Higher-Age US Mortality By Education and Cause of Death: Trends, Inequality and Controllable Risk Factors

Andrew J.G. Cairns[a], [b] and Cristian Redondo Lourés[c]

This version: May 27, 2022

Abstract

We present an empirical analysis of US mortality over the period 1989 to 2017 by gender, age, education level, and by cause of death. The use of 51 causes of death, rather than a much smaller number, allows us to focus on the roles of controllable, preventable and non-preventable risk factors as determinants of mortality inequality.

A wide range of graphical diagnostics is used to highlight particular features of the data that might not be clear if we were to rely on a small number of standard mortality plots. These are used in the group of vignettes that draw out our observations on conclusions regarding various groups of causes of death. In broad terms we find: considerable variation in improvement rates by cause; considerable variation in levels of mortality inequality by education level linked to the presence of controllable risk factors; and generally increasing levels of inequality over time.

Keywords

Cause of death mortality, education level, mortality inequality, graphical diagnostics, controllable risk factors, cohort effects.

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

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

[c]APR LLP

Acknowledgements

This work forms part of the research programme "Modelling, Measurement and Management of Longevity and Morbidity Risk" funded by the Actuarial Research Centre of the Institute and Faculty of Actuaries, the Society of Actuaries and the Canadian Institute of Actuaries. The authors gratefully acknowledge feedback from the respective steering and oversight groups.

Contents

1 Introduction

In this paper, we present a detailed, empirical analysis of mortality by cause of death. Typically, in actuarial work (life and pensions) we are primarily interested in current and future levels of all-cause mortality. Forecasting future mortality rates requires careful statistical analysis of historical mortality rates by calendar year and by age, with popular models including, for example, the Lee and Carter (1992) model and the CBD family of models (Cairns et al., 2006, 2009). However, it is sometimes argued that the purely-statistical approach that underpins these models leads to forecasts that lack a convincing narrative[1] and, therefore, cannot be relied upon.

This paper seeks to address this issue by looking in detail at mortality by cause of death. In doing so, we seek to gain a better understanding of historical changes in all-cause mortality. In recent decades, mortality has mostly declined but not in all years and not at all ages.

- Which causes of death have given rise to the biggest reductions in mortality, and at which ages?
- Which causes of death have been rising in recent years and which have had a significant impact on all-cause mortality?
- To what extent are these changes driven by medical advances and public-health campaigns, and what is the complementary impact of changes in people's lifestyles and habits?

This paper builds on a number of other papers on cause of death mortality in the actuarial literature including the IAA Mortality Forum report (Arık et al., 2021a) a series of papers by Arnold et al. (2013, 2015, 2016, 2021), and Gutterman (2021). Some of these are comprehensive in the sense of covering all causes of death while others (for example, Gutterman, 2021, and Arık et al., 2021b) focus more on specific causes with a link to a particular risk factor such as alcohol consumption or smoking. Additionally, Arık et al. (2021a) and Arnold and Glushko (2021) cover multiple countries while others (e.g. Redondo Lourés and Cairns, 2020, 2021) cover multiple sub-populations within a single country. Also, within the wide range of papers, there is considerable variation in the granularity of the breakdown of causes of death. In some cases, the number of cause-of-death groupings is relatively small and this is often driven by the underlying model being fitted and the application. For example, Arnold and Glushko (2021) group causes of death into five categories (infectious and parasitic diseases, cancers, circulatory, respiratory and external). In that case, the need for a relatively-small number is driven by the objective to fit a multivariate

[1] For example, in medical terms, the past cannot repeat itself: if we have found a cure for cancer, this cannot happen again in the future.

time-series model as the dependencies between causes and countries needs to be modelled and estimated.

In this paper, we use the relatively large number of 51 causes of death. Our reason for this is that we seek to understand the causal link between specific risk factors and all-cause mortality. In particular, what are the drivers of mortality inequality by education level? If we were to consider all cancers as one group then trends and inequalities would be linked to multiple risk factors, and it is difficult to know which risk factors are important. But, by subdividing, we can identify specific cancers that have well-researched links to specific risk factors: for example, the link between smoking and lung cancer. Additionally, the *relative risk* linking risk factors and causes of death varies considerably and this, in turn leads to significantly varying levels of inequality. As part of this discussion, we choose to distinguish between *controllable, preventable and non-preventable risk factors.* Controllable risk factors cover all risk factors that are in the control of the individual such as smoking and diet, and this group of risk factors is the main cause of mortality inequality that we discuss later in this paper. However, we also infer from our empirical analyses that some degree of inequality by education level exists even when there are no controllable risk factors (such as prostate and breast cancer). This might be due to differences (inequalities) in the availability and quality of healthcare following onset of a disease. But it might also be due to non-preventable risk factors such as *conscientiousness.*[2] People who are more conscientiousness, are more likely to get an early diagnosis following onset and then to follow a prescribed course of treatment.

A further reason for taking a granular approach is that it allows us to link mortality improvements to specific medical advances and then consider at the level of individual causes what the prospects are for future improvements and their consequence impact on all-cause mortality. Many treatments will be quite specific to individual causes rather than higher level groups. If we can understand how these have affected death rates from thse specific causes we can also then judge what the potential is for further reductions from that cause. Equally, we might see that other causes have seen little in the way of improvements with there being no obvious medical advances. We can then consider if the cause of death is simply very hard to treat (or to prevent the onset) or if there are reasonable prospects for a significant advance in the future.

1.1 Outline of the paper

The outline of the rest of the paper is as follows. The data used in the remainder of the paper is introduced in Section 2 including a list of the 51 causes of death. In Section 3 we summarise the key features of all-cause mortality for different groups. We then discuss cohort effects – a feature of both all-cause and cause-of-death mor-

[2]For example, Notfle and Robins (2007) conclude that conscientious is correlated with education achievement.

tality – in Section 4: statistically, what is a cohort effect and what causes cohort effects? In Section 5 we introduce the notion that identifiable individual risk factors might be classified as either controllable, preventable or non-preventable, and, in Section 6 we discuss why it is helpful in an investigation into mortality inequalities needs risk factors to be classified in this way.

Analysis and discussion of the cause-of-death data begins in Section 7. At this point we are zoomed out and consider 8 cause-of-death groups rather than the 51 we use later on in order to make some initial observations about the pattern of mortality. In Section 8 we focus on ischaemic heart disease and introduce the range of graphical diagnostics that help us to identify as many interesting nuances as we can from the available data. We then briefly digress in Section 9 and demonstrate how much the level of mortality inequality varies from one cause of death to another, linking the discussion to controllable risk factors. We also discuss in this section how each cause of death has experienced quite different improvement rates and what the reasons for this might be. We then continue our analysis of individual causes of death in Sections 10 to 14 with each section having a theme: smoking, alcohol, deaths of despair, accidents, neurological diseases, and causes with no significant risk factors. Each of these themes has its own story to tell and makes use of different graphical diagnostics.

Section 16 concludes and looks ahead to future work.

2 Data

This paper discusses death rates $m(g, e, c, y, x)$ where

- g is **sex** (male or female)
- e is the **education level**, with a low education meaning any level of education up to and including a high-school dimploma, and high education meaning high school plus any further college or university education (even if this did not lead to e.g. a bechelor's degree)
- c is the **cause of death**
- y is the **calendar year** (single years)
- x is the **age** (single ages)

For each cell (g, e, c, y, x) the death rate is defined as

$$m(g, e, c, y, x) = \frac{D(g, e, c, y, x)}{E(g, e, c, y, x)}$$

where $D(g, e, c, y, x)$ represents the death count for the cell and $E(g, e, y, x)$ represents the corresponding mid-year population estimate (an approximation to the central exposed to risk; exposures) with the cause of death excluded.

The data described below are available on request from the corresponding author.

2.1 Exposures

Exposures have been derived using data from two sources: the Human Mortality Database (HMD; www.mortality.org) for the total population; and the Current Population Survey to subdivide into low and high-educated. (See Redondo Lourés and Cairns, 2020, 2021 for further details of the split between low and high education.) Additionally, the methods of Cairns et al. (2014) were used to smooth out perceived anomalies in the US population data. Specifically, $E(g, e, y, x) = R(g, e, y, x)E(g, \text{all}, y, x)$, where $R(g, e, y, x)$ is the proportion of the (g, y, x) population with education level e, and $E(g, \text{all}, y, x)$ is the total population (all levels of education).

2.2 Deaths

Death counts by cause of death have been derived using data from the Centers for Disease Control and Prevention (CDC). (Again, see Redondo Lourés and Cairns, 2020, 2021.)

2.3 Range of years, ages and cohorts

Our data cover the period 1989 to 2017, and ages 40 to 84. However, within this range the oldest (born in 1913 or earlier) and the youngest (born in or after 1971) are excluded. This is because we have too few observations of these cohorts to be able to get reliable estimates of the split between low and high educated.

2.4 Causes of death

Deaths, $D(g, e, c, y, x)$, within each (g, e, y, x) cell have been subdivided by cause of death, c. In this study, we have used 51 causes of death (giving complete coverage of all causes). These are listed in Table 1, and a precise mapping between the 51 causes of death and the current ICD-10 codes can be found in Appendix A. The reason for this level of granularity is that we wish to link, as much as possible, specific causes of death to controllable risk factors. For example, lung cancer has smoking as the key controllable risk factor, whereas breast cancer has no significant controllable risk

factors. With larger groupings (e.g. all cancers) the impact of specific controllable risk factors becomes much less clear.

1	Infectious diseases				
2	Cancer: mouth, gullet	3	Cancer: oesophageal		
4	Cancer: stomach	5	Cancer: colon	6	Cancer: rectum, anus
7	Cancer: liver	8	Cancer: pancreas	9	Cancer: other digestive system
10	Cancer: larynx	11	Cancer: lung, bronchus, trachea	12	Cancer: skin
13	Cancer: breast	14	Cancer: cervix	15	Cancer: uterus
16	Cancer: ovary	17	Cancer: other female genital	18	Cancer: prostate
19	Cancer: other male genital	20	Cancer: bladder	21	Cancer: urinary organs
22	Cancer: lymphatic etc.	23	Benign tumours	24	Cancer: other locations
25	Blood diseases	26	Diabetes		
27	Vascular dementia	28	Other mental illness	29	Parkinson's disease
30	Alzheimer's	31	Other diseases of nervous system		
32	Blood pressure + rheumatic fever	33	Ischaemic heart diseases	34	Non-rheumatic valve disorders
35	Other heart diseases	36	Cerebrovascular diseases	37	Circulatory diseases
38	Influenza	39	Pneumonia	40	Other acute respiratory infections
41	Chronic Obstructive Pulmonary Disease	42	Other respiratory diseases		
43	Liver cirrhosis	44	Other liver diseases	45	Other digestive diseases
46	Diseases: skin, bone, tissue	47	Diseases: urine, kidney,...		
48	Suicide	49	Road/other accidents	50	Accidental Poisonings
51	Other causes				

Table 1: List of causes of death used in this analysis.

The resulting death rates by cause of death, $m(g, e, c, y, x)$, can then give us insight into the emerging patterns in all-cause mortality. For example, can notable features in the all-cause mortality data be linked to specific causes of death or specific controllable risk factors?

2.5 Estimated proportions of low/high educated

In Figure 1, we plot the proportions that have a low education, $R(g, \text{low}, y, x)$, by year of birth, $y - x$.

In general, levels of education have been improving cohort by cohort (i.e. higher proportions are now entering higher education). However, for males, since 1950 the proportion entering some form of higher education has fallen slightly. As an individual cohort ages, we see that the proportion who have a low education falls slowly (e.g. the dark-blue line for age 70 is below the green line for age 60). This is because the low-educated group experiences higher mortality than the high-educated group, and so the size of the low-educated cohort falls more quickly.

The gradual decline could manifest itself as a cohort effect in national population death rates. For example, death rates might have fallen faster than they would have done if education proportions had stayed constant as the result of a one-off education-linked mortality 'dividend'. We discuss this further in Section 4.

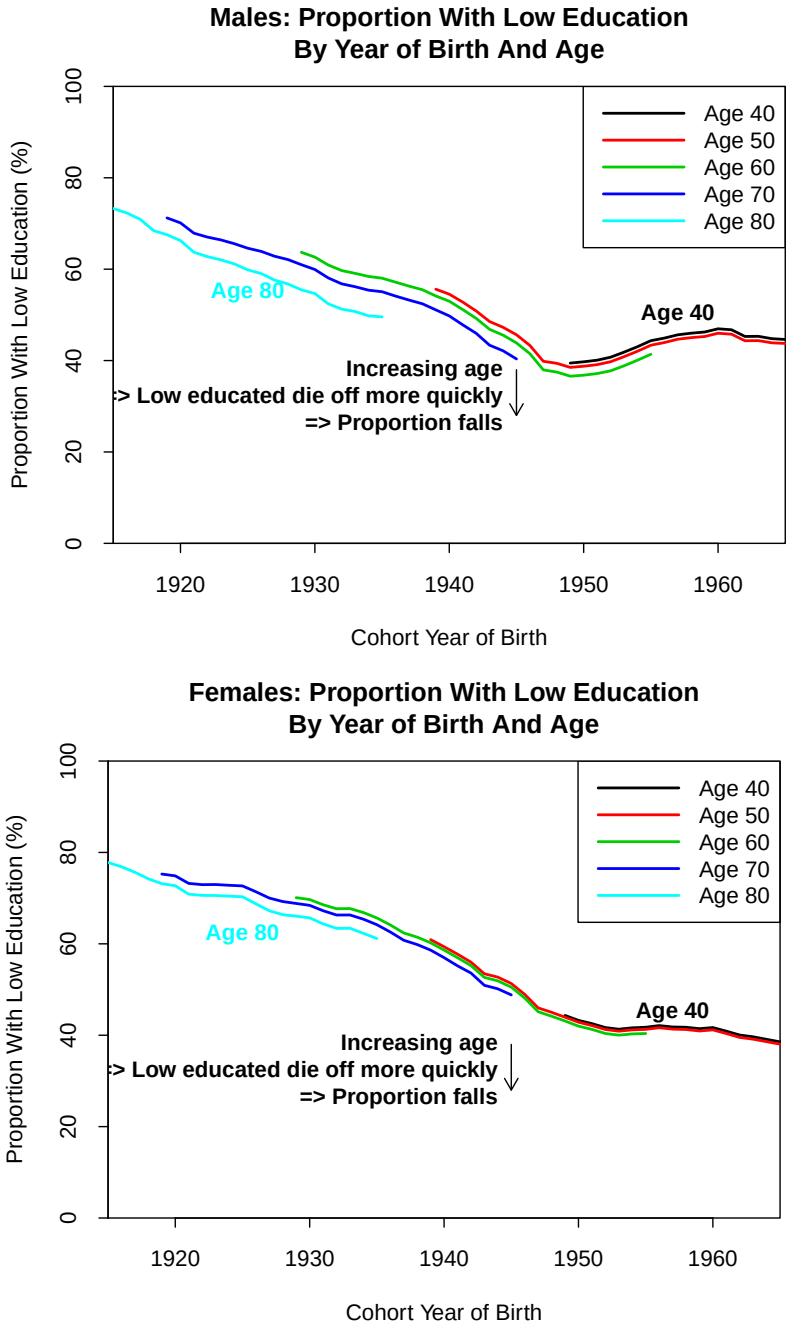

Figure 1: The populations of low-educated males (top) and females (bottom) as proportions of the populations all males and females at a given age by year of birth.

3 Introductory comments on all-cause mortality

Figure 2 shows how death rates have changed over time at three ages: 55, 65 and 75.

- All six of the sub-plots show a significant widening of the gap between low and high-educated males and females at the three ages, with the gap widening at the rate of about 1% to 1.5% per annum over the 28 year period.
- The commencement of the much-discussed stagnation of mortality for the low-educated groups (see, for example, Case and Deaton, 2015) is very much dependent on age. It starts early when we look at age 55, but relatively late when we focus on age 75. This suggests a cohort effect with those born in the 1950's being most badly affected.

Figure 3 shows smoothed mortality improvement rates for all-cause mortality for low and high-educated males and females. The patterns of improvement are generally quite complex but we can pick out some specific features:

- The colours are generally warmer for the high-educated groups indicating stronger mortality improvements on average and, therefore, a widening inequality gap between low and high educated.
- Low-educated males have a reasonably clear cohort effect associated with individuals born around 1950 (green diagonal). The green colouring indicates that this cohort has tended to have higher mortality than cohorts born 5 years earlier or later.

  This same cohort effect is perhaps just visible in the high-educated males plot on the right.

  For females, it is less obvious that there is a cohort effect. But, if there is, it might be that the female cohorts born around 1955 have done less well (i.e. later than the 1950 for males) than cohorts on either side.
- There were significant improvements in male mortality below age 50 in the late 1990s: the result of new treatments for HIV and AIDS.
- The late 1990s and 2000s saw strong improvements at high ages: a feature that we will explore later.
- We see generally higher improvement rates in the 2000s compared to the 2010s.

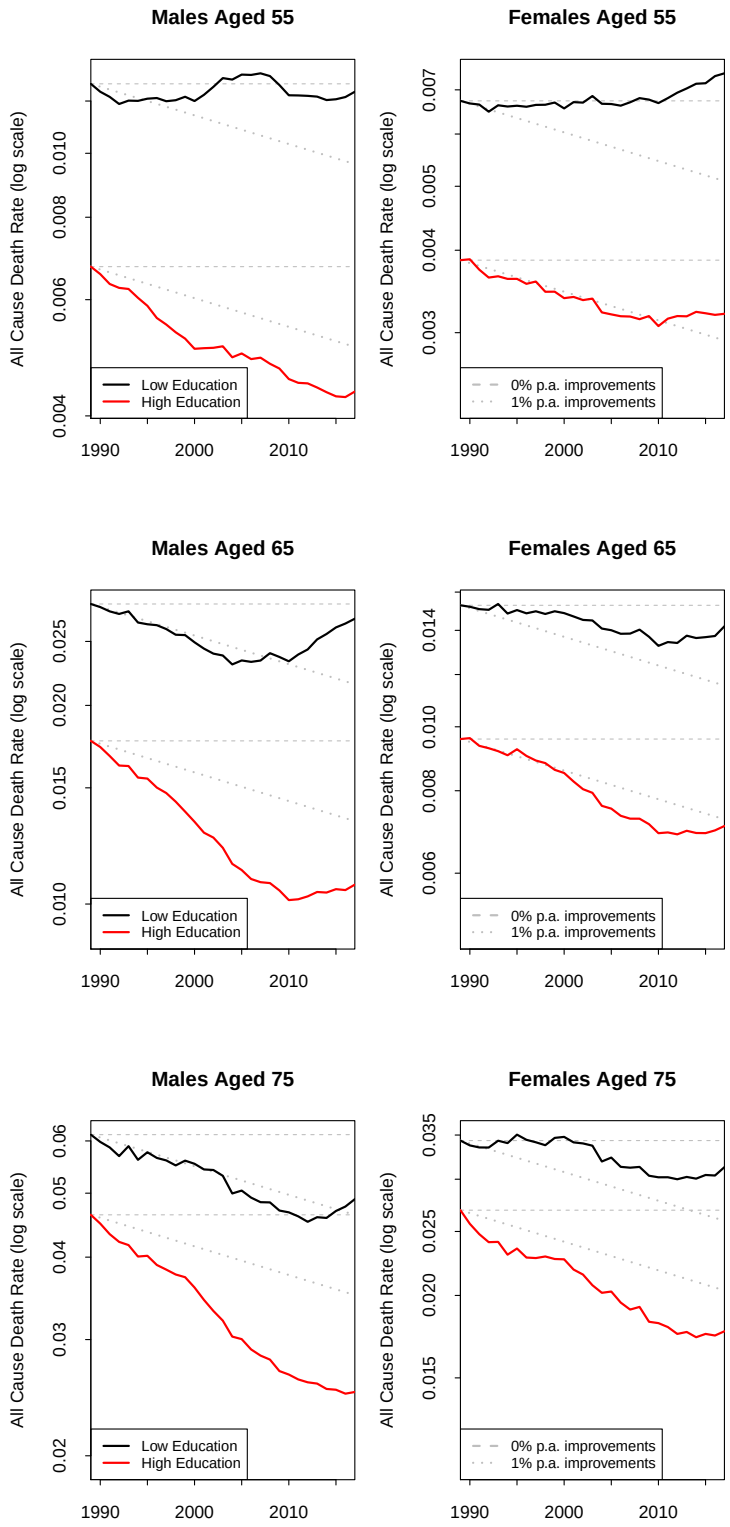

Figure 2: All-cause death rates (log scale) for low and high educated males and females at ages 55, 65 and 75. Dotted and dashed lines indicate the path that death rates would have taken if there was a constant 0% (dashed) or 1% (dotted) improvement rate.

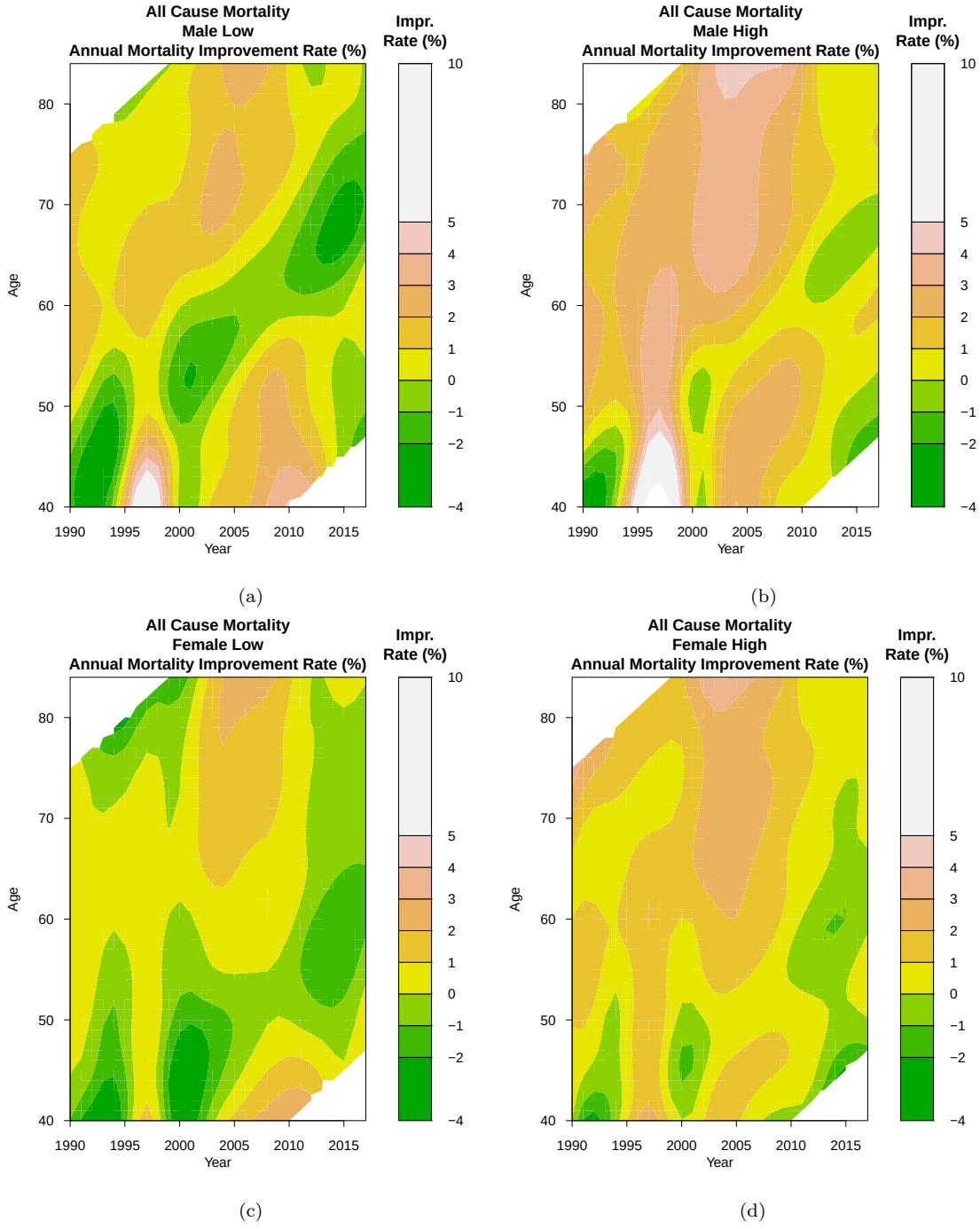

Figure 3: Heat maps showing smoothed mortality improvement rates for all-cause mortality for males ((a), (b)) and females ((c), (d)), low education ((a), (c)) and high education (((b), (d)). The bar to the right of each plot gives the colour scale for the improvement rates.

3.1 All-cause summary

From Figures 2 and 3 we can summarise the pattern of all-cause mortality as follows:

- We see different improvement rates at different ages, by age and by education level.
- We can observe significant levels of mortality inequality by education level, especially at younger ages.
- The inequality gap between high and low educated groups has been widening.
- We have seen bigger improvements in mortality at high ages in the late 1990s and 2000s.
- But we also see a general slowdown in improvements since around 2010, although the picture is complex and somewhat dependent on age.

