end{aligned} \tag{5}$$

Choosing constraint (5) as opposed to the more common $\sum_t \kappa_t = 0$ gives the interpretation that α_x represents the fitted mortality rates in the last year of observation, t_n. This, coupled with the chosen definition of the set of coding changes times, $\mathcal{S}$, ensures that the coding system in force in the period $[s_h, t_n]$ (i.e. the most recent one) is the reference for the fitted and projected mortality rates.

It is desirable that the underlying coding-adjusted "jump-free" mortality trend from the cause of death of interest is captured by κ_t and that parameters $\delta_x^{(i)}$ capture the discontinuities in mortality induced by the changes in the cause of death coding system. This goal can be accomplished by carefully choosing constants a_i, $i = 1, \ldots, h$, in the family of transformations defined by (4). To do so, we use a similar approach to the one introduced by Rey et al. (2011) for the removal of discontinuities in age-specific mortality rates by cause of death. Specifically, we set the constants a_i, $i = 1, \ldots, h$, by fitting the model

$$\kappa_t = g(t) + \sum_{i=1}^{h} a_i f^{(i)}(t) + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma) \qquad \text{i.i.d.}, \tag{6}$$

where $g(t)$ is a continuous function and ϵ_t is a Gaussian error term. Model (6) decomposes the time trend κ_t into:

i. a smooth function $g(t)$ representing the underlying"jump-free"mortality trend;

ii. the jumps in mortality $\sum_{i=1}^{h} a_i f^{(i)}(t)$ induced by data production changes; and
iii. the noise ϵ_t around the"jump-free"trend.

In order to estimate the smooth function $g(t)$ we follow Rey et al. (2011) and use a thin plate penalised regression spline with the smoothness parameter derived automatically using generalised cross-validation. As noted by Rey et al. (2011), this can be easily accomplished using the **MGCV R** package (Wood 2020).

Finally, given constants a_i, $i = 1, \ldots, h$, estimated from model (6) and using the family of transformations defined by Equation (4), we can recover the "jump-free" trend using the expression

$$\tilde{\kappa}_t = g(t) + \epsilon_t = \kappa_t - \sum_{i=1}^{h} a_i f^{(i)}(t), \tag{7}$$

with revised data production adjustments

$$\tilde{\delta}_x^{(i)} = \delta_x^{(i)} + a_i \beta_x \quad i = 1, \ldots, h.$$

The procedure defined by Equations (6) and (7) is schematised in Fig. 1. This approach can also be adapted for mortality models other than the Lee-Carter model. For instance, in Villegas et al. (2023), we use an extension to the age-period-cohort mortality model, incorporating coding changes.

2.2 Modelling differentials: The Three-way Lee-Carter model with coding adjustments

In order to allow for the quantification of socio-economic differences in cause-specific mortality, we now embed under a multipopulation setting the Lee-Carter model with coding adjustments introduced in the previous section. To this end, we use the Three-way Lee-Carter model proposed by Russolillo et al. (2011) to deal with mortality data disaggregated according to a third criterion besides age and time. Specifically, we assume that the central death rate for a given cause of death for age x at time t in subpopulation g, denoted μ_{xtg}, is given by

$$\log \mu_{xtg} = \alpha_x + \alpha_{xg} + \beta_x \lambda_g \kappa_t + \sum_{i=1}^{h} \delta_x^{(i)} f^{(i)}(t), \tag{8}$$

where α_x captures the general age-specific mortality pattern, α_{xg} captures age-subpopulation-specific deviations from the general age pattern α_x, κ_t represent the overall level of mortality at time t, β_x measures the age-specific responses to changes in the general level κ_t, λ_g captures trend mortality differences across subpopulations, and $\delta_x^{(i)}$ and $f^{(i)}(t)$ have the same interpretation as in model (1). We note that in model (8) it is assumed that cause of death coding changes have a uniform impact across subpopulations. Section 2.4 elaborates further on the interpretation of the parameters of the model.

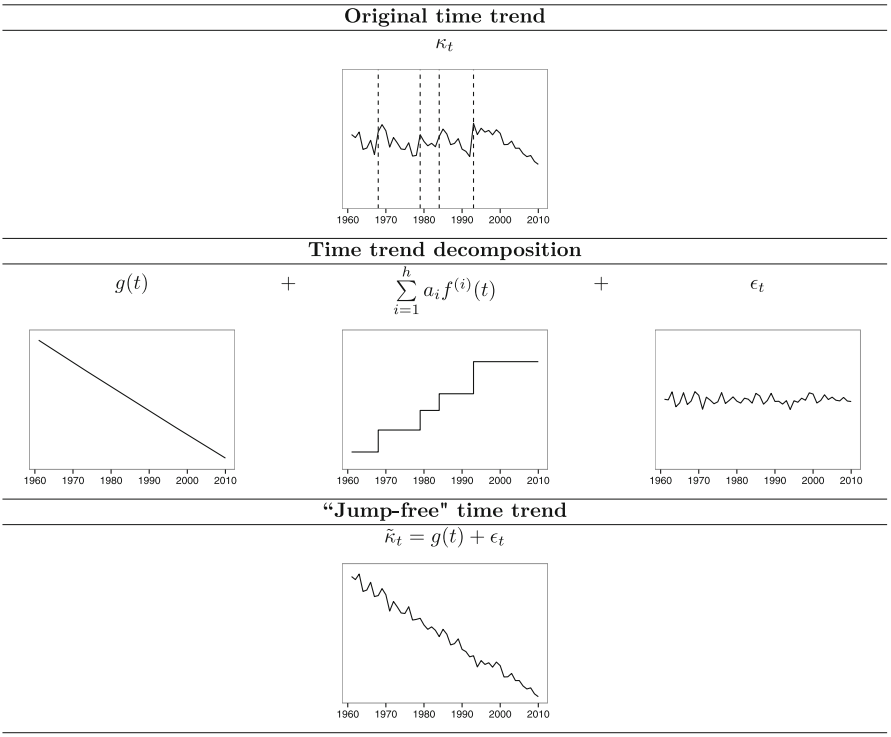

Fig. 1 Schematic representation of the decomposition of the mortality trend κ_t

2.3 Parameter estimation

In order to estimate the parameters of the Three-way Lee-Carter model with coding adjustments, we resort to a two-stage estimation strategy using data from a reference population. More specifically, in a first stage the subpopulation independent parameters $\alpha_x, \beta_x, \kappa_t$ and $\delta_x^{(i)}$ are estimated using the reference population data. Then, in a second stage and conditional on the estimated values of α_x, β_x, κ_t and $\delta_x^{(i)}$, the remaining subpopulation dependent parameters α_{xg} and λ_g are estimated. Thus, our estimation procedure entails the following two steps:

i. Fit the Lee-Carter model with code adjustments defined by Equation (1) to the reference population to obtain parameter estimates $\hat{\alpha}_x$, $\hat{\beta}_x$, $\hat{\kappa}_t$ and $\hat{\delta}_x^{(i)}$.
ii. Given parameter estimates $\hat{\alpha}_x$, $\hat{\beta}_x$, $\hat{\kappa}_t$ and $\hat{\delta}_x^{(i)}$, obtain parameters estimates $\hat{\alpha}_{xg}$ and $\hat{\lambda}_g$ by fitting the generalised linear model

$$D_{xtg} \sim Poisson(e_{xtg}\mu_{xtg})$$

$$\log \mu_{xtg} = \hat{\alpha}_x + \alpha_{xg} + \hat{\beta}_x \lambda_g \hat{\kappa}_t + \sum_{i=1}^{h} \hat{\delta}_x^{(i)} f^{(i)}(t),$$

where, D_{xtg} and e_{xtg} are, respectively, the number of deaths for the cause of death of interest and the population at age x and time t for subpopulation g.

Such a two-stage estimation strategy has several advantages when the subpopulations represent different socio-economic strata of a national population. These advantages include:

- National mortality data are normally available for a longer period than mortality data disaggregated by socio-economic stratification, permitting a more precise estimation of the long-run mortality trend of the different causes of death.
- Modelling the subpopulations alongside the national population will ensure the consistency of the subpopulation-specific mortality forecasts with the national mortality forecasts.

2.4 Mortality level differential and mortality trend differentials

In general, we can distinguish between two type of mortality differentials among subgroups: mortality level differentials and mortality trend differentials. The former refers to differentials in the average level of mortality, whereas the latter refers to differentials in the pace of mortality change. In the modelling framework defined by Equation (8) parameters α_x and α_{xg} relate to mortality level differentials, whilst parameters β_x, λ_g and κ_t determine mortality trend differentials.

The term $\exp(\alpha_x)$ measures the general level of mortality in the reference population and the term $\exp(\alpha_{xg})$ quantifies the percentage deviations of each subgroup from this reference pattern of mortality. That is, if $\exp(\alpha_{xg}) > 1$ then at age x subgroup g has a higher mortality than the reference population and if $\exp(\alpha_{xg}) < 1$ then at age x subgroup g has a lower mortality than the reference population.

To understand how our modelling approach assesses trend differentials in mortality we can compare for each age x the rate of change over t of the log-mortality rates in the reference population,

$$\frac{d\ \log \mu_{xt}}{dt} = \beta_x \frac{d\ \kappa_t}{dt},$$

with the rate of change over t of the log-mortality rates in each subpopulation,

$$\frac{d\ \log \mu_{xtg}}{dt} = \beta_x \lambda_g \frac{d\ \kappa_t}{dt} = \lambda_g \frac{d\ \log \mu_{xt}}{dt}$$

From this last equation it is clear that λ_g can be interpreted as the percentage deviation of mortality change in each subgroup relative to the mortality change in the reference population. That is, if mortality is improving in the reference population (i.e. $d\ \log \mu_{xt} \big/ dt\ < 0$) then $\lambda_g > 1$ means that for subpopulation g mortality improves at a faster pace than it does in the reference population, whereas $\lambda_g < 1$ means that it decreases at a slower pace. Similarly, if mortality from a specific cause in the reference population is worsening (i.e. $d\ \log \mu_{xt} \big/ dt > 0$) then $\lambda_g > 1$ means that for subpopulation g mortality is worsening at a faster pace than in the reference and $\lambda_g < 1$ means that it is worsening at a slower pace.

2.5 Forecasting

Under the proposed model the time evolution of mortality and of mortality differentials is represented by the univariate time index κ_t. Therefore, forecasts of age-subgroup-specific mortality rates by cause of death can be obtained by modelling and forecasting κ_t using autoregressive moving average (ARIMA) models. In general, we assume that κ_t follows an ARIMA(p, q, d) with drift:

$$\Delta^d \kappa_t = \theta_0 + \phi_1 \Delta^d \kappa_{t-1} + \cdots + \phi_p \Delta^d \kappa_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \cdots + \theta_q \xi_{t-q} \qquad (9)$$

where Δ is the difference operator; θ_0 is the drift parameter, $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients with $\phi_p \neq 0$; $\theta_1, \ldots, \theta_q$ are the moving average coefficients with $\theta_q \neq 0$; and ξ_t is a Gaussian white noise process with variance σ_ξ.

Model (9) can then be used to obtain projected values of the time index $\dot{\kappa}_{t_n+r}$, $r = 1, 2, \ldots$, which are then inserted into equation (8), omitting the coding adjustment terms, to provide forecasted age-group-specific mortality rates corrected for cause of the coding adjustments:

$$\dot{\mu}_{x,t_n+r,g} = \exp\left(\hat{\alpha}_x + \hat{\alpha}_{xg} + \hat{\beta}_x \hat{\lambda}_g \dot{\kappa}_{t_n+r}\right)$$

3 Case study: Mortality by cause of death and deprivation in England

In this section we present an empirical application that illustrates the previous discussion on the modelling of mortality by cause of death and socio-economic stratification. Specifically, we examine the relationship between deprivation and mortality for the main causes of death in England for the period 2001-2020.

3.1 Data

Similar to other mortality studies in England (see e.g., Lu et al. (2014), Villegas and Haberman (2014), and Wen et al. (2023)), we use the English Index of Multiple Deprivation (IMD) to measure socio-economic conditions at a small area level (Noble et al. 2007). The IMD is a composite index of deprivation made up of seven domains each with an associated weight.[1] The IMD is based on small geographies called Lower Layer Super Output Areas (LSOAs). There are 32,482 LSOAs in England with an average population of approximately 1,500 people. Between 2001 and 2020 there have been five versions: IMD 2004, IMD 2007, IMD 2010, IMD 2015, and IMD 2019. In our analysis LSOAs are grouped into deprivation quintiles based on their IMD score as

[1] The seven deprivations domains with their percentage participation in the index are: i) Income deprivation (22.5%), ii) Employment deprivation (22.5%), iii) Health deprivation and disability (13.5%), iv) Education, skills and training deprivation (13.5%), v) Barriers to housing and services (9.3%), vi) Crime (9.3%), and vii) Living Environment deprivation (13.5%).

extant in the given year,[2] with Q1 denoting the most deprived quintile and Q5 the least deprived quintile. Accordingly, in any given year, the entire population of England (and their corresponding deaths) is divided into five subpopulations according to the LSOAs in each IMD quintile. This approach addresses potential issues of credibility and consistency that may arise with alternative socio-economic stratification methods that do not encompass the entire English population.

