

# Dynamic hazards modelling for predictive longevity risk assessment

Elena Kulinskaya<sup>a,\*</sup>, Lisanne Andra Gitsels<sup>a,b</sup>, Ilyas Bakbergenuly<sup>a</sup>, Nigel R. Wright<sup>a,c</sup>

<sup>a</sup> School of Computing Sciences, University of East Anglia, Norwich Research Park, NR47TJ, Norwich, UK

<sup>b</sup> University College London Great Ormond Street Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK

<sup>c</sup> Life & Health Data Science, Aviva

## ARTICLE INFO

### Article history:

Received May 2020

Received in revised form October 2020

Accepted 2 November 2020

Available online 21 November 2020

### JEL classification:

C18

C13

C46

I13

### Keywords:

Hazard function

Health data

Mortality

Population health

Landmark analysis

Gompertz distribution

## ABSTRACT

Predictive risk assessment and risk stratification models based on geodemographic postcode-based consumer classification are widely used in the pension and life insurance industry. However, these are static socio-economic models not directly related to health information. Health information is increasingly used for annuity underwriting in the UK, using health status when the annuity is purchased. In real life, people develop new health conditions and lifestyle habits and can start and stop a certain treatment regime at any time. This requires the ability to dynamically classify clients into time-varying risk profiles based on the presence of evolving health-related conditions, treatments and outcomes. We incorporate landmark analysis of electronic health records (EHR), in combination with the baseline hazards described by Gompertz survival distributions, for dynamic prediction of survival probabilities and life expectancy. We discuss a case-study based on landmark analysis of the survival experience of a cohort of 110,243 healthy participants who reached age 60 between 1990–2000.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Life expectancy (LE) and longevity projections are of the greatest importance both to the pension and insurance industry and to their clients. Longevity-trend projections are used for managing longevity risk in pricing and reserving for insurance and annuity products as well as for costing of public and private pensions. Changes in mortality projections directly affect annuities costs, especially in the decreasing interest rates environment. LE is also a paramount consideration to individuals planning their financial goals and retirement strategies.

Predictive risk assessment and risk stratification models based on geodemographic postcode-based consumer classification are widely used for life insurance/annuities underwriting (Richards, 2008; Villegas and Haberman, 2014). These methods are based on the anticipated differences in longevity among groups of people, such as by sex, age, and deprivation. However, these socio-economic models are not directly related to health information.

Health information is increasingly used for annuity underwriting in the UK, using health status when the annuity is purchased. In medicine, availability of electronic health records (EHR) and other big health data and the accompanying rapid development

of analytical means to interpret such data paves the way to the emergence of precision medicine (Hulsen et al., 2019). Similar to precision medicine, precision actuarial products should aim to tailor underwriting and reserving to the individual and be changing over time with the health characteristics of each client.

In real life, people develop new health conditions and lifestyle habits and can start and stop a certain treatment regime at any time. Additionally, clinical guidelines are regularly updated with new evidence, resulting in new eligibility criteria and treatment courses. Moreover, in the UK, the introduction of the pension freedoms in 2015 resulted in an emergence of a variety of flexible retirement options, from drawdown to fixed-term annuities, necessitating dynamic decision-making with varying time horizon both by the individuals and by the companies.

All this requires the ability to dynamically classify individuals into time-varying subgroups with predictable life expectancy based on the presence of evolving health-related conditions, treatments and outcomes. The next paragraphs set out examples of this for individuals (or their financial advisers) and insurance and pensions providers.

For an individual approaching retirement, using their pension savings pot to purchase an annuity is no longer the default option. Pension freedoms mean that, guided by financial advice, each individual must make decisions about their own investment and longevity risk. These decisions are made not just once at

\* Corresponding author.

E-mail address: [e.kulinskaya@uea.ac.uk](mailto:e.kulinskaya@uea.ac.uk) (E. Kulinskaya).

retirement but dynamically thereafter, allowing for emerging investment performance, changing lifestyle, changing attitudes to risk, and developing health conditions. For example, an individual might choose not to purchase an annuity when they retire, using an income drawdown product and retaining their investment and longevity risk. Over the next years, their changing health status will mean that their life expectancy will change. Based on this, they might decide to use some of their pensions savings to buy an annuity, or they might feel able to increase the level of income that they take from their savings, or they might feel that they have to decrease the income that they take.

When pricing and reserving for their longevity risks, it is becoming increasingly important that insurance and pensions providers allow for heterogeneity of health conditions and lifestyle as well as socioeconomic status, how these change over time and the impact of changes to treatment regimes. The premium calculation would account for possible future health trajectories of the individual, appropriately weighted according to the results of a dynamic model. In simple terms, this will involve pricing to allow for an individual's medical conditions and lifestyle as with underwritten annuities. Pricing for 'healthy' individuals with no serious medical conditions will be 'select' and will have longer life expectancy than an aggregate level. This 'select' effect will also have to allow for the fact that the individual buying an annuity has made an active decision to do so, as explained above. Life expectancies used in pricing annuities will need to allow for possible future health trajectories. Premiums based on historic mortality experience will reflect historic trajectories and this will need adjusting where expected future trajectories are different (e.g. due to changes in treatment regimes, promotion of smoking cessation, obesity prevalence, etc.).

There are implications for both pricing and reserving assumptions used by annuity providers. As an example, if a higher annuity rate is offered to a smoker at age 60, this price will need to allow for future changes to smoking habits, health trajectories and treatment regimes. Ten years later, the longevity risk element of the reserve for this individual, now 70 years old, will need to allow for the possible trajectory of that individual over the last ten years, which is unknown to the annuity provider. Unless the annuity provider requests updates from clients regarding their health and lifestyle, the provider would not know the longevity trajectory of the client. The individual might have given up smoking or developed serious health problems. This longevity risk element will be different from the longevity risk of a 70-year-old smoker who is applying for a new annuity because, for the latter, changes in health, etc. over the previous 10-year period will be known to the annuity provider. The individual did not give up smoking between the ages of 60 and 70, and health status is known, for example. The landmark analysis presented in this paper provides a way of allowing for this dynamically.

In medicine, the Cox proportional hazards model is the most popular method of time-to-event analysis or survival analysis. The vast majority of clinical trials and observational studies that analyse survival outcomes use this model. In our previous work (Kulinskaya et al., 2020) we developed a method to incorporate proportional hazards modelling of EHRs into actuarial modelling of hypothetical changes in population or group life expectancy due to medical advances and health interventions. To do this successfully, some parametric assumptions about the shapes of survival distributions are necessary. We demonstrated that both Gompertz and Weibull survival distributions in combination with the Cox model can be successfully used for actuarial calculations. We illustrated our methodology on the important example of the survival benefits of statins.

Statin therapy for primary prevention of cardiovascular disease (CVD) has been reported to improve life expectancy (Mihaylova et al., 2012) and is widely available in the UK (NICE,

2016). In Kulinskaya et al. (2020) we estimated the effect of statin prescription on longevity at ages 70 and 75. We also calculated the hypothetical changes in national life expectancy if all eligible people were to be prescribed a statin. However, these calculations were based on the health and lifestyle characteristics and statin prescription or lack thereof only at the baseline (Gitsels et al., 2016). Additionally, an assumption of proportional hazards means that a treatment such as statins or a risk factor such as smoking has the same effects at any age, but this does not appear to be a realistic assumption.

In this paper, we incorporate landmark analysis, a dynamic method of survival analysis, in combination with the baseline hazards described by the Gompertz survival distributions, to translate time-varying medical and lifestyle hazards into dynamic predictions of survival and life expectancy. We discuss a case study based on our landmark analysis on the use of statins, Gitsels et al. (2021). This case study considers the EHRs of a cohort of 110,243 participants who reached age 60 between 1990–2000 with no previous history of cardiovascular disease or statin prescription at baseline. Participants' medical history was updated at 'landmark' time points occurring every six months. Individual life expectancy depends on the individual time-varying health trajectory. Changes in overall and group life expectancy depend on the composition of the population and these attributes. As an application of our methodology, we developed a life expectancy calculator that is available on <https://mylongevity.org>.

## 2. Basics of landmark analysis

This section briefly reviews the methodology of landmark analysis. Cox proportional hazards model is introduced in Section 2.1, and the main concepts of landmark analysis are outlined in Section 2.2.