4 Cohort effects

We introduced earlier, in the context of education levels, the concept of a cohort effect which we now discuss a bit further.

4.1 Definition of a cohort effect

Relative to the general time trend, if a birth cohort experiences higher/lower mortality than otherwise anticipated (for example, out of line with cohorts before and after) then this is called a cohort effect. Separating out cohort effects from general period trends and trend changes can be challenging: sometimes what looks like a cohort effect can be (at least partly) the product of a combination of age and period effects (see, for example, Cairns et al., 2009).

Cohort effects were first observed in an actuarial context by Willets (2004) in British data, but they can also bee seen in US data: for example, low-educated males in Figure 3. However, when we see a cohort effect in a heat map of improvement rates it is a statistical feature of the data: we don't initially have an obvious cause of that cohort effect.

4.2 Possible causes of cohort effects

As a first possibility, we have already introduced the idea that the level of educational attainment (specifically the changing split between high and low educated) might introduce a cohort effect.

The data clearly indicate that higher-educated people have lower mortality (e.g. Figure 2). Better-educated people *on average* follow a healthier lifestyle and look after their health better. They have access to better jobs with better pay. And (especially in the US) they have access to better healthcare.

We should also acknowledge that the low and high education groups are not homogeneous in terms of their mortality rates. As a stylised example, suppose we educate 5% more to higher levels. On average, this 5% might have: lower mortality than the average for original the low educated group; higher mortality than the average for the original high educated group. So the mortality of both groups goes up! As a consequence, the increase over time in the percentage in the high-education group might be artificially pushing down the improvement rates that we estimate for both the high and low-educated groups.)

But there are alternative and more compelling reasons for cohort effects (see, for example, Holford, 1991). Even if the proportions of low and high educated stay the same there will be significant variations over time (cohort to cohort) in lifestyle-related *controllable* risk factors (which we discuss further in the next section). For example, smoking prevalence is known to vary considerably by cohort with different variations by cohort by education group and by sex. There might also be cohort-related variation in behaviour related to diet, exercise, excessive alcohol consumption etc. Additionally, the changing economic environment is thought to have impacted more severely on middle-aged, lower-educated workers (Case and Deaton, 2015): for example, redundancy at age 55 might lead to long-term unemployment which, in turn, can lead to a decline in self-esteem and in health. So low points in the economic cycle can have a greater impact on certain cohorts.

In places we will estimate the magnitude of a cohort effect using the CBD-X3 model (Hunt and Blake, 2014; Dowd et al., 2020). This model is outlined in Appendix B.

5 Controllable risk factors

Medical and related journals contain thousands of articles that analyse the link between specific risk factors and the risk of death from specific causes. From our perspective, we seek to establish what the difference is between baseline death rates, where the risk factor of interest is absent, and death rates for individuals who have a specific risk factor. The ratio of the two is referred to as the *relative risk*. A relative risk of 1 indicates that a risk factor has no impact on mortality from a specific cause. If the relative risk is bigger than 1 then there is an increased risk of death from that cause. Often the relative risk is significant but only a little bit bigger than 1. But, in other cases, the relative risk is very much bigger than 1 (for example, the relative risk for a smoker to die from lung cancer versus a benchmark non-smoker). Depending on the benchmark, relative risks can, of course, be less than 1: for example, individuals

who take statins versus those who don't; individuals who follow a particular healthy lifestyle; or particular racial groups that are less susceptible to particular diseases.

We choose to classify risk factors into three groups:

- **Controllable risk factors:** these are risk factors that are in the control of each individual with examples being smoking, poor diet, lack of exercise, excess alcohol consumption etc.
- **Preventable risk factors:** these are risk factors which are not controllable or easily controllable by the individual but which can be *prevented* or reduced through other means. A specific example is Human Papilloma Virus (HPV). Almost all cases of cervical cancer occur in females who have been previously infected with HPV, and it is extremely rare amongst females who have never been infected by HPV. It is considered to be preventable as young females can now be vaccinated before they become sexually active and this much reduces the chances of developing cervical cancer much later in life. (See, for example, the Cancer Research UK website, `www.cancerresearchuk.org`, for further details.)
- **Non-preventable risk factors:** these are risk factors that cannot be easily controlled or prevented. Examples, include genetic or racial factors (e.g. prostate cancer; Cheng et al., 2009, and Taitt, 2018)[3] and personality traits such as conscientiousness (see, for example, Kern and Friedman, 2008, and Deary, Weiss and Batty, 2010).

The classification of a specific risk factor is not always clear. For example, educational attainment might be classified as non-preventable but arguably, something can be done, at least early in life, to improve or change this characteristic.

A table of risk factors for the 51 causes of death can be found in Appendix A, Figure 36.[4] For cancers, the table also gives an indication of what proportion of deaths (in the UK) are due to specific risk factors such as smoking. These then give an indication of how many deaths are avoidable.

In this paper, our focus is on controllable risk factors as these are the most likely to lead to inequalities between different socio-economic groups.

[3] See, also, `www.cancer.org/cancer/prostate-cancer/causes-risks-prevention/risk-factors.html`.

[4] The table of risk factors in the Appendix is a work in progress and should not be regarded as definitive or complete.

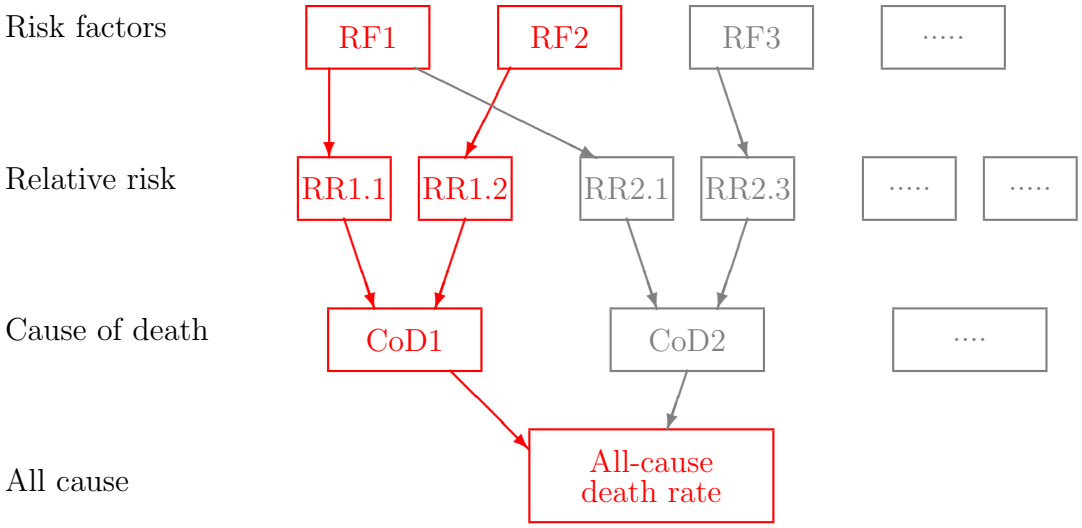

Figure 4: Stylised representation of how individual risk factors combine with relative risks to affect death rates from specific causes. Cause of death 1 (CoD1) has two significant risk factors (RF1 and RF2) with associated relative risks $RR1.1$ and $RR1.2$.

6 Why do mortality inequalities exist?

We have observed mortality inequalities at the all-cause level in Section 3. It follows that these inequalities must be present for at least some causes of death. In the following sections we will see that, in fact, significant inequalities can be seen for almost all of the 51 causes of death (the exception being Parkinson's Disease). Before that, we will discuss how, from a purely statistical point of view, inequalities arise.

In Figure 4 we illustrate in a stylised way how the all-cause death rate at either an individual or group level depends on various factors. The all-cause death rate is a combination of death rates from specific causes. Each cause of death has a set of controllable, preventable and non-preventable risk factors (top row) that have an impact on mortality from that specific cause (third row), and the impact comes through the relative risks (second row) attached to each risk factor for that specific cause of death.[5] Each cause-specific death rate has its own baseline rate, but, at the group level, this is then adjusted to reflect the prevalence of each risk factor combined with the relative risk for each of these risk factors.

In our setting, if we observe high levels of cause-specific mortality *inequality*, the most likely explanation is that we have a combination of

- one or more significant controllable risk factors for this cause of death;

[5]The same risk factor will have different relative risks for each cause of death. Hence, in the diagram, Risk Factor 1 has relative risk $RR1.1$ for cause 1, and $RR2.1$ for cause 2 etc.

- significant differences in the prevalences of these risk factors between the low and high-educated groups;
- high values for the relative risks for at least some of these risk factors.

Other explanations for inequality are possible, although these are likely to have a modest impact on inequality. First, low-educated people are less likely to have access to high-quality easy-to-access health care, and this will push up death rates. We find indirect evidence for this in Section 15. Second, some racial groups are over-represented in the low-educated group (e.g. African Americans). It is also known that specific racial groups have significantly higher death rates for specific causes after adjustment for other socio-economic factors (e.g. African American men are more than twice as likely to die from prostate cancer than non-Hispanic whites: see Section 15). Third, the personality trait of conscientiousness is a relevant factor which can affect the propensity to get an early diagnosis, and the level of engagement with a treatment process following diagnosis. Trautwein et al. (2009) demonstrate that there is a significant correlation between educational attainment and conscientiousness, and, in our case, this might lead to higher death rates for the low-educated groups due to the association between conscientiousness and mortality rates (Kern and Friedman, 2008, and Deary et al., 2010).

7 Analysis of cause-of-death data: higher-level groupings

To begin with we will look at eight higher-level cause-of-death groups rather than the full 51 to get an initial feel for the key headlines. The eight groupings are:

- Circulatory and cardiovascular diseases (Table 1, causes 32-37)
- Cancers that have significant controllable risk factors (2-11, 20, 21)
- Other cancers (12-19, 22-24)
- Respiratory diseases (38-42)
- Other causes (25, 26, 43-47, 51)
- Diseases of the mental and nervous system and Alzheimer's disease (27-31)
- Infectious diseases (1)
- Death due to external causes (48-50)

Figures 5 to 10 show death rates for these eight groups for low and high-educated males and females and for ages 45, 65 and 80. Each figure displays both stacked and unstacked death rates. Stacked plots place each cause-of-death group on top of the previous ones so that the cumulative rates add up to the all-cause death rate. Unstacked plots simply show the death rate from a particular cause group and on a logarithmic scale. Each pair of low and high-education plots use the same scale to facilitate a straightforward comparison of the two education levels.

We can make the following observations:

- The trends that we see in all of the unstacked plots will be the result of:
  - improvements in the treatment of the underlying diseases;
  - changes in the prevalence of controllable risk factors linked to particular causes of death;
  - jumps resulting from the change from the ICD-9 to ICD-10 classifications;
  - the secondary effects that result from changes elsewhere (for example: the increasing numbers of survivors of cardiac events might increase the prevalence of certain dementias later in life as prior cardiac events increase the risk of onset of these diseases; see, for example, Ng et al., 2013, and Gottesman et al., 2017).
- In general, we see that deaths from circulatory and cardiovascular diseases are the leading cause of death for males and older females, with the combined cancers a close second. But we can also see that circulatory and cardiovascular diseases have experienced the biggest improvements, so that cancers are now a more significant cause of death in some cases (especially high-educated males and females). This means that there is less room for improvements in mortality at the all-cause level due to further reductions in circulatory and cardiovascular diseases. In fact, we see that the improvements in circulatory and cardiovascular diseases have stalled in recent years at some ages.
- In most of the plots we see a big difference between death rates for the controllable cancers for low versus high educated. In contrast, the non-controllable cancers ('cancers: other') are much closer in magnitude. The former is a clear indication that lower-educated groups tend to have a higher prevalence of harmful controllable risk factors such as smoking.

  We can also see in most plots that the high-educated groups have made proportionally stronger reductions (from an already lower starting point) in controllable cancers.
- The unstacked plots allow us to identify more easily which groups have falling, flat or rising mortality. Of particular interest are cases where the curve is rising for the low-educated group and falling for the high-educated group:

for example, deaths from respiratory diseases for females aged 65. These are observations that require a closer look by making use of the data with 51 cause-of-death groups.

- At age 80 we can see a steady increase in death rates for the mental/nervous/ Alzheimer's group (Table 1 causes 27-31). We will consider the underlying causes in a later section.

- At age 45 we can see a high degree of inequality (and rising rates) for deaths from external causes (causes 48-50: suicide, accidents and accidental poisoning). Again the detail will be discussed later.

In Figure 11 we plot the age profile for the breakdown of all-cause mortality by age in 2015. We have chosen here to express the stacked death rates as a percentage of all-cause mortality excluding external causes. The reason for the exclusion (although not critical) is to allow us to focus on the underlying physical health and frailty of people in the different groups.

- All four sub-plots exhibit the same general pattern across ages. In particular, cancers, as a percentage of all-cause deaths peak in the 60s (except for high-educated females). At older ages, dementias and respiratory diseases increase significantly, and cardiovascular diseases increase slightly.

- We also see that death rates due to external causes as a percentage of all-cause mortality are quite similar at all ages for the low and high-educated groups, but different for males and females.

- In all cases, the "other cancers" (mostly non-controllable) feature more prominently (as a percentage) amongst high-educated males and females.

- Deaths from external causes, as a percentage, stablise in all four sub-plots above around age 70 and at about the same level. This requires some further investigation but suggests that fatal accidents in later life are proportional to the underlying frailty of each individual.

- By age 85, differences in percentage terms between the low and high-educated groups are relatively small for each of the 8 cause-of death groups. This is consistent with the observation that mortality inequalities tend to diminish with age.

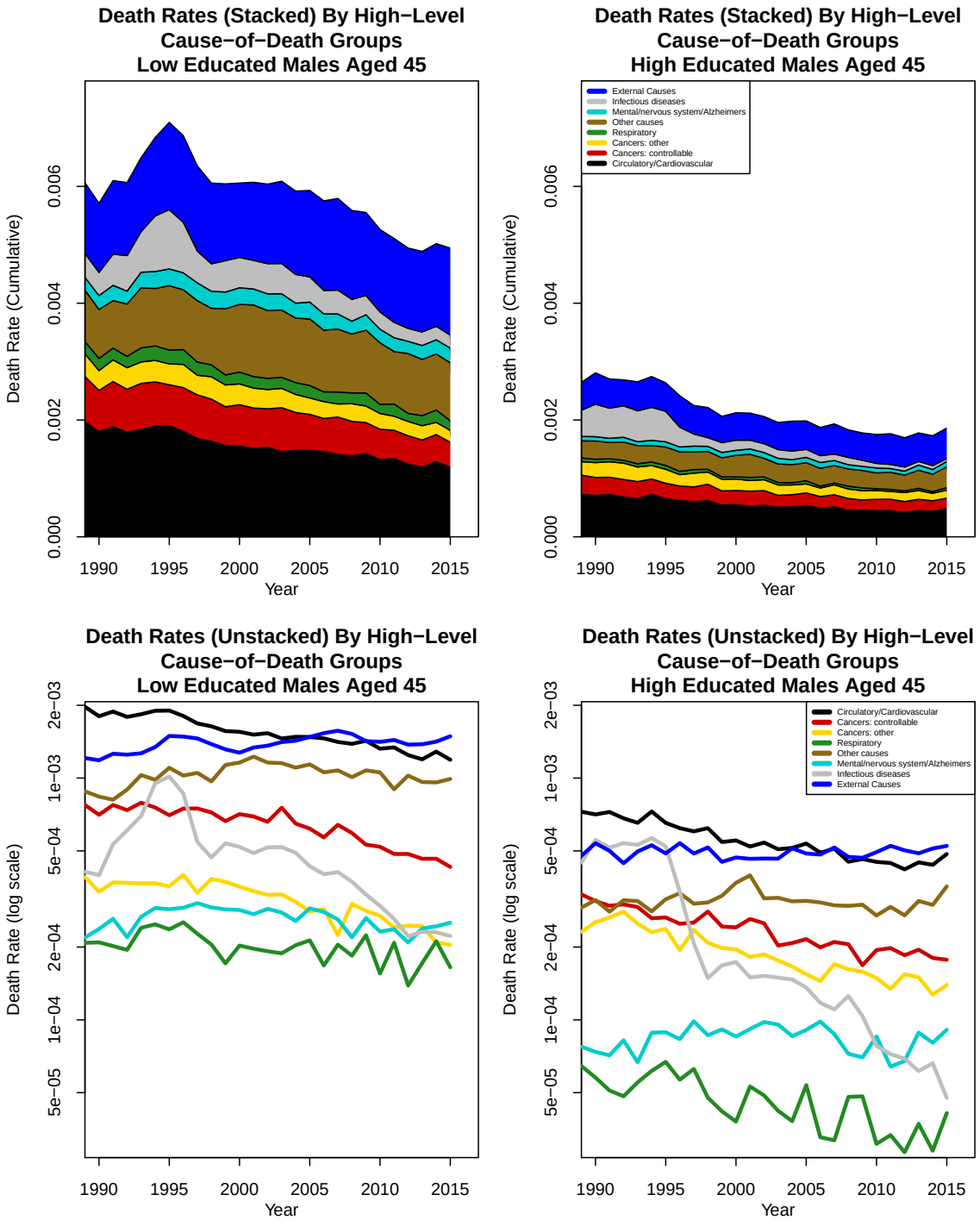

Figure 5: Evolution of death rates over time by higher-level cause-of-death groups for males aged 45. Top: stacked or cumulative death rates adding up to all-cause death rates. Bottom: unstacked or individual death rates for each cause-of-death group. Left: low-educated. Right: high educated.

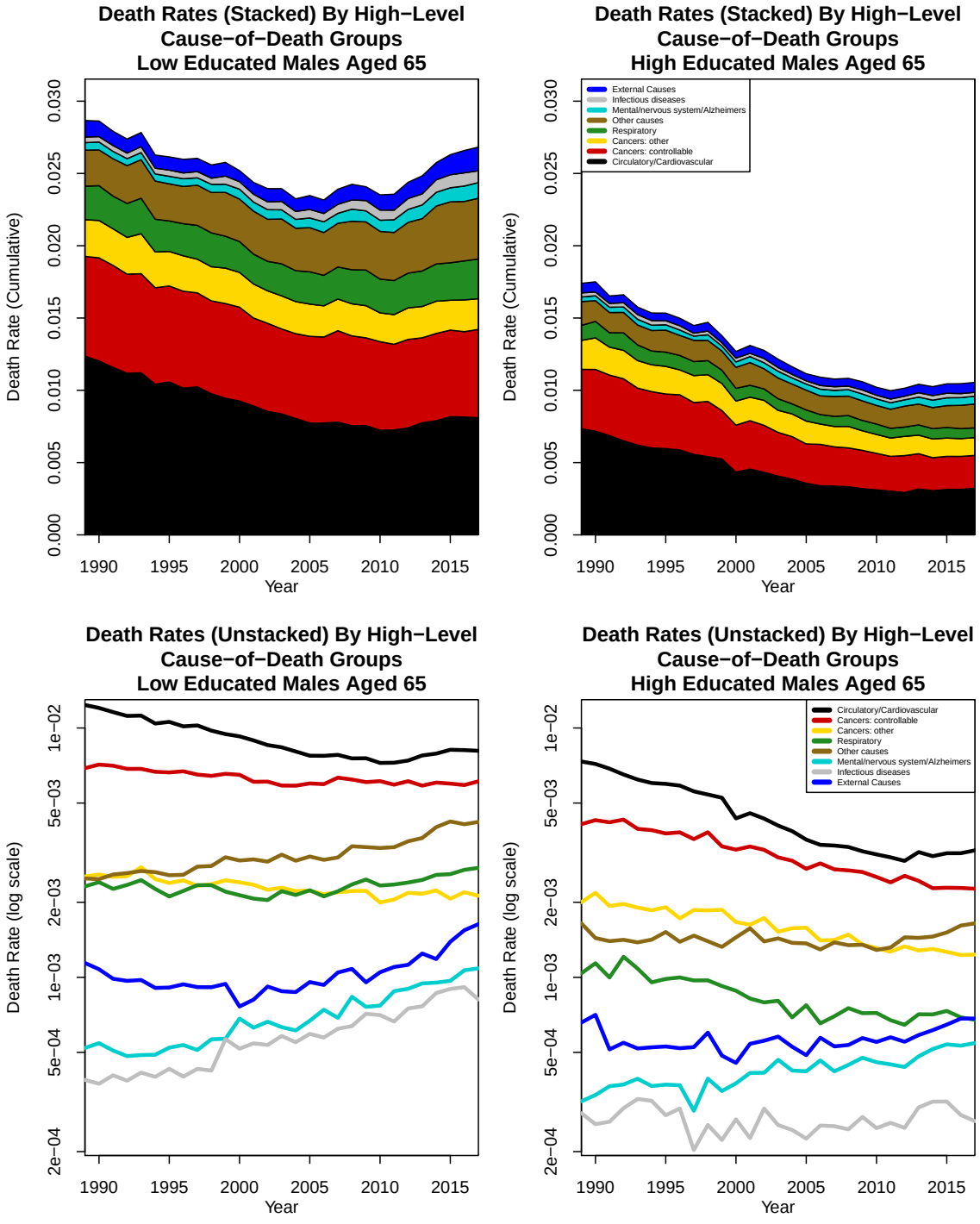

Figure 6: As Figure 5 but for males aged 65.

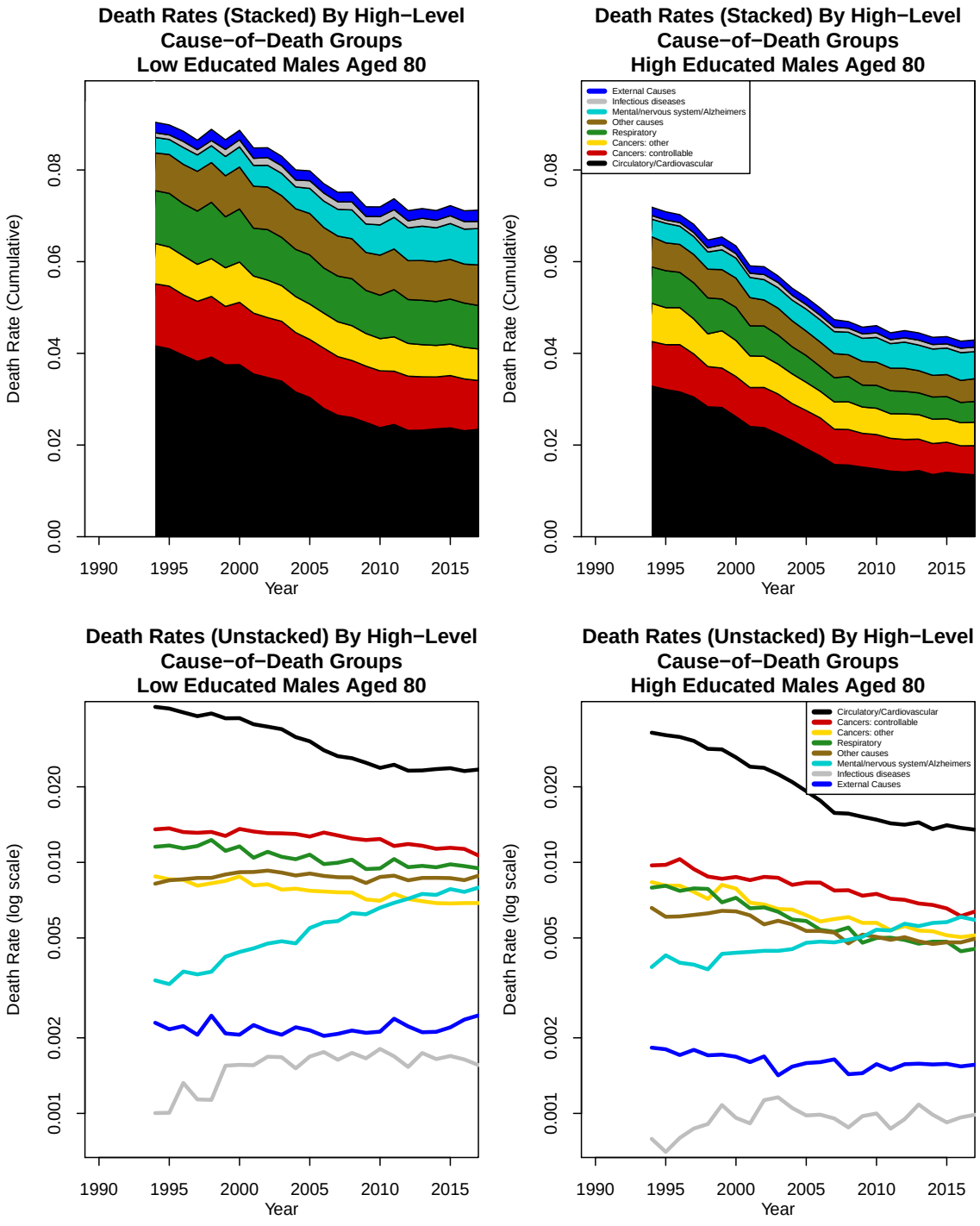

Figure 7: As Figure 5 but for males aged 80.