We use publicly available data from ONS comprising death counts for 11 major causes of death for the English population disaggregated by IMD quintiles for the period 2001 to 2020 and age groups, 50-54, …, 80-84, 85+.[3] We exclude younger ages and concentrate on ages above 50 as these are the more relevant ages for pension and annuity products. We consider 10 groups of causes of death derived from the International Classification of Diseases (ICD) plus a group "Rest of causes" including the remaining deaths. The resulting 11 group of causes together with ICD codes included in each group are presented in Table 1. In 2019, the 10 selected causes represent around 80% of the total number of deaths in England and Wales between ages 50 to 84 with cancers and circulatory diseases constituting, respectively, 37% and 24% of the deaths.

As for the reference population, we use publicly available cause of death data for England and Wales obtained from the 20th Century Mortality Files produced by ONS.[4] From these data we obtained number of death by cause of death for ages 50-54, …, 80-84, 85+ and the period 1968-2020. Over this 50 year period cause of death coding in England and Wales has undergone five major changes which are summarised in Table 2.

In order to examine the magnitude and trends of socio-economic differentials in mortality for the main causes of death, for each cause we have fitted the Three-way Lee-Carter model with coding adjustments introduced in Sect. 2.2 to England mortality data for the period 2001-2020 with IMD quintile as covariate and England and Wales data for the period 1968-2020 as reference population data. We have excluded age category 85+ from the analysis as it groups many ages into a single category, having a distortionary effect on the results. In addition, we have assumed that mortality trends need to be adjusted for cause of death coding changes in years $s_1 = 1979$, $s_2 = 1984$, $s_3 = 1993$ and $s_4 = 2001$, corresponding to the introduction of ICD-9, the changes in the application of Rule 3, the introduction of the automated cause coding system (ACCS), and the introduction of ICD-10, respectively (Moriyama et al. 2011; Rooney and Devis 1996).

Figures 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13 present for the 11 groups of causes of death the estimated parameters for the England Wales reference population, while Fig. 14, 15 and 16 depict the mortality differential parameters for the deprivation subpopulations. We defer, however, the analysis of these fitted parameters until Sects. 3.3 and 3.4 and

[2] IMD 2004 for data in years 2001-2003, IMD 2007 for data in years 2004-2006, IMD 2010 for years 2007-2010, IMD 2015 for years 2011-2015, and IMD 2019 for years 2016-2010.

[3] https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/1324deathregistrationsandpopulationsbysexageandimdquintileenglandandwales2001to2020.

[4] https://www.data.gov.uk/dataset/2548e46b-873e-4668-968c-25d6c155dd73/the-20th-century-mortality-files.

Table 1 Grouping of causes, ICD codes and percentage of deaths in 2019 for ages 50 to 84 in England and Wales

Cause group	% Deaths 2019	ICD-10	ICD-9	ICD-8
Ischaemic heart diseases	12%	I20-I25	410-414	410-414
CVD and stroke	5%	I60-I69	430-438	430-438
Other Circulatory Diseases	7%	I00-I16, I26-I28, I30-I52, I70-I99	390-409, 415-429, 439-459	390-409, 415-429, 439-459
Lung cancer	8%	C33, C34	162	162
Breast and prostate cancer	5%	C50, C61	174,185	174,185
Digestive cancers	7%	C00-C21	140-154	140-154
Other cancers	17%	C22-C32, C37-C49,C51-C58,C60,C62-C97,D00-D49	155-161,163−173.175-184. 186-239	155-161,163−173.175-184. 186-239
Respiratory diseases	13%	J00-J99	460-519	460-519
Alcohol related liver disease and cirrhosis	2%	K70, K73-K74	571	571
Alzheimers and Dementia	3%	G30-G31	290,294,298,331	290,294,298
Rest of causes	21%	All other codes	All other codes	All other codes

Table 2 Changes in cause of death coding implemented in England and Wales over the period 1968-2020

Change	ICD-8	ICD-9	Rule 3	ACCS	ICD-10
Year	1968	1979	1984	1993	2001

concentrate first on the quality of the correction for coding adjustments performed by the models.

3.2 Evaluation of coding adjustments

In order to examine the performance of the fitted models with regards to the quality of the coding adjustments, we plot in Fig. 2 observed, fitted and code-adjusted cause-specific age standardised death rates (ASDR) for the age range 50-84. These ASDRs have been computed using the expression

$$ASDR_t = \frac{\sum_x \omega_x \mu_{xt}}{\sum_x \omega_x}$$

where ω_x represents the 2013 European Standard Population (Pace et al. 2013) and μ_{xt} corresponds to

- d_{xt}/e_{xt}, in the case of the observed ASDRs, with d_{xt} being the observed number of deaths from a given cause;
- $\exp\left(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t + \sum_{i=1}^{4} \hat{\delta}_x^{(i)} f^{(i)}(t)\right)$ in the case of the fitted ASDRs; and
- $\exp\left(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t\right)$ in the case of the code-adjusted ASDRs.

Referring to Fig. 2 we highlight the following features:

- The reassuring close alignment of the fitted and observed ASDRs for all cause groups and both genders.
- The satisfactory adjustment for the marked trend breaks induced by the broadening in the application of Rule 3 of the ICD-9 introduced in England after 1984. The change in the application of Rule 3 meant that some "terminal conditions" were not coded as the underlying cause of death if some other "major condition" was reported in the death certificate (Rooney and Devis 1996; Brock et al. 2006; Goldacre et al. 2003). One of the main effects of this was a sudden fall in the numbers of deaths from respiratory diseases with mental and behavioural conditions (among other diseases) being the other side of the change in the application of Rule 3 and experiencing a sudden increase in the numbers of deaths (Rooney and Devis 1996; Griffiths and Rooney 2006).
- The possible "false-positive" trend adjustment performed in 1994 for ischaemic heart disease in men and in 1980 for lung cancer in men, where the model may have mistakenly identified genuine changes in trend as changes in coding.

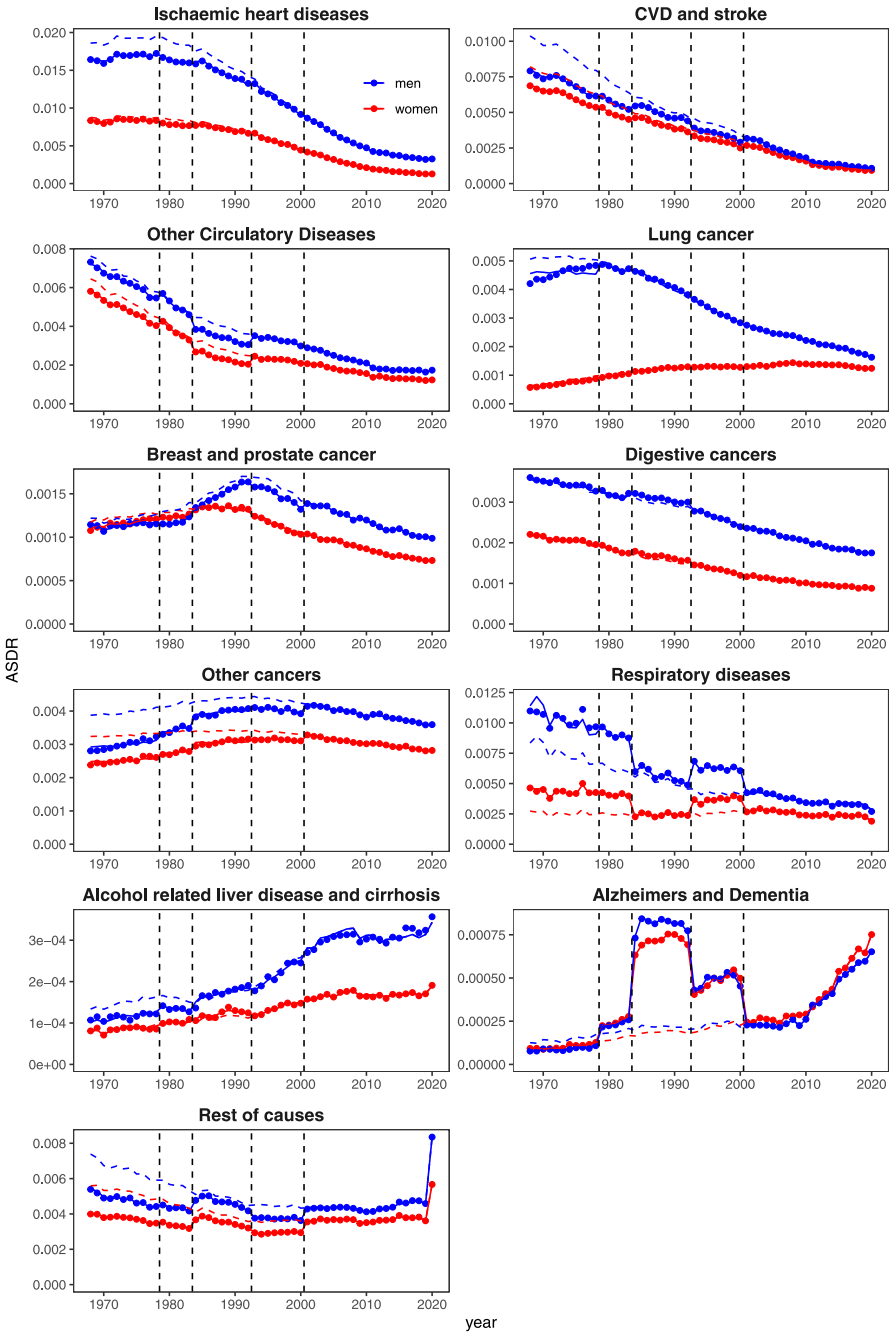

Fig. 2 Age standardised death rates (ASDR) for the age range 50-84 for selected causes of death in England and Wales reference population. Dots show observed ASDRs, solid lines fitted ASDRs and dashed lines coding-adjusted ASDRs. Vertical dashed lines indicate coding-adjustment years

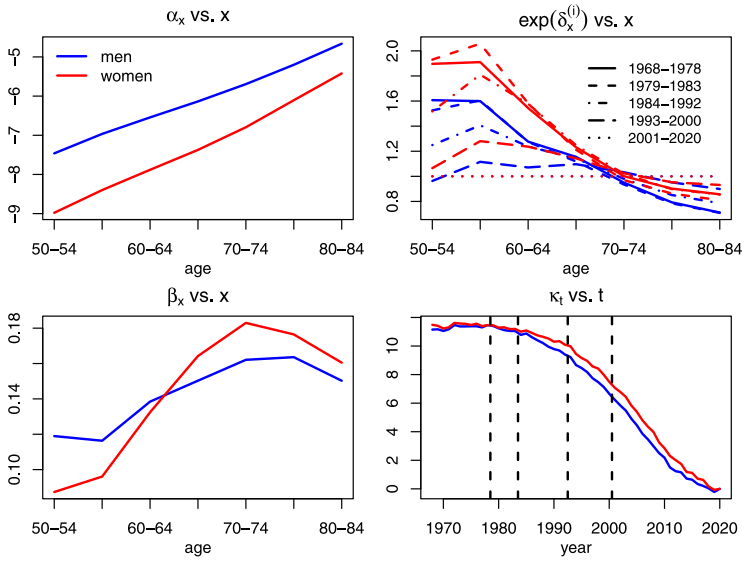

Fig. 3 Fitted parameters for the Lee-Carter model with code adjustments fitted to ischaemic heart diseases in England for the period 1968–2020

Overall, the proposed Lee-Carter model with cause of death coding adjustments performs very well in recovering the underlying trend of mortality for the different causes of death. This will be crucial for the satisfactory forecasting of these trends as performed in Sect. 3.5.

3.3 Historical cause-specific mortality patterns: England and Wales 1968–2020

Having seen that the Three-way Lee-Carter model with coding adjustments performs a satisfactory adjustment for cause of death coding changes, we now use this model to analyse the historical pattern of cause-specific mortality for the period 1968-2020 in the England and Wales population and to analyse the differentials in mortality across deprivation quintiles in England for the period 2001–2020.