### 2.1. Cox proportional hazards model

Cox proportional hazards model is a semi-parametric method of survival analysis which is widely used in medical applications. In Cox regression, given a vector of covariates  $Y$ , the hazard function or force of mortality at time  $t$  is factorised as  $\mu(t|Y) = \mu_0(t) \exp(Y^T \beta)$  for a vector of parameters  $\beta$ . The baseline hazard  $\mu_0(t)$  is not specified, and interest is centred on the hazard ratios  $\mu(t|y_1)/\mu(t|y_2) = \exp((y_1 - y_2)^T \beta)$ , which do not depend on the baseline hazard and are constant over time. This is termed the proportional hazards assumption. However, some parametric assumptions about the shape of the baseline hazard are necessary to estimate a survival function or a life expectancy. The three common parametric distributions easily combined with the proportional hazards assumption are exponential, Weibull and Gompertz distributions. Denote the baseline log-hazard by  $\lambda_0(t) = \log \mu_0(t)$ . Then  $\lambda_0(t) = a$  corresponds to the exponential,  $\lambda_0(t) = a + bt$  to Gompertz  $G(a, b)$ , and  $\lambda_0(t) = a + b \log t$  to Weibull baseline hazards. The respective proportional hazards models are called Cox–exponential, Cox–Gompertz and Cox–Weibull survival models (Bender et al., 2005).

We shall concentrate on Cox–Gompertz model, as it is well accepted that the Gompertz distribution provides a good description of human mortality between ages 50 and 95 (Brenner et al., 1993; Spiegelhalter, 2016; Vaupel, 2010). Under this model, the increase in the annual hazard of mortality associated with ageing one year is approximately constant. For England and Wales in 2010, this increase was 1.103 for men and 1.111 for women.

We are aware of an ongoing debate on the validity of Gompertz law at high ages. Recent publications arguing for an extension of Gompertz law to 106 years and beyond include Newman (2018) and Gavrilov and Gavrilova (2019a,b). A competing view

is that of mortality deceleration – distributions of this nature can result from the heterogeneity of Gompertz or Makeham distributions across sub-populations, modelled by a gamma frailty, and can be represented by logistic models, see [High Age Mortality Working Party \(2015\)](#) and [Feehan \(2018\)](#) for a comprehensive discussion. The impact of a Gompertz mortality shape in comparison with S2PML tables which assume mortality deceleration at high ages is  $-2.0\%$  to  $-0.4\%$  of annuity value at age 90 male, and  $-0.2\%$  to  $-0.0\%$  at age 65 ([High Age Mortality Working Party, 2015](#)).

### 2.2. Landmark analysis

In the landmarking approach, dynamic predictions for the conditional survival after  $t = t_{LM}$  are based on current information for all patients still alive just prior to  $t_{LM}$  ([van Houwelingen and Putter, 2011](#)). The sliding landmark model for a prediction window  $w$ , at each landmark point  $t_{LM}$ , is the simple Cox model

$$\mu(t|Y, t_{LM}, w) = \mu_0(t|t_{LM}, w) \exp(Y^T \beta_{LM}), \quad s \leq t \leq s + w,$$

for the data set obtained by truncation at  $s = t_{LM}$  and administrative censoring at  $t_{LM} + w$ . Here  $\mu_0(t|t_{LM}, w)$  is the baseline hazard or force of mortality within the window  $w$ . This is a convenient way to obtain a dynamic prediction without fitting a complicated model with time-varying effects.

Such a prediction data set, called “strata” is created for each of a set of prediction time points  $\{s_1, \dots, s_L\}$ ,  $20 \leq L \leq 100$ , keeping the window width  $w$  fixed. We address the choice of window width in Section 4. All the strata are stacked into the so-called super-prediction data set comprising the full data for landmark analysis. Passing from stratum  $s$  to  $s + 1$  corresponds to sliding the window over the time range. A set of Cox models fitted at each prediction data set is the so-called “crude model” of landmark analysis.

More sophisticated modelling is required to analyse the full super-prediction data set. In the pseudo-partial log-likelihood landmark model  $ipl^*$ , the regression coefficients  $\beta_{LM}(s)$  are assumed to be the polynomials of time  $s$ , and the baseline hazard is modelled as  $\mu_0(t|t_{LM}, w) = \mu_0(t) \exp(\Theta(s))$ , resulting in a smooth time-varying hazard

$$\mu(t|Y, t_{LM} = s, w) = \mu_0(t) \exp(\Theta(s) + Y^T \beta_{LM}(s)), \quad s \leq t \leq s + w, \tag{1}$$

where  $\beta_{LM}(s)$  and  $\Theta(s)$  are the  $k$ th degree polynomials in  $s$ .

The risk set  $R(t_i)$  for an event time  $t_i$  is present in all strata with  $s \leq t_i \leq s + w$ . An individual at risk at  $t_i \geq s$  has  $n_{is} = \#\{s \leq t_i\}$  copies in the risk set  $R(t_i)$ . For an individual with an event at  $t_i$  in the original data set, there are  $n_{is}$  tied events in the stacked data. The parameters of the  $ipl^*$  model are estimated by maximising the integrated (over  $s$ ) pseudo-partial log-likelihood introduced by [van Houwelingen \(2007\)](#).

It is worth noting, that in the  $ipl^*$  model only the baseline hazard  $\mu_0(t)$  changes within a window, but the intercept values  $\Theta(s)$  and the covariate effects  $\beta_{LM}(s)$  are fixed at their starting values at  $x_{LM} = s$ . Therefore, predictions for all  $s \in \{s_1, \dots, s_L\}$  are obtained from estimated cumulative hazards

$$M(s+w|Y, x_{LM} = s) = \exp(\Theta(s) + Y^T \beta_{LM}(s))(M_0(s+w) - M_0(s-)),$$

where  $M_0(s) = \int_0^s \mu_0(t)$  are the baseline cumulative hazards at  $s$ .

### 3. Linking landmark analysis results to a life table

In this Section, we develop a method to translate the results from a landmark analysis to prediction of survival and life expectancy for sub-populations with varying risk profiles, i.e. with

particular combinations of health, demographic and lifestyle factors. Our approach exploits the proportional hazards assumption within a window  $[s, s + w]$  to combine a population-based baseline hazard with a covariates-specific hazard ratio terms estimated by landmark analysis on a different population or a subpopulation of a relevant population.

For a population, the period life table supposes 100,000 live births and shows the number  $l_x$  surviving to the exact age of  $x$ . The survival function at age  $x$  is  $S(x) = l_x/100000$ . There are typically separate life tables for males and females but otherwise this survival function is a weighted average of the survival functions of the people with different risk profiles within the population. Suppose that the population at age  $x_0$  consists of multiple risk groups  $j = 1, \dots, J$ , each corresponding to a particular set of covariates  $Y_j(x_0)$ . Let  $f_j(x_0) = P(Y_j(x_0))$  be the prevalence of risk group  $j$  at age  $x_0$  ( $\sum_j f_j(x_0) = 1$ ).

Then the value of the overall population survival function  $S(x) = P(T \geq x)$  for the random lifetime  $T$ , at age  $x \geq x_0$  is the weighted mean of the survival functions  $S_j(x) = P(t \geq x|Y_j(x_0))$  in the individual risk groups:

$$S(x) = \frac{\sum_j f_j(x_0) S_j(x)}{\sum_j f_j(x_0)}. \tag{2}$$

The sum of weights in the above equation is 1, but we kept the denominator as, in real data, the estimated prevalences  $\hat{f}_j$  are subject to rounding and perhaps other errors.

We use methodology developed in [Kulinskaya et al. \(2020\)](#) to find survival functions  $S_j(x)$  for each risk group  $j$  at age  $x$ .

#### 3.1. Survival functions under Gompertz–Landmark model

Given a window  $[s, s + w]$ , the hazards in the  $ipl^*$  model of landmark analysis (1) are very similar to those in the Cox model. Landmark analysis also allows estimating the cumulative baseline hazards  $M_0(x)$ . The goodness-of-fit of a particular parametric survival distribution to the cumulative hazards estimated by the  $ipl^*$  model needs to be evaluated. In this Section we assume that these hazards are well described by a Gompertz distribution  $G(a, b)$ . Substituting the Gompertz baseline hazard  $\mu_0(x) = \exp(a + bx)$  into (1), the hazards are

$$\mu(x|Y, x_{LM} = s, w) = \exp(a(s) + bx + \Theta(s) + Y^T \beta_{LM}(s)), \quad s \leq x \leq s + w, \tag{3}$$

where  $a(s)$  is the baseline value which may depend on the landmark  $s = x_{LM}$ , and the effects  $\beta_{LM}$  are centred at the baseline, so that  $\beta_{LM} = 0$  provides the baseline risk. The log-hazards in various risk subgroups  $Y = Y_j$  differ by intercept but have the same slope  $b$ .

The cumulative hazards, and the survival functions are obtained by integrating the baseline hazards within the window. The survival function can be written as

$$S(x|Y, x_{LM} = s) = \exp(-e^{a(s) + \Theta(s) + Y^T \beta_{LM}} b^{-1} (e^{bx} - 1)), \quad s \leq x \leq s + w. \tag{4}$$

#### 3.2. Calibration of a predictive survival model

The hazard ratio term’s contribution to the survival model needs to be calibrated to provide correct population hazard or survival function at a specific age. One possible approach is to centre the hazard (1) to ensure that the baseline hazards function represents, in some sense, an average risk. [Royston \(2012\)](#) suggests that this may be achieved by centring the linear predictor  $Y^T \beta$  to have zero mean over the individuals in the dataset, i.e. adding up all these values over the dataset should be zero. Instead, similar to [Royston and Altman \(2013\)](#) we average survival

functions across all risk subgroups and ensure that this equals the population survival function.