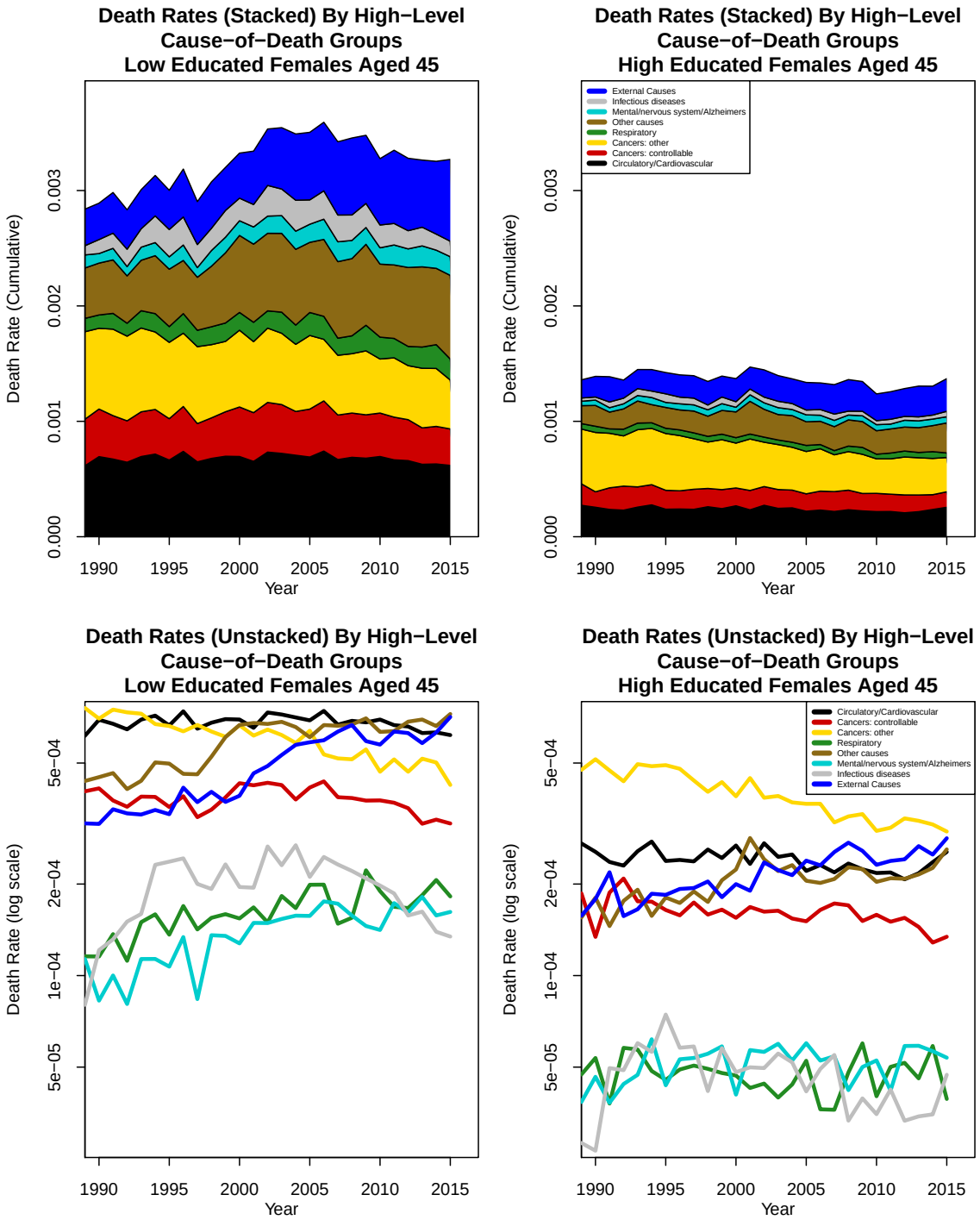

Figure 8: As Figure 5 but for females aged 45.

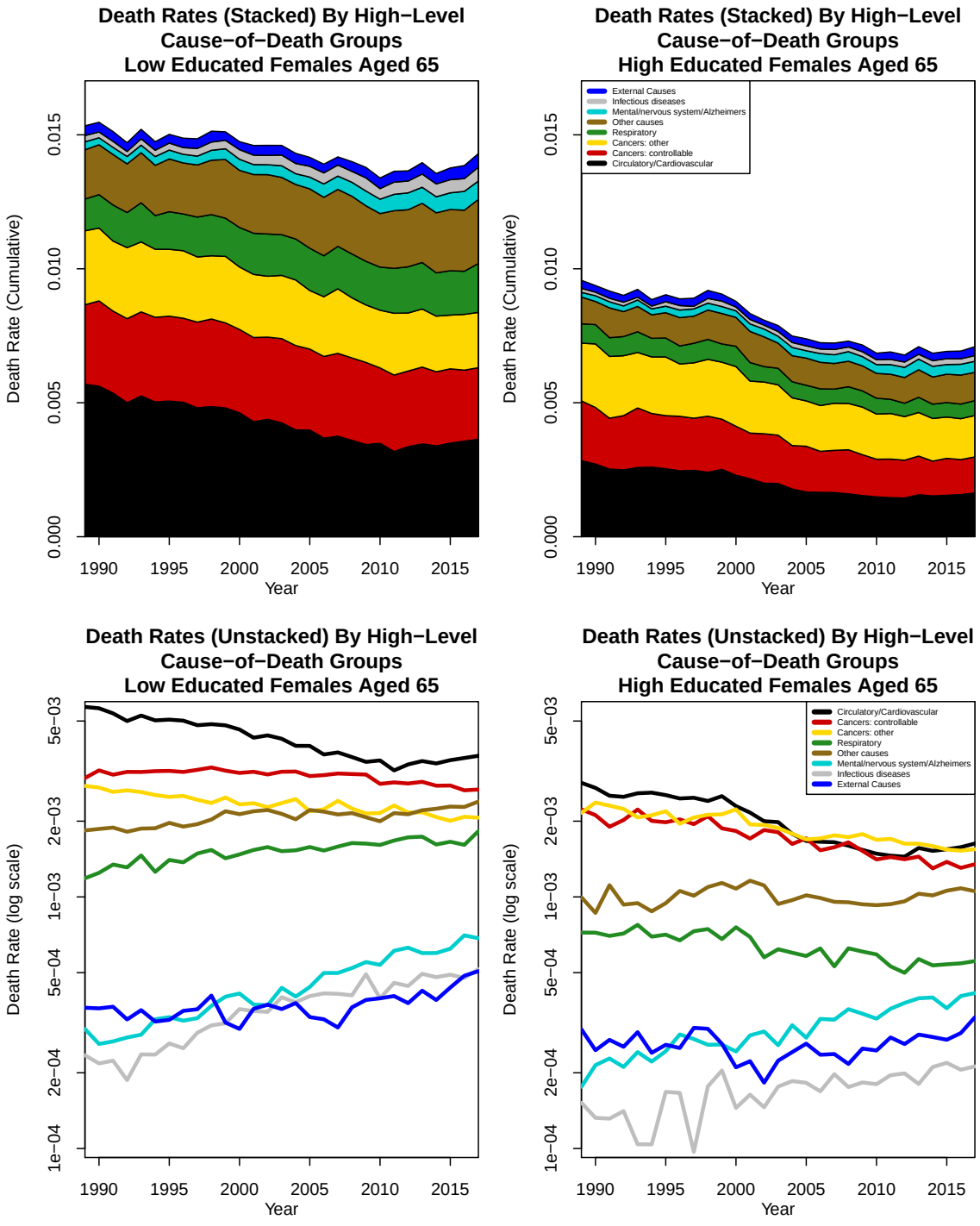

Figure 9: As Figure 5 but for females aged 65.

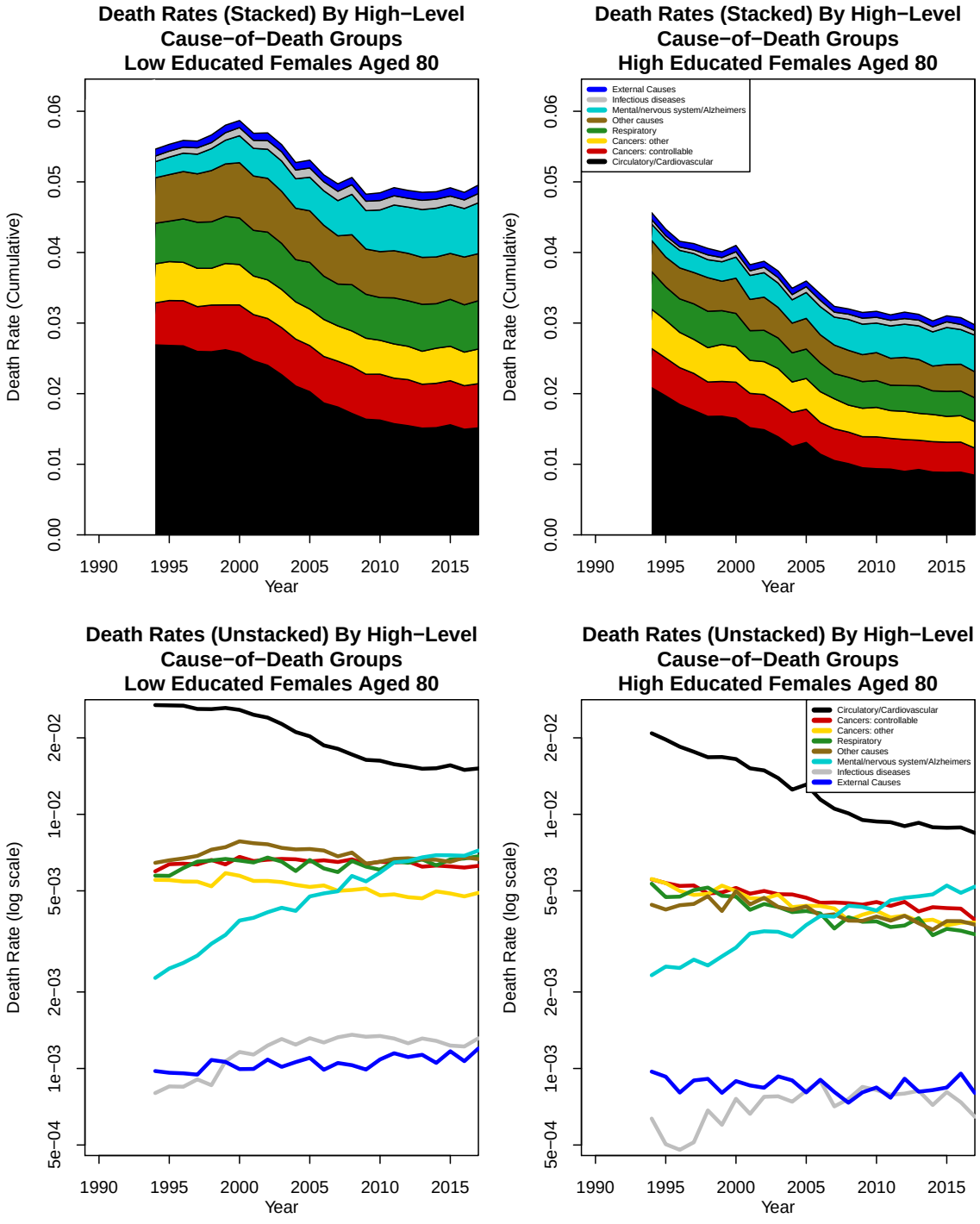

Figure 10: As Figure 5 but for females aged 80.

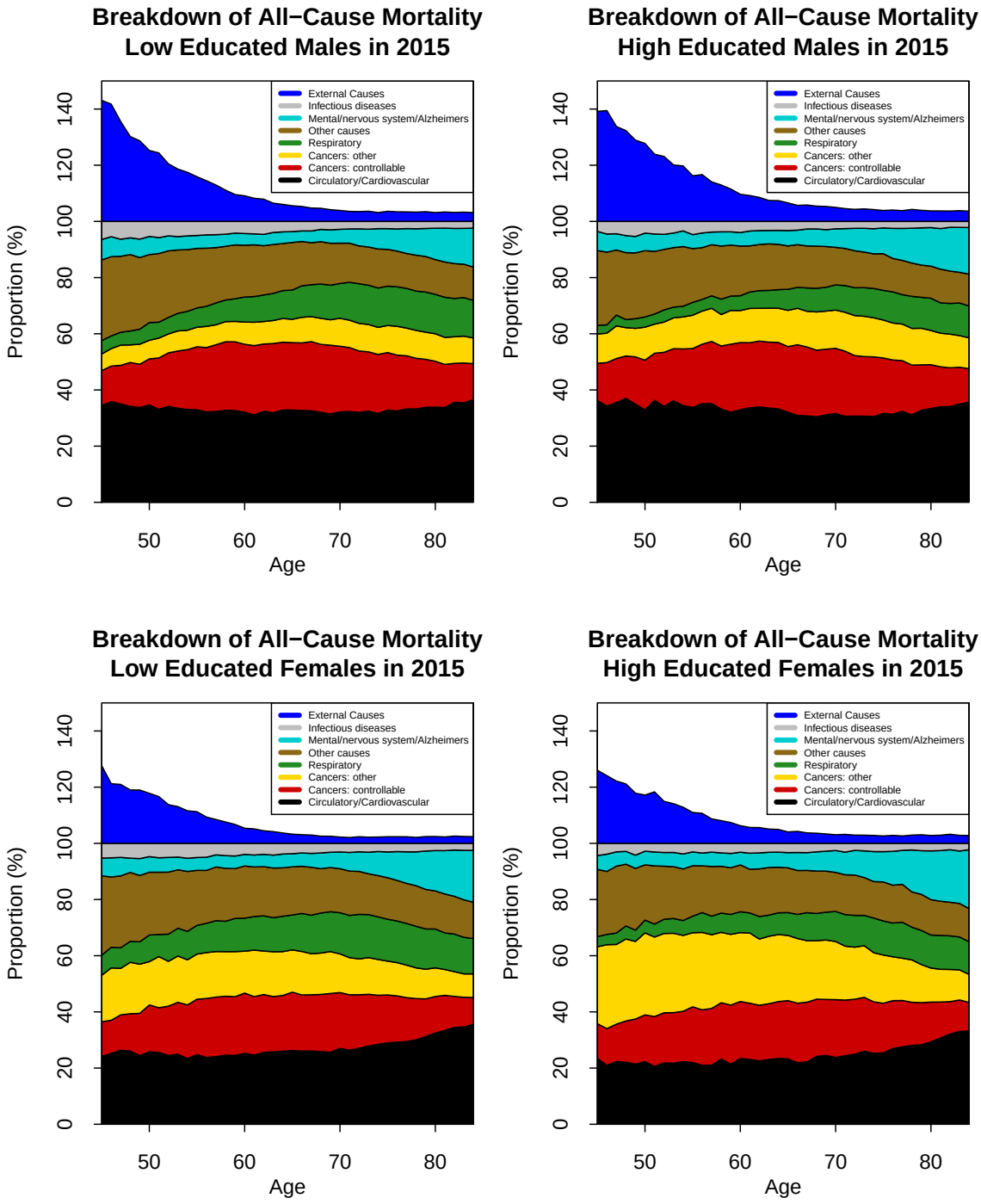

Figure 11: Death rates by age in 2015 as a cumulative percentage of all-cause mortality excluding external causes. Top: males. Bottom: females. Left: low educated. Right: high educated.

8 Ischaemic heart disease (IHD)

In this section we will illustrate the range of graphical diagnostics that can be used to identify patterns in the data and to make inferences about the underlying drivers and impacts on all-cause mortality. As we shall see, the same data plotted in different ways can bring to light additional insights that might not be obvious when we look at the more common single plots.

A standard starting point is to plot death rates $m(g, e, c, y, x)$ against time and age in Figure 12. Apart from the plotted death rates, the sub-plots also include horizontal gray lines to help interpret the numbers: each line is $2\times$ the line below. The top two rows show death rates from ischaemic heart disease (IHD) for ages 55 and 75. Each has a small jump in 1998 due to the change from ICD-9 to ICD-10. In this empirical analysis, this is a feature that we note, but do not need to remedy in this empirical study. More importantly we can see the following. First, death rates declined steadily from 1989 until some time in between 2010 and 2015 with some variation between sub-groups and ages. Death rates are now about half of what they were in 1989, with the introduction and widespread use now of statins at higher ages significantly reducing the incidence of heart disease (Orkaby et al., 2020). Second, there are significant levels of inequality and the inequality gap has been widening. The gap is wider at younger ages (bottom two rows). For example, at age 55, the low-education group death rate is about 3 times the high-education death rate from IHD.

The lower two rows show mortality against age in 2000 and 2015. These highlight that inequality is greater at younger ages for IHD, gradually narrowing with age. Nevertheless, we also see that there has been a more significant widening of the inequality gap at these high ages.

In Figure 13 we show deaths from IHD as a proportion of deaths from all causes: $D(g, e, c, y, x)/D(g, e, \text{all}, y, x)$. Results are shown as a heat map. We can make the following observations:

- The change from ICD-9 to ICD-10 results in a pre- and post-1998 split in each heat map (consistent with Figure 12).
- As a percentage, IHD can be seen to be very sigificant at all ages, but it also features more prominently at higher ages. In part, this age gradient is exaggerated by deaths due to external causes (causes 48-50). Figure 14 mitigates this by removing deaths due to external causes with a noticeable impact below age 55.
- Above age 55, in each year the proportions due to IHD are relatively constant, but, more importantly, we can see how much IHD, in percentage terms, has declined: falling from over 25% of all deaths at higher ages down to aound

16%. This gives a clear indication of how effective preventative medication and other medical interventions have been in the case of IHD relative to other causes of death.

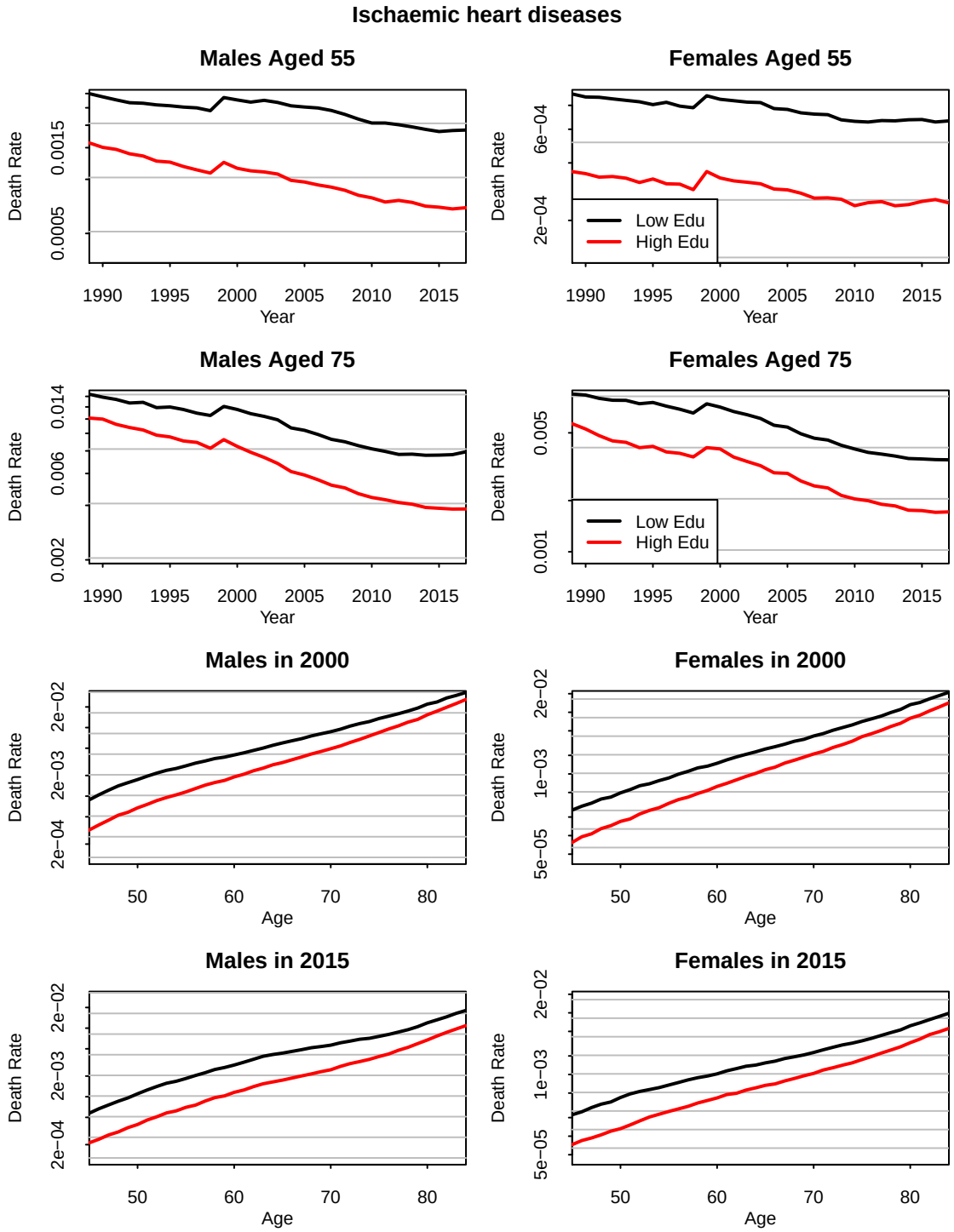

Figure 12: Death rates for ischaemic heart disease. Each sub-plot shows low and high-educated mortality. Left: males. Right: females. Top row: age 55 against time. Second row: age 75 against time. Third row: year 2000 against age. Bottom row: year 2015 against time. Horizontal grey lines: each line is 2× the line below.

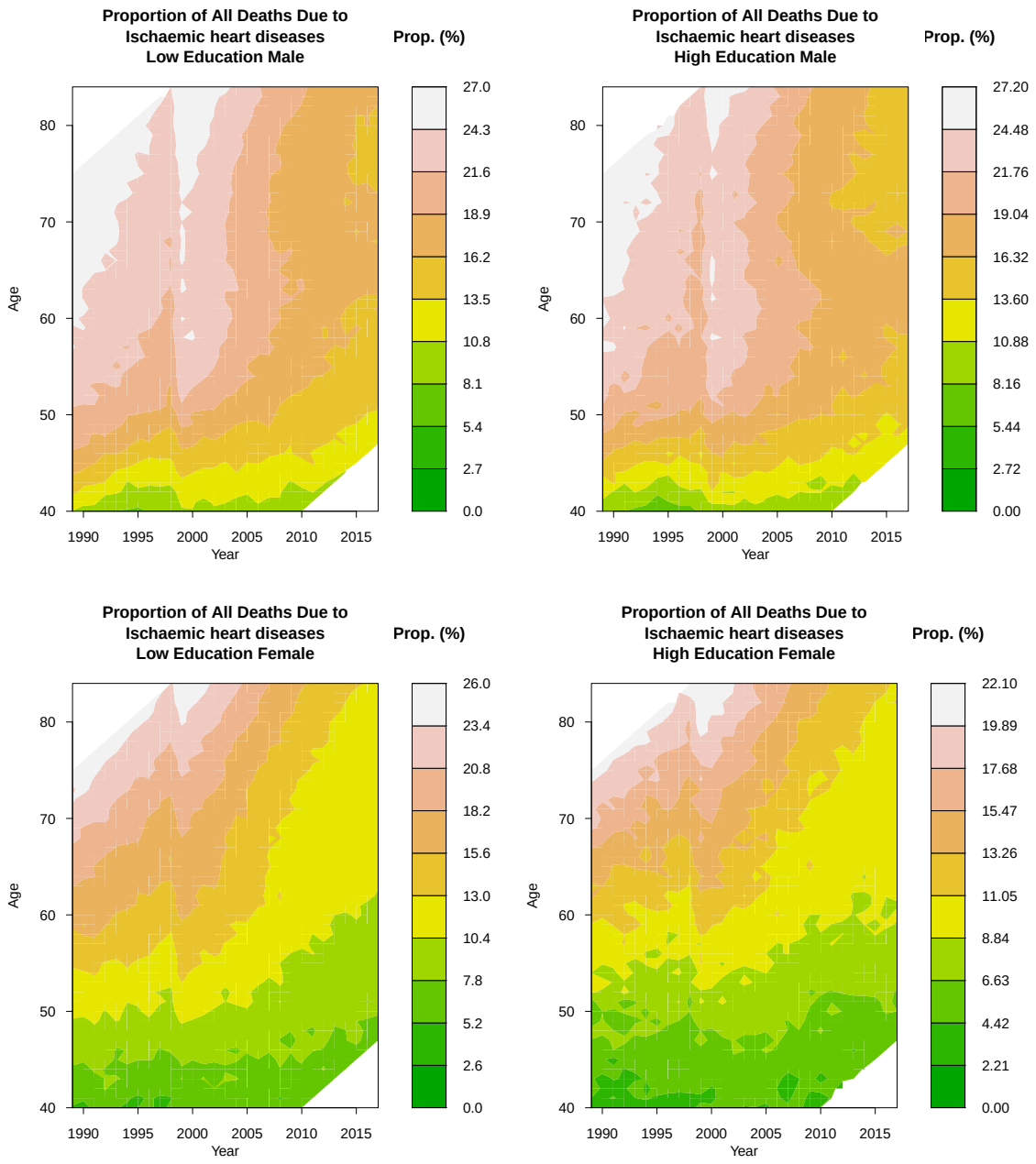

Figure 13: Ischaemic heart disease. Heat maps showing the death rate as a proportion (%) of all cause mortality by year and age for each sub-group.