Figures 3 to 13 depict for the 11 group of causes of death the parameter estimates of the Lee-Carter model with coding adjustments fitted to mortality data for men and women in the England and Wales reference population. From these figures we note the following important features:

- The pattern of increasing mortality with age, represented by α_x, is consistent across all cause groups, with the notable exception of alcohol-related mortality (see Fig. 11), which exhibits a decrease in mortality with advancing age. Additionally, several causes, including Ischemic Heart disease (IHD), Cardiovascular disease (CVD) and stroke, other circulatory diseases, respiratory diseases, and Alzheimer's and Dementia, demonstrate an almost linear increase in log death rates with age. These age-related patterns contribute to the exponential growth in mortality rates with age, a characteristic feature of all-cause mortality.

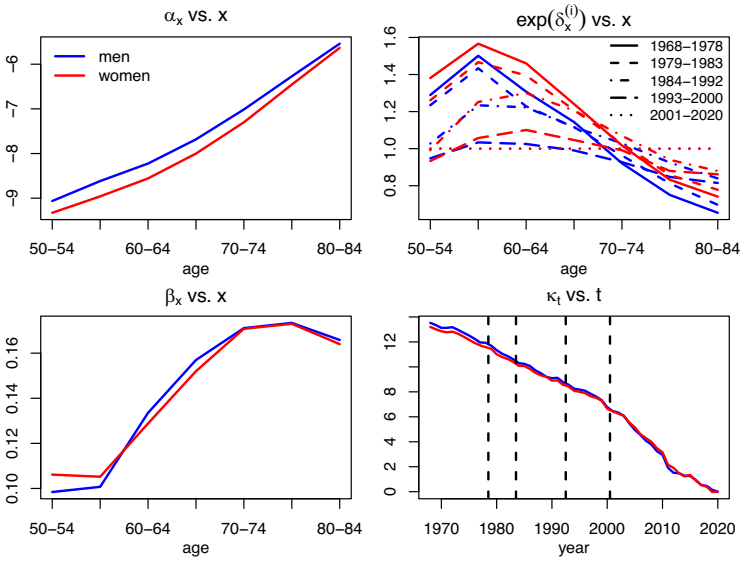

Fig. 4 Fitted parameters for the Lee-Carter model with code adjustments fitted to CVD and stroke in England for the period 1968–2020

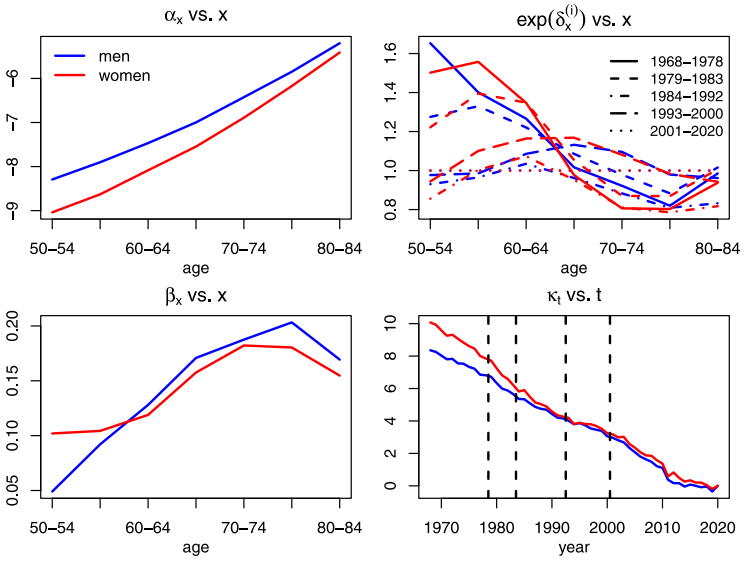

Fig. 5 Fitted parameters for the Lee-Carter model with code adjustments fitted to other circulatory diseases in England for the period 1968–2020

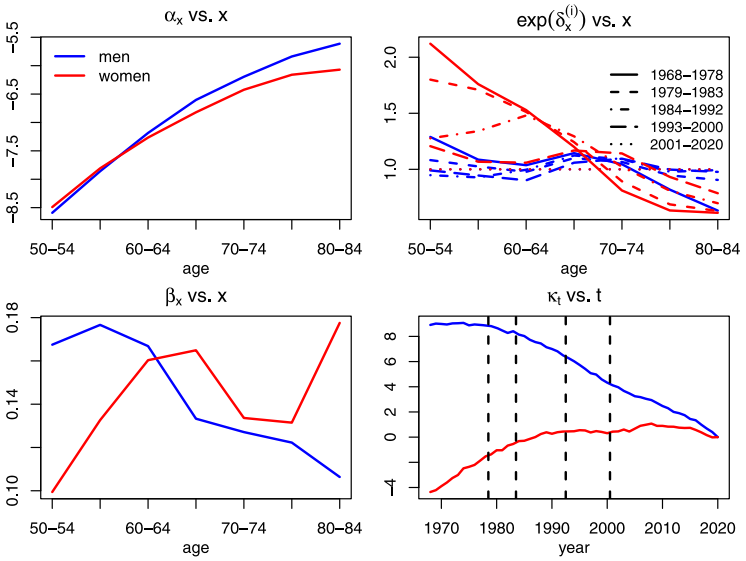

Fig. 6 Fitted parameters for the Lee-Carter model with code adjustments fitted to lung cancer in England for the period 1968–2020

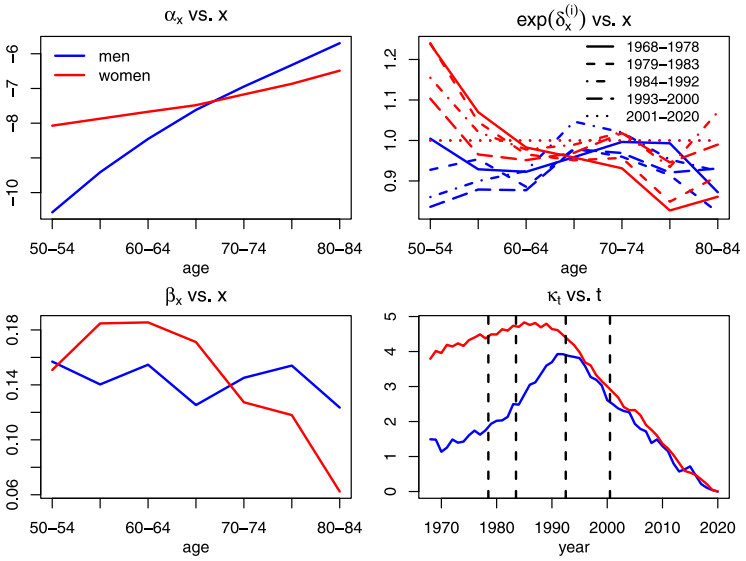

Fig. 7 Fitted parameters for the Lee-Carter model with code adjustments fitted to Breast and prostate cancer in England for the period 1968–2020

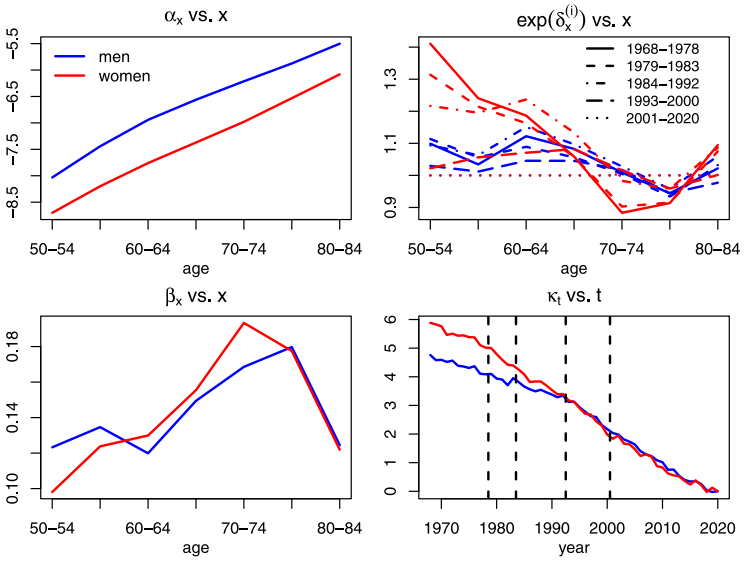

Fig. 8 Fitted parameters for the Lee-Carter model with code adjustments fitted to Digestive cancers in England for the period 1968–2020

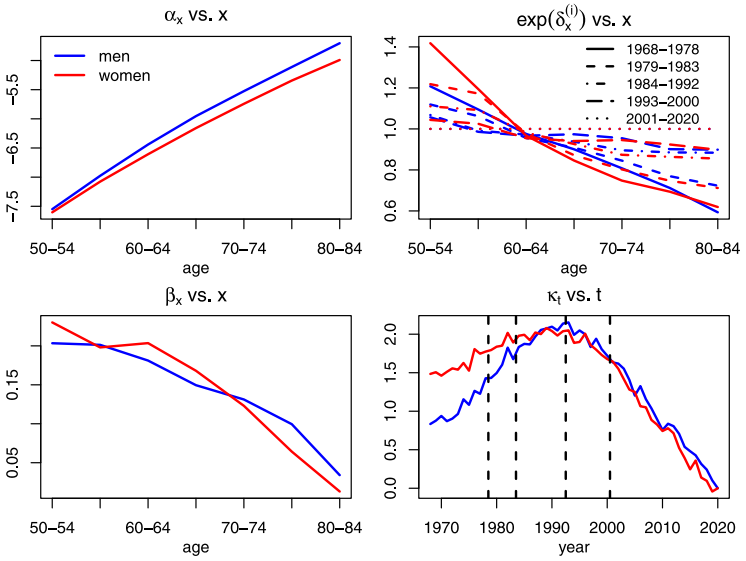

Fig. 9 Fitted parameters for the Lee-Carter model with code adjustments fitted to other cancers in England for the period 1968-2020

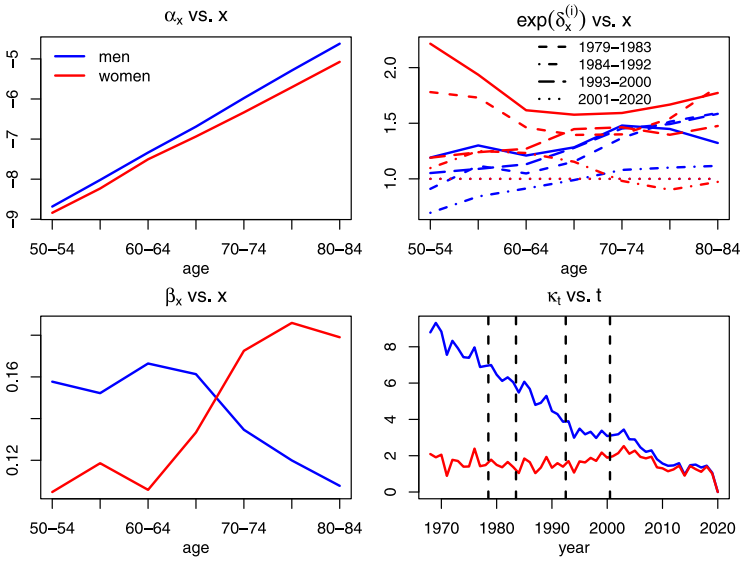

Fig. 10 Fitted parameters for the Lee-Carter model with code adjustments fitted to respiratory diseases in England for the period 1968–2020

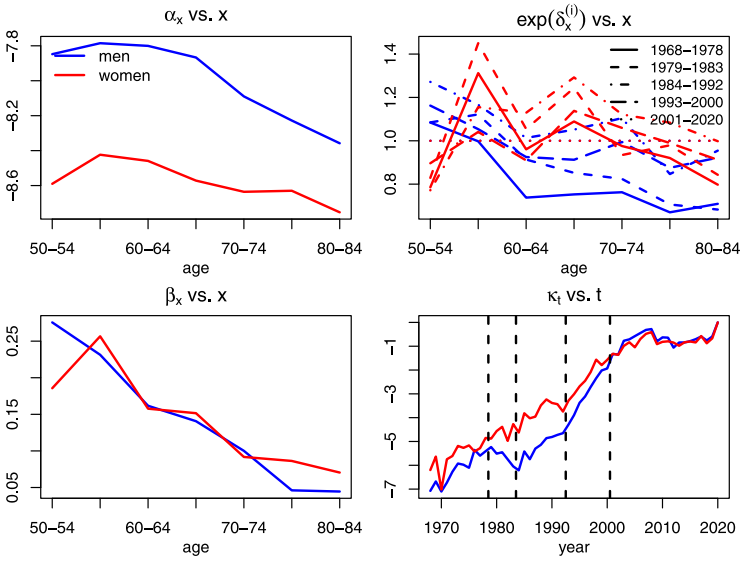

Fig. 11 Fitted parameters for the Lee-Carter model with code adjustments fitted to Alcohol related liver disease and cirrhosis in England for the period 1968–2020