Following Kulinskaya et al. (2020), for a set of the risk groups  $j = 1, \dots, J$ , substitute the survival functions (4) at age  $x \geq x_{LM}$  into Eq. (2) to obtain the overall survival function as

$$S(x|x_{LM}) = \sum_j f_j(x_{LM}) \exp(-e^{a(x_{LM}) + \theta(x_{LM}) + Y_j^T \beta_{LM}} b^{-1}(e^{bx} - 1)). \quad (5)$$

At  $x = x_{LM}$ , this is a non-linear equation with one unknown,  $a(x_{LM})$ . The left-hand side is given by the period life-table and the slope  $b$  should be determined for a particular population of interest. As  $S(x)$  is a decreasing function of  $a$ , Eq. (5) has a unique solution.

Substituting a set of estimated prevalences  $\{\hat{f}_j(x_{LM})\}$ , estimated landmark parameters  $\hat{\theta}$  and  $\hat{\beta}_{LM}$ , and the estimated Gompertz slope  $\hat{b}$ , and solving Eq. (5) for  $a_0(x_{LM})$ , the component estimated survival functions  $S_j(x|x_{LM} = s) = \hat{S}(x|Y_j, x_{LM} = s)$  are found from (4).

### 3.3. Life expectancy at a landmark age

For a particular population, the parameters  $a = a(x_{LM})$  and  $b$  of the underlying Gompertz distribution  $G(a, b)$  can be estimated from a period life table. Next, the component survival functions  $S_j(x|x_{LM} = s)$  are estimated from (5). Note, that these survival functions estimate the true survival for the risk group  $j$  only within a window  $[s, s + w]$ . However, this is the best estimate of the survival function available at age  $x_{LM}$  for the people with  $Y = Y_j(x_{LM})$  even beyond the window, without knowledge of their future health trajectories.

By definition, the remaining life expectancy at age  $z$ ,  $e(z) = \int_z^\infty S(t)dt/S(z)$ . Missov and Lenart (2013) state that the remaining LE at age  $z$  for a Gompertz distribution  $G(a, b)$  can be written as  $\int_z^\infty G_{(a,b)}(x)dx = b^{-1} \exp(b^{-1}e^a)E_1(b^{-1}e^{a+bx})$ , where  $E_1(z) = \int_z^\infty t^{-1} \exp(-t)dt$  denotes the exponential integral. However, this expression should be divided by the survival  $S(z)$ , to provide a proper remaining LE at  $z$ . Thus, the remaining life expectancy at age  $z$  for a Gompertz distribution is obtained as

$$e_{G(a,b)}(z) = \frac{b^{-1} \exp(b^{-1}e^a)E_1(b^{-1}e^{a+bz})}{\exp(-e^a b^{-1}(e^{bz} - 1))}. \quad (6)$$

Similar to Kulinskaya et al. (2020), the component remaining life expectancies  $e_j(z)$  for each risk group  $j$  at age  $z \geq x_{LM}$  are obtained from (4) substituting a component distribution  $G(a_j, b)$ , with  $a_j = a(x_{LM}) + \theta(x_{LM}) + Y_j^T \beta_{LM}$ .

To calculate the population remaining life expectancy, consider the survival function of the overall population at age  $x \geq x_{LM}$ , which is a finite mixture of subpopulations,  $S(x) = \sum_j f_j(x_{LM})S_j(x)$ ,  $\sum_j f_j(x_{LM}) = 1$ . Then the remaining life expectancy at age  $z \geq x_{LM}$  is

$$e(z) = \frac{\int_z^\infty S(x)dx}{S(z)} = \frac{\sum_j f_j(x_{LM})S_j(z) \int_z^\infty S_j(x)dx/S_j(z)}{\sum_j f_j(x_{LM})S_j(z)} = \frac{\sum_j f_j(x_{LM})S_j(z)e_j(z)}{\sum_j f_j(x_{LM})S_j(z)}. \quad (7)$$

Using (7), we can estimate a hypothetical impact of changing prevalences of an intervention or lifestyle at a landmark age. Consider one covariate (“intervention”), denoted by  $y_1$ , and coded as 0 or 1. By specifying  $f_j(x_{LM}) = 0$  for all risk groups with  $y_1(x_{LM}) = 1$ , we obtain a hypothetical remaining life expectancy  $e_0(z)$ ,  $z \geq x_{LM}$  if there was no intervention of interest, and, by specifying  $f_j(x_{LM}) = 0$  for all risk groups with  $y_1 = 0$ , a hypothetical remaining life expectancy  $e_1(z)$ ,  $z \geq x_{LM}$  with full uptake of the intervention.

### 3.4. Dynamic estimation of survival and life expectancy

The estimated survival function  $\hat{S}_j(x) = \hat{S}(x|Y_j, x_{LM} = x)$  and the estimated remaining life expectancy  $\hat{e}_j(x)$  for a particular risk profile  $Y_j(x)$  given by Eqs. (4) and (6) with coefficients  $\hat{\theta}(x)$  and  $\hat{\beta}_{LM}(x)$  estimated from a landmark analysis and with the Gompertz parameters  $\hat{a}(x)$  and  $\hat{b}$  estimated from a population life table, are the best estimates available at the age  $x$ . These estimates can be easily updated for the next landmark point  $x = s + 1$  given that the risk profiles are updated to  $Y_j(s + 1)$  and their prevalences  $f_j(s + 1)$  are available. As discussed in the Introduction, this dynamic recalculation of the survival and remaining life expectancy may be of benefit both to institutions (such as pension schemes or insurance companies) and to individuals planning a future course of their retirement.

It is also possible to estimate a hypothetical survival function and remaining life expectancy for a particular health/lifestyle trajectory over the life-course, such as the healthiest people with particular lifestyle choices for the rest of their lives, or people taking up a particular intervention for some time interval. To do this, we only need to choose the covariate values  $Y(x)$  for every  $x \geq x_0$ , where  $x_0$  is the landmark age of interest. The hypothetical survival function at any age  $x \geq x_0$  is then defined as  $S_H(x|Y(x))$ ,  $x \geq x_0$ , i.e. as in (4) with  $x = s$ . Note, that this survival function, conditional on the future health/lifestyle trajectory, differs from the unconditional survival function  $S(x|Y(s))$ ,  $s \leq x \leq s + w$  given in (4) that is based on the risk profile at  $s$  and is averaged over all possible risk trajectories which start from  $Y(s)$ .

To estimate the hypothetical remaining life expectancy  $e_H(z|Y(x), x \geq z)$  from a landmark analysis between ages  $s_1 = x_0$  and  $s_L = x_T$ , based on the survival function  $S_H(x|Y(x))$ , we simply use the trapezoidal rule from age  $z$  to age  $x_T$ , and continue with the Gompertz residual LE at age  $x_T$ . Assume, for simplicity, that the landmark points are  $\Delta$  apart. Then

$$e_H(z) = \Delta[S_H(z)/2 + \sum_{i=1}^{(x_T-z)/\Delta-1} S_H(z+i\Delta) + S_H(x_T)/2] + e(x_T|Y(x_T)). \quad (8)$$

## 4. Case study: survival benefits of statins

Statins are a class of lipid-lowering drugs that are prescribed to prevent cardiovascular disease (CVD). Statins became widely available from 2000 onward and first-line treatment for patients with CVD in 2007 (NICE, 2013). For primary prevention, the eligibility criterion in the UK is based on the 10-year risk of a first cardiac event calculated using QRISK2, which incorporates information on multiple demographic, medical, and lifestyle factors (<https://www.qrisk.org/>). These eligibility criteria were consistently lowered over time, from  $\geq 40\%$  initially, to  $\geq 20\%$  in 2006 (Hippisley-Cox et al., 2008); and in 2014, when the cardiac risk threshold was further lowered to a QRISK2 score  $\geq 10\%$  (NICE, 2016). In 2017, 11.8 million people in England, almost all men over 60 and all women over 75, were eligible for statin prescription (Ueda et al., 2017).

Statins are life-long prescriptions in the prevention of cardiac events, yet little is known about the overall survival benefit of long-term prescription. This is because clinical trials are expensive to carry out and therefore tend to be of relatively short duration, with statins trials having on average two to five years of treatment exposure (Armitage et al., 2019). Observational studies contribute to the statin research by assessing the effects of prolonged exposure (Collins et al., 2016). The availability of electronic health records (EHR) makes it easier to follow-up patients' health information for an extended period.