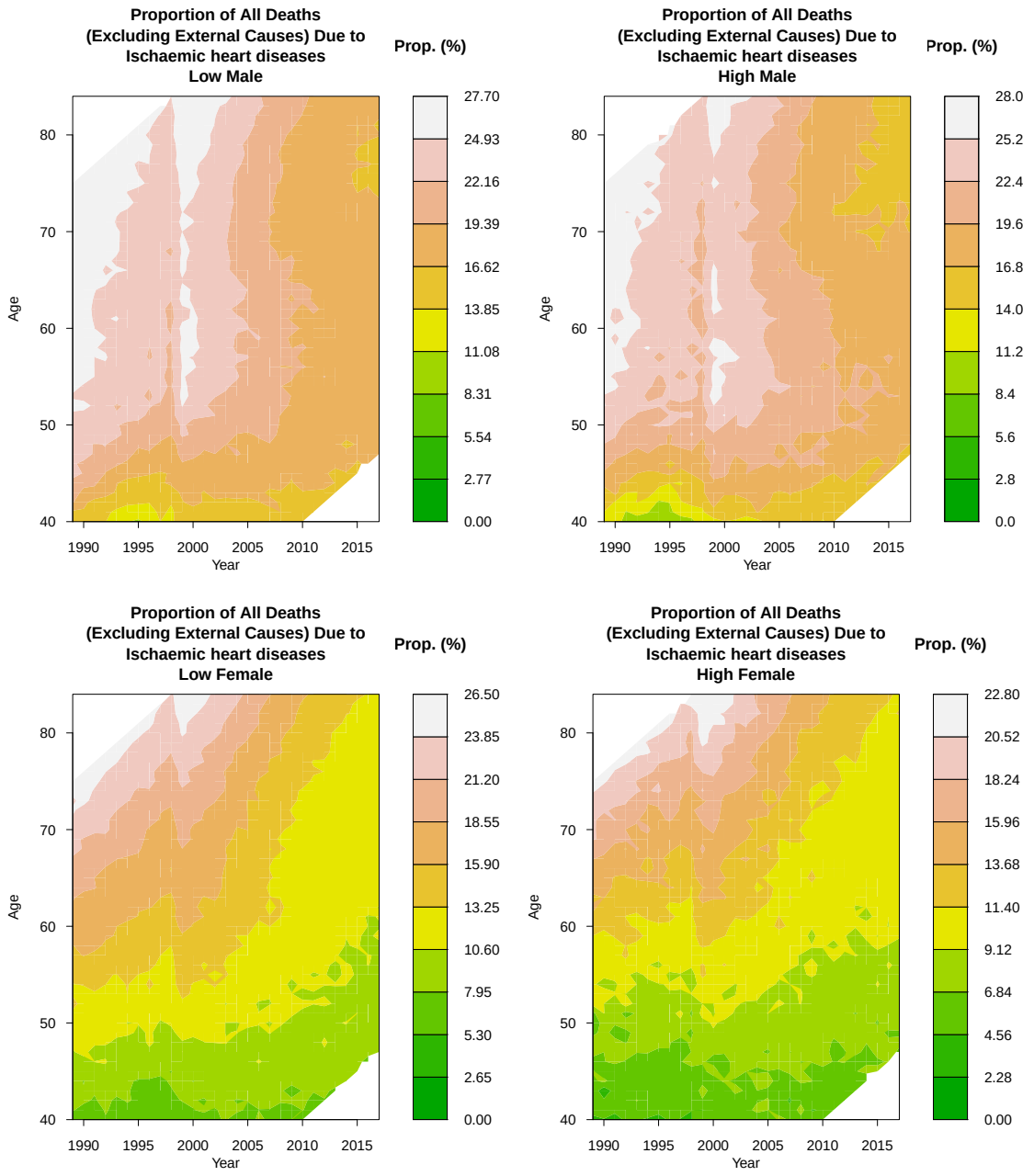

Figure 14: Ischaemic heart disease. Heat maps showing the death rate as a proportion (%) of all cause mortality *excluding external causes* by year and age for each sub-group.

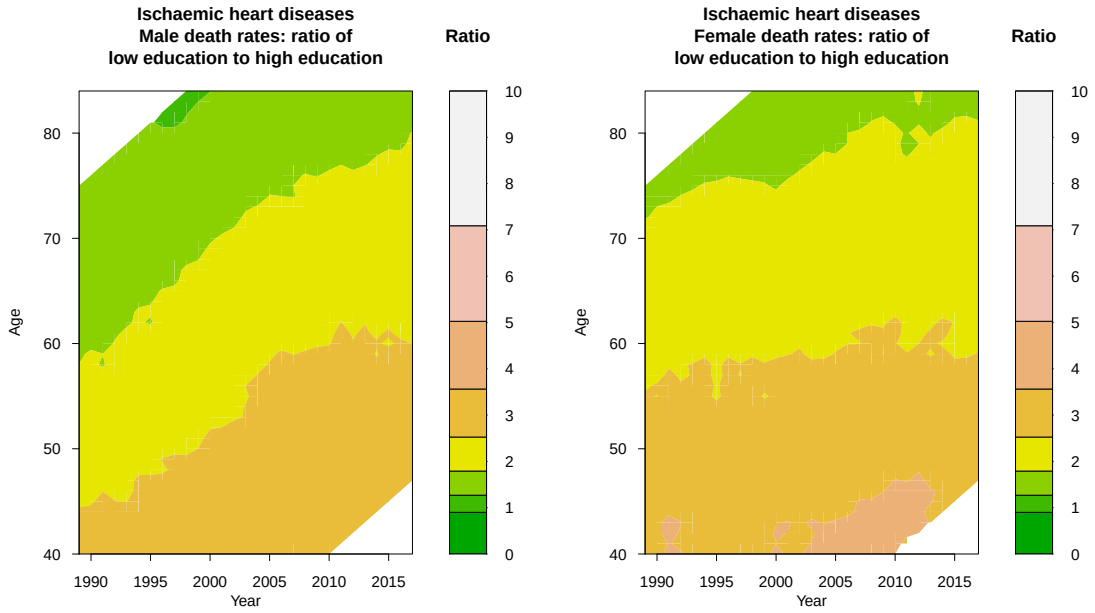

Figure 15: Ischaemic heart disease. Heat maps showing ratio of the death rates for the low-educated to the high-educated sub-groups.

In Figure 15 we plot the ratio of the death rate in the low-educated group to the high-educated group: $m(g, \text{low}, c, y, x)/m(g, \text{high}, c, y, x)$. Higher values (yellows, oranges, pinks) indicated higher levels of mortality inequality. Both plots show higher levels of inequality (mortality about $3\times$ (males) to $4\times$ (females)) at younger ages with a narrowing inequality gap with age. Females exhibit greater levels of inequality than males over the whole period. For females we see a modest shift to greater inequality over the period 1989 to 2017. For males, there is a more dramatic change with inequality in 2017 finally matching the previously higher levels of inequality between the two female populations.

Widening inequality might be the result of a combination of factors:

- a widening gap in the prevalence of the underlying controllable risk factors;
- differing and diverging levels of engagement with or access to preventative measures such as statins;
- a widening gap in access to healthcare and/or the latest medical advances.

Annual mortality improvement rates $(-\log m(g,e,c,y,x)/m(g,e,c,y-1,x))$ for IHD are illustrated as a heat map in Figure 16. Results plotted here use smoothed outputs from the CBD-X3 model (Dowd et al., 2020, and Appendix B; see, also, Hunt and Blake, 2014) which incorporates three period effects and a cohort effect. Smoothing allows us to see more clearly the underlying trends: heat maps for unsmoothed improvement rates (even for IHD) contain too much sampling noise to present a clear picture.

Key features are as follows. First, we see the most significant improvements happened between 2000 and 2010 at the higher ages. Second, for low-educated males, we can see a modest cohort effect affecting (negatively) people born in the mid 1940s. Other cohort effects and in other groups are only just detectable and certainly not strong. Third, high-educated males and females have exhibited consistently higher rates of improvement (consistent with the widening inequality gap noted in previous figures).

Salami et al. (2017) report steadily increasing statin use over the period 2002-2013. Statins were already in common use at the beginning of that period so, potentially the acceleration observed here in the 2000-2010 period is the delayed impact of earlier uptake in the 1990s, especially as the early receivers of statins might have been those at higher risk of death from heart disease. Unsurprisingly, statin use in the under 65s is much lower than the over 75s, perhaps explaining the lower improvements at younger ages, although, equally, statins might be less effective at these younger ages. The analysis of Salami et al. (2017) also found that there was little difference in uptake between low and high-educated groups after adjusting for other factors.

8.1 Ischaemic heart disease summary

IHD is one of the most significant causes of death in all four groups and at all ages. We have seen significant reductions in mortality since 1989, but the reasons for this are potentially complex. Several controllable risk factors influence IHD death rates resulting in significant and growing levels of mortality inequality. Additionally, IHD has seen significant improvements due to the increased use of statins. With this mixture it is tricky to establish how much of the changes in the level of mortality are due to specific risk factors, medical advances or other factors.

This section has introduced a range of graphical diagnostics that can be used to investigate the dynamics of a specific cause of death. We will return to other individual causes of death in later sections and in a more selective way. Before then we illustrate how the different causes of death compare in more general terms.

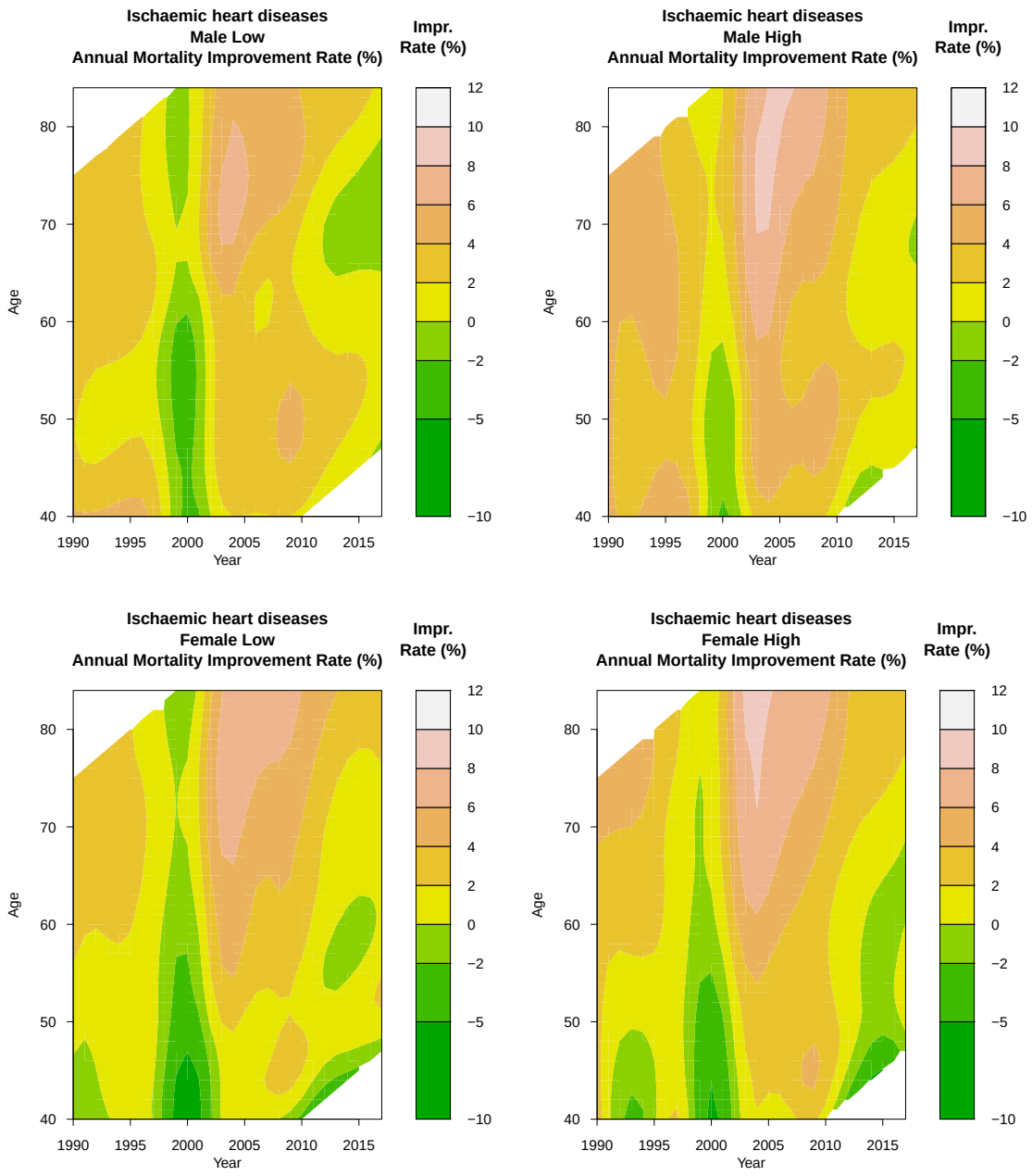

Figure 16: Ischaemic heart disease. Heat maps showing the mortality improvement rates (smoothed) by year and age for each sub-group.

9 A broader picture of inequality and improvements

We now zoom out and look at some summary statistics for all 51 causes of death. The aim here is to identify which causes of death help explain changes in all-cause mortality since 1989. Equally, are there also causes of death (especially the larger ones) that exhibit quite different patterns.

In Figure 17 we compare Age Standardised Mortality Rates (ASMRs) for ages 50 to 74 for low versus high-educated males and females in 2001 and 2017.[6] Within each plot, the dots represent one of the 51 causes of death, colour-coded by disease group. Causes of death with little inequality lie close to the 1× line, whereas those with greater inequality lie more to the southeast, with a few causes such as ischaemic heart disease, lung cancer and COPD (#'s 33, 11 and 41) close to the 3× line. Additional analysis shows that death rates or ASMRs for younger ages tend to have greater levels of inequality than these averages over ages 50 to 74. We can see that the controllable cancers (green dots) typically, but not always, exhibit greater inequality than other cancers (black dots).

Shifting from 2001 to 2017, most of the dots move further away from the 1× diagonal indicating growing inequality across a range of causes of death, and not just IHD. This general shift towards greater inequality is consistent with some mixture of:

- a widening gap in the prevalence of a small number of controllable risk factors in the two populations (for example, if smoking prevalence has been falling at a faster rate in the high-educated group than the low-educated).
- a widening gap in the availability and quality of healthcare.

In Figures 18 and 19 we compare improvement rates over distinct time periods 1999-2008 and 2008-2017 at ages 55 and 75 for each subgroup.[7] In each plot the cluster of dots (one dot for each cause of death) show the improvement rate per annum for 1999-2008 on the x-axis and for 2008-2017 on the y-axis. Dot sizes are proportional to the cause-specific death rate in 2008 to help focus attention on causes that make a bigger difference in the all-cause improvement rates. Each plot also includes the equivalent all-cause improvement rates (purple triangle) to allow comparison with individual cause of death improvement rates.

A number of features can be highlighted in these plots:

[6] The use, here, of ASMRs helps to reduce the noise or sampling variation in causes with relatively few deaths. ASMRs are based on the European Standard Population 2013. Other standard populations would give very similar results.

[7] We choose to start at 1999 to exclude the impact of the ICD-9 to ICD-10 change, and 2008 is simply the midpoint.

- In both figures, along each low-high pair, the cluster for the high-educated group tends to be more towards the top left than the low-educated group. This indicates that for each gender-age pairing mortality has been improving at a faster rate for all causes in the high-educated group and not just at the all-cause level illustrated in Figure 2.

- In all plots there are dots that indicate negative improvement rates in one or other period. This might be because of increasing prevalence of controllable risk factors such as poor diet. It can also happen as a side-effect of improvements elsewhere. For example, people might, on average, be more frail as a result of increased numbers surviving cardiac events. These survivors might be more susceptible to death from other causes such as vascular dementia.

- Dots that fall below the main diagonal indicate a slowdown in mortality improvements for a particular cause. There is a more complex mixture in the sub-plots. At age 55, low-educated males, the cluster is a little above the diagonal but with few dots below.[8] For high-educated males the cluster is more spread out with more dots below. For the high-educated, this could be due to more significant improvements in controllable risk factors in the first time period than the second. For females aged 55, more points lie below the diagonal indicating a more significant slowdown since 2008: consistent with what we see for all-cause mortality in Figure 2.

- At age 75 (Figure 19) the clusters are more clearly below the diagonal indicating a general slowdown in improvements over the second period. One exception amongst the more-significant causes of death (larger dots) is lung cancer where we see faster improvements in the second period. We discuss this further in the next section.

- Also at age 75, we can see the strong improvements experienced by all groups for ischaemic heart disease: stronger up to 2008 than after.

  Interesting, also, we can see that the pattern of red dots for cardiovascular diseases is quite similar in each plot (similar spacing and almost linear arrangement). These causes of death have smoking, exercise and obesity as common risk factors (see Appendix A). These will vary over time in different ways from one gender-education group to the next, pushing the red dots up or down, left or right in a similar way within each sub-plot. But then the consistent alignment or pattern of the dots might be consistent with differences in the (medical) prevention (e.g. statins) and treatment of each of the causes of death. For example, ischaemic heart disease has seen strong improvements in prevention and treatment, while other heart diseases less so.

[8] The plots of all-cause mortality in Figure 2 suggest a peak in mortality at age 55 around our split point here of 2008 for low-educated males. This is consistent with the balance of dots being above the diagonal in Figure 18.

- There is a reasonably consistent relationship between lung cancer (green) and COPD (blue) in each plot. As with to cardiovascular diseases, lung cancer and COPD have smoking as the common risk factor, pushing the pair of dots around the plot. But the common relationship between the green and blue dots would be consistent with different improvements in the treatment of each disease. As remarked above, we discuss this further in the next section.

- Where there has been an increase in all-cause mortality there is no single big cause of death that explains this. Instead the plots indicate that there are several causes of death that contribute.

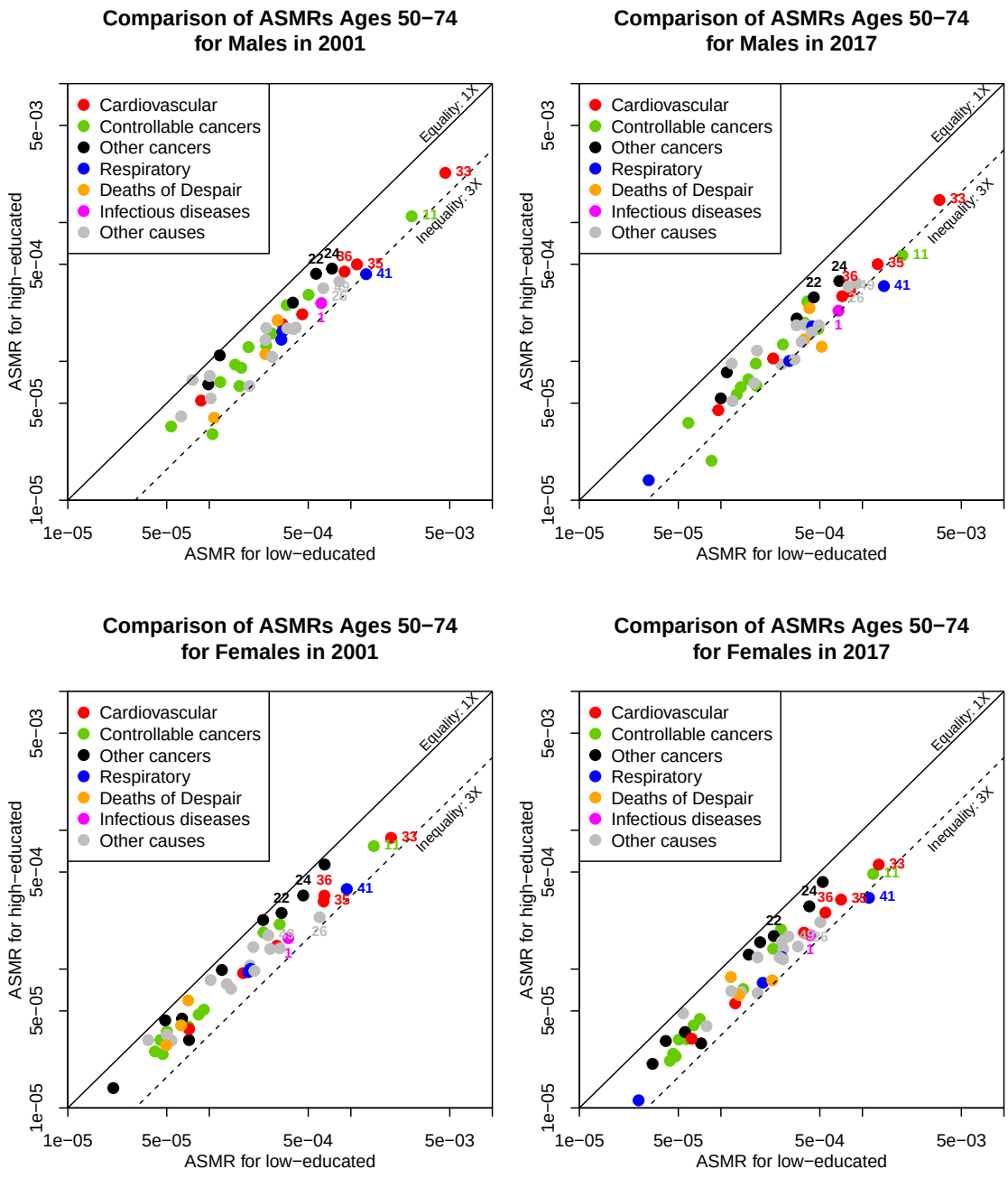

Figure 17: Comparison of ASMRs for low-educated versus high-educated males and females. ASMRs are calculated for the age range 50-74. Dashed line: ASMR for the low-educated is three times the ASMR for the high-educated. Numbered causes of death: see Table 1.

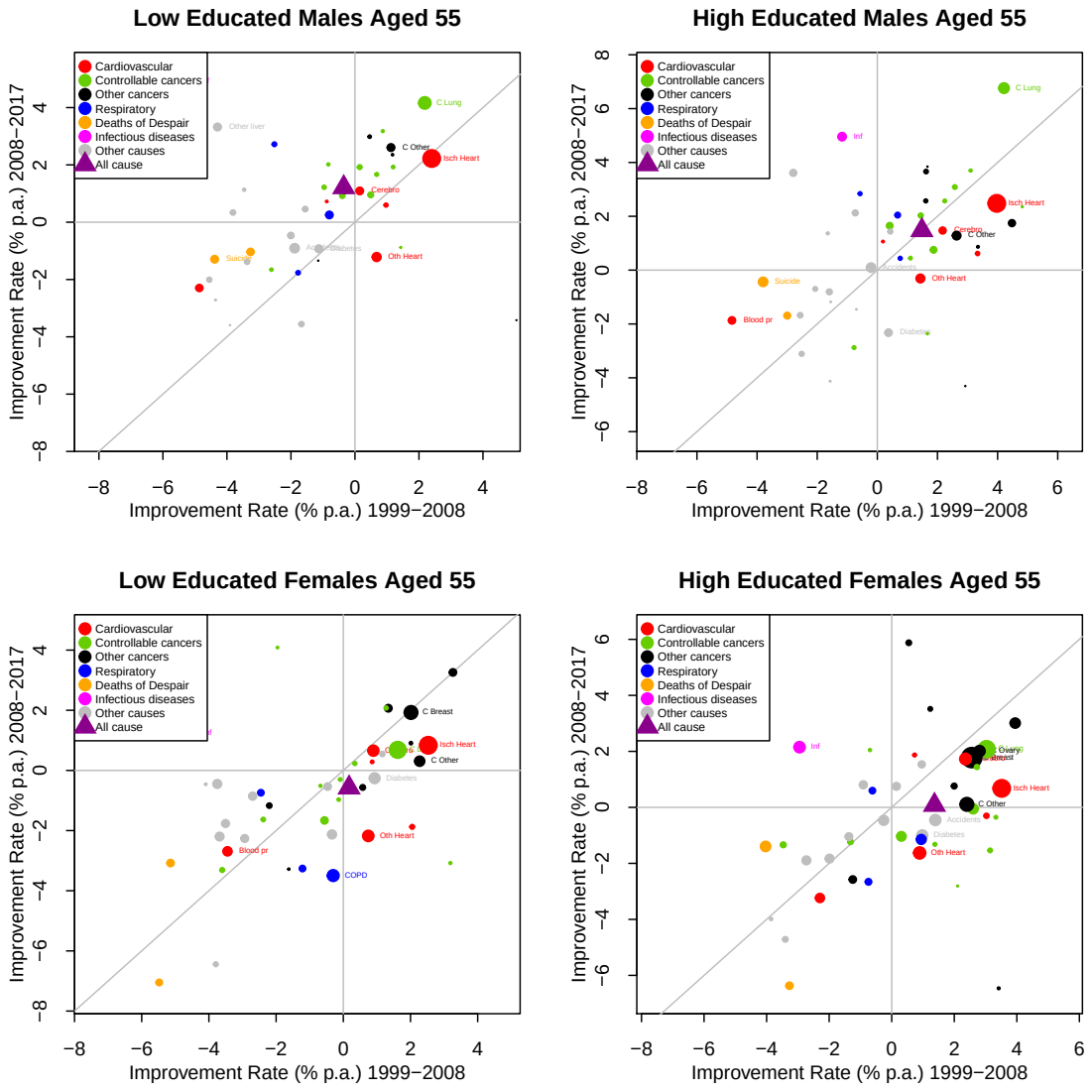

Figure 18: Annual improvement rates at age 55 for 1999-2008 versus 2008-2017. Dots are coloured by cause of death group. Dot size is proportional to the death rate. Larger causes of death have abbreviated labels. Purple triangle: all-cause mortality improvements.