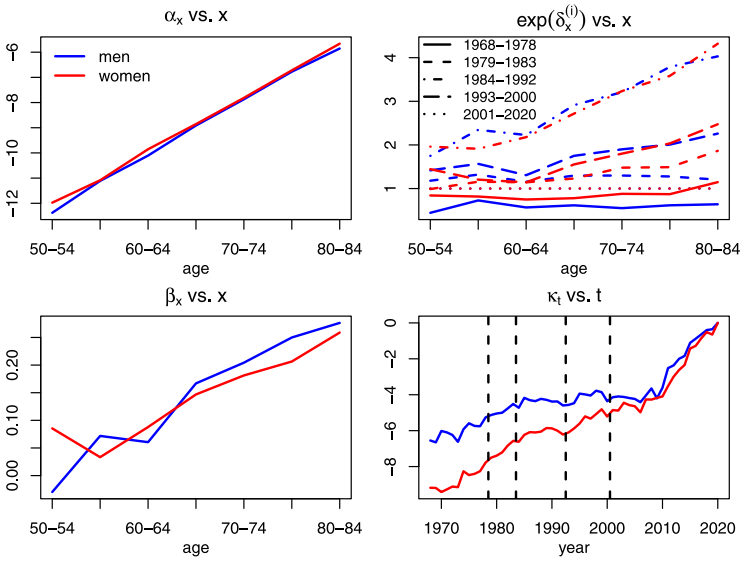

Fig. 12 Fitted parameters for the Lee-Carter model with code adjustments fitted to Alzheimers and Dementia in England for the period 1968–2020

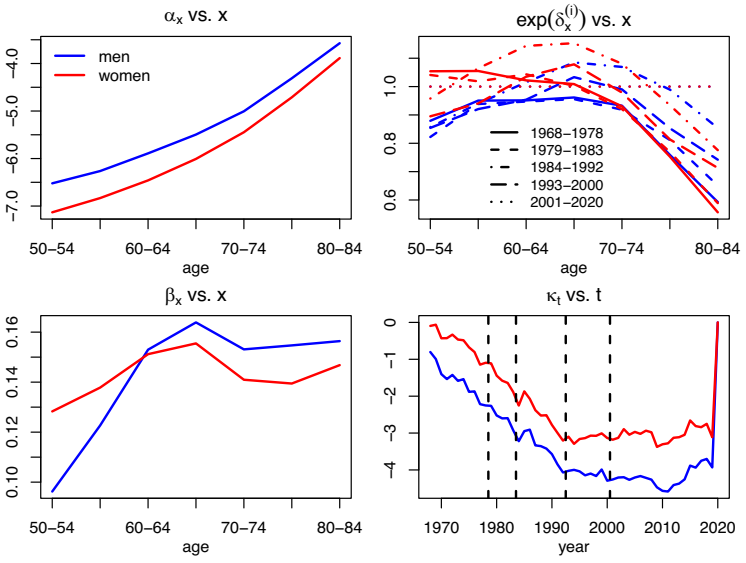

Fig. 13 Fitted parameters for the Lee-Carter model with code adjustments fitted to Rest of causes in England for the period 1968–2020

- The steady and significant decline in death rates from IHD, CVD and stroke and, other circulatory diseases captured by the κ_t parameters, in Figs. 3, 4 and 5, respectively. This considerable reduction in mortality has resulted in circulatory causes passing from being the main cause of death in 1968 to being in 2019 the second cause of death after Neoplasms for both sexes. However, a notable shift occurred in 2010 in the trends of κ_t for the three circulatory disease groups. This shift aligns with patterns reported for the USA (Villegas et al. 2023) and other developed countries (Djeundje et al. 2022). Possible reasons for this deceleration in mortality improvement from cardiovascular diseases include an increase in the prevalence of diabetes and obesity, the leveling off of benefits from reduced smoking prevalence, and the fact that recent advancements in the treatment and prevention of circulatory diseases have been more incremental (Mensah et al. 2017).

- There are noticeable changes in the trends of the κ_t parameters for lung, breast, prostate, and other cancers, which exhibit a decline after a period of increase (see Figs. 6, 7 and 9). A notable exception is digestive cancers, which have shown a steady mortality decline (see Fig. 8). While the patterns for digestive and other cancers are consistent between men and women, lung cancer presents significant differences. Lung cancer mortality in men began to decline in the late 1970 s, whereas for women, a decline in lung cancer mortality has only recently started to emerge. These gender disparities in cancer mortality can be partially attributed to differences in smoking behavior; cigarette consumption peaked in the 1950 s for men and in the 1960 s for women (Di Cesare and Murphy 2009). For both prostate and breast cancers, mortality rates began to decline in the early 1990 s. This is likely due to the widespread introduction of prostate-specific antigen screening (Giona 2021) for prostate cancer, and for breast cancer, an increase in the detection of smaller, palpable tumors and the use of adjuvant chemotherapy (Jatoi and Miller 2003).

- The trends observed in κ_t for respiratory diseases show a contrasting pattern between genders. Men exhibit a steady decline in mortality rates, while women experience a stagnation in these rates, as illustrated in Fig. 10. These differences in trends can also be partly attributed to the variations in smoking prevalence between genders, providing a potential explanation for the recent convergence in life expectancy between men and women reported in Mayhew and Smith (2014),

- There was a sharp increase in mortality related to alcohol until the late 2000 s, as depicted by κ_t in Fig. 11, with this trend stabilising in the subsequent years.

- Mortality from Alzheimer's and Dementia, as indicated by κ_t in Fig. 12, has been steadily increasing, with a noticeable acceleration after 2010. However, as highlighted by Griffiths and Rooney (2006), mortality trends for mental conditions have been significantly influenced by changes in the coding rules for causes of death. Therefore, these trends should be interpreted with caution.

- In Fig. 13, we observe a significant spike in κ_t in 2020 for the "rest of causes" group which includes deaths from COVID-19. Notably, aside from this surge in κ_t for the "rest of causes" group, there are no major deviations in mortality trends for

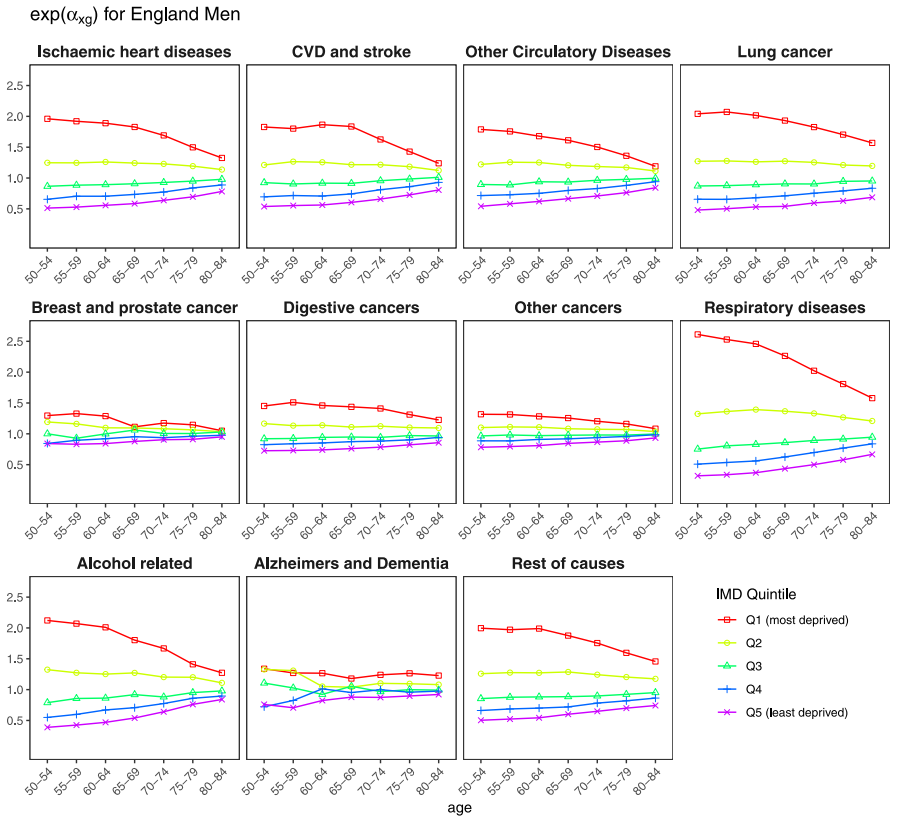

Fig. 14 Mortality level differentials for the different causes of death across deprivation quintiles of England Men ($\exp(\alpha_{xg})$)

the other groups of causes in 2020. The only exceptions are a slight increase for alcohol-related mortality and a slight decrease in mortality for respiratory diseases.

3.4 Historical cause-specific mortality differentials: England 2001-2020

In the Three-Way Lee-Carter model we can assess level and trend differentials in mortality by examining parameters α_{xg} and λ_g, respectively. Figures 14 and 15 plot the estimated values of $\exp(\alpha_{xg})$ for the different causes of death in the England deprivation subpopulations. Recalling the discussion in Sect. 2.4 and since in we are using year 2020 as the reference year in our modelling (i.e. $\kappa_{2020} = 0$), $\exp(\alpha_{xg})$ can be interpreted as the estimated percentage deviation in 2020 of mortality at age x in subpopulation g relative the England and Wales reference population. Similarly, the values of $\exp(\alpha_{x,Q1}) \big/ \exp(\alpha_{x,Q5})$, which are reported in Table 3, represent the estimated mortality rate ratio at age x between the most and least deprived quintiles of England. These quantities give an indication of relative mortality differentials in

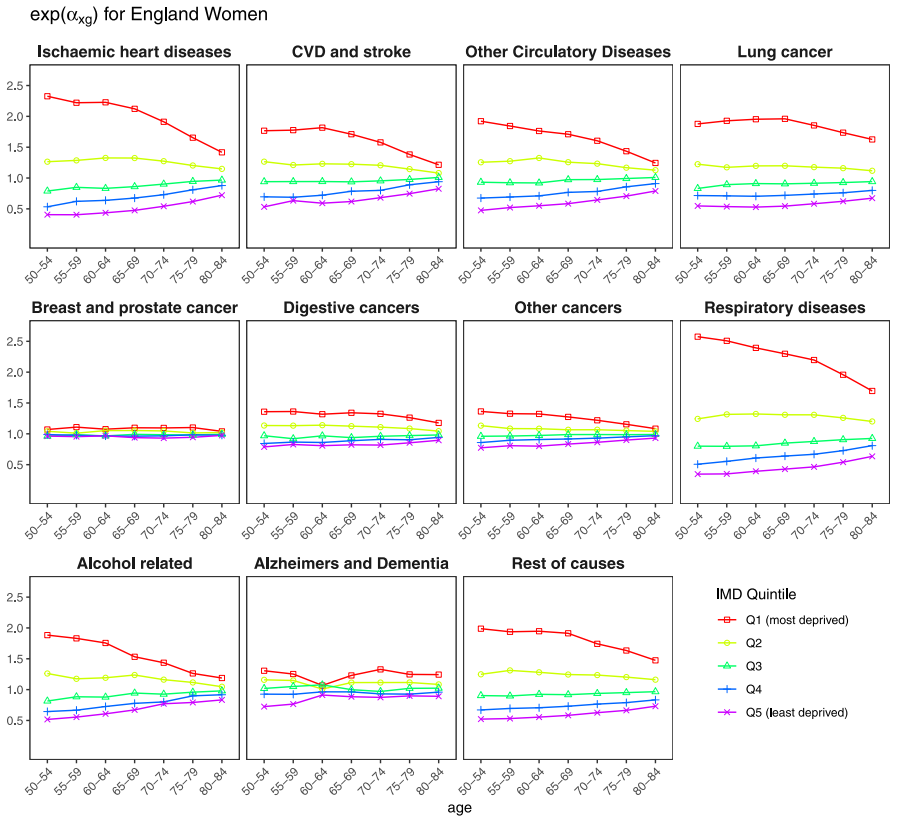

Fig. 15 Mortality level differentials for the different causes of death across deprivation quintiles of England Women ($\exp(\alpha_{xg})$)

2020 for the different causes of death. From Figs. 14 and 15 and Table 3 we note the following:

- There is a consistent pattern of mortality level differentials across all eleven groups of causes, with more deprived quintiles exhibiting significantly higher death rates than less deprived ones. However, these differentials vary notably across causes, genders, and ages.