**Table 1**

Hazard ratios (HR) of smoking, diabetes type II and obesity for all-cause mortality obtained from landmark analyses with 5-, 10- and 30-years windows.

| Age (years) | Window (years) | Smoking |        |      | Diabetes |        |      | Obesity |        |      | Statin (1930–35) |        |      | Statin (1936–40) |        |      |
|-------------|----------------|---------|--------|------|----------|--------|------|---------|--------|------|------------------|--------|------|------------------|--------|------|
|             |                | HR      | 95% CI |      | HR       | 95% CI |      | HR      | 95% CI |      | HR               | 95% CI |      | HR               | 95% CI |      |
| 65          | 5              | 2.95    | 2.77   | 3.15 | 1.57     | 1.45   | 1.70 | 0.95    | 0.88   | 1.03 | 0.82             | 0.68   | 0.99 | 0.75             | 0.68   | 0.82 |
|             | 10             | 2.66    | 2.54   | 2.78 | 1.55     | 1.45   | 1.64 | 1.01    | 0.96   | 1.07 | 0.87             | 0.76   | 0.99 | 0.76             | 0.71   | 0.81 |
|             | 30             | 2.34    | 2.26   | 2.42 | 1.58     | 1.50   | 1.66 | 1.07    | 1.02   | 1.13 | 0.92             | 0.84   | 1.00 | 0.79             | 0.74   | 0.84 |
| 70          | 5              | 2.58    | 2.46   | 2.71 | 1.39     | 1.32   | 1.47 | 0.93    | 0.88   | 0.98 | 0.84             | 0.78   | 0.90 | 0.68             | 0.64   | 0.72 |
|             | 10             | 2.40    | 2.31   | 2.50 | 1.40     | 1.34   | 1.47 | 0.99    | 0.95   | 1.04 | 0.86             | 0.82   | 0.91 | 0.71             | 0.68   | 0.75 |
|             | 30             | 2.26    | 2.18   | 2.35 | 1.42     | 1.36   | 1.48 | 1.02    | 0.98   | 1.07 | 0.86             | 0.82   | 0.90 | 0.73             | 0.70   | 0.76 |
| 75          | 5              | 2.24    | 2.13   | 2.36 | 1.32     | 1.25   | 1.39 | 0.90    | 0.86   | 0.95 | 0.77             | 0.73   | 0.81 | 0.67             | 0.63   | 0.71 |
|             | 10             | 2.14    | 2.03   | 2.25 | 1.32     | 1.25   | 1.39 | 0.95    | 0.90   | 1.00 | 0.78             | 0.74   | 0.82 | 0.68             | 0.65   | 0.72 |
|             | 15             | 2.11    | 2.01   | 2.22 | 1.32     | 1.25   | 1.38 | 0.96    | 0.91   | 1.01 | 0.78             | 0.75   | 0.82 | 0.69             | 0.65   | 0.73 |
| 80          | 5              | 1.84    | 1.69   | 2.01 | 1.27     | 1.19   | 1.36 | 0.89    | 0.82   | 0.96 | 0.73             | 0.68   | 0.77 | 0.61             | 0.50   | 0.73 |
|             | 10             | 1.80    | 1.64   | 1.98 | 1.25     | 1.16   | 1.35 | 0.90    | 0.83   | 0.98 | 0.73             | 0.69   | 0.78 | 0.63             | 0.53   | 0.73 |
|             | 30             | 1.82    | 1.65   | 2.00 | 1.25     | 1.16   | 1.34 | 0.90    | 0.83   | 0.98 | 0.74             | 0.70   | 0.79 | 0.65             | 0.56   | 0.75 |

Furthermore, in clinical practice, patients are not fixed on a certain treatment regime as during a randomised control trial but instead sequential treatment decisions are made in managing their changing cardiac risk and emerging morbidities, resulting in time-varying statin use in individual patients. Therefore there is a need for dynamic survival prediction of long-term time-varying statin therapy.

In our previous work [Kulinskaya et al. \(2020\)](#), we used the hazard ratios associated with statin prescription for primary and secondary prevention of CVD obtained from Cox regressions at key retirement ages ([Gitsels et al., 2016, 2017](#)) to evaluate survival benefits of statins for an individual and their effect on the period life expectancy in England and Wales.

In this Section, we use the results of our recent analysis of the survival benefits of statins ([Gitsels et al., 2021](#)) to demonstrate the use of landmark analysis for individual and population life expectancy.

#### 4.1. Landmark analysis of the survival benefits of statins

Our retrospective cohort study ([Gitsels et al., 2021](#)) of the survival benefits of statins used primary care records of The Health Improvement Network (THIN) UK database. The cohort included 110,243 patients who turned 60 between 1990 and 2000, were neither diagnosed with cardiovascular disease nor prescribed statins and were residential in England or Wales. The cohort was followed up until January 2017 (16.6 years on average), where the medical history was updated every half a year. Landmark analyses were carried out by fitting Cox proportional hazards regressions of all-cause mortality associated with current statin prescription at each landmark from age 60 to 85 (51-time points), adjusted for medical history.

38% of patients entered the study in 1990–95 and the remaining 62% entered in 1996–2000. Most patients (98%) were at the lowest cardiac risk with a QRISK2 score of <20%. During follow-up, the cardiac risk of the study population increased, which in general is largely driven by age.

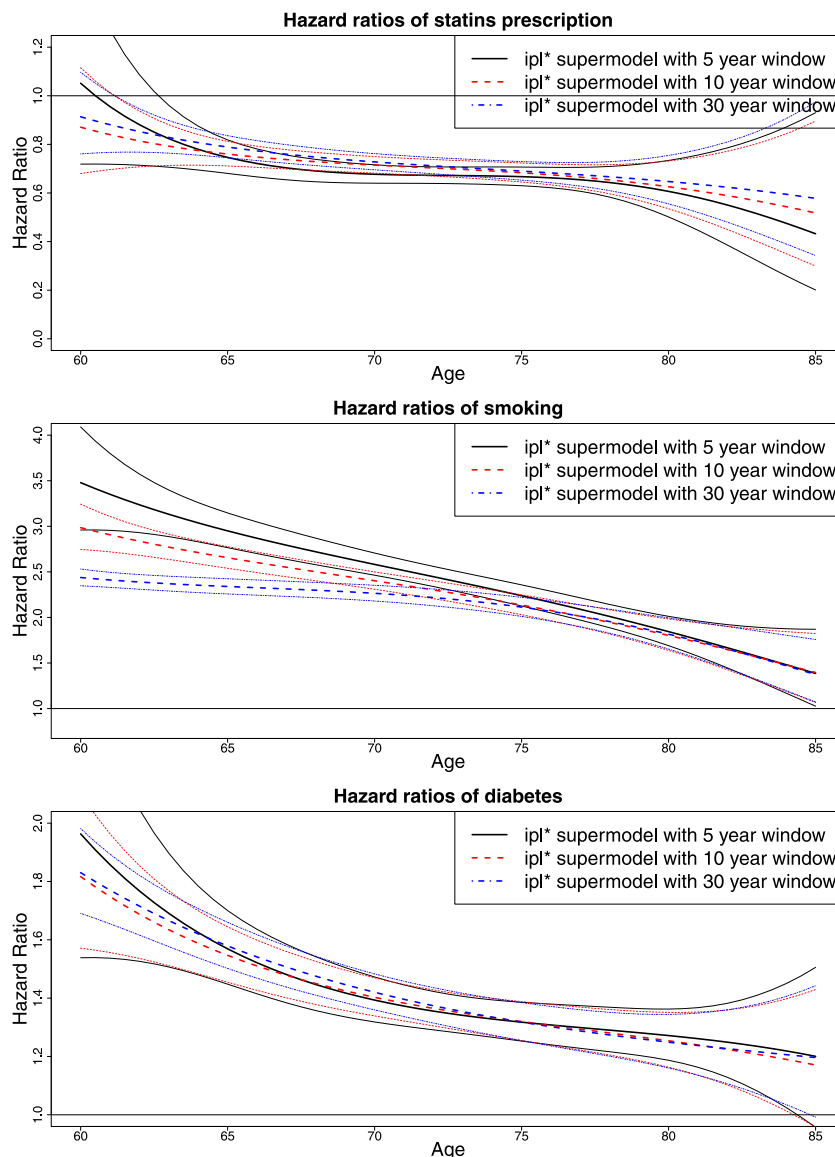
The median age at which the study population had its first statin prescription was at 70 (interquartile range 66–74) years old. The prevalence of a history of statin prescription by ages 65, 70, 75, 80 and 85, was 8%, 27%, 47%, 54% and 57%. In 2015, current statin prescription in 75-year olds at low (QRISK2 <20%, medium (20%–40%), or high ( $\geq 40\%$  or presence of CVD) cardiac risk was 10%, 35%, and 75%, respectively. Statin prescription was less common among older patients, where 80- and 85-year olds were prescribed approximately 10 and 30 percentage points lower, respectively, compared to the younger patients in the highest cardiac risk group. Adherence of statin prescription defined as a continuing prescription at least 75% of follow-up time differed by

birth cohort, from approximately 90% at any age in patients born in 1936–40, to age-dependent and somewhat lower adherence in patients born in 1930–35, increasing from 75% at age 61 to 90% at age 65, after which it slowly levelled to 80% by age 75 and dropped down after age 82.