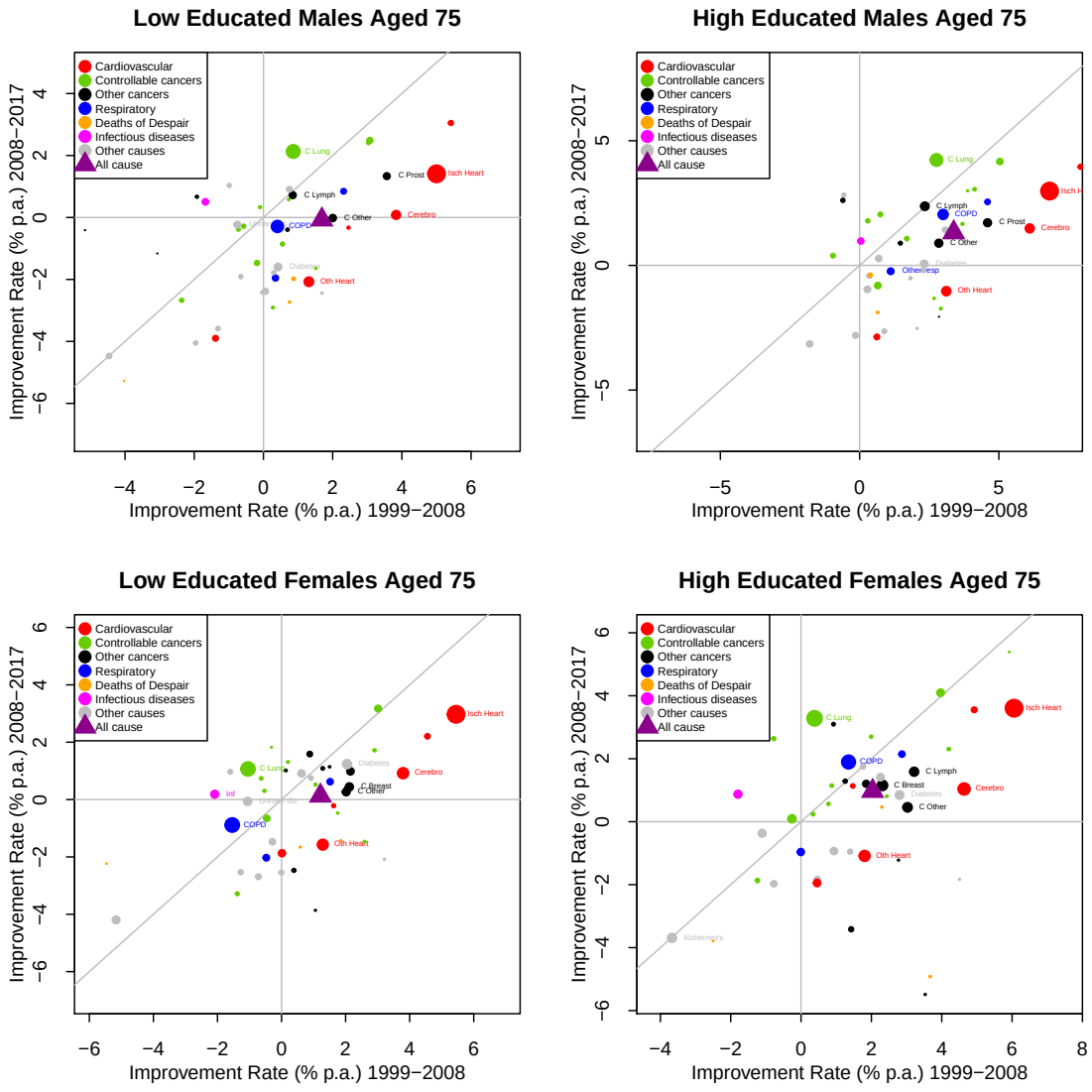

Figure 19: Improvement rates at age 75 for 1999-2008 versus 2008-2017. Dots are coloured by cause of death group. Dot size is proportional to the death rate. Larger causes of death have abbreviated labels. Purple triangle: all-cause mortality improvements.

10 Lung cancer and COPD mortality: the link with smoking

We now return to individual causes of death and focus on causes with specific controllable risk factors. In this section we will look at smoking which is known to be the dominant risk factor for lung cancer (cause 11) and chronic obstructive pulmonary disease (COPD), which is the main cause of death in cause 41 (Appendix A). Smoking has a very high relative risk for both diseases (see, for example, Inoue-Choi et al. 2018, and Forey et al., 2011). Additionally, the relative risk depends on smoker status (never smoked, current smoker, quitter), years since quitting (where relevant), years since commencement of smoking and smoking intensity (e.g. number of cigarettes per day) (see, for example, Anderson et al., 2012, Jemal, et al., 2018, Choi et al., 2018, Guo et al., 2009, and Liu et al., 2015). So, while smoking prevalence is a key factor at the group level this is not the sole determinant of the death rates. So, although death rates might be expected to go up and down with prevalence, over time, the average smoking intensity and duration within the smokers group might gradually change with an additional impact on lung cancer and COPD death rates.

Figure 20 gives a stylised representation of the main transitions. The smoking-related risk factors primarily influence the onset of lung cancer and COPD (transitions A and B). Aggregate group-specific diagnosis rates will, therefore, go up and down almost in lockstep in response to changes in the average smoking prevalance and the other smoking-related risk factors. Once the disease is established, individual death rates (C and D) are mainly dependent on the latest medical advances with relatively little dependence on the risk factors (although giving up smoking will help). So, when we see below death rates diverging, this will be mainly due to changes in smoking prevalence. However, we need to be mindful that, in the US, differing levels of access to healthcare might also lead to some degree of divergence. Education level might also have an impact on the timing of diagnosis and, therefore, the potential for successful treatment (Arık et al., 2021). Everything else being

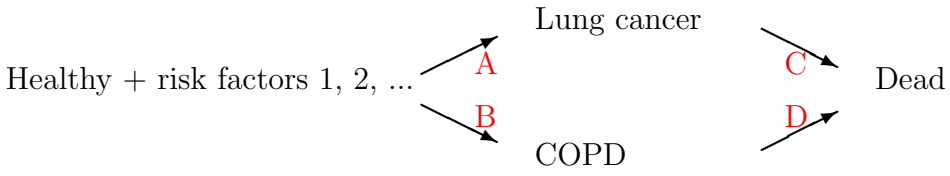

Figure 20: Stylised representation of the links between lung cancer and COPD and smoking-related risk factors. Transitions A and B are very dependent on the underlying risk factors and the associated relative risks. Transitions C and D are much more dependent on the treatments and their improvements over time, and much less dependent on the risk factors.

equal, education level is not considered to be predictive of higher or lower mortality assuming access to the same treatments (Herndon et al., 2008).

Figures 21 and 22 show death rates from lung cancer and COPD respectively against time and against age. The lower plots show that lung cancer (Figure 21) has a much flatter curve of death rates by age (plateauing in the late 70s) than COPD (Figure 22) which is more steadily rising.

Both figures for ages 55 and 75 show a similar pattern of growing inequality. This is particularly so for females aged 75 where rates are initially quite similar and end up more than 3 times different. The initial small gap is consistent with the oldest generations in our sample being unaware in their earlier years of the harmful effects of smoking, and that, in the absence of this fact, low and high-educated females were equally likely to be smokers.

The plots for ages 55 and 75 also reveal an interesting comparison between the two causes of death. In all eight cases (sex $\times$ education $\times$ age), from a given starting point, the death rates from lung cancer show a significant improvement relative to COPD: for example, for low-educated males aged 55, lung cancer mortality has been falling while COPD has been rising slightly. This difference in improvement rates is more significant at age 55 than age 75.

This feature is explored further in Figure 23. This shows heat maps of the ratio of lung cancer to COPD death rates. The colour gradient from top to bottom reflects the different age profiles for the two causes (lower plots in Figures 21 and 22) remarked on above. But as we move from left to right in each heat map we see a gradual colour change that indicates that the ratio is falling at each age. At the higher ages, this decline only begins around 2005. At younger ages the decline starts as far back as 1989.

Given that the two causes of death only have one major risk factor (smoking) a possible inference from these plots is that medical advances in the treatment of lung cancer (see, for example, Howlader et al., 2020, who detail improvements since 2002 in the 2-year survival probability following diagnosis of non-small-cell lung cancer) have had a more significant effect than other advances (if any) in the treatment of COPD. The heat maps present a consistent picture of faster improvements in lung cancer mortality compared to COPD across the four gender-education subgroups: larger relative improvements at younger ages of between 3 to 4% per annum and smaller relative improvements at higher ages of around 1 to 1.5%.

A further possibility is that the differences in improvement rates might be due to the cumulative impact of smoking on the two death rates. For example, suppose younger cohorts of smokers are progressively smoking less intensively (fewer cigarettes per day) on average. This might have a more significant impact on lung cancer than COPD mortality if the cumulative impact of smoking has a stronger influence on lung cancer onset and mortality than COPD. This would then manifest itself in the

form of a slightly different cohort effect for COPD compared to lung cancer.

In Figure 24, for low-educated females, we show heat maps of the proportion of all deaths due to lung cancer (left) and COPD (right). The patterns are slightly different reflecting the different age-profiles for lung cancer and COPD (lower plots in Figures 21 and 22 again). But both plots reveal a strong peak that suggests a cohort effect linked to higher smoking prevalence amongst those born in the 1940s. For high-educated females there is less evidence of a peak of this type. The equivalent heat maps for low and high-educated males suggest a peak in smoking prevalence around 1930.

Figure 25, show heat maps of the ratio of low to high-educated female mortality for lung cancer (left) and COPD (right). These plots show a very clear diagonal pattern and show very strong evidence for cohort effects. Most likely, these cohort effects reflect the changing proportions by cohort of smokers in each group. We see near parity in the top left of each heat map indicating equal proportions of smokers in each education group initially. As we move towards the bottom right the contours indicate a widening gap in the proportion of smokers in the two groups which suggests that successive generations of high-educated females are listening much more than low-educated females to the public-health advice about the harmful effects of smoking. The increase in the ratio from top left to bottom right is bigger for COPD. This might be the result of differences between the two causes of death in the relative risk for smoking: death rates for high-educated females COPD would fall faster if the relative risk is higher than the relative risk for lung cancer.

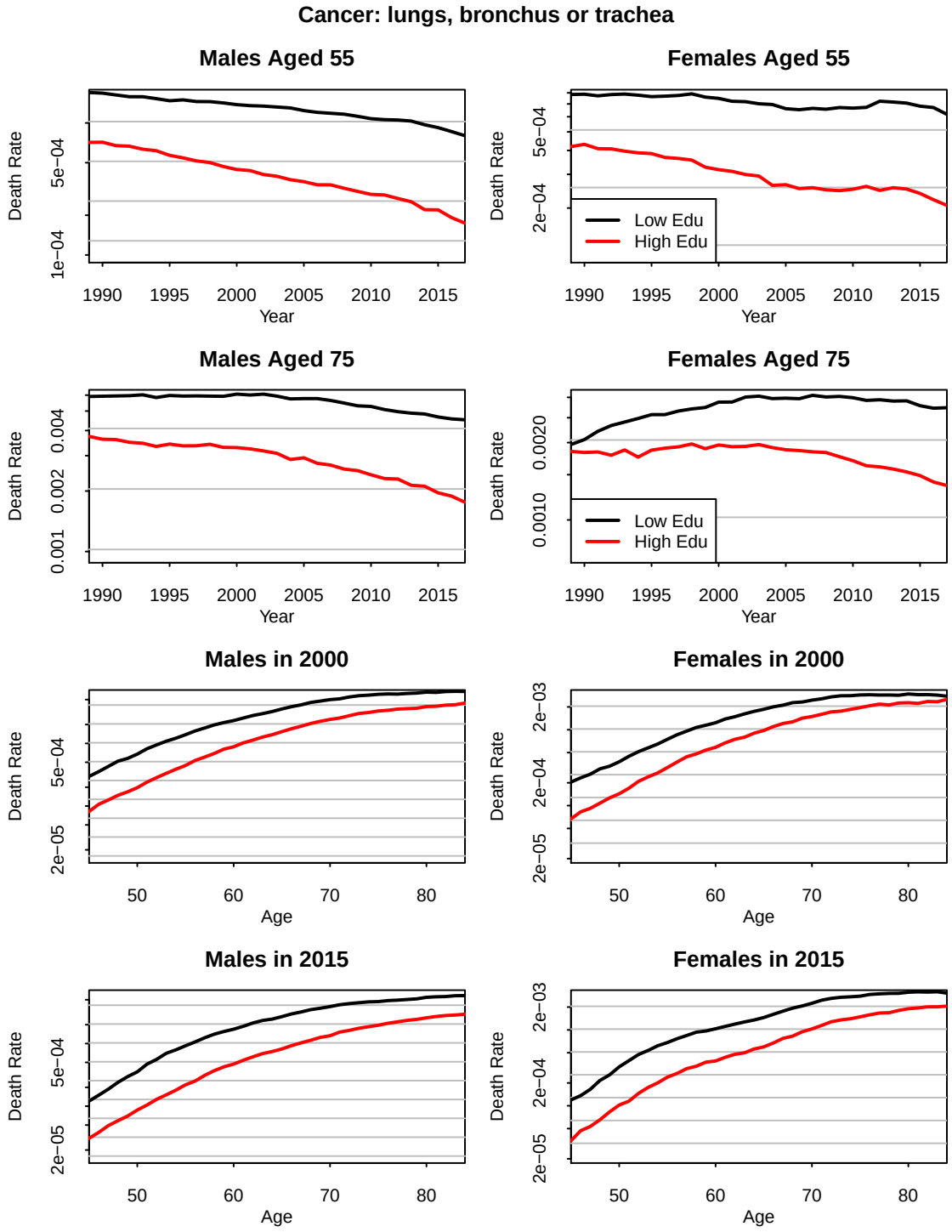

Figure 21: Death rates for lung (and related) cancer. Each sub-plot shows low and high-educated mortality. Left: males. Right: females. Top row: age 55 against time. Second row: age 75 against time. Third row: year 2000 against age. Bottom row: year 2015 against time. Horizontal grey lines: each line is 2× the line below.

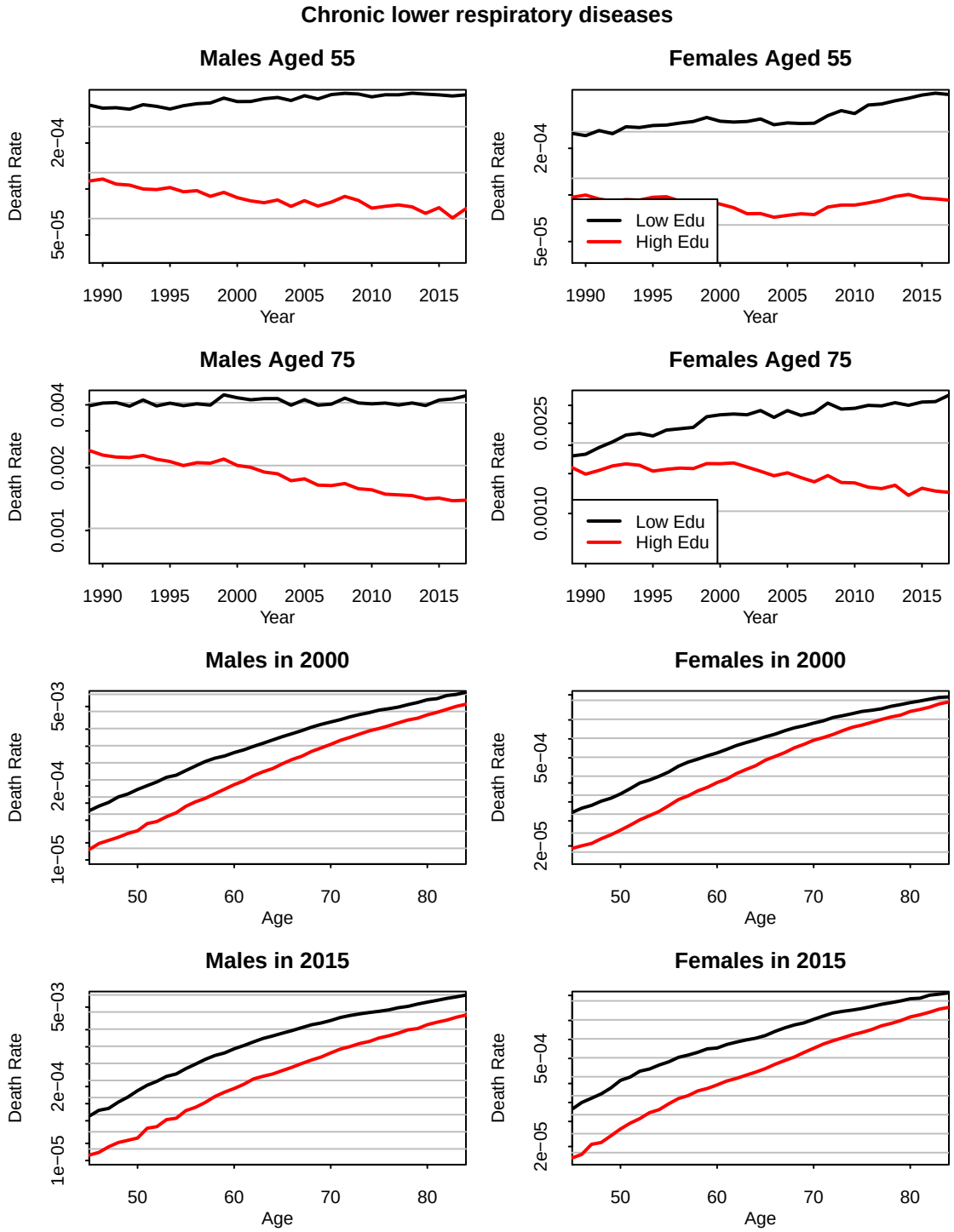

Figure 22: Death rates for chronic lower respiratory diseases (mainly COPD). Each sub-plot shows low and high-educated mortality. Left: males. Right: females. Top row: age 55 against time. Second row: age 75 against time. Third row: year 2000 against age. Bottom row: year 2015 against time. Horizontal grey lines: each line is 2× the line below.

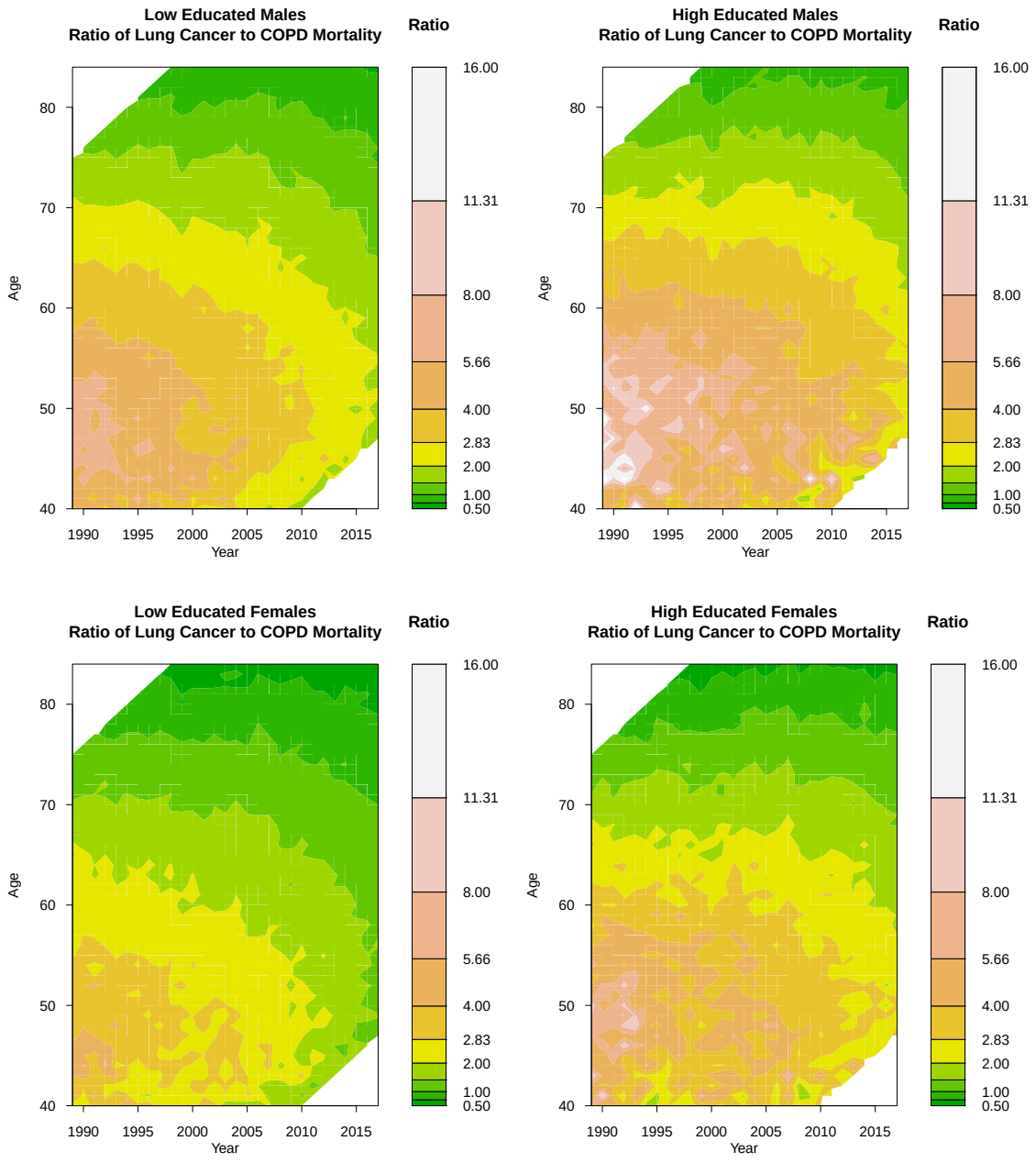

Figure 23: Ratio of lung cancer death rate to COPD death rate for each sex and education level.

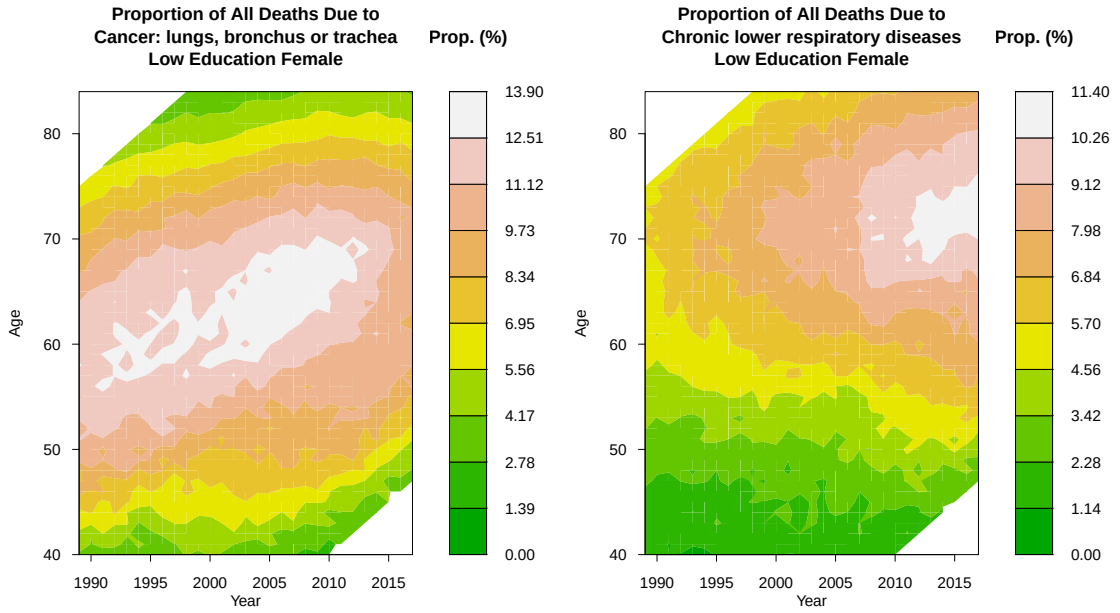

Figure 24: Comparison of lung cancer (left) and COPD (right) mortality. Cause-specific mortality as a proportion of all cause mortality for low-educated females.