- IHD, lung cancer, respiratory diseases, and alcohol-related causes exhibit the most pronounced relative mortality differentials, while breast and prostate cancers show the least variation across deprivation quintiles. Notably, the mortality differentials for alcohol-related and respiratory causes in the male population are significant. For instance, men aged 50-54 in the most deprived quintile have over five times the mortality rate for these two causes of death compared to their counterparts in the least deprived quintile. In contrast, relative differentials in breast cancer mortality are minimal, with women in the most deprived quintile experiencing only 7% to 18% higher mortality than those in the least deprived quintile. These findings align

Table 3 Relative mortality level differentials for the different causes of death between the most and least deprived quintiles of England ($\exp(\alpha_{x,Q1}) / \exp(\alpha_{x,Q5})$)

	50–54	55–59	60–64	65–69	70–74	75–79	80–84
Men							
Ischaemic heart diseases	3.82	3.64	3.39	3.12	2.65	2.15	1.69
CVD and stroke	3.4	3.25	3.32	3.03	2.47	1.97	1.54
Other Circulatory Diseases	3.3	3.02	2.7	2.42	2.12	1.78	1.41
Lung cancer	4.24	4.12	3.79	3.57	3.06	2.71	2.29
Breast and prostate cancer	1.54	1.6	1.53	1.27	1.3	1.26	1.1
Digestive cancers	2	2.07	1.97	1.89	1.8	1.6	1.43
Other cancers	1.68	1.65	1.59	1.49	1.39	1.31	1.16
Respiratory diseases	8.18	7.51	6.65	5.2	4.04	3.13	2.37
Alcohol related	5.46	4.87	4.28	3.34	2.6	1.85	1.52
Alzheimers and Dementia	1.76	1.8	1.54	1.34	1.42	1.41	1.33
Rest of causes	3.97	3.77	3.66	3.12	2.71	2.28	1.96
Women							
Ischaemic heart diseases	5.73	5.49	5.11	4.46	3.53	2.67	1.96
CVD and stroke	3.32	2.81	3.07	2.76	2.32	1.85	1.47
Other Circulatory Diseases	4.04	3.54	3.2	2.93	2.49	2.03	1.58
Lung cancer	3.43	3.59	3.7	3.6	3.18	2.79	2.42
Breast and prostate cancer	1.11	1.16	1.11	1.17	1.18	1.17	1.07
Digestive cancers	1.72	1.65	1.63	1.63	1.62	1.48	1.31
Other cancers	1.76	1.64	1.65	1.53	1.41	1.28	1.16
Respiratory diseases	7.43	7.16	6.1	5.37	4.73	3.62	2.67
Alcohol related	3.65	3.3	2.89	2.28	1.86	1.59	1.43
Alzheimers and Dementia	1.8	1.64	1.17	1.39	1.52	1.39	1.39
Rest of causes	3.81	3.65	3.52	3.28	2.78	2.46	2.02

with those reported by Arık et al. (2021), who suggest that deprivation is not a significant factor in explaining changes in breast and prostate cancer mortality rates.

- In terms of age, there is a general decrease in relative mortality differentials as people get older. For instance, men and women aged 50-54 in the most deprived quintile have, respectively, 3.82 and 5.73 higher mortality from IHD than persons of the same sex and age in the least deprived quintile. By contrast, at age 80-84 the mortality ratio from IHD between the most deprived quintile and the least deprived quintile reduces to 1.69 and 1.96 for men and women, respectively.
- In terms of gender, there are generally bigger socio-economic mortality differences in men than in women for all causes with the notable exception of IHD, where women show significantly higher differentials than men.

Figure 16 depicts 95% confidence intervals of parameters λ_g for the different causes of deaths. The corresponding central estimates are reported in Table 4. Since the

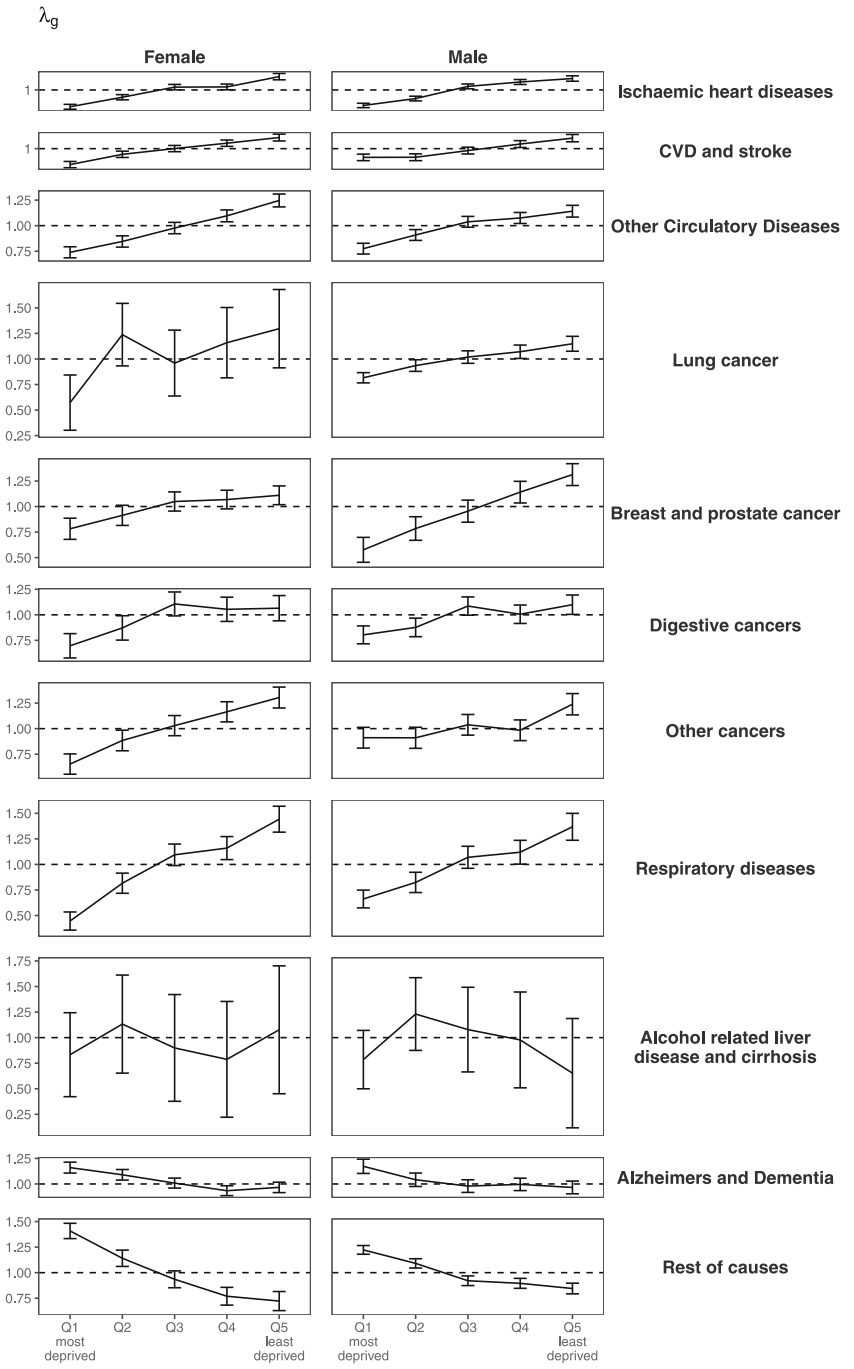

Fig. 16 Mortality trend differentials for the different causes of death for the deprivation quintiles of England (λ_g)

Table 4 Mortality trend differentials for the different causes of death for the deprivation quintiles of England (λ_g). AI: average percentage annual improvement in England and Wales

	λ_{Q1}	λ_{Q2}	λ_{Q3}	λ_{Q4}	λ_{Q5}	$AI_{2001-2020}$	$AI_{2010-2020}$
Men							
Ischaemic heart diseases	0.85	0.92	1.03	1.08	1.11	4.69	3.12
CVD and stroke	0.91	0.92	0.98	1.05	1.10	4.85	4.23
Other Circulatory Diseases	0.78	0.91	1.04	1.08	1.14	2.23	1.58
Lung cancer	0.82	0.94	1.02	1.07	1.15	3.09	3.53
Breast and prostate cancer	0.58	0.78	0.95	1.14	1.31	1.88	1.85
Digestive cancers	0.80	0.88	1.09	1.01	1.10	1.53	1.45
Other cancers	0.91	0.91	1.04	0.98	1.24	1.23	1.08
Respiratory diseases	0.66	0.82	1.07	1.12	1.37	2.35	2.24
Alcohol related	0.79	1.23	1.08	0.98	0.65	- 1.00	−0.90
Alzheimers and Dementia	1.17	1.04	0.98	1.00	0.97	−3.11	−5.15
Rest of causes	1.22	1.09	0.92	0.90	0.85	−3.20	−6.52
Women							
Ischaemic heart diseases	0.84	0.93	1.03	1.03	1.13	5.32	4.07
CVD and stroke	0.84	0.94	1.00	1.05	1.11	4.84	4.51
Other Circulatory Diseases	0.74	0.85	0.98	1.10	1.25	2.39	1.96
Lung cancer	0.57	1.24	0.96	1.16	1.30	0.33	1.28
Breast and prostate cancer	0.78	0.91	1.05	1.07	1.11	2.14	1.95
Digestive cancers	0.70	0.87	1.11	1.05	1.06	1.38	1.19
Other cancers	0.65	0.88	1.03	1.16	1.30	1.24	1.06
Respiratory diseases	0.45	0.82	1.09	1.16	1.44	1.48	1.87
Alcohol related	0.83	1.13	0.90	0.79	1.08	−0.99	−1.17
Alzheimers and Dementia	1.16	1.09	1.01	0.93	0.97	−3.65	−5.85
Rest of causes	1.41	1.14	0.94	0.77	0.72	−2.39	−4.72

interpretation of λ_g is closely linked to the direction and pace of mortality change for a specific cause (see Sect. 2.4), Table 4 also includes average annual cause-specific improvement rates in England and Wales for the periods 2001-2020 and 2010-2020. This latter period is of interest as several causes of death seem to have experienced a trend change after 2010. The average improvement (AI) rates have been computed as $AI_{2001-2020} = \frac{1}{k} \times \frac{\kappa_{2001}}{2020-2001}$ and $AI_{2010-2020} = \frac{1}{k} \times \frac{\kappa_{2010}}{2020-2010}$, where $k = 7$ is the number of age bands used in the fitting of the models.

Figure 16 and Table 4 reveal the following in relation to mortality trend differentials across deprivation quintiles of England:

- Trend differentials are not as clear and consistent as level differentials, with the plots of λ_g showing a variety of patterns for the different causes of death.
- For circulatory causes, for instance, there is a clear socio-economic gradient in mortality improvements for both genders, with least deprived quintiles experiencing significantly faster mortality improvements than most deprived quintiles

across the three associated cause groups: IHD, CVD and stroke and other circulatory diseases. Since circulatory diseases were the main cause of death over the 2001-2020 period, improvement differentials from this cause are the main determinant of trend differentials in all-cause mortality. Therefore, it is not surprising to see that the magnitude of improvement differentials in mortality from the three circulatory disease groups of causes are in close agreement with the all-cause trend differentials across deprivations quintiles we have reported before in Villegas and Haberman (2014).

- The trend differentials among various types of cancers are diverse. For men's lung cancer and women's other cancers, there is a pronounced gradient, with the least deprived quintiles showing much faster improvements compared to the most deprived quintiles. In contrast, for women's lung cancer, men's other cancers, and digestive cancers in both genders, the gradient is less distinct. In many cases, the differences in improvement are not significant, as indicated by the error bars for λ_g crossing 1. Finally, while breast and prostate cancers exhibit only mild socio-economic differentials in mortality levels, there are clear differences in trends across groups. Specifically, mortality from prostate cancer in men and breast cancer in women, although to a lesser extent, is improving much more rapidly in the least deprived quintile compared to the most deprived quintile. This suggests that mortality differentials for these two cancers may become more pronounced in the future.
- For alcohol-related liver disease and cirrhosis, the estimates of λ_g are highly uncertain, with error bars crossing 1 for all quintiles. This indicates non-significant trend differentials. This finding stands in stark contrast to the level differentials, where the most deprived quintiles display significantly higher mortality than the least deprived quintiles
- The significant increase in mortality from Alzheimer's and Dementia for both genders since 2000 has been accompanied by pronounced socio-economic disparities. Notably, the mortality rates for the most deprived quintiles have been worsening at a much faster pace compared to the other quintiles.
- Likewise, for the rest of causes group, the modest increase in mortality from 2000 to 2019, followed by the COVID-19 mortality spike in 2020, has been accompanied by a pronounced socio-economic gradient. The most deprived quintiles have experienced a more significant worsening in mortality compared to the least deprived quintiles,
- The very minor differences between the λ_g values for the two least deprived quintiles, Q4 and Q5, across several causes of death suggest that these quintiles have experienced similar mortality improvements. This observation aligns with findings reported by Villegas and Haberman (2014), where it was noted that the differences in all-cause mortality improvements for these two quintiles during the 1981-2007 period were minimal.