Landmark analyses with 5-, 10- and 30-years window were carried out to dynamically predict the survival effects associated with statin therapy. Potential confounders consisted of sex, year of birth, Townsend deprivation quintile, chronic kidney disease, diabetes, treated hypertension, hypercholesterolaemia, aspirin, BMI, alcohol consumption status and smoking status. The final *ipl\** landmark model included the medical history and the significant interaction between statin prescription and year of birth. This interaction was defined as no statin prescription (reference level), statin prescription in patients born in 1930–35, and statin prescription in patients born in 1936–40. Our landmark model also produced time-varying estimated hazards for all other risk factors in the model, of which high deprivation, the QRISK2 score of  $\geq 40\%$  or CVD diagnosis, smoking and diabetes resulted in the most pronounced survival harms ([Gitsels et al., 2021](#)). The models were assessed on the proportional hazards assumption and discrimination ([Antolini et al., 2005](#)).

The adjusted hazards of all-cause mortality associated with current statin prescription, smoking and type II diabetes at each age are presented in [Fig. 1](#) and (for key ages) in [Table 1](#), which also includes obesity. Each of the plots depicts hazard ratios from landmark analyses with 5-, 10- and 30-year window. For statins and diabetes, the results of analyses from all three window widths are not significantly different. However, for smoking the hazards decrease with the increase in the window width. For ages 60 to 70, the hazards are significantly higher for a 5-year window than for a 10- or a 30-year window (at 2.95 (95% confidence interval 2.77–3.15), 2.66 (2.54–2.78) and 2.34 (2.26–2.42), respectively, at age 65), though they converge by age 75. This is because in the landmark analysis, only the risk profile at the start of a window is used regardless of possible future changes. Diabetes type II is a life-long chronic condition, and we mentioned high adherence to statins, but people may change their lifestyle choices, such as stop smoking, in the longer term. This means that the hazard of smoking in a 30-year window is averaged over persistent smokers and people who stopped smoking at different points over 30 years, and is therefore lower than the hazard for a 5-year window.

These differences in hazards at different window widths are only noticeable at ages up to age 75. The oldest patients, born in 1930, would be 86 by the end of the study, so the width of 30 years is, in fact 26 years for this cohort at age 60, and just 21 years for the oldest people in the younger birth cohort at age 60. When the oldest participants are 75, they are observed at



**Fig. 1.** Adjusted 5-, 10-, and 30-year hazards of all-cause mortality associated with current statin prescription (top), smoking status (middle) and diagnosed diabetes type II (bottom) by landmark age in the 1936–40 birth cohort. All models included cardiac risk, sex, deprivation, statin prescription, chronic kidney disease, diabetes, hypertension, hypercholesterolaemia, aspirin, body mass index, alcohol consumption and smoking.

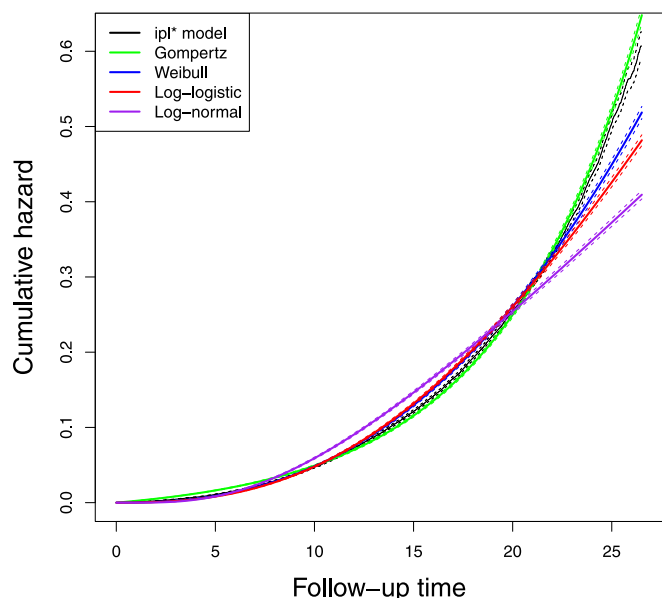
most for 11 years, therefore the results from the 30-year window and 10-year window would be very similar for this birth cohort by age 75. For the 1936–40 birth cohort, the 30-year window and the 5-year window almost coincide at age 75.

There was a clear trend of improved survival prospects associated with statins prescribed at increasing age with a significant survival benefit from approximately age 63 onward. The survival prospects associated with statin prescription differed by birth cohort, where patients born in 1936–40 had better prospects compared to patients born in 1930–35. Compared to no statin prescription, statin prescription in patients born in 1936–40 at ages 65, 70, 75 and 80 was associated with a hazard of mortality of 0.80 (95% CI 0.75–0.85), 0.73 (0.69–0.76), 0.68 (0.65–0.73), and 0.63 (0.55–0.74), respectively. Similarly, for patients born in 1930–35, statin prescription at ages 65, 70, 75, 80 and 85 was associated with a hazard of mortality of 0.92 (0.83–1.01), 0.87 (0.82–0.91), 0.79 (0.75–0.83), 0.74 (0.70–0.78) and 0.76 (0.65–0.89), respectively. No other interactions with current statin prescription were found. Furthermore, among patients with a history of statin prescription, the survival prospects at any landmark age did not differ by how long ago the first prescription was.

#### 4.2. Estimating survival and life expectancy

In our previous article [Kulinskaya et al. \(2020\)](#) we demonstrated that the Gompertz distribution provides an adequate model for all-cause mortality, for the England and Wales population. We used the period life table by Townsend score (TS) quintiles centred at 2010 provided by [Office of National Statistics \(2017\)](#) (ONS) to estimate parameters  $a(x)$  and  $b(x)$  of the Gompertz distributions for each gender by TS quintile combination. The hazards clearly differ among the TS quintiles, from the highest at TS quintile 5 (most deprived) to the lowest at TS quintile 1 (least deprived), and by gender (higher for males).

In this case study, we established that the baseline hazards from the *ipl\** landmark model were also well described by the Gompertz distribution. [Fig. 2](#), created using the R package *flexsurv*, demonstrates the fit of several popular survival distributions to the cumulative baseline hazards. The Gompertz distribution clearly provides the best fit, also confirmed by the AIC values, 221298.2, 221467.9, 221930.3, and 225109.3 for Gompertz, Weibull, log-logistic and log-normal distributions, respectively.



**Fig. 2.** Cumulative baseline hazards (solid lines) with 95% confidence intervals (dashed lines) estimated from the ipl\* model (black) with the superimposed cumulative hazards and their confidence intervals from the best-fitting Gompertz (green), Weibull (blue), log-normal (aqua) and log-logistic (red) distributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Total number of people in Townsend score (TS) quintile 1 (least deprived) and 5 (most deprived area) by sex (M/F) and cardiac risk (QRISK2) for birth cohort 1936–40 at ages 65–75, and for combined birth cohorts 1930–35 and 1936–40 at ages 75–85.

| TS quintile | Sex | QRISK2 category | Age  |      |      |      |     |
|-------------|-----|-----------------|------|------|------|------|-----|
|             |     |                 | 65   | 70   | 75   | 80   | 85  |
| 1           | M   | <20             | 5895 | 2004 | 0    | 0    | 0   |
| 1           | M   | 20–40           | 1431 | 3675 | 3829 | 1439 | 11  |
| 1           | M   | 40+             | 534  | 1095 | 1916 | 2558 | 500 |
| 1           | F   | <20             | 8209 | 5959 | 1579 | 0    | 0   |
| 1           | F   | 20–40           | 171  | 1135 | 4048 | 3197 | 34  |
| 1           | F   | 40+             | 293  | 612  | 1144 | 1800 | 688 |
| 5           | M   | <20             | 1896 | 217  | 0    | 0    | 0   |
| 5           | M   | 20–40           | 1401 | 2038 | 1237 | 417  | 1   |
| 5           | M   | 40+             | 373  | 774  | 1094 | 998  | 170 |
| 5           | F   | <20             | 3419 | 1755 | 102  | 0    | 0   |
| 5           | F   | 20–40           | 454  | 1221 | 1961 | 1055 | 56  |
| 5           | F   | 40+             | 293  | 618  | 897  | 1176 | 239 |

The next step is to use Eq. (5) to estimate the component survival functions  $S_j(x|x_{LM} = s)$  for all risk profiles  $j$  at 51 landmark ages  $s$ . To do this, we need to estimate the prevalences  $f_j(x)$ .