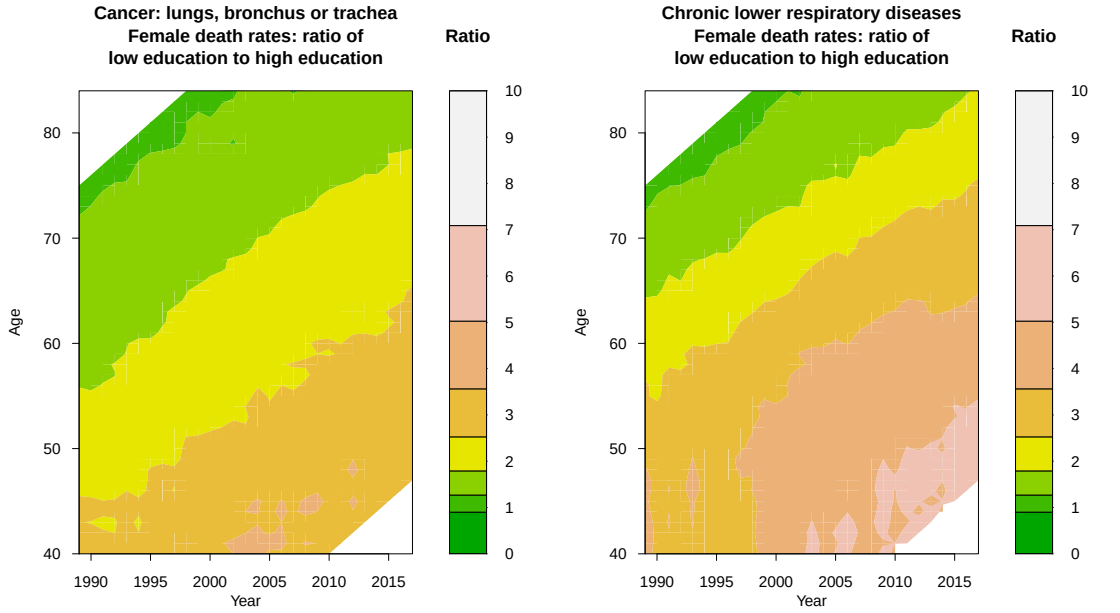

Figure 25: Comparison of lung cancer (left) and COPD (right) mortality. Ratio of mortality for low-educated to high-educated females.

10.1 Lung-cancer, COPD and smoking summary

This section has shown how it is possible to make inferences about smoking habits and relative improvements in disease treatment on the basis of death-rate data alone. We have found:

- that there has been a growing gap between smoking prevalence in the low and high-educated subgroups;
- that smoking prevalence amongst females has been rising relative to males;
- there have been potential peaks in smoking prevalence for males born around 1930, and born in the early 1940s for low-educated females;
- evidence that medical advances have been more significant for lung cancer than COPD;
- evidence that COPD has a higher relative risk for smoking than lung cancer.

We again stress that lung-cancer and COPD death rates depend not just on smoking prevalence, but also on smoking intensity, duration, cessation and time since cessation. As a consequence (Redondo Lourés and Cairns, 2021), we do not see an exact match between smoking prevalence and death rates.

11 Liver diseases and alcohol

As with smoking, excessive alcohol consumption is well known to be associated with increased mortality from a variety of causes, some of which we highlight here. Our discussion here complements the more detailed overview by Gutterman (2021) and references therein.

Figure 26 (panels (a) to (c)) shows heat maps of the proportions of deaths due to liver cancer, alcoholic liver disease (cirrhosis) and other liver diseases for low-educated males. Panel (a) for liver cancer has a very striking hot-spot towards the right and its orientation suggests that this is due to a significant cohort effect, peaking for those born in the early 1950s. This same cohort effect is also visible for alcoholic and other liver diseases (panels (b) and (c)) but at different ages. Panel (b) also shows what might be a later cohort effect below age 50 after 2005. However, further investigation suggests that this is the result of a more general increase in deaths from alcoholic liver disease at all ages after 2005. The different positions of the peaks for the early 1950's cohort is consistent with the different age-profile of deaths from the three causes. Death rates from liver cancer rise steadily with age (although more steeply up to around age 60). In contrast, death rates from alcoholic

liver diseases peak in the 50's, and the curve for other liver diseases lie between the two. These three contrasting curves then influence the position of the peak in the proportion-of-all-cause heat maps.

This commonality is consistent with known links between liver cancer and liver disease,[9] Liver cirrhosis and other liver diseases are known risk factors for liver cancer, and these, in turn have excessive alcohol consumption as a key risk factor.

The suggested common cohort effect is confirmed when we fit the CBD-X3 model (Dowd et al., 2020; see Appendix B). The model is fitted independently to each cause of death and the fitted cohort effects, $\gamma(c)$, are plotted against year of birth in Figure 26 panel (d). These are clearly well aligned and are, therefore, consistent with the existence of a strong alcohol-specific cohort effect: that is, the fitted cohort effect picks up cohort-specific variation in excessive alcohol consumption. The swings in the cohort effect indicate that, everything else being equal, death rates from liver cancer are about 50% higher for those born in the early 1950s compared to those born 10 years before or after. The magnitude of the swings reflects a combination of the relative risk for excessive alcohol consumption for these causes and swings in the prevalence of excessive alcohol consumption. The magnitude of the swing in the cohort effect suggests that relative risks are likely to be high for these diseases (see, for example, Hart et al., 2010), while the magnitude of a swing in the prevalence of excessive alcohol consumption needs further investigation.

The cohort effects in Figure 26 panel (d) are, as remarked above, very well aligned, but they are of different magnitudes. One possible implication is, that if excessive alcohol consumption is the key driver, then the relative risk associated with that behaviour is highest for liver cancer and lowest (but still high) for alcoholic liver disease. However, some studies suggest that, for liver cancer, the relative risk associated with excessive alcohol consumption is up to 2 times (see, for example, Bagnardi et al., 2015, and Petrick et al., 2018), whereas, for other liver diseases the relative risk is higher (see, for example, Im et al., 2021, and Roerecke et al., 2019, who focus on the onset of liver cirrhosis).

Another possibility is that the level of excess alcohol consumption also makes a difference: for example, moderate excess alcohol consumption (perhaps via a non-fatal case of liver disease) might lead to a greater risk of liver cancer. However, the meta-analysis of Turati et al. (2014) indicated that moderate alcohol consumption has only a modest impact on liver cancer mortality.

Similar patterns can be found for high-educated males and low and high-educated females. The size of the effect is a bit smaller, though for both high-educated subgroups and females. The relative magnitudes of the cohort effects in Figure 26 panel (d) for low-educated males is quite similar to high-educated males (that is, liver cirrhosis less strong than liver cancer). For females, the cohort effects are

[9] See, for example, `www.cancer.org/cancer/liver-cancer/causes-risks-prevention/risk-factors.html`.

more noisy (fewer cases) but it also seems that relative differences in magnitude of the cohort effects are smaller than for males. This is consistent with the analysis of Roerecke et al. (2019) who estimate a higher relative risk for excessive alcohol consumption for females than for males.

11.1 Liver diseases summary

Data for three liver-related causes of death point to a very strong cohort effect. Either directly or indirectly, excessive alcohol consumption is the key controllable risk factor allowing us to interpret this as being the alcohol-specific cohort effect.

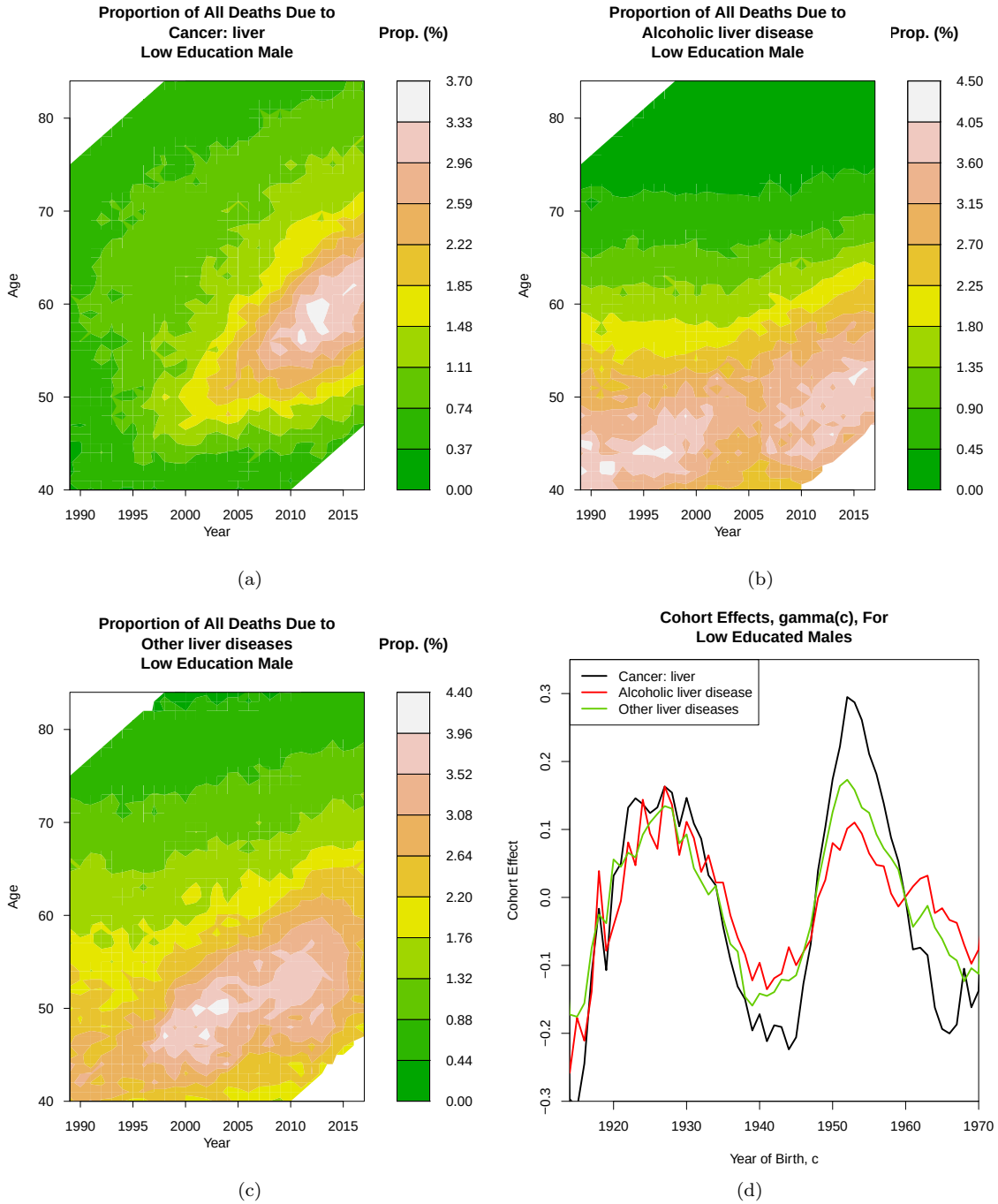

Figure 26: Low-educated males. (a)-(c): Proportions of deaths due to liver cancer, alcoholic liver disease, other liver disease and other digestive diseases. (d) Cohort effects, $\gamma(c)$, for the three causes of death.

12 Deaths of despair

Deaths of despair (Case and Deaton, 2015) with our groupings are a group of three causes of death (alcoholic liver disease, suicide and accidental poisoning: numbers 43, 48 and 50) that are strongly associated with people experiencing a general feeling of despair due to their economic and social circumstances (current and perceived future prospects).

We have already considered alcoholic liver disease but we now look at this alongside suicide and accidental poisoning. In Figures 27, 28 and 29 we plot death rates against year and age for these three causes. Deaths of despair constitute only a small percentage of deaths at high ages, but are much more significant in the 40s and 50s.

The three plots indicate a significant worsening in death rates at age 55 (and other middle ages) but the degree of worsening varies quite considerably between the three causes. Alcoholic liver disease and suicide have moderate increases, with bigger increases for females. But the third group of plots for accidental poisoning presents a shocking picture: at age 55 we see that death rates have risen by as much as 20 times (low-educated females) since 1989.

Further investigation points to accidental poisoning having a cohort effect that is somewhat similar to those for alcoholic liver disease (Figure 26 (d)). So, although we referred previously to this as being the alcohol-specific cohort effect it might equally be referred to as the deaths-of-despair cohort effect. It is an indication that excessive alcohol consumption goes hand in hand with a general feeling of despair. In the bottom two rows of Figure 29 the bulge that pushes out from the left for both males and females is exactly the cohort effect for the cohorts born in the early 1950s.

Beyond the three specific deaths of despair (causes 43, 48 and 50), we have already identified other liver disease and liver cancer as being strongly correlated with alcoholic liver disease. But when we look at fitted cohort effects using the CBD-X3 model we can also detect a similarly-shaped cohort effect for the modelfit to the pneumonia data. This could reflect the recognised link between excessive alcohol consumption (see, for example, Kornum et al., 2012) and pneumonia hospitalisations. But pneumonia also has a strong connection to smoker status (Kornum et al., 2012) so the pneumonia cohort effect is not uniquely driven by alcohol consumption and despair.

12.1 Deaths of despair summary

So-called deaths of despair (alcoholic liver disease, suicide and accidental poisoning) have increased significantly since 1989, especially for accidental poisoning. But we

have also found that these increases are stronger still amongst cohorts born in the early 1950s (males) to late 1950s (females).

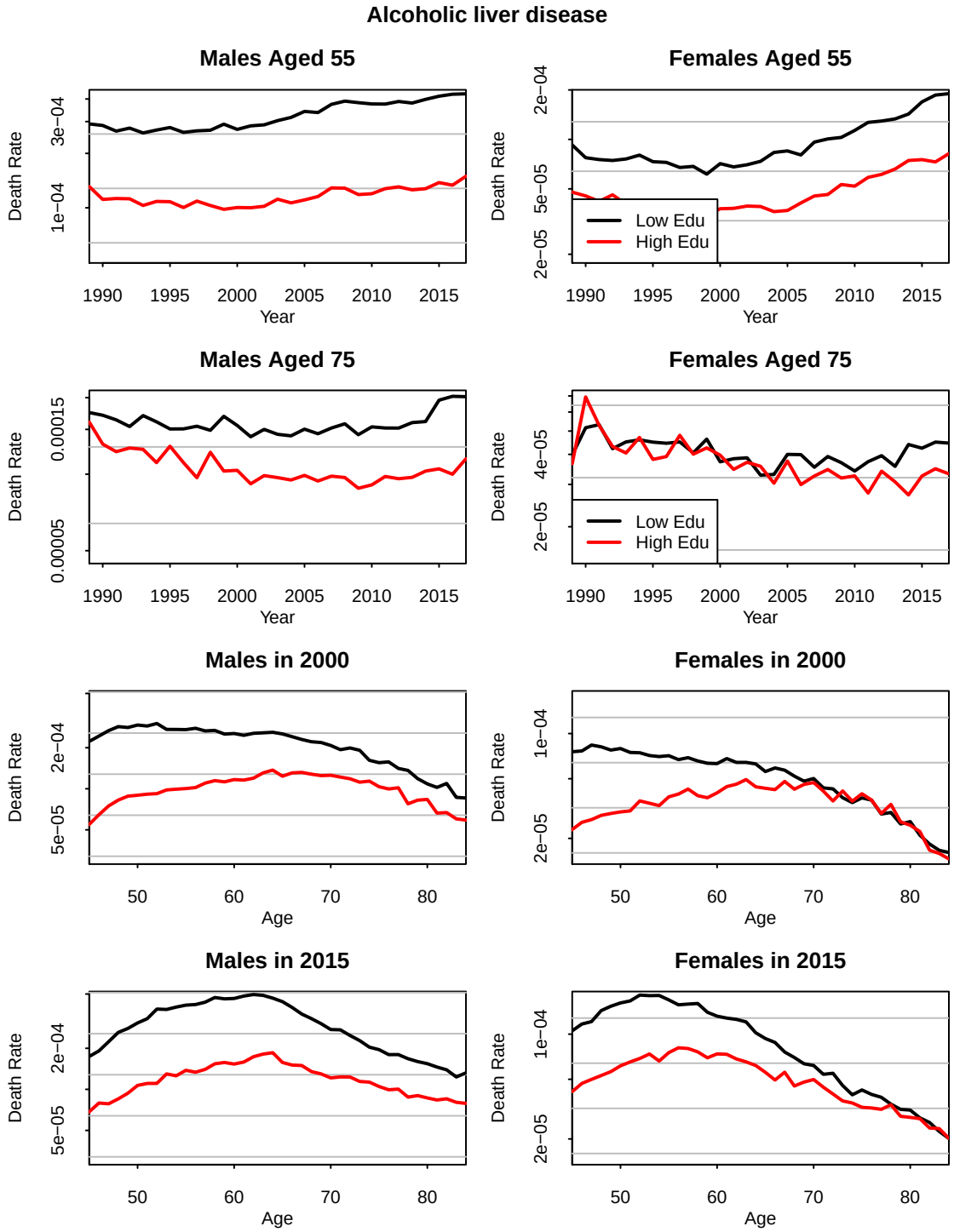

Figure 27: Death rates for alcoholic liver disease. Each sub-plot shows low and high-educated mortality. Left: males. Right: females. Top row: age 55 against time. Second row: age 75 against time. Third row: year 2000 against age. Bottom row: year 2015 against time. Horizontal grey lines: each line is 2× the line below.

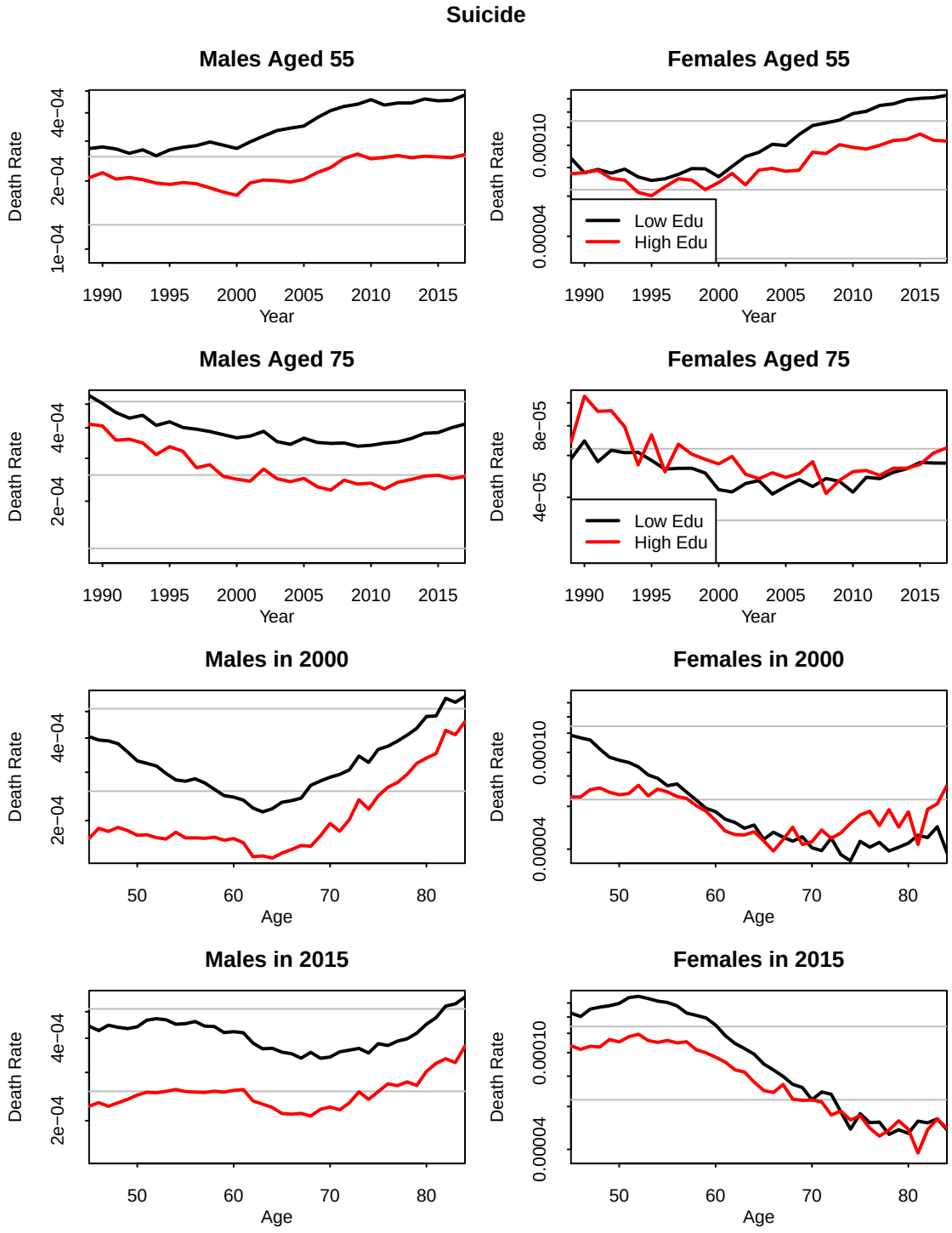

Figure 28: Death rates for suicide. Detail as in Figure 27.

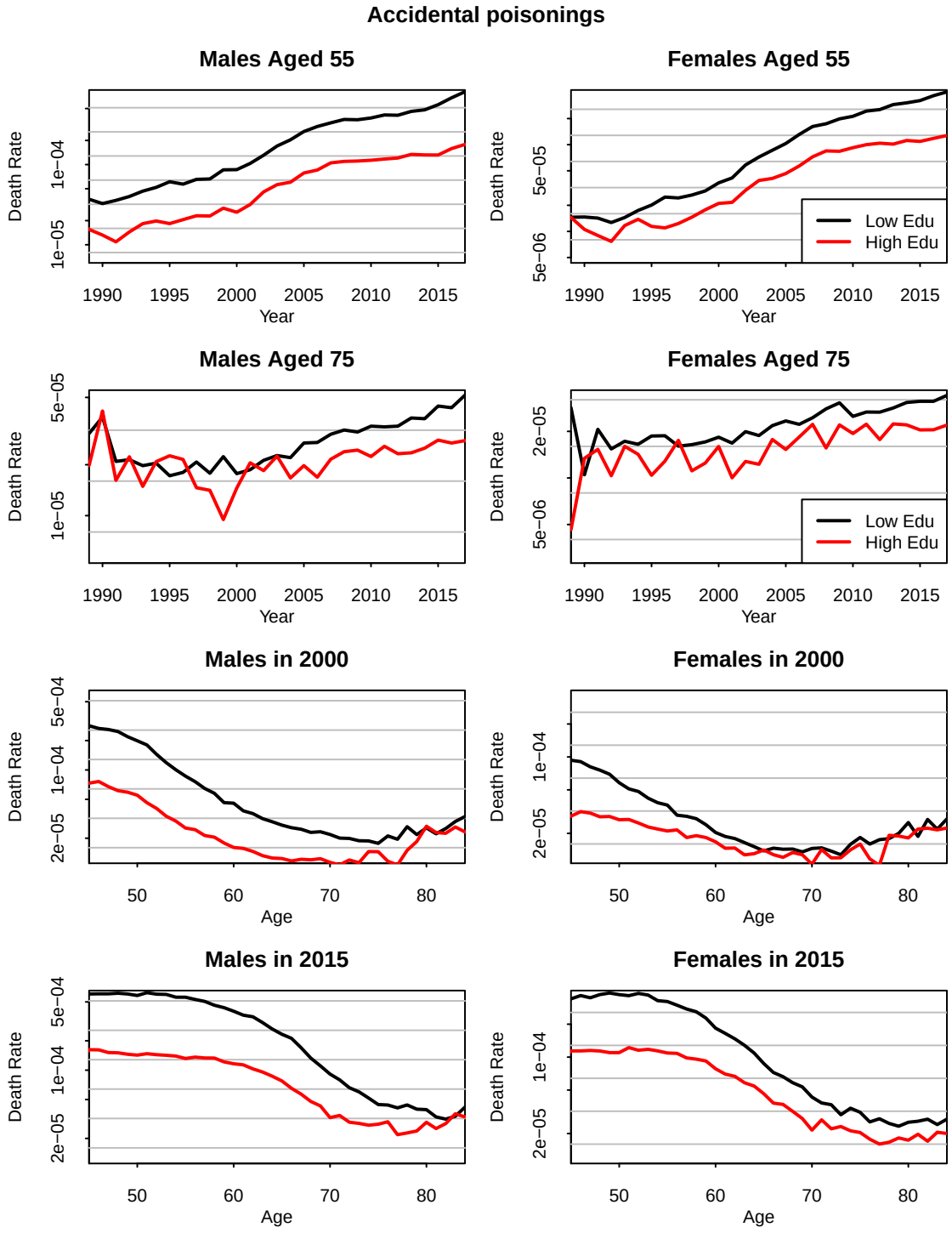

Figure 29: Death rates for accidental poisoning. Detail as in Figure 27.