3.5 Cause-specific mortality differential projections

To project cause-specific mortality rates and examine the possible trends in mortality differentials, we use ARIMA models to extrapolate the time index κ_t for each cause of death. Table 5 specifies the ARIMA(p, d, q) model used for each of the 11 causes of death. These models have been selected among all possible combinations of $p, d, q \in \{0, 1, 2\}$ to optimise the BIC of the model. It is important to note that for the rest of causes group, we excluded the data point from 2020, κ_{2020}, due to the unusual spike in mortality rates attributed to COVID-19. Similarly, in subsequent projections for this group, we continue the trends observed from 1968 to 2019 and disregard the 2020 data point.

Figures 17 and 18 depict for men and women in England, respectively, the projected evolution of deprivation-specific mortality rates at selected ages for the different causes of death. Both figures clearly show that the Three-way Lee-Carter model with code-adjustments is in general able to fit and forecast satisfactorily the mortality trends across deprivation subgroups and causes of death. However, lung cancer in women is a notable exception, with the age- and deprivation-specific fitted values failing to capture the patterns observed in historical rates. This discrepancy may be due to our model not accounting for cohort effects, which are known to be a significant factor in lung cancer mortality. Therefore, caution is advised when interpreting the results and projections for this cause.

Figures 19 and 20 show predictions of age standardised death rates for the different causes of death for the England and Wales male and female reference populations, along with matching ASDRs by deprivation quintile in England. These figures show that forecasts for most causes of death and deprivation quintiles look visually consistent with the historical evolution of mortality.

The exceptional performance of the Lee-Carter model with code adjustments in capturing the variety of mortality trends observed among the different causes of death is noteworthy, especially considering the work of Di Cesare and Murphy (2009) who suggest that different modelling approaches may be required for different causes of death. However, in line with Di Cesare and Murphy (2009), we have found that the Lee-Carter model is unable to capture the inverted U-Shape observed in lung cancer mortality for women.

In order to examine the possible future evolution of mortality differentials, we report in Tables 6 and 7 the historical and projected values of relative differences and absolute differences in ASDRs between the most and least deprived quintiles for the 11 groups of causes of death. From these tables we can distinguish different patterns in the trends of relative and absolute mortality differentials among the different causes:

- Firstly, some causes have exhibited a reduction in absolute mortality differentials, but simultaneously, due to faster mortality improvements in the least deprived subpopulations, they have experienced a widening of mortality differentials in relative terms. This group includes IHD, CVD and stroke, and other circulatory diseases for both genders, as well as lung cancer for men. For example, the ASDR for CVD and stroke among men in the most deprived quintile increased from 1.51 times the ASDR of men in the least deprived quintile in 2001 to 1.81 times in

Table 5 ARIMA model used to forecast the time index κ_t for the different causes of death

Cause	Men	Women
Ischaemic heart diseases	ARIMA(1,2,2)	ARIMA(0,2,2)
CVD and stroke	ARIMA(0,2,1)	ARIMA(0,2,1)
Other Circulatory Diseases	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift
Lung cancer	ARIMA(0,2,2)	ARIMA(0,2,1)
Breast and prostate cancer	ARIMA(2,2,2)	ARIMA(2,2,2)
Digestive cancers	ARIMA(1,1,0) with drift	ARIMA(0,1,1) with drift
Other cancers	ARIMA(0,2,2)	ARIMA(0,2,2)
Respiratory diseases	ARIMA(2,1,2) with drift	ARIMA(1,0,1) with non-zero mean
Alcohol related liver disease and cirrhosis	ARIMA(0,1,0) with drift	ARIMA(1,1,0) with drift
Alzheimers and Dementia	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift
Rest of causes	ARIMA(0,2,1)	ARIMA(0,2,1)

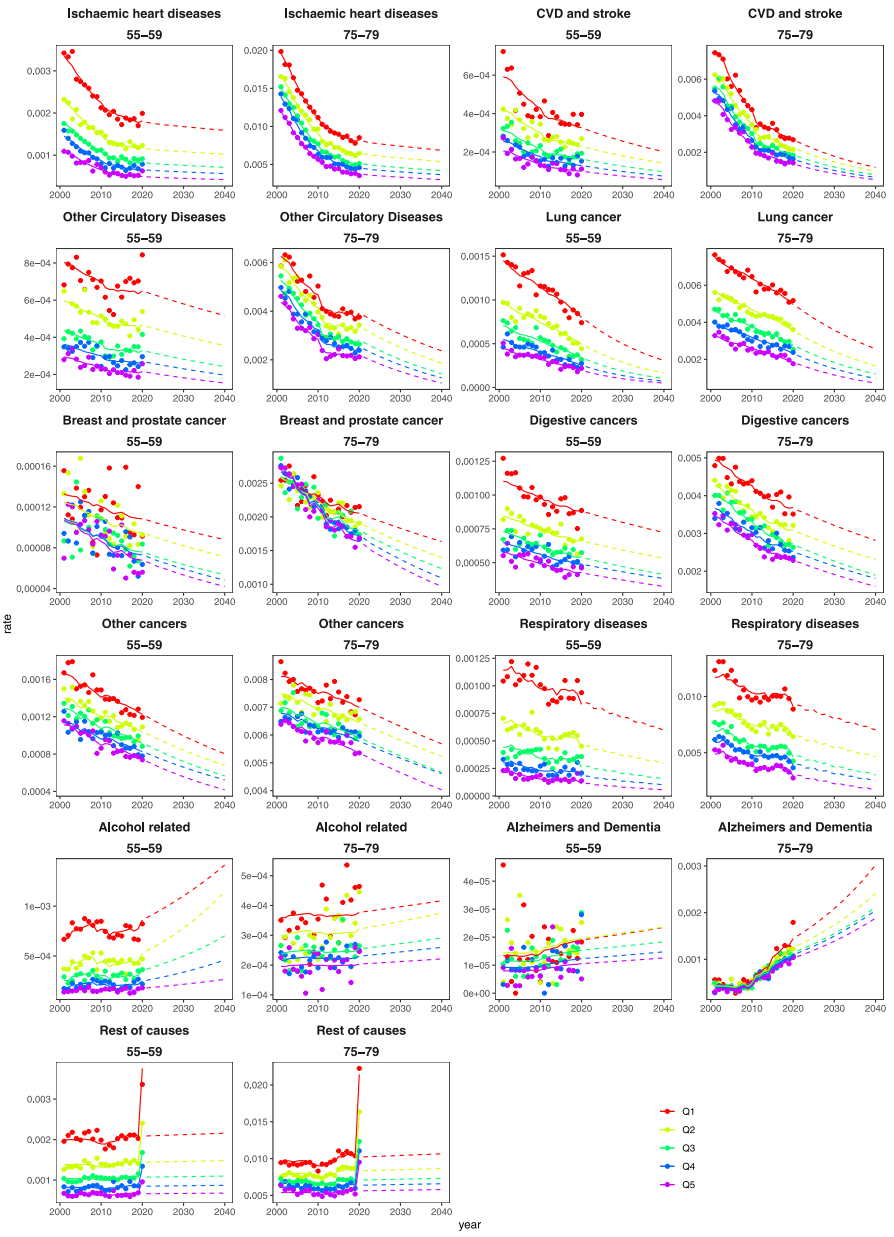

Fig. 17 Time series of fitted and forecasted cause-specific mortality rates μ_{xtg} at selected ages for the deprivation subpopulation of England males. The dots show observed rates for the period 2001-2020

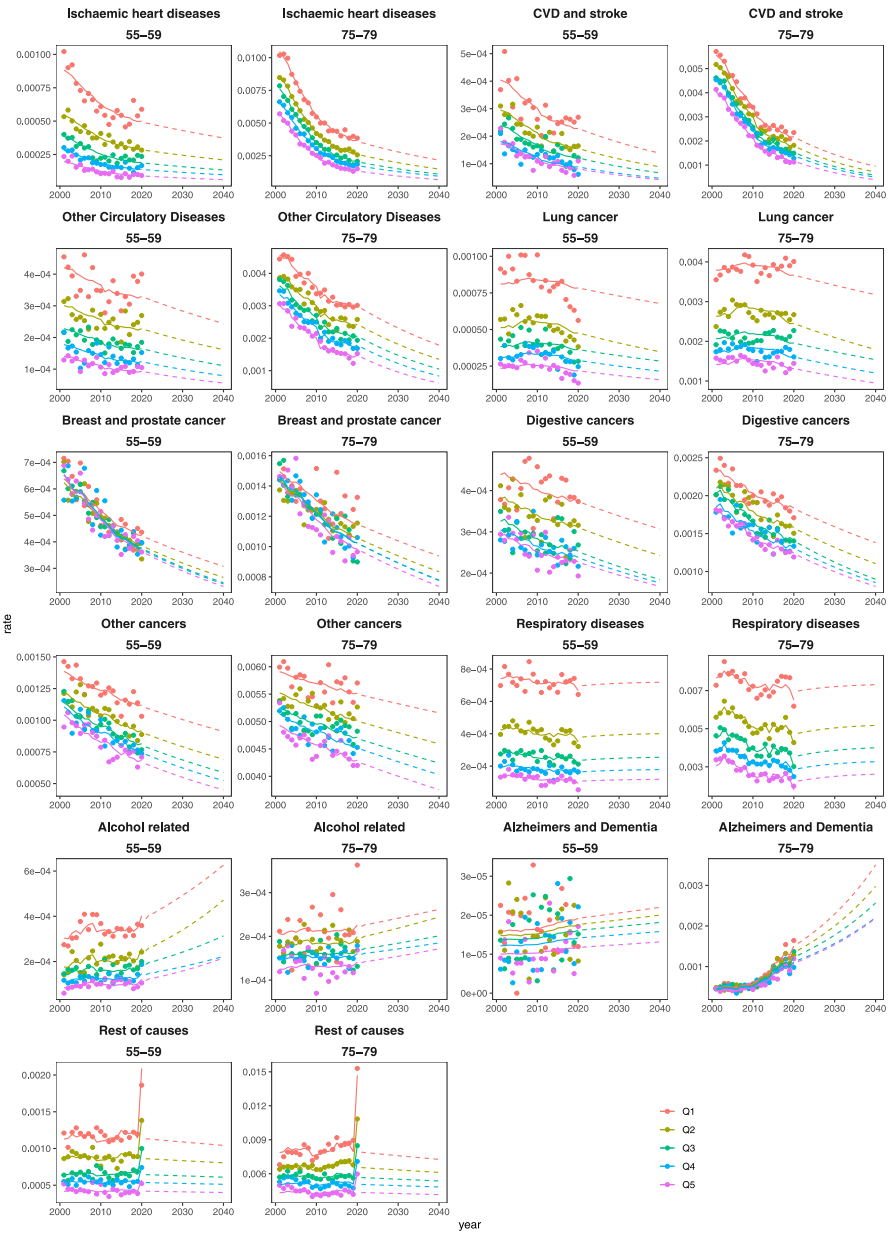

Fig. 18 Time series of fitted and forecasted cause-specific mortality rates μ_{xtg} at selected ages for the deprivation subpopulation of England females. The dots show observed rates for the period 2001-2020

Table 6 Ratio of age standardised mortality rates between the most and least deprived populations for the different causes of death. Values up to 2020 are observed ratios and post 2020 are projected ratios

	2001	2005	2010	2015	2020	2025	2030	2035	2040
Men									
Ischaemic heart diseases	1.72	1.85	2.00	2.24	2.41	2.28	2.28	2.23	2.11
CVD and stroke	1.51	1.65	1.74	1.81	2.00	2.00	2.07	2.13	2.17
Other Circulatory Diseases	1.40	1.55	1.67	1.69	1.88	1.87	1.97	2.07	2.17
Lung cancer	2.26	2.48	2.62	2.83	2.87	3.02	3.20	3.38	3.55
Breast and prostate cancer	0.97	0.97	1.09	1.20	1.20	1.28	1.31	1.26	1.12
Digestive cancers	1.52	1.50	1.56	1.70	1.62	1.70	1.73	1.77	1.80
Other cancers	1.30	1.21	1.27	1.30	1.33	1.34	1.36	1.37	1.38
Respiratory diseases	2.29	2.56	2.49	2.65	3.19	3.14	3.43	3.58	3.86
Alcohol related liver disease and cirrhosis	2.67	3.07	3.07	3.02	3.23	3.39	3.51	3.63	3.76
Alzheimers and Dementia	1.16	1.13	1.12	1.13	1.48	1.43	1.49	1.56	1.64
Rest of causes	1.60	1.75	1.70	2.04	2.37	1.89	1.91	1.94	1.98
Women									
Ischaemic heart diseases	1.84	1.98	2.19	2.59	2.83	2.75	2.84	2.87	2.84
CVD and stroke	1.31	1.46	1.45	1.64	1.89	1.88	1.99	2.09	2.18
Other Circulatory Diseases	1.48	1.54	1.74	1.81	2.03	2.10	2.27	2.45	2.64
Lung cancer	2.50	2.63	2.58	2.96	2.93	3.11	3.22	3.27	3.25
Breast and prostate cancer	0.97	1.01	1.07	1.12	1.22	1.15	1.18	1.19	1.20
Digestive cancers	1.31	1.42	1.44	1.53	1.44	1.52	1.57	1.62	1.67
Other cancers	1.20	1.24	1.25	1.36	1.36	1.37	1.39	1.41	1.43
Respiratory diseases	2.19	2.43	2.52	2.89	3.28	2.88	2.79	2.73	2.70
Alcohol related liver disease and cirrhosis	2.29	2.19	2.69	2.40	2.17	2.32	2.29	2.25	2.22
Alzheimers and Dementia	1.23	1.04	1.23	1.28	1.38	1.47	1.54	1.62	1.70
Rest of causes	1.47	1.58	1.75	1.99	2.36	1.74	1.74	1.75	1.77