Due to the study recruitment period (1990–2000), the 1936–40 birth cohort participants were 76 to 80 years old by the end of the study in January 2017. Therefore, we used the observed prevalences in this cohort for ages 60–75, and the prevalences in the two combined birth cohorts for ages 76 to 85, where the number of participants considerably reduced with age due to death or attrition; see Table 2 for the total numbers for this selection.

We considered, for each sex and TS quintile, all combinations of statin use (2 levels), smoking (3), hypertension (3), diabetes (2), hypercholesterolaemia (2), BMI category (3) and cardiac risk (3), 648 combinations in total, where Table 3 provides the numbers of diabetics, smokers and obese patients at key ages for TS quintiles 1 and 5, by cardiac risk. A large number of combinations

**Table 3**

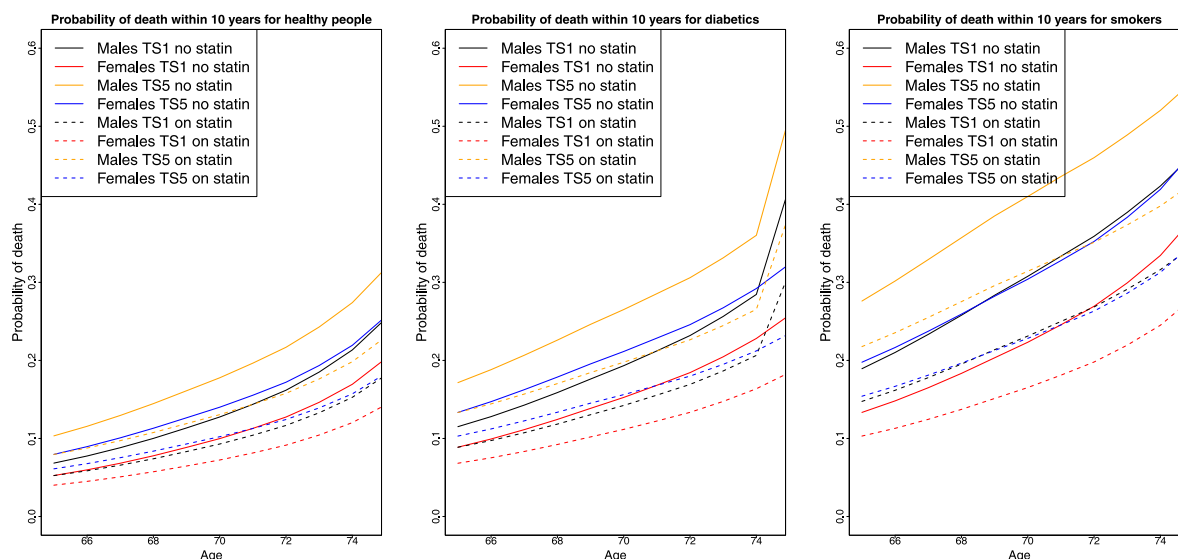
Counts of diabetics, smokers and obese patients in Townsend score (TS) quintile 1 (least deprived) and 5 (most deprived area) by sex (M/F) and cardiac risk (QRISK2) for birth cohort 1936–40 at ages 65–75 and for combined birth cohorts 1930–35 and 1936–40 at ages 75–85.

| TS quintile | Sex | QRISK2 category | Health   | Age  |      |     |     |     |
|-------------|-----|-----------------|----------|------|------|-----|-----|-----|
|             |     |                 |          | 65   | 70   | 75  | 80  | 85  |
| 1           | M   | <20             | Diabetic | 18   | 0    | 0   | 0   | 0   |
| 1           | M   | 20–40           | Diabetic | 443  | 473  | 144 | 0   | 0   |
| 1           | M   | 40+             | Diabetic | 88   | 293  | 726 | 656 | 85  |
| 1           | F   | <20             | Diabetic | 177  | 7    | 0   | 0   | 0   |
| 1           | F   | 20–40           | Diabetic | 135  | 447  | 338 | 10  | 0   |
| 1           | F   | 40+             | Diabetic | 32   | 100  | 358 | 624 | 91  |
| 5           | M   | <20             | Diabetic | 0    | 0    | 0   | 0   | 0   |
| 5           | M   | 20–40           | Diabetic | 248  | 167  | 35  | 0   | 0   |
| 5           | M   | 40+             | Diabetic | 96   | 310  | 466 | 335 | 42  |
| 5           | F   | <20             | Diabetic | 53   | 0    | 0   | 0   | 0   |
| 5           | F   | 20–40           | Diabetic | 217  | 293  | 172 | 6   | 0   |
| 5           | F   | 40+             | Diabetic | 64   | 209  | 388 | 462 | 54  |
| 1           | M   | <20             | Smoker   | 242  | 0    | 0   | 0   | 0   |
| 1           | M   | 20–40           | Smoker   | 829  | 521  | 166 | 0   | 0   |
| 1           | M   | 40+             | Smoker   | 57   | 115  | 223 | 204 | 22  |
| 1           | F   | <20             | Smoker   | 758  | 68   | 0   | 0   | 0   |
| 1           | F   | 20–40           | Smoker   | 51   | 440  | 276 | 5   | 0   |
| 1           | F   | 40+             | Smoker   | 39   | 58   | 100 | 224 | 25  |
| 5           | M   | <20             | Smoker   | 5    | 0    | 0   | 0   | 0   |
| 5           | M   | 20–40           | Smoker   | 968  | 552  | 157 | 0   | 0   |
| 5           | M   | 40+             | Smoker   | 139  | 249  | 348 | 225 | 12  |
| 5           | F   | <20             | Smoker   | 726  | 3    | 0   | 0   | 0   |
| 5           | F   | 20–40           | Smoker   | 284  | 623  | 305 | 6   | 0   |
| 5           | F   | 40+             | Smoker   | 95   | 187  | 233 | 292 | 23  |
| 1           | M   | <20             | Obese    | 938  | 226  | 0   | 0   | 0   |
| 1           | M   | 20–40           | Obese    | 320  | 818  | 652 | 127 | 0   |
| 1           | M   | 40+             | Obese    | 112  | 259  | 510 | 517 | 63  |
| 1           | F   | <20             | Obese    | 1493 | 1130 | 228 | 0   | 0   |
| 1           | F   | 20–40           | Obese    | 79   | 344  | 983 | 568 | 1   |
| 1           | F   | 40+             | Obese    | 62   | 151  | 348 | 454 | 126 |
| 5           | M   | <20             | Obese    | 426  | 20   | 0   | 0   | 0   |
| 5           | M   | 20–40           | Obese    | 311  | 502  | 281 | 64  | 0   |
| 5           | M   | 40+             | Obese    | 93   | 243  | 367 | 274 | 32  |
| 5           | F   | <20             | Obese    | 919  | 481  | 14  | 0   | 0   |
| 5           | F   | 20–40           | Obese    | 174  | 386  | 605 | 282 | 10  |
| 5           | F   | 40+             | Obese    | 112  | 239  | 366 | 420 | 65  |

were absent or very scarce in the data, for example, cardiac risk increases with age and the presence of morbidities and is strongly associated with lifestyle factors. Even though there were approximately 2000 male patients with QRISK2 < 20% at age 70 in TS quintile 1 (Table 2), they had no diabetes and did not smoke. There were also no diabetics or smokers among men at medium cardiac risk (QRISK2 20%–40%) from age 80 onward (Table 3). Females are somewhat better represented in these categories at 70, but not at older ages.

Typically, at all ages, a set of 15–30 risk profiles with more than 50 participants includes from 50% (at age 80) to 89% (at age 60) of all people. And there are only 5 to 15 risk profiles with more than 100 participants, which include from 25% (at age 80) to 82.5% (at age 60) of all people. As an example, for males in TS quintile 1, these typical profiles include only non- and ex-smokers from age 75, and people without untreated hypertension from age 70, whereas the QRISK2 increases with age. For the numerous low-count risk profiles, the estimated prevalences  $\hat{f}_j$ , and therefore the respective survival functions and LEs may not be robust.

We estimated the component survival functions (for all window widths) and the life expectancies (based on the 30-year window) for non-zero count risk profiles at each landmark age, using Eqs. (5) and (6). Obtained LEs at age 65 by sex, QRISK2



**Fig. 3.** Probabilities of all-cause death within 10 years for the three health trajectories: healthy people, diabetics and smokers by statin prescription, sex and deprivation (TS1 least deprived and TS5 most deprived quintile), obtained from landmark analysis with window width of 10 years.

category, TS quintile and statin use for healthy people, diabetics and smokers without further morbidities are given in Table 4. As an application of our methodology, we developed a life expectancy calculator that is available on <https://mylongevity.org>. This LE calculator uses LEs based on the 30-year window.