13 Accidental deaths

In Figure 30 we investigate how accidental deaths in 2000 and 2015 (cause 49) vary by age as a proportion of all causes excluding external causes (see also Figure 11). The general downwards slope from left to right reflects the well-known observation that those who die at younger ages are much more likely to die from an external cause than at older ages. However, an expected and quite striking feature of these plots is that for each sub-group, the curve is quite flat above age 70. At the same time, we also see an increase between 2000 and 2015. At high ages, 'accidental falls' predominate in this group and the flat curves here indicate that the chances of, for example, having a life-threatening fall are proportional to an individual's probability of death, particularly if this quantity is linked to physical frailty. It might be, though, that physical frailty is more closely linked to some causes of death than others, a line of work that needs further investigation: for example, dementia-related deaths are likely to be preceded by a longer period of physical frailty and greater risk of a fall than some other causes. So, rather than consider accidental deaths as a proportion of all-cause mortality, it might be better to compare with a suitably weighted average of other death rates, with greater weight attached to causes of death that are preceded by longer periods of physical frailty.

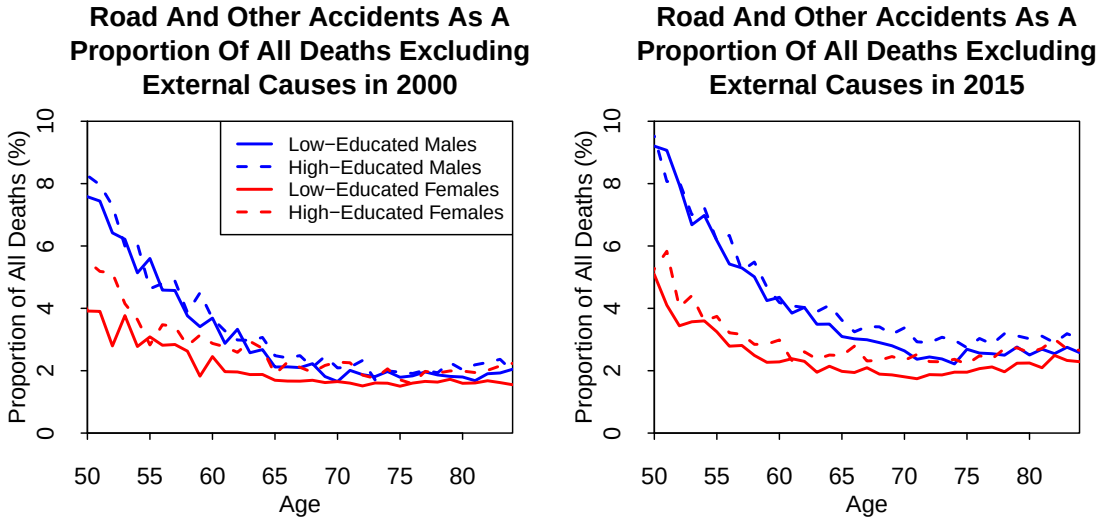

Figure 30: Deaths in 2017 due to road and other accidents as a proportion of all-cause mortality excluding external causes against age.

The increase in the proportion from 2000 to 2015 might suggest that, while all-cause death rates are falling, underlying physical frailty and the risk of death from an accidental fall are either static or falling less quickly. In other words, through other medical advances we are, perhaps, keeping physically frail people alive for longer (meaning that improvements in healthy life expectancy are slower than im-

provements in (total) life expectancy). A related explanation is that death rates from dementias and related neurological diseases have been rising: again an indirect pointer that older people are spending more years in a state of physical frailty. A third, potential reason is a gradual trend towards recognising a fall as the cause of death.

Given the flatness on the curve at high ages and the striking similarity between sex and education levels, a further, open question is what underpins the remaining small differences: higher proportions for high educated and for males, especially in 2015? One possibility is that higher educated males and females spend longer in a physically frail state. This needs further investigation, but it can be noted (see, for example, Figure 11) that the proportion at high ages of deaths due to dementias and related neurological diseases is slightly higher for high-educated males and females than low-educated males and females. Equally, though, the proportions of deaths due to neurological diseases in Figure 11 are higher for females while accidental deaths (Figure 30 for females are lower).

13.1 Road and other accidents summary

Within each group we have found that, above age 70, the proportion of all deaths that are due to accidents within each gender-education group with small variation by gender, education level and over time. This leveling off is consistent with the incidence of potentially life-threatening falls and other accidents being approximately proportional to an individual's underlying physical frailty. The increase in proportion from 2000 to 2015 is consistent with the observation that, although life expectancies are increasing, they are also spending more years in a physically frail state or poor health.

14 Neurological Diseases

In Figure 31 we have plotted observed death rates for the three main groups of neurological diseases that affect older age groups: vascular and other dementias (cause 27); Parkinson's disease (29); and Alzheimer's and related degenerative diseases (30). For comparison, we have added ischaemic heart disease (IHD; 33) and cerebrovascular diseases (36). IHD has been chosen as being typical of the large number of older-age causes of death that are approximately Gompertz with a growth rate of about 0.1 (that is, a gradient of about 0.1 on a logarithmic scale).[10] The three neurological diseases stand out very clearly as having a very much steeper gradient than all other causes of death (about 0.2), and this feature is consistent across all four gender-education groups. Parkinson's flattens slightly at the higher ages, but is still steeper than most other causes of death. Cerebrovascular disease as included here as mini-strokes feature prominently in the symptoms of vascular dementia. However, as can be seen, cerebrovascular diseases are more-closely aligned with the Gompertz(0.1) than the Gompertz(0.2).

We do not have any specific biological explanation for this very striking observation, but it points to two conclusions:

- mortality from a large group of causes of death grow in line with what we might postulate as a general (and, probably, unavoidable) ageing process (Gompertz(0.1));
- but there is a distinct process of neurological deterioration that grows at a much faster rate (at least up to age 85 in our data).

For comparison, we also plot death rates in 2001 for the same causes in Figure 32, and we can note that the steepness of the neurological death-rate curves is the about same. Building on our observations in Section 15, we can also see that death rates from the three neurological disease groups are all higher (to varying degrees) in 2017 than 2001. As remarked before, this might be the result of several factors including greater numbers surviving previous cardiovascular events.

Closer inspection of Figures 31 and 32 (as well as plots not included here) allows us to make some further observations:

- low-educated males and females have quite similar death rates by age for each of vascular dementia and Alzheimer's for both 2001 and 2017 (and, indeed, years in between);

[10] A smaller, but significant number of causes exhibit either flatter or non-linear shapes than Gompertz(0.1). Some causes of death, such as alcoholic liver diseases, peak in the 50's or 60's and then decline. Other curves (mainly cancers, with lung cancer as a good example) increase with age but deviate significantly from linearity on a logarithmic scale and, in particular, get less steep with age.

- high-educated males and females have quite similar death rates by age for each of vascular dementia and Alzheimer's for both 2001 and 2017 (and, indeed, years in between);
- but high-educated males and females have significantly lower death rates by age than low-educated males and females for each of vascular dementia and Alzheimer's.

The equality between males and females at a given level of education is unusual given that most other causes of death exhibit significant differences, although the reasons behind this equality are not clear. The differences between low and high educated groups might be linked to differing levels of access to healthcare following the onset of one of these neurological diseases, or the likelihood of obtaining an earlier diagnosis.

14.1 Neurological diseases summary

We have identified that all three of the neurological disease groups that affect older ages (vascular dementia, Parkinson's and Alzheimer's) have a common statistical property: the death-rate curve is very steep compared to all other causes of death. The reasons for this are not known but it suggests a different biological mechanism from other older-age causes of death.

Age-specific death rates from all three are on the rise, possibly due to increased numbers of survivors of prior cardiovascular events.

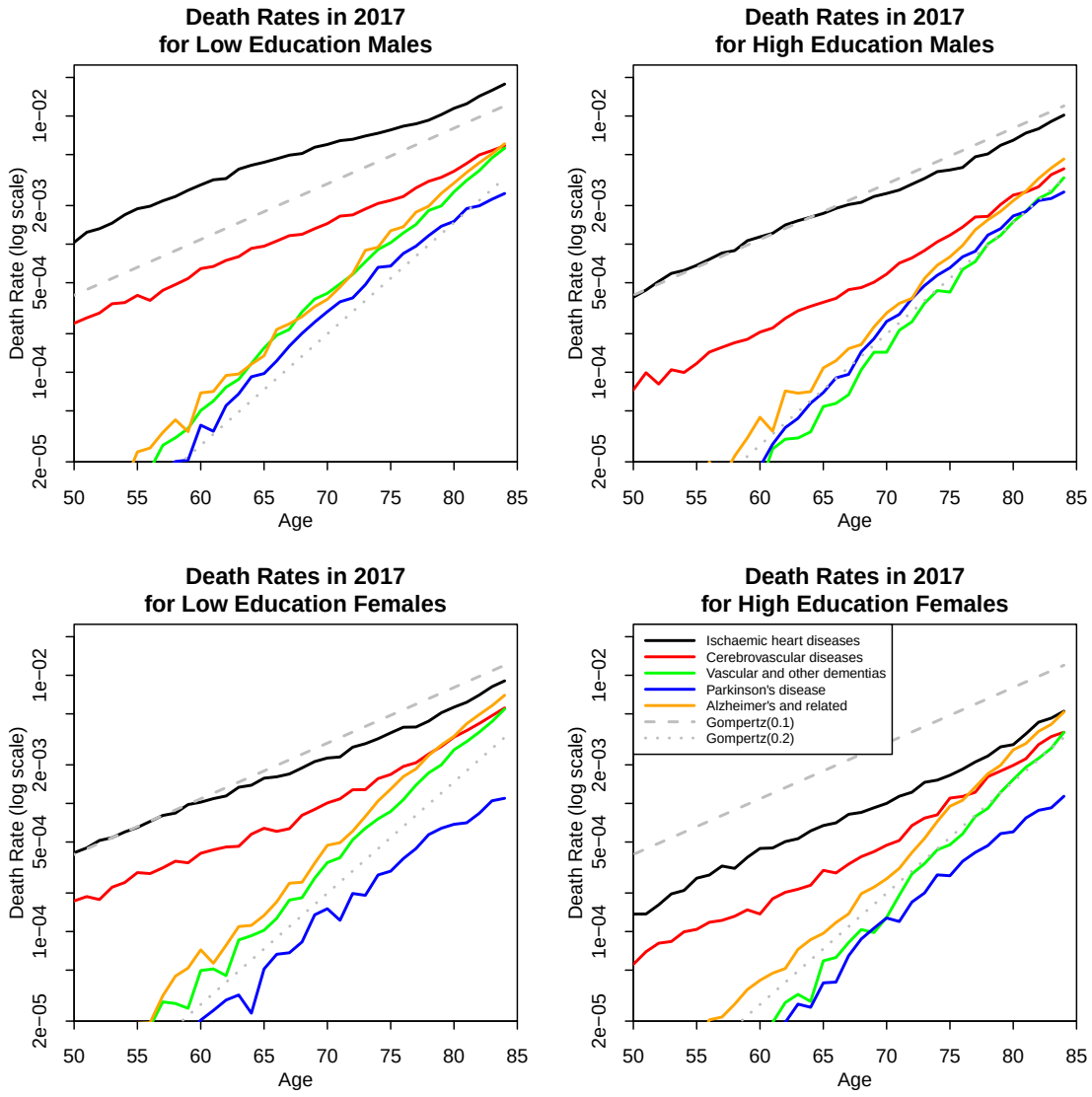

Figure 31: Death rates in 2017 against age for selected causes of death by gender and education level. Gompertz curves with growth-rate parameters of 0.1 and 0.2 (dashed and dotted lines) are added for reference.

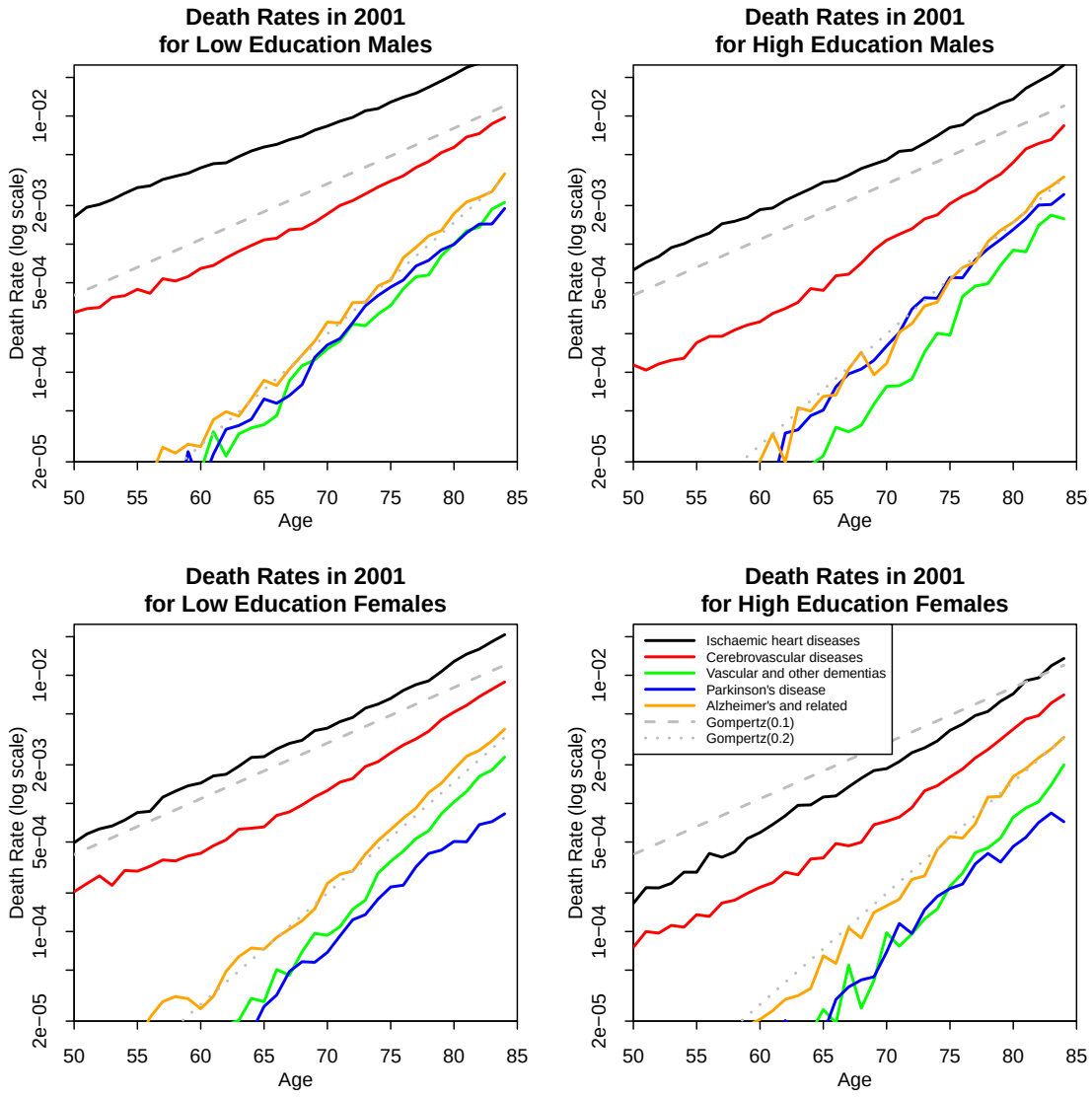

Figure 32: Death rates in 2001 against age for selected causes of death by gender and education level. Gompertz curves with growth-rate parameters of 0.1 and 0.2 (dashed and dotted lines) are added for reference.

15 Causes of death with no significant controllable risk factors

Balancing some of the earlier sections, we now look at selected causes of death where there are either no known *controllable* risk factors or where a link to a controllable risk factor is established but the relative risk is fairly close to 1. These are mainly certain cancers and some neurological diseases (Parkinson's and Alzheimer's).

In Figures 33 and 34 we plot Age Standardised Mortality Rates (ASMRs) for six causes for males and females respectively. For males, for prostate and lymphatic cancer, there are no significant controllable risk factors. In spite of this, we can clearly see a significant widening of the inequality gap. For prostate cancer, the gap is initially quite narrow. But, by 2017, the ASMR for low-educated males is about 60% higher. Some of the gap might be due to *conscientiousness*, a "big five" personality trait that is known to have a significant impact on mortality (see, for example, Kern and Friedman, 2008, and Deary et al., 2010). In this context, conscientiousness might have an impact on how quickly an individual gets a positive diagnosis following onset, and it might affect how well an individual engages with the subsequent treatment process. Possible explanations for the growing gap are:

- growing inequality in access to good quality treatments following diagnosis;
- interaction between a shrinking proportion with a low education (Figure 1) and conscientiousness;
- changes in the racial mix of the low and high-educated groups.

The racial mix is certainly likely to contribute to the underlying gap. Cheng et al. (2009) indicate that mortality rates from prostate cancer for African Americans after adjusting for socio-economic status are more than double equivalent rates for non-Hispanic whites. Aditionally, African American males at all age groups are more likely to fall in the low-educated group (US Census Bureau, 2020, using data based on the Current Population Survey). In combination these would produce some degree of separation between the low and high-educated groups. However, the same data by age group (US Census Bureau, 2020) indicate that the education gap between African Americans and the rest of the population is *narrowing*. Genetic/racial differences cannot, therefore, provide an easy explanation for the widening mortality inequality gap in deaths from prostate cancer.

Figure 33 also shows ASMRs for ages 75-84 for Parkinson's and Alzheimer's diseases: both causes of death with very modest or no significant controllable risk factors. For Alzheimer's we see a growing gap: the reason for which is not obvious given the absence of risk factors. The increases might be the result of a suggested risk factor that is yet to be established. Specifically, the growing numbers of survivors

of prior cardiovascular events might be more susceptible to Alzheimer's than other, healthier people. Parkinson's is the only cause of death with a reverse inequality, but even this gap has disappeared by 2017. The reverse inequality might be due to smoking prevalence in the older groups given that smoking is thought to reduce the risk of Parkinson's disease (Mappin-Kasirer et al., 2020). Data discussed in Section 10 indicate lower levels of smoking amongst high-educated males than low-educated males. However, this does not explain the narrowing gap over time. Smoking prevalence has (Section 10) has also declined faster amongst high-educated than low-educated males and so the reverse mortality gap should be getting wider rather than the narrowing gap that we see in Figure 33. A possible reason for the closing of the gap is, again, changing levels of access to and engagement with the treatment process following diagnosis.

For females (Figure 34) there is less of a difference for Parkinson's which might reflect the near parity in smoking prevalence noted in Section 10. Alzheimer's presents a similar picture to males. For the four cancers we see a mixed picture in terms of the gap. The gap is widening slowly for breast cancer, ovarian cancer and lymphatic cancer, but less quickly than for males prostate cancer. For breast cancer there are relatively few non-controllable risk factors (unlike race for prostate cancer) so that the observable and growing gap is more likely to be due to access to healthcare and, perhaps, conscientiousness. Cancer of the uterus has obesity as a risk factor,[11] so the increasing ASMRs are likely to be linked to generally-increasing levels of obesity in both the low and high-educated groups. Nevertheless, inequality for cancer of the uterus remains fairly constant and at a similar level to breast, ovarian and lymphatic cancers in 2017. Improvements in the treatment of cancer of the uterus will, of course, mitigate the impact of increasing obesity, but the balance between the two with these data is unclear: our data simply show us the net effect.

Figure 33 also includes pancreatic cancer and cancer of the urinary organs (mainly cancer of the kidney) as they have relatively low levels of inequality compared to other causes. However, they are known to have some cases that are linked to smoking and obesity[12]. For cancer of the urinary organs and pancreatic cancer the gap is initially small but steadily widens. Recalling Section 10, this widening is could be due to a widening gap in smoking prevalence in the two groups. It could also be due to a widening gap in obesity rates in the two groups. But, given the discussion above about prostate cancer, we now have to consider access to, and quality of, cancer treatments as a potential cause of the widening gap.

[11] See `www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk`.

[12] See, for example, `www.cancer.org/cancer/pancreatic-cancer/causes-risks-prevention/risk-factors.html` and Appendix A, Figure 36.

15.1 No significant controllable risk factors summary

For this group of causes of death, arguably, there should be very little mortality inequality. But the data do reveal an inequality gap and that this gap is getting wider. This might the result of a combination of factors including: access to and quality of healthcare following diagnosis; conscientiousness impacting on engagement with the healthcare system; non-preventable risk factors such as race.

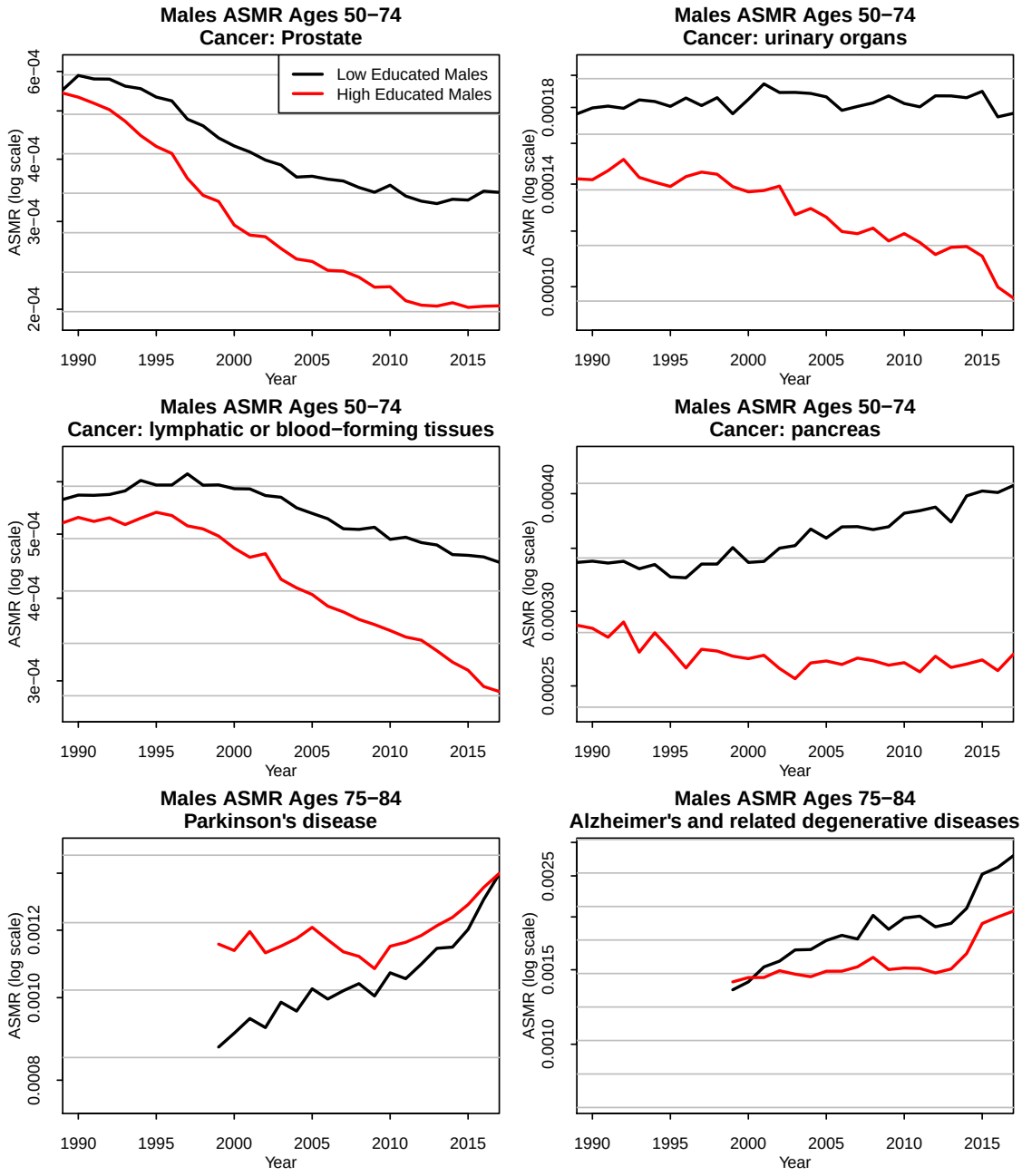

Figure 33: Age Standardised Mortality Rates (ASMRs) for either ages 50-74 (top and middle rows) or ages 75-84 (bottom row; data are incomplete up to 1998) for males for selected causes of death. Horizontal gray lines are spaced at multiples of 1.2.