Table 7 Absolute difference in age standardised mortality rates between the most and least deprived populations for the different causes of death. Values are reported in deaths per 1000 persons.. Values up to 2020 are observed ratios and post 2020 are projected ratios

	2001	2005	2010	2015	2020	2025	2030	2035	2040
Men									
Ischaemic heart diseases	4.84	4.21	3.44	3.14	3.10	2.74	2.68	2.67	2.70
CVD and stroke	1.35	1.31	1.04	0.86	0.79	0.65	0.56	0.49	0.43
Other Circulatory Diseases	0.99	1.08	1.12	0.94	1.14	0.99	0.94	0.89	0.84
Lung cancer	2.39	2.37	2.32	2.21	1.92	1.62	1.41	1.23	1.08
Breast and prostate cancer	- 0.04	- 0.04	0.11	0.20	0.19	0.23	0.24	0.20	0.11
Digestive cancers	1.01	0.91	0.95	1.03	0.88	0.91	0.89	0.87	0.84
Other cancers	1.11	0.79	0.95	1.01	1.06	1.00	0.99	0.97	0.95
Respiratory diseases	3.69	4.08	3.44	3.49	3.37	3.36	3.18	3.08	2.92
Alcohol related liver disease and cirrhosis	0.30	0.38	0.37	0.36	0.44	0.48	0.54	0.61	0.68
Alzheimers and Dementia	0.04	0.02	0.03	0.06	0.27	0.30	0.42	0.58	0.79
Rest of causes	2.10	2.47	2.28	3.59	7.93	3.11	3.22	3.39	3.62
Women									
Ischaemic heart diseases	2.63	2.23	1.71	1.57	1.46	1.20	1.10	1.03	0.97
CVD and stroke	0.73	0.83	0.59	0.60	0.62	0.49	0.44	0.38	0.34
Other Circulatory Diseases	0.82	0.80	0.86	0.81	0.89	0.82	0.78	0.74	0.69
Lung cancer	1.27	1.38	1.45	1.61	1.48	1.43	1.40	1.37	1.33
Breast and prostate cancer	- 0.03	0.01	0.06	0.10	0.15	0.10	0.11	0.11	0.11
Digestive cancers	0.32	0.38	0.38	0.41	0.33	0.36	0.36	0.36	0.35
Other cancers	0.61	0.69	0.67	0.91	0.91	0.87	0.89	0.90	0.91
Respiratory diseases	2.17	2.62	2.36	2.71	2.48	2.61	2.60	2.59	2.58
Alcohol related liver disease and cirrhosis	0.13	0.14	0.16	0.15	0.16	0.18	0.19	0.20	0.22
Alzheimers and Dementia	0.05	0.01	0.06	0.13	0.26	0.39	0.57	0.81	1.14
Rest of causes	1.39	1.71	2.00	2.81	5.28	2.03	2.01	2.03	2.09

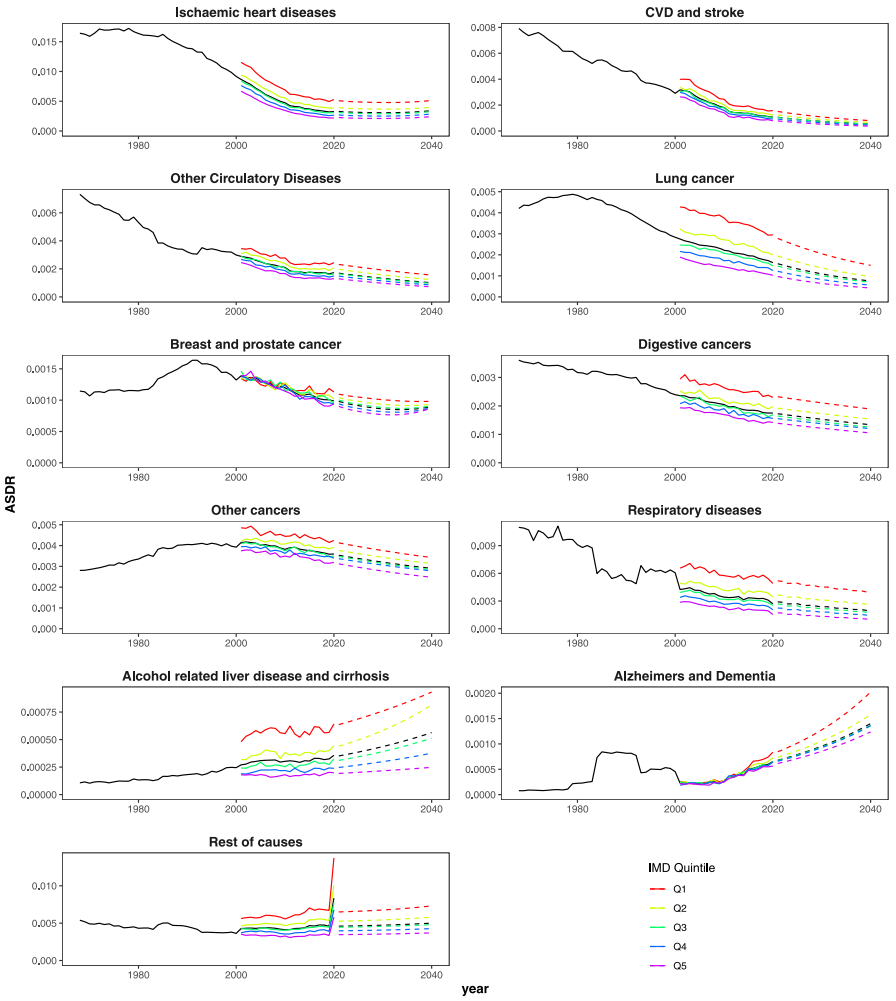

Fig. 19 Age standardised death rates (ASDR) projections for the age range 50-84 for selected causes of death in England for Males by deprivation quintile. Solid lines show observed ASDRs and dashed lines projected ASDRs

2015. Moreover, if this trend continues at the same rate observed from 2001 to 2020, the ratio between the ASDR for CVD and stroke among men in the most deprived quintile and the least deprived quintile could reach 2.17 by 2040.

- Secondly, other causes have shown an increase in relative differentials while absolute differentials remained mostly unchanged. This category is predominantly cancer-related, with digestive cancer being a significant example. For instance, the ratio of ASDR for digestive cancer among women in the most deprived quintile compared to the least deprived quintile increased from 1.31 in 2001 to 1.53 in 2015, and it is projected to rise to 1.67 by 2040. Meanwhile, the absolute differ-

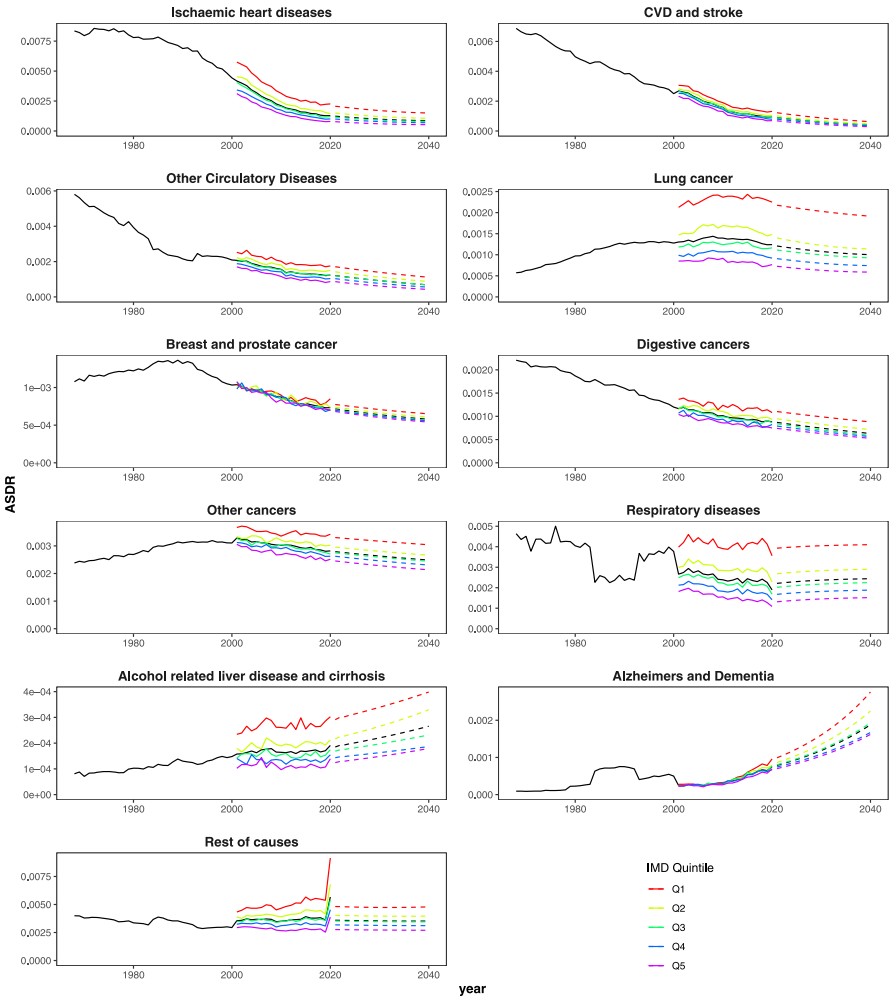

Fig. 20 Age standardised death rates (ASDR) projections for the age range 50-84 for selected causes of death in England for females by deprivation quintile. Solid lines show observed ASDRs and dashed lines projected ASDRs

ence in ASDR between the two extreme quintiles has remained (and is projected to remain) stable at around 0.36 deaths per 1000 women.

- Thirdly, Alzheimer's and Dementia, along with the rest of the causes of death, have experienced a widening of mortality differentials in both absolute and relative terms during the 2001-2020 period. These differentials are also projected to continue widening in the future.

4 Discussion

In this paper we have developed Lee-Carter based techniques for the modelling of mortality by cause of death and socio-economic stratification. An application to mortality data for deprivation subpopulations in England showed that these modelling techniques offer a simple, yet effective, approach for the assessment of the magnitude of historical mortality differentials for the main causes of death and for the projection of the possible future evolution of these differentials. Key to the success of our approach was the integration into the Lee-Carter framework of modelling tools to control for data production changes in cause of death data which, if not taken into account appropriately, could hinder the validity of any analysis of trends in cause-specific mortality.

The empirical part of this paper revealed a clear association between area deprivation and mortality rates, with people living in more deprived areas of England consistently having higher mortality rates for the leading causes of death than those living in less deprived areas. The mortality differentials shown in this study are substantial although with varying magnitude across causes, sexes and ages. IHD, lung cancer, respiratory diseases and alcohol related causes of death showed the highest differences over the 2001-2020 period of study, with mortality in the most deprived quintile for some ages and causes of death being more than five times the mortality rates of the least deprived quintile. Breast and prostate cancer, by contrast, showed the smallest relative mortality differences.

We have observed that mortality differentials are widening in relative terms across all cause groups for men, and across all but alcohol-related liver disease for women. However, in absolute terms, the trends vary among causes of death. Cardiovascular diseases have shown a reduction in absolute differentials, cancer-related groups have mostly stable absolute differentials, while Alzheimer's and Dementia, along with the rest of causes group, have experienced a widening in absolute differentials. This variation in trends highlights the need for a nuanced approach in policy formulation to address mortality inequalities, as emphasised by Mackenbach (2015). Relative differentials, by underlining proportional disparities, draw attention to systemic inequalities and the necessity for broader social interventions. Conversely, absolute differentials, by revealing the actual number of lives impacted, can inform targeted health policies and resource allocation, particularly addressing critical areas such as the increasing burden of Alzheimer's and Dementia.