As an example, for a healthy female aged 65-years old and resident in the least deprived postcode, Table 4 (and our calculator) provides a LE of 87.8 vs 89.5 years with/without statins, and this is 86.4 vs 88.1 years for the male equivalent. Diabetes decreases LE by more than 3 years; among diabetics, statins increase the LE from 84.6 to 86.2 years for females and from 83.1 to 84.8 years for males. Smoking decreases LE by approximately 6 years; among smokers, statins increase the LE from 81.9 to 83.5 years for females and from 80.4 to 82.0 years for males.

We also considered in more detail three hypothetical risk trajectories over the life-course: the “healthy” people with a healthy weight, no morbidities, and non-smokers; and the groups of “healthy diabetics” and “healthy smokers” who differ only by a current diagnosis of diabetes type II and keep this diagnosis, or are current smokers and do not quit. We assume no further morbidities over the life-course. We also assume the best possible cardiac health for all participants, i.e. the lowest realistically QRISK2 category for this group at age 65, and a switch to a higher cardiac risk category only when the majority in this cohort moved to the higher QRISK2 score. This happens at different ages for different risk cohorts. See the Appendix for more details.

Probabilities of death within 10 years for these three trajectories, by statin prescription, sex and TS quintiles 1 and 5 are illustrated in Fig. 3. These probability curves are not smooth because the cardiac risk is changed step-wise once or twice at particular ages over the life-course for each risk trajectory. In these plots, the absolute differences in survival probabilities with/without statins are visibly higher in diabetics and smokers than in healthy people.

We also calculated hypothetical LEs (8) for our three risk trajectories at the three window widths, which are given in Table 5. Comparing these LEs to the LEs from Table 4, the hypothetical LEs for all three cohorts are considerably higher than the static LEs obtained from the risk profile at age 65. Similarly, the LEs calculated at the window width of 5 years are noticeably higher than the LE at larger window widths. This is because we considered the healthiest possible life-course, and the static LEs

**Table 4**

Life expectancy at age 65 for healthy people, diabetics and smokers by statin prescription, sex and deprivation (TS1 least and TS5 most deprived area) and sex calculated at window width of 30 years.

| TS | Sex | Healthy   |        | Diabetes  |        | Smokers   |        |
|----|-----|-----------|--------|-----------|--------|-----------|--------|
|    |     | No statin | Statin | No statin | Statin | No statin | Statin |
| 1  | F   | 87.79     | 89.52  | 84.56     | 86.21  | 81.90     | 83.49  |
| 1  | M   | 86.35     | 88.11  | 83.09     | 84.76  | 80.45     | 82.02  |
| 2  | F   | 87.75     | 89.55  | 84.42     | 86.13  | 81.69     | 83.32  |
| 2  | M   | 86.11     | 87.92  | 82.77     | 84.48  | 80.08     | 81.68  |
| 3  | F   | 87.50     | 89.35  | 84.06     | 85.82  | 81.28     | 82.94  |
| 3  | M   | 85.67     | 87.55  | 82.23     | 83.99  | 79.49     | 81.12  |
| 4  | F   | 87.22     | 89.18  | 83.62     | 85.46  | 80.73     | 82.45  |
| 4  | M   | 85.40     | 87.38  | 81.80     | 83.63  | 78.95     | 80.64  |
| 5  | F   | 87.07     | 89.11  | 83.34     | 85.24  | 80.37     | 82.13  |
| 5  | M   | 84.78     | 86.87  | 81.01     | 82.92  | 78.09     | 79.81  |

are averaged across all possible health trajectories within each window, including those in ill health. The 5-year window is more relevant for this calculation of hypothetical LE.

### 5. Discussion

Kulinskaya et al. (2020) proposed a methodology to evaluate the potential impact of recent medical advances and/or public health decisions on issues of actuarial interest. This methodology incorporated hazard ratios, obtained in medical studies by the use of the Cox proportional hazards model, into underwriting individual lives, into pricing and reserving. This was achieved by the use of the Gompertz–Cox model for modelling potential changes in human life expectancy.

In this study, we extended this methodology to incorporate the results of the landmark analysis, a statistical method which allows dynamic modelling of the force of mortality. Combining age-dependent health and lifestyle hazard ratios obtained from landmark analysis, with the baseline Gompertz hazards, we obtain dynamic predictions of the survival and life expectancy for particular risk trajectories. We illustrated our methodology on a case-study based on the landmark analysis of the use of statins, (Gitsels et al., 2021). In this analysis, we differentiated between 648 possible risk groups within each deprivation quintile and sex.



**Table 5**

Hypothetical life expectancies for the healthiest people, diabetics and smokers by statin prescription, sex and deprivation (TS1 least to TS5 most deprived area) calculated at window widths of 5, 10 and 30 years assuming the health/lifestyle stays the same over time.

| 5 year window  |   | Sex   | Healthy   |        | Diabetes  |        | Smokers   |        |
|----------------|---|-------|-----------|--------|-----------|--------|-----------|--------|
| TS             |   |       | No statin | Statin | No statin | Statin | No statin | Statin |
| 1              | F | 88.40 | 93.23     | 87.22  | 92.01     | 85.66  | 90.66     |        |
| 1              | M | 86.37 | 91.18     | 85.14  | 89.92     | 83.37  | 88.39     |        |
| 2              | F | 88.17 | 93.16     | 86.95  | 91.90     | 85.33  | 90.48     |        |
| 2              | M | 86.03 | 90.98     | 84.77  | 89.68     | 82.92  | 88.08     |        |
| 3              | F | 87.63 | 92.76     | 86.38  | 91.46     | 84.67  | 89.97     |        |
| 3              | M | 85.26 | 90.33     | 83.97  | 88.99     | 82.04  | 87.31     |        |
| 4              | F | 86.94 | 92.28     | 85.64  | 90.92     | 83.83  | 89.35     |        |
| 4              | M | 84.48 | 89.79     | 83.13  | 88.37     | 81.05  | 86.55     |        |
| 5              | F | 86.54 | 92.10     | 85.21  | 90.69     | 83.31  | 89.04     |        |
| 5              | M | 83.59 | 89.08     | 82.26  | 87.65     | 80.14  | 85.78     |        |
| 10 year window |   | Sex   | Healthy   |        | Diabetes  |        | Smokers   |        |
| TS             |   |       | No statin | Statin | No statin | Statin | No statin | Statin |
| 1              | F | 88.43 | 92.23     | 87.38  | 91.19     | 85.75  | 89.71     |        |
| 1              | M | 86.50 | 90.33     | 85.40  | 89.25     | 83.60  | 87.62     |        |
| 2              | F | 88.19 | 92.11     | 87.11  | 91.04     | 85.41  | 89.50     |        |
| 2              | M | 86.16 | 90.11     | 85.03  | 88.99     | 83.16  | 87.30     |        |
| 3              | F | 87.72 | 91.78     | 86.61  | 90.66     | 84.83  | 89.05     |        |
| 3              | M | 85.43 | 89.48     | 84.27  | 88.33     | 82.32  | 86.56     |        |
| 4              | F | 87.07 | 91.31     | 85.91  | 90.15     | 84.03  | 88.44     |        |
| 4              | M | 84.72 | 88.99     | 83.50  | 87.77     | 81.41  | 85.85     |        |
| 5              | F | 86.63 | 91.03     | 85.43  | 89.82     | 83.47  | 88.03     |        |
| 5              | M | 83.92 | 88.36     | 82.70  | 87.13     | 80.56  | 85.14     |        |
| 30 year window |   | Sex   | Healthy   |        | Diabetes  |        | Smokers   |        |
| TS             |   |       | No statin | Statin | No statin | Statin | No statin | Statin |
| 1              | F | 88.65 | 91.90     | 87.52  | 90.78     | 85.93  | 89.40     |        |
| 1              | M | 86.79 | 90.10     | 85.63  | 88.95     | 83.83  | 87.39     |        |
| 2              | F | 88.43 | 91.80     | 87.28  | 90.65     | 85.60  | 89.19     |        |
| 2              | M | 86.46 | 89.87     | 85.27  | 88.70     | 83.39  | 87.05     |        |
| 3              | F | 88.00 | 91.48     | 86.81  | 90.30     | 85.03  | 88.75     |        |
| 3              | M | 85.76 | 89.28     | 84.55  | 88.08     | 82.55  | 86.33     |        |
| 4              | F | 87.39 | 91.05     | 86.16  | 89.83     | 84.25  | 88.15     |        |
| 4              | M | 85.09 | 88.82     | 83.84  | 87.57     | 81.67  | 85.64     |        |
| 5              | F | 86.93 | 90.73     | 85.68  | 89.47     | 83.66  | 87.70     |        |
| 5              | M | 84.35 | 88.26     | 83.11  | 87.01     | 80.81  | 84.95     |        |

The calculation of LE at such a fine scale will not only be useful for individuals for improving life expectancy by healthy lifestyle changes and/or exploring their retirement options, but also for independent financial advisors and the insurance industry for financial planning and reserving of actuarial products. To facilitate calculation of LE, we developed a life expectancy calculator that is available on <https://mylongevity.org>. Our R software which will enable bulk calculation of LE is due to be released later in the year.