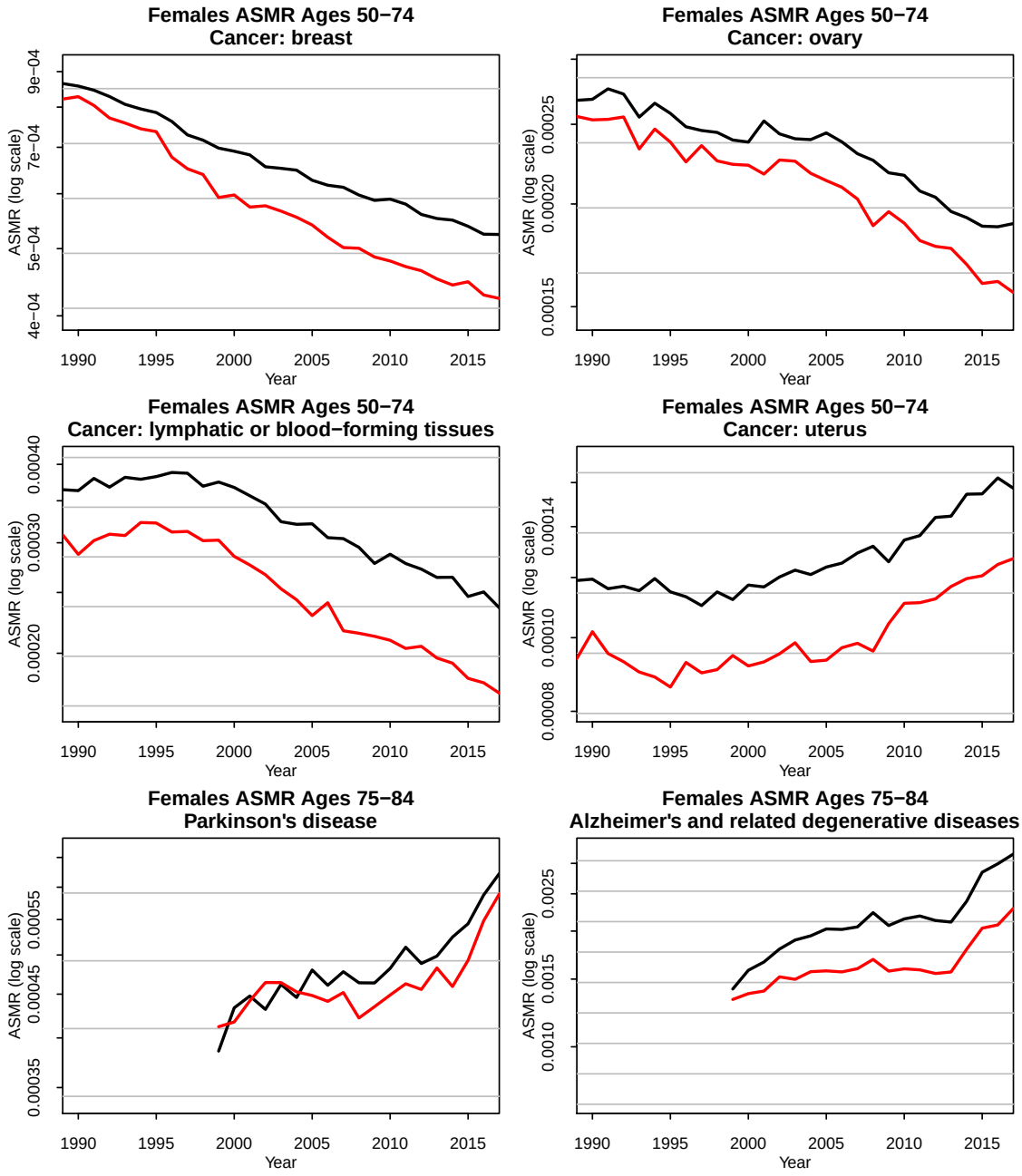

Figure 34: Age Standardised Mortality Rates (ASMRs) for either ages 50-74 (top and middle rows) or ages 75-84 (bottom row; data are incomplete up to 1998) for females for selected causes of death. Horizontal gray lines are spaced at multiples of 1.2.

16 Conclusions

In this paper we have carried out a wide ranging analysis of 51 causes of death for the older US population by gender and education level, with a particular focus on 7 groups with common features: ischaemic heart disease; smoking-related diseases; alcohol-related diseases; deaths of despair; accidents; and causes with no significant risk factors.

We have made use of a wide range of graphical diagnostics and have demonstrated how different visualisations of the same data can reveal different features of the data: mortality trends; improvement rates; inequality; and cohort effects. A feature of some of our analysis is how the presence of a strong cohort effect or high levels of inequality can be linked to the existence of a large and significant relative risk for a specific *controllable* risk factor (for example, the effect of smoking on lung-cancer death rates). But, when we look at causes of death with no significant risk factors, we can observe smaller, but still significant, levels of inequality. This points to the possibility that non-controllable risks factors such as conscientiousness and access to healthcare as having relative risks larger than 1.

In future work, we seek to analyse any remaining significant causes of death along similar lines. We also have modelling work in progress building on the CBD-X3 model used here to identify a small number of cohort effects that can be linked to specific controllable risk factors such as smoking and excessive alcohol consumption. Finally, we hope to build on these combined results to assess the potential impact of, for example, a 50% reduction in smoking prevalence, or reductions in excessive alcohol consumption.

References

Anderson, C.M., Burns, D.M., Dodd, K.W., Feuer, E.J. (2012) Birth-cohort-specific estimates of smoking behaviors for the U.S. population. Risk Analysis 32, S14–S24.

Arık, A., Babad, Y., Raphael, D. (2021a) Mortality by causes of death. *IAA Mortality Forum Paper*. `https://www.actuaries.org/IAA/Documents/Publications/Papers/MWG_MortalityByCauseOfDeath.pdf`

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

Arnold, S., and Glushko, V. (2021) Short- and Long-Term Dynamics of Cause-Specific Mortality Rates Using Cointegration Analysis. *North American Actuarial Journal* e-publication, 23 pages, DOI: 10.1080/10920277.2021.1874421

Arnold, S., and Sherris, M. (2013) Forecasting mortality trends allowing for cause-

of-death mortality dependence. *North American Actuarial Journal* 17, 273–282.

Arnold, S., and Sherris, M. (2015) Causes-of-death mortality: What do we know on their dependence? *North American Actuarial Journal* 19, 116–128.

Arnold, S., and Sherris, M. (2016) International cause-specific mortality rates: New insights from a cointegration analysis. *ASTIN Bulletin* 46, 9–38.

Bagnardi, V., Rota, M., Botteri, E., Tramacere, I., Islami, F., Fedirko, V., Scotti, L., Jenab, M., Turati, F., Pasquali, E., Pelucchi, C., Galeone, C., Bellocco, R., Negri, E., Corrao, G., Boffetta, P., La Vecchia, C. (2015) Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *British Journal of Cancer* 112, 580-593.

Cairns, A.J.G., Blake, D., and Dowd, K., (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73, 687-718.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* 13, 1–35.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Khalaf-Allah, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin* 41, 29–59.

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

Cairns, A.J.G., Kallestrup-Lamb, M., Rosenskjold, C.P.T., Blake, D., and Dowd, K. (2019) Modelling socio-economic differences in the mortality of danish males using a new affluence index. *ASTIN Bulletin* 49, 555–590.

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

CDC, 2019. National health interview survey. Accessed: November 2019, `https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm`

Cheng, I., Witte, J.S., McClure, L.A., Shema, S.J., Cockburn, M.G., John, E.M., and Clarke, C.A., (2009) Socioeconomic status and prostate cancer incidence and mortality rates among the diverse population of California. *Cancer Causes & Control* 20, 1431-1440.

Deary, I.J., Weiss, A., and Batty, G.D. (2010) *Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities.*

Monograph, University of Edinburgh.

Dowd, K., Cairns, A.J.G., and Blake, D. (2020) CBDX: A workhorse mortality model from the Cairns-Blake-Dowd family. *Annals of Actuarial Science* 14, 445–460.

Forey, B., Thornton, A.J., and Lee, P.N. (2011) Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema *BMC Pulmonary Medicine* 11: 36.

Gottesman, R.F., Albert, M.S., Alonso, A., Coker, L.H., Coresh, J., Davis, M., Deal, J.A., McKhann, G.M., Mosley, T.H., Sharrett, A.R., Schneider, A.L.C., Windham, B.G., Wruck, L.M., Knopman, D.S. (2017) Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort. *JAMA Neurology* 74(10): 1246-1254.

Guo, N.L., Tosun, K., and Horn, K. (2009) Smoking duration, respiratory symptoms, and COPD in adults aged $\geq$45 years with a smoking history. *Lung Cancer* 66, 386–392.

Gutterman, S. (2021) Alcohol and Mortality: An Actuarial Issue, *North American Actuarial Journal*, e-publication, 21 pages, DOI: 10.1080/10920277.2021.1946660

Hart, C.L., Morrison, D.S., Batty, G.D., Mitchell, R.J., and Davey Smith, G. (2010) Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies *British Medical Journal*, 340: c1240.

Herndon, J.E., Kornblith, A.B., Holland, J.C., and Paskett, E.D. (2008) Patient Education Level As a Predictor of Survival In Lung Cancer Clinical Trials *Journal of Clinical Oncology* 26, 4116–4123.

Holford, T.R. (1991) Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual Review of Public Health* 12, 425–457.

Howlader N, Forjaz G, Mooradian MJ, et al. (2020) The effect of advances in lung-cancer treatment on population mortality. *New England Journal of Medicine*, 383, 640-649.

Human Mortality Database (2021) University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `www.mortality.org`. Accessed 15/11/2021.

Hunt, A., and Blake, D. (2014) A general procedure for constructing mortality models. *North American Actuarial Journal* 18, 116–138.

Im, P.K., Millwood, I.Y., Kartsonaki, C., Gui, Y., Chen, Y., Turnbull, I., Yu, C., Du, H., Pei, P., Lv, J., Walters, R.G., Li, L., Yang, L., and Chen, Z. (2021) Alcohol drinking and risks of liver cancer and non-neoplastic chronic liver diseases in China: a 10-year prospective study of 0.5 million adults *BMC Medicine* 19, 216.

Inoue-Choi, M., Hartge, P., Liao, L.M., Caporaso, N., and Freedman, N.D. (2018)

Association between long-term low-intensity cigarette smoking and incidence of smoking-related cancer in the national institutes of health-AARP cohort *International Journal of Cancer* 142, 271–280.

Jemal, A., Miller, K.D., Ma, J., Siegel, R.L., Fedewa, S.A., Islami, F., Devesa, S.S., and Thun, M.J. (2018) Higher lung cancer incidence in young women than young men in the United States. *New England Journal of Medicine* 378, 1999–2009.

Jemal, A., Ward, E., Anderson, R.N., Murray, T., and Thun, M.J. (2008) Widening of socioeconomic inequalities in U.S. death rates, 1993–2001. *PLOS ONE* 3, e2181.

Kern, M.L., and Friedman, H.S. (2008) Do conscientious individuals live longer? A quantitative review. *Health Psychology* 27, 505-512.

Kornum, J.B., Due, K.M., Nørgaard, M., Tjønneland, A., Overvad, K., Sørensen, H.T., and Thomsen, R.W. (2012) Alcohol drinking and risk of subsequent hospitalisation with pneumonia *European Respiratory Journal* 39, 149-155

Larsson, S.C., and Wolk, A. (2007) Overweight, obesity and risk of liver cancer: a meta-analysis of cohort studies. *British Journal of Cancer*, 97, 1005-1008.

Lee, R.D., and Carter, L.R., 1992. Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87, 659–671.

Liu, Y., Pleasants, R.A., Crift, J.B., Wheaton, A.G., Heidari, K., Malarcher, A.M., Ohar, J.A., Kraft, M., Mannino, D.M., and Strange, C. (2015) Smoking duration, respiratory symptoms, and COPD in adults aged $\geq$45 years with a smoking history *International Journal of Chronic Obstructive Pulmonary Disease* 10, 1409–1416.

Mappin-Kasirer, B., Pan, H., Lewington, S., Kizza, J., Gray, R., Clarke, R., and Peto, R. (2020) Tobacco smoking and the risk of Parkinson disease: A 65-year follow-up of 30,000 male British doctors, *Neurology* 94, e2132-e2138.

Ng, J.B., Turek, M., and Hakim, A. M. (2013) Heart disease as a risk factor for dementia. *Clinical epidemiology* 5, 135–145.

Noftle, E.E., and Robins, R.W. (2007) Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology* 93, 116–130.

Orkaby, A.R., Driver, J.A., Ho, Y., Lu, B., Costa, L., Honerlaw, J., Galloway, A., Vassy, J.L., Forman, D.E., Gaziano, J.M., Gagnon, D.R., Wilson, P.W.F., Cho, K., and Djousse, L. (2020) Association of Statin Use With All-Cause and Cardiovascular Mortality in US Veterans 75 Years and Older. *Journal of the American Medical Association* 324, 68–78.

Petrick, J.L., Campbell, P.T., Koshiol, J., Thistle, J.E., Andreotti, G., Beane-Freeman, L.E., Buring, J.E., Chan, A.T., Chong, D.Q., Doody, M.M., Gapstur, S.M., Gaziano, J.M., Giovannucci, E., Graubard, B.I., Lee, I.M., Liao, L.M., Linet, M.S., Palmer, J.R., Poynter, J.N., Purdue, M.P., and McGlynn, K.A. (2018) To-

bacco, alcohol use and risk of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: The Liver Cancer Pooling Project. *British Journal of Cancer* 118, 1005–1012.

Redondo Lourés, C., and Cairns, A.J.G. (2020) Mortality in the US by education level. *Annals of Actuarial Science* 14, 384–419.

Redondo Lourés, C., and Cairns, A.J.G. (2021) Cause of Death Specific Cohort Effects in U.S. Mortality. *Insurance: Mathematics and Economics* 99, 190-199.

Renshaw, A., Haberman, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance Mathematics and Economics* 38, 556–570.

Richards, S.J., Kirkby, J.G., and Currie, I.D. (2006) The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal* 12, 5–38.

Roerecke, M., Vafaei, A., Hasan, O.S.M., Chrystoja, B.R., Cruz, M., Lee, R., Neuman, M.G., and Rehm, J. (2019) Alcohol Consumption and Risk of Liver Cirrhosis: A Systematic Review and Meta-Analysis. *American Journal of Gastroenterology* 114, 1574-1586.

Salami, J.A., Warraich, H., Valero-Elizondo, J., Spatz, E.S., Desai, N.R., Rana, J.S., Virani, S.S., Blankstein, R., Khera, A., Blaha, M.J., Blumenthal, R.S., Lloyd-Jones, D., and Nasir, Khurram (2017) National Trends in Statin Use and Expenditures in the US Adult Population From 2002 to 2013: Insights From the Medical Expenditure Panel Survey *JAMA Cardiology* 2, 56-65.

Schlenger, W.E., Corry, N.H., Williams, C.S., Kulka, R.A., Mulvaney-Day, N., DeBakey, S., Murphy, C.M., and Marmar, C.R. (2015) A prospective study of mortality and trauma-related risk factors among a nationally representative sample of Vietnam veterans. *American Journal of Epidemiology* 182, 980–990.

Taitt, H.E. (2018) Global Trends and Prostate Cancer: A Review of Incidence, Detection, and Mortality as Influenced by Race, Ethnicity, and Geographic Location. *American Journal of Men's Health* 12, 1807–1823.

Trautwein, U., Ludtke, O., Roberts, B. W., Schnyder, I., and Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology* 97, 1115-1128.

Turati, F., Galeone, C., Rota, M., Pelucchi, C., Negri, E., Bagnardi, V., Corrao, G., Boffeta, P., La Vecchia, C. (2014) Alcohol and liver cancer: a systematic review and meta-analysis of prospective studies *Annals of Oncology* 25, 1526-1535.

United States Census Bureau (2020) Educational Attainment in the United States: 2019. `www.census.gov/content/census/en/data/tables/2019/demo/educational-attainment/cps-detailed-tables.html`

U.S. Department of Health and Human Services (2004) The health consequences of smoking: A report of the surgeon general. www.ncbi.nlm.nih.gov/books/NBK44695/

Villegas, A.M., Kaishev, V.K., and Millossovich, P. (2018) StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software* 84, 1–38.

Willets, R. (2004) The cohort effect: Insights and explanations. *British Actuarial Journal* 10, 833–877.

A ICD10 codes and significant risk factors

	Informal description: primary cause of death	ICD10 Codes
1	Infectious diseases	A00-B99
2	Cancer in lip, oral cavity, pharynx	C00-C14
3	Oesophageal cancer	C15
4	Stomach cancer	C16
5	Cancer of colon	C18
6	Cancer of rectum or anus	C19-C21
7	Cancer of liver	C22
8	Cancer of pancreas	C25
9	Cancer: other digestive system	C17, C23-C24, C26
10	Cancer in larynx	C32
11	Cancer in lungs, bronchus or trachea	C33-C34
12	Skin cancer	C43-C44
13	Breast cancer	C50
14	Cancer in female genital organs: cervix	C53
15	Cancer in female genital organs: uterus	C54-C55
16	Cancer in female genital organs: ovary	C56
17	Cancer in female genital organs: other	C51-C52, C57-C58
18	Cancer in male genital organs: prostate cancer	C61
19	Cancer in male genital organs: other	C60, C62-63
20	Bladder cancer	C67
21	Cancer in urinary organs	C64-C66,C68
22	Cancer in lymphatic or blood-forming tissues	C81-C96
23	Benign tumours or tumours without specification	D10-D49
24	Cancer in other locations	C30-C31, C37-C39, C40-C42, C45-C49, C69-C80, C97, D00-D09
25	Diseases in blood and blood-forming organs	D50-99, E00-E07, E15-E90
26	Diabetes	E08-E14
27	Vascular and other dementias	F01-F03
28	Other mental illnesses	F04-F99
29	Parkinson's disease	G20-G25
30	Alzheimers and related degenerative diseases	G30-G31
31	Other nervous system	G00-G19, G32-G44, G46-G99, H00-H99
32	Increased blood pressure or rheumatic fever	I00-I16
33	Ischaemic heart diseases	I20-I25
34	Non-rheumatic valve disorders	I34-I38
35	Other heart diseases	I26-I28, I30-I33, I39-I52
36	Cerebrovascular diseases	G45, I60-I69
37	Circulatory diseases	I70-I99
38	Influenza	J09-J11
39	Pneumonia	J12-J18
40	Other acute respiratory infections	J00-J06, J19-J22
41	Chronic lower respiratory diseases	J40-J47
42	Other respiratory diseases	J30-J39, J60-J99
43	Alcoholic liver disease	K70
44	Other liver diseases	K71-K77
45	Other digestive diseases	K00-K69, K80-K99
46	Diseases in skin, bones, and connective tissue	L00-L99, M00-M99
47	Diseases of the genitourinary system	N00-N99
48	Suicide	X60-X84, Y10-Y34
49	Road and other accidents	V00-V99, W00-W99, X00-X39, X50-X59, Y40-Y86, Y87-Y89, Y90-Y99
50	Accidental poisonings	X40-X49
51	Other causes of death	O00-O99, P00-P99, Q00-Q99, R00-R99, U00-U99, X85-X99, Y00-Y09
52	Other factors not used as the primary cause of death	S00-S99, T00-T99, Z00-Z99

Figure 35: Table of ICD codes for the 52 cause of death groupings. Group 52 is for completeness, but not used as a primary cause of death in our data.

Figure 35 lists the 51 causes of death used in this report and provides a mapping to the ICD-10 codes used to allocate individual deaths to a specific cause.

Figure 36 can be considered to be a work in progress[13] that gives an indication of which risk factors have a significant impact on cause-specific mortality. For example, oesophageal cancer has smoking, alcohol, and obesity as controllable or preventable risk factors. The Cancer Research UK website (`cancerresearchuk.org.uk`) adds detail by providing estimates of what proportions of all deaths from specific cancers are likely to be due to specific risk factors.

Risk factors are separated into two groups. The first few columns are mainly controllable or preventable risk factors. The second group lists selected co-morbidities that lead to increased risk of death from specific causes.

[13]Some causes of death (apart from cancers) have yet to be fully reviewed: for example, *other liver diseases*, which has a blank row. Other cause groups are quite mixed and so are also left blank: for example, *infectious diseases*.

Significant Risk Factors

Cause of death	Cancer Research UK: Cases per annum	Cancer Research UK: Preventable Cases (%)	HMD England & Wales deaths 2016 (cause of death database)	Smoking	Diet	Exercise	Alcohol	Obesity	Sun (UV)	Infections	Pollution	Hormones	Workplace/Environment	Other	History of cardiovascular events	High blood pressure (*)	High cholesterol	Diabetes	Stress and anger	Pre-existing respiratory disease	Pre-existing heart disease	Prior head trauma	Liver cirrhosis
Infectious diseases			5846																				
Cancer in lip, oral cavity, pharynx	?	?/46/88	2607	X																			
Oesophageal cancer	9101	59	7044	34			13	27															
Stomach cancer	6697	54	3970	15				6		41													
Cancer of colon ("bowel")	42042	54	8399	7	41	5	6	11			2												
Cancer of rectum ("bowel") or anus	?/1438	?/91	6497							91													
Cancer of liver	5906	49	4745	20			7	23		10								x					x
Cancer of pancreas	9972	31	8403	22				12															
Cancer: other digestive system	1040	20	4288					20															
Cancer in larynx	?	73	729	64																			
Cancer in lungs, bronchus, or trachea	47235	79	30824	72							8		13										
Skin cancer	15920/150000	86/?	3343						86														
Breast cancer	55213	23	10441				8	8				2		5									
Cervical cancer	3192	99.8	733	21						99.8													
Cancer of the Uterus	9314	34	2098					34															
Ovarian cancer	7470	11	3733					7															
Cancer: other female genital organs	?	?	834(M+F)																				
Prostate cancer	47740	0	10592																				
Cancer: other male genital organs	3000?	12	834(M+F)							12													
Bladder cancer	10187	49	4815	45									6	2									
Cancer in urinary organs	?	?	4057	13				24															
Cancer in lymphatic or blood-forming tissues	15972?	40/3/?	11943							40/3/-													
Benign tumours or tumours without specification	?	?	3415																				
Cancer in other locations	?	?	17760																				
Diseases in blood and blood-forming organs, as well as endocrine sufferings and deficiency diseases			3345																				
Diabetes			5818			x		x							x	x	x						
Vascular and other dementias			49204	x				x							x	x	x	x					
Other mental illnesses			1075																				
Parkinson's disease			6689	inverse									x									x	
Alzheimers and related degenerative diseases			18184												?	?	?	?					
Other diseases in the nervous system as well as sense organs			6287																				
Increased blood pressure or rheumatic fever			7703	x	x	x		x															
Ischaemic heart diseases			58928	x		x		x								x	x	x	x				
Non-rheumatic valve disorders			5692																				
Other heart diseases			19794	x		x		x								x	x	x	x				
Cerebrovascular diseases			33611	x		x	x									x	x	x					
Circulatory diseases			11400	x				x								x	x	x	x				
Influenza			436																				
Pneumonia			28100	x																x	x		
Other acute respiratory infections			4409																				
Chronic lower respiratory diseases			31874	x									x										
Other respiratory diseases			9332																				
Alcoholic liver disease			4884				x	?															
Other liver diseases			3411																				
Other digestive diseases			17325		x	x													x				
Diseases in skin, bones, and connective tissue			6006																				
Urine, kidney, genital organs, and breast gland			9467	x				x								X		X			x		
Suicide			5018																				
Road and other accidents			13000																				
Accidental poisonings			3233																				
Other causes of death			3638																				

Figure 36: Significant risk factors for each cause of death. Risk factors with numbers indicate what percentage of deaths (source: `cancerresearchuk.org.uk`) that are considered to be caused by the risk factor. For other causes of death, an X indicates that the risk factor is significant. Columns 2 and 3 indicate, for cancers only, numbers of UK cases and preventable deaths. Column 4 gives total deaths in 2016 from each cause (Human Mortality Database, 2021).

B The CBD-X Model

Dowd et al. (2020) describe the CBD-X family of models as follows:

$$\log m(t,x) = \alpha(x) + \sum_{i=1}^{N} \beta_i(x)\kappa_i(t) + \gamma(t-x),$$

where

- $N = 1, 2, 3$ is the number of age-period effects (we use $N = 3$ in this paper)
- x is the age last birthday,
- t is the calendar year,
- $t - x$ is the year of birth,
- $\alpha(x)$ is a non-parametric age effect,
- the $\beta_i(x)$ are parametric age effects with
  - $\beta_1(x) = 1$ for all x
  - $\beta_2(x) = x - \bar{x}$ where $\bar{x}$ is the mean age
  - $\beta_3(x) = (x - \bar{x})^2 - \sigma_X^2$ where $\sigma_X^2 = \frac{1}{n_x}\sum_x (x - \bar{x})^2$ is the variance of the ages
- $\gamma(c)$ is a non-parametric age affect for year of birth $c = t - x$.

With $N = 3$, we require 7 identifiability constraints:

- $\sum_t \kappa_i(t) = 0$ for $i = 1, 2, 3$,
- $\sum_c \gamma(c)(c - \bar{c})^j = 0$ for $j = 0, 1, 2, 3$ where $\bar{c}$ is the mean year of birth.

(Other constraints can be used.)