With the noticeable exception of ischaemic heart disease, we have also found that mortality differentials are higher among men than among women. Furthermore, clearer trend mortality differentials were observed in the male population. These results are consistent with previous studies using area-based deprivation measures in the UK (Romeri et al. 2006; Villegas and Haberman 2014) and other developed countries (Singh 2003; Turrell and Mathers 2001; Windenberger et al. 2011) and with other UK (White et al. 2003) and international (Saurel-Cubizolles et al. 2009) studies using alternative markers of socio-economic conditions (e.g. occupation and educational attainment).

There are a number of possible limitations with the modelling approach introduced in this paper. The first and most important one is that, both in the fitting and fore-

casting of the models, we have assumed independence among the different causes of death. This is clearly an unrealistic assumption as, for instance, the same risk factors can affect several causes at the same time. An alternative approach to allow for the dependence between the causes of death is offered by Vector Autoregression (VAR) and Vector Error Correction Models (VECMs) as suggested by Arnold (-Gaille) and Sherris (2013); Arnold Gaille and Sherris (2015). Hence, one could use such multivariate time series techniques to model and forecast simultaneously the time series indices, κ_t, for the different causes of death, as opposed to the independent univariate ARIMA models used in this paper.

Second, our modelling approach is an age-period model which assumes that the mortality evolution for each cause of death is driven exclusively by a single period index, κ_t. Although this has proven to be sufficient for the restricted age range we have considered in this study, additional period factors may be required if looking at a wider age range encompassing younger ages whose mortality change may be driven by different risk factors. Furthermore, some causes of death characterised by clear cohort effects may require the inclusion of a cohort term and could be better modelled using an age-period-cohort approach (Di Cesare and Murphy 2009; Cairns 2023). We note, however, that in spite of the lack of an explicit cohort term, the Lee-Carter model with coding adjustment showed a reasonable goodness-of-fit and forecasting performance for circulatory diseases and most cancers, causes of death for which Willets (2004) has reported clear cohort effect among the England and Wales population. The surprising good performance of the Lee-Carter with code adjustments when applied to these two group of causes is explained by an unexpected side effect of the inclusion of the coding adjustments parameters, $\delta_x^{(i)}$, which seem to be partially capturing some of the cohort effects associated to these causes of death. These encouraging results also suggests that the piece-wise technique we have introduced to account for coding adjustments may be useful in other setting not related to cause-specific mortality such as when modelling all-cause mortality in populations which have undergone structural changes in their mortality trend (see e.g. Van Berkum et al. (2016)).

Finally, the Three-way Lee-Carter model used to capture socio-economic differences in mortality assumes, for the sake of parsimony, that mortality trend differentials are constant across ages. This is despite the fact that trend differentials may show important variations by age as reported, for instance, in Bajekal et al. (2013) for coronary heart diseases. Hence, a subject of further investigation is the development of modelling approaches which allow for a more complex age-structure of mortality trend differentials.

We also recognise additional limitations of our study stemming from the dataset used in the analysis of socio-economic difference in cause-specific mortality in England. First, we have used an ecological measure of socio-economic deprivation instead of individual level socio-economic measures. This leads to the usual problem of any ecological study whereby outcomes obtained at the group level do not necessarily apply at the individual level (Greenland 2001). This ecological design also limits the ability of associating a causal pathway between socio-economic conditions and mortality, for which individual level information would be required.

Second, we note that our data measure mortality rates for those living in a particular area at a particular point in time, ignoring any past exposure to risk factors whilst living in other areas. It is plausible that healthier people will tend to migrate from more deprived areas to less deprived ones and that less healthy people remain at home, resulting in a potential bias towards higher mortality inequalities (Norman et al. 2005).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability This research uses publicly available data.

Declarations

Conflict of interest Steven Haberman is a guest editor for the special issue in which this paper appears. Madhavi Bajekal is an honorary member of staff at UCL, funded by Legal & General Assurance Society Limited (L&G) as part of its wider research collaboration with UCL on longevity research. L&G did not have any role in the study design, data collection, analysis, interpretation, writing, approval or submission of the manuscript. The authors do not have other Conflict of interest to declare that are relevant to the content of this article.

References

Arık, A., Dodd, E., Cairns, A., Streftaris, G.: Socioeconomic disparities in cancer incidence and mortality in England and the impact of age-at-diagnosis on cancer mortality. PLoS One **16**(7), e0253854 (2021)

Arnold (-Gaille), S., & Sherris, M: Forecasting Mortality Trends Allowing for Cause-of-Death Mortality Dependence. N. Am. Actuar. J. **17**, 273–282 (2013). https://doi.org/10.1080/10920277.2013.838141. (**ISSN 1092-0277.**)

Arnold (-Gaille), S., & Sherris, M: Causes-of-Death Mortality: What Do We Know on Their Dependence? N. Am. Actuar. J. **19**(2), 116–128. https://doi.org/10.1080/10920277.2015.1011279

Bajekal, M., Scholes, S., O'Flaherty, M., Raine, R., Norman, P., Capewell, S.: Unequal trends in coronary heart disease mortality by socioeconomic circumstances, England 1982-2006: an analytical study. PLoS ONE, 8, (2013). https://doi.org/10.1371/journal.pone.0059608

Brock, A., Griffiths, C., Rooney, C.: The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales. Health Stat Q. 9–17 (2006)

Cairns, A.J.G.: The common cohort effect model for cause of death data (2023)

Continuous Mortality Investigation. In: Working Paper 3 – Projecting future mortality: A discussion paper (2004)

Cesare, D., Mariachiara, M., Michael, J.: Forecasting mortality, different approaches for different cause of deaths? The cases of lung cancer; influenza, pneumonia, and bronchitis; and motor vehicle accidents. Brit. Actuar. J. **15**, 185–211 (2009)

Djeundje, V.B., Haberman, S., Bajekal, M., Lu, J.: The slowdown in mortality improvement rates 2011–2017: a multi-country analysis. Europ. Actuar. J. **12**(2), 839–878 (2022)

GAD. National population projections : review of methodology for projecting mortality, 2001

Giona, S. : The epidemiology of prostate cancer. Exon Publications, pp 1–15 (2021)

Goldacre, M.J., Duncan, M.E., Cook-Mozaffari, P., Griffith, M.: Trends in mortality rates comparing underlying-cause and multiple-cause coding in an English population 1979–1998. J. Public Health **25**, 249–253 (2003). https://doi.org/10.1093/pubmed/fdg058

Greenland, S.: Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. Int. J. Epidemiol. **30**, 1343–50 (2001). (**ISSN 0300-5771**)

Griffiths, C., Rooney, C.: Trends in mortality from Alzheimer's disease, Parkinson's disease and dementia, England and Wales, 1979-2004. Health Statistics Quarterly, (2006)

Jatoi, I., Miller, A.B.: Why is breast-cancer mortality declining? Lancet Oncol. **4**, 251–254 (2003). https://doi.org/10.1016/S1470-2045(03)01037-4. (**ISSN 14702045.**)

Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**, 659–671 (1992)

Lu, J.L.C., Wong, W., Bajekal, M.: Mortality improvement by socio-economic circumstances in England (1982 to 2006). Brit. Actuar. J. **19**(1), 1–35 (2014)

Mackenbach, J.P.: Should we aim to reduce relative or absolute inequalities in mortality? Europ. J. Public Health **25**(2), 185–185 (2015)

Mayhew, L.: Smith, David: Gender convergence in human survival and the postponement of death. North Am. Actuar. J. **18**, 194–216 (2014). https://doi.org/10.1080/10920277.2013.863140. (**ISSN 1092-0277.**)

Mensah, G.A., Wei, G.S., Sorlie, P.D., Fine, L.J., Rosenberg, Y., Kaufmann, P.G., Mussolino, M.E., Hsu, L.L., Addou, E., Engelgau, M.M., et al.: Decline in cardiovascular mortality: possible causes and implications. Circulat. Res. **120**, 366–380 (2017)

Meyricke, R., Sherris, M.: The determinants of mortality heterogeneity and implications for pricing annuities. Insur. Math. Econom. **53**, 379–387 (2013). https://doi.org/10.1016/j.insmatheco.2013.06.002

Moriyama, I.M., Loy, R.M., Robb-Smith, A.H.T.: History of the statistical classification of diseases and causes of death. National Center for Health Statistics, (2011). ISBN 9780840606440

Noble, M., Mclennan, D., Wilkinson, K., Whitworth, A., Exley, S., Barnes, H., Dibben, C.: The English indices of deprivation 2007. Department of Communities and Local Government, (2007)

Norman, P.: Boyle, Paul, Rees, Philip: selective migration, health and deprivation: a longitudinal analysis. Soc. Sci. Med. **60**, 2755–71 (2005). https://doi.org/10.1016/j.socscimed.2004.11.008. (**ISSN 0277-9536.**)

Pace, M., Lanzieri, G., Glickman, M., Zupanič, T.: Revision of the European Standard Population: report of Eurostat's task force. Publications Office of the European Union, (2013)

Rey, G., Aouba, A., Pavillon, G., Hoffmann, R., Plug, I., Westerling, R., Jougla, E., Mackenbach, J.P.: Cause-specific mortality time series analysis: a general method to detect and correct for abrupt data production changes. Populat. Health Metrics **9**, 1–11 (2011). https://doi.org/10.1186/1478-7954-9-52. (**ISSN 1478-7954.**)

Richards, S.J.: Selected issues in modelling mortality by cause and in small populations. Brit. Actuar. J. **15**, 267–283 (2009)

Ridsdale, B.: Gallop, p. 2010. Adrian, Mortality by cause of death and by socio-economic and demographic stratification (2010)

Romeri, E.: Baker, Allan, Griffiths, Clare: mortality by deprivation and cause of death in 1999–2003. Health Statist. Quart. **32**, 19–34 (2006)

Rooney, C., Devis, T.: Mortality trends by cause of death in England and Wales 1980-94: the impact of introducing automated cause coding and related changes in 1993. Population Trends, pp 29–35, (1996)

Russolillo, M., Giordano, G., Haberman, S.: Extending the Lee-Carter model: a three-way decomposition. Scandinav. Actuar. J. **346**, 96–117 (2011). https://doi.org/10.1080/03461231003611933

Saurel-Cubizolles, M.-J., Chastang, J.-F., Menvielle, G., Leclerc, A., Luce, D.: Social inequalities in mortality by cause among men and women in France. J. Epidemiol. Commun. Health **63**, 197–202 (2009). https://doi.org/10.1136/jech.2008.078923

Singh, G.K.: Area deprivation and widening inequalities in US mortality, 1969–1998. Am. J. Public Health **93**, 1137–1143 (2003)

The Marmot Review. Fair society, healthy lives: strategic review of health inequalities in England post 2010, (2010). ISSN 1476-5616

Turrell, G.: Mathers, Colin: Socioeconomic inequalities in all-cause and specific-cause mortality in Australia: 1985–1987 and 1995–1997. Int. J. Epidemiol. **30**, 231–239 (2001). https://doi.org/10.1093/ije/30.2.231. (**ISSN 0300-5771.**)

Berkum, V.: Frank, Antonio, Katrien, Vellekoop, Michel: the impact of multiple structural changes on mortality predictions. Scandin. Actuar. J. **2016**(7), 581–603 (2016)

Villegas, A.M., Bajekal, M., Haberman, S., Zhou, L.: Key drivers of long-term rates of mortality improvements in the united states: period, cohort, and cause of death analysis, 1959–2016. North Am. Actuar. J. **20**, 1–31 (2023)

Villegas, A.M.: Haberman, Steven: On the modeling and forecasting of socioeconomic mortality differentials: an application to deprivation and mortality in England. North Am. Actuar. J. **18**, 168–193 (2014). https://doi.org/10.1080/10920277.2013.866034. (**ISSN 1092-0277.**)

Wen, J.: Cairns, Andrew JG, Kleinow, Torsten: modelling socio-economic mortality at neighbourhood level. ASTIN Bull. J. IAA **53**(2), 285–310 (2023)

White, C., van Galen, F., Chow, Y.H.: Trends in social class differences in mortality by cause, 1986 to 2000. Health Statistics Quarterly, 25–37, (2003)

Willets, R.C.: The cohort effect: insights and explanations. British Actuar. J. **10**, 833–877 (2004)

Windenberger, F., Rican, S., Jougla, E., Rey, G.: Spatiotemporal association between deprivation and mortality: trends in France during the nineties. European Journal of Public Health, 1–7, (2011). https://doi.org/10.1093/eurpub/ckr029

Wood, S.N.: Package mgcv, 2020. http://cran.r-project.org/web/packages/mgcv/index.html