There are, however, some limitations to our current implementation of this methodology. The calculation of component survival functions and LEs requires estimated prevalences  $\hat{f}_j$  for each risk group. In general, these prevalences are calendar time-dependent, for example, the prevalence of smoking decreased while that of obesity increased in the last decade. Our estimation procedure was constrained by the study design and period, and the prevalences used are up-to-date only for the oldest birth cohort. The up-to-date prevalences for all risk profiles at all ages are required to better evaluate current LEs.

Additionally, for the low-count risk profiles, the estimated prevalences  $\hat{f}_j$ , and therefore the respective survival functions and LEs might not be robust. A better option may be to model the prevalences and to reduce the number of estimated parameters. We intend to pursue this modelling in our further research.

**Acknowledgements**

This work was supported by the Economic and Social Research Council [grant number ES/L011859/1 to L.A.G and E.K.]; by the Institute and Faculty of Actuaries (IFoA) commissioned through the Actuarial Research Centre [grant to E.K., I.B and N.R.W].

**Appendix. QRISK2 changes for the three health trajectories over the life-course**

*Healthy people*

- Age 65–74, QRISK2 less than 20.
- Age 75–84, QRISK2 between 20 and 40.
- Age 85+, QRISK2 40+ or CVD.

*Diabetics*

- Males:
  - Age 65–74, QRISK2 20–40.
  - Age 75+, QRISK2 40+ or CVD
- Females:
  - Age 65–79, QRISK2 20–40.
  - Age 80+, QRISK2 40+ or CVD.

*Smokers*

- Males:
  - Age 65–79, QRISK2 20–40.
  - Age 80+, QRISK2 40+ or CVD
- Females:
  - Age 65–74, QRISK2 <20.
  - Age 75–79, QRISK2 20–40.
  - Age 80+, QRISK2 40+ or CVD

**References**

Antolini, L., Boracchi, P., Biganzoli, E., 2005. A time-dependent discrimination index for survival data. *Stat. Med.* 24 (24), 3927–3944, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2427>.

Armitage, J., Baigent, C., Barnes, E., et al., 2019. Efficacy and safety of statin therapy in older people: a meta-analysis of individual participant data from 28 randomised controlled trials. *Lancet* 293 (10170), 407–415.

Bender, R., Augustin, T., Blettner, M., 2005. Generating survival times to simulate Cox proportional hazards models. *Stat. Med.* 24 (11), 1713–1723, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2059>.

Brenner, H., Gefeller, O., Greenland, S., 1993. Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. *Epidemiology (Cambridge, Mass)* 4 (3), 229–236, URL <http://www.ncbi.nlm.nih.gov/pubmed/8512987>.

Collins, R., Reith, C., Emberson, J., et al., 2016. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* 388 (10059), 2532–2561.

Feehan, D., 2018. Separating the signal from the noise: Evidence for deceleration in old-age death rates. *Demography* 55 (6), 2025–2044, URL <https://doi.org/10.1007/s13524-018-0728-x>.

Gavrilov, L.A., Gavrilova, N.S., 2019a. Late-life mortality is underestimated because of data errors. *PLoS Biol.* 17 (2), 1–7, URL <https://doi.org/10.1371/journal.pbio.3000148>.

Gavrilov, L.A., Gavrilova, N.S., 2019b. New trend in old-age mortality: Gompertzianization of mortality trajectory. *Gerontology* 65 (5), 451–457, URL <https://doi.org/10.1159/000500141>.

Gitsels, L., Bakbergenuly, I., Steel, N., Kulinskaya, E., 2021. Do statins reduce mortality in older people? Findings from a longitudinal study using primary care records. *Family Med. Commun. to appear*.

Gitsels, L.A., Kulinskaya, E., Steel, N., 2016. Survival benefits of statins for primary prevention: A cohort study. *PLoS One* 79, 1–14, URL <https://doi.org/10.1371/journal.pone.0166847>.

Gitsels, L.A., Kulinskaya, E., Steel, N., 2017. Survival prospects after acute myocardial infarction in the UK: A matched cohort study 1987–2011. *BMJ Open* 7, 1, URL <https://doi.org/10.1136/bmjopen-2016-013570>.

- High Age Mortality Working Party, 2015. Initial report on the features of high age mortality. Continuous Mortality Investigation, Working paper 85. URL <https://www.actuaries.org.uk/system/files/field/document/cmiwp85-v2.pdf>.
- Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., Brindle, P., 2008. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ (Clin. Res. Ed.)* 336 (7659), 1475–1482, URL <https://doi.org/10.1136/bmj.39609.449676.25>.
- van Houwelingen, H.C., 2007. Dynamic prediction by landmarking in event history analysis. *Scand. J. Stat.* 34 (1), 70–85, URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2006.00529.x>.
- van Houwelingen, H., Putter, H., 2011. *Dynamic Prediction in Clinical Survival Analysis*, first ed. CRC Press, Inc., USA.
- Hulsen, T., Jamuar, S.S., Moody, A.R., Karnes, J.H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D.A., McKinney, E.F., 2019. From big data to precision medicine. *Front. Med.* 6, 34, URL <https://www.frontiersin.org/article/10.3389/fmed.2019.00034>.
- Kulinskaya, E., Gitsels, L., Bakbergenuly, I., Wright, N., 2020. Calculation of changes in individual and period life expectancy based on proportional hazards model of an intervention. *Insurance Math. Econom.* 93, 27–35, URL <https://doi.org/10.1016/j.insmatheco.2020.04.006>.
- Mihaylova, B., Emberson, J., Blackwell, L., Keech, A., Simes, E., Baigent, C., 2012. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet* 380 (9841), 581–590, URL [https://doi.org/10.1016/S0140-6736\(12\)60367-5](https://doi.org/10.1016/S0140-6736(12)60367-5).
- Missov, T.I., Lenart, A., 2013. Gompertz–Makeham life expectancies: Expressions and applications. *Theor. Popul. Biol.* 90 (C), 29–35, URL <https://ideas.repec.org/a/eee/thpobi/v90y2013icp29-35.html>.
- Newman, S., 2018. Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biol.* 16 (12), e2006776, URL <https://doi.org/10.1371/journal.pbio.2006776>.
- NICE, 2013. Myocardial infarction: cardiac rehabilitation and prevention of further MI. Clinical guideline [CG172]. URL <https://www.nice.org.uk/guidance/CG172>.
- NICE, 2016. Cardiovascular disease: risk assessment and reduction, including lipid modification. Clinical guideline [CG181]. URL <https://www.nice.org.uk/guidance/cg181>.
- Office of National Statistics, 2017. Life table by single year of age and sex, by Townsend deprivation quintiles in England and Wales, 2009 to 2011. URL <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/adhocs/007714lifetablebysingleyearofageandsexbytownsenddeprivationquintiles/inenglandandwales2009to2011>.
- Richards, S., 2008. Applying survival models to pensioner mortality data. *Br. Actuar. J.* 14, 257–303, URL <https://doi.org/10.1017/S1357321700001720>.
- Royston, P., 2012. Estimating a Smooth Baseline Hazard Function for the Cox Model. URL <https://pdfs.semanticscholar.org/2f32/9b48f674a74253eb428b71ff237365fd4051.pdf>.
- Royston, P., Altman, D., 2013. External validation of a Cox prognostic model: principles and methods. *BMC Med. Res. Methodol.* 13, 257–303, URL <https://doi.org/10.1186/1471-2288-13-33>.
- Spiegelhalter, D., 2016. How old are you, really? Communicating chronic risk through “effective age” of your body and organs. *BMC Med. Inf. Decis. Mak.* 16 (1), 104, URL <https://doi.org/10.1186/s12911-016-0342-z>.
- Ueda, P., Wai-Chun Lung, T., P., C., Danaei, G., 2017. Application of the 2014 NICE cholesterol guidelines in the English population. *Br. J. Gen. Pract.* 67 (662), e598–e608.
- Vaupel, J.W., 2010. Biodemography of human ageing. *Nature* 464 (7288), 536–542, URL <https://doi.org/10.1038/nature08984>.
- Villegas, A.M., Haberman, S., 2014. On the modeling and forecasting of socio-economic mortality differentials: An application to deprivation and mortality in England. *N. Am. Actuar. J.* 18 (1), 168–193, URL <https://doi.org/10.1080/10920277.2013.866034>.