NDRS

NATIONAL DISEASE REGISTRATION SERVICE

NHS Digital

Method for creating smoothed sub-national life tables

About the NDRS

The National Disease Registration Service (NDRS) is part of NHS Digital (NHSD). Its purpose is to collect, collate and analyse data on patients with cancer, congenital anomalies, and rare diseases. It provides robust surveillance to monitor and detect changes in health and disease in the population. NDRS is a vital resource that helps researchers, healthcare professionals and policy makers make decisions about NHS services and the treatments people receive.

The NDRS includes:

- the National Cancer Registration and Analysis Service (NCRAS) and
- the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)

Healthcare professionals, researchers and policy makers use data to better understand population health and disease. The data is provided by patients and collected by the NHS as part of their care and support. The NDRS uses the data to help:

- understand cancer, rare diseases, and congenital anomalies
- improve diagnosis
- plan NHS services
- improve treatment
- evaluate policy
- improve genetic counselling

NATIONAL DISEASE REGISTRATION SERVICE

National Disease Registration Service
NHS Digital (NHSD)
The Leeds Government Hub
7&8 Wellington Place
Leeds
LS1 4AP

For queries relating to this document, please contact:
NDRSenquiries@nhs.net

Contents

1. Rationale

The National Disease Registration Service (NDRS) produces estimates of adult cancer survival for people aged 15 to 99.

We limit the upper age to 99 to reflect the difficulty in generating reliable life table estimates for people aged 100 or over. The age range 15 to 99 matches the range used in the International Cancer Survival Standard (ICSS) which we use for age standardisation of adult cancer survival estimates. A significant number of adult patients will die from causes unrelated to their cancer diagnosis. To show only the effect of cancer deaths on patient survival, we produce net survival estimates.

We calculate net survival estimates by comparing the survival of cancer patients with that expected based on the general population. This comparison matches the same profile of age, gender, socio-economic status and region. Population life tables provides the mortality experience for the general population.

Diagnoses in those aged 0 to 14 years are considered in the childhood cancer survival estimates. The clinical consensus is that children are unlikely to die from an unrelated cause within ten years of their cancer diagnosis. Because of this view, life tables are not used for childhood cancer estimates.

It is important to use life tables which use the most up to date mortality data available, to reflect changes in mortality rates. Office for National Statistics (ONS) currently produces life tables by single year of age at a national level and life tables by age group at a sub-national level. The ONS summarises the range of life tables they publish in this explainer article.

The life tables created for cancer survival will look at single year of age at a sub-national level. We smooth the data to overcome issues around high volatility and risk of disclosure.

This life table method was developed with advice from the Biostatistics Research Group, University of Leicester. The Cancer Survival Group, London School of Hygiene and Tropical Medicine created life tables with their method using the same data. Comparisons between the two sets of life tables and the survival estimates were consistent.

2.Smoothed life tables

This paper includes an explanation on the method used to create a different type of life table – smoothed life tables. These smoothed life tables provide estimated (modelled) life expectancy by single year of age and gender over a 3-year period for each Index of Multiple Deprivation (IMD) quintile within each region of England – a total of 45 area estimates per year.

A three-year period is used to smooth fluctuations due to exceptional events, for example, a “flu” epidemic. Smoothed life tables contain estimated mortality rates based on a statistical model, rather than rates calculated directly from the deaths registered and population estimates.

Using a smoothed life table approach allows estimated complete life tables to be produced for areas whose population is too small for conventional complete life tables, for applications where abridged life tables are not appropriate.

If you would like to provide feedback on these new life tables, please contact ndrsenquiries@nhs.net

3.Method for creating smoothed life tables

3.1. Data specification

Table 1 shows a breakdown of all the data included in the creation of smooth life tables. Populations and deaths data were aggregated into Index of Multiple Deprivation (IMD) quintiles. The Lower Super Output Areas (LSOAs) are weighted allocations, so there may not be an equal population size per quintile.

These life tables are designed for use in producing the cancer survival estimates for England by NCRAS. These life tables may be used for other epidemiological calculations in England. Results may not be comparable across other UK countries since each country has its own IMD.

Table 1: Input data specification

Data	Notes
Populations and number of deaths by:	Mid-year population estimates and number of deaths registered
Gender	Males, Females
Age	Single year of age 0, 1, … 89, 90+
Rolling 3-year period	From “2001 to 2003” to “2017 to 2019” (as of September 2021)
Country	England
Regions	Regions of England (formally known as Government Office Regions): North East, North West, Yorkshire and The Humber, South West, South East, East, East Midlands, West Midlands, London
Deprivation quintile	Index of Multiple Deprivation: 1 (most deprived), 2, 3, 4, 5 (least deprived) IMD over time: 1999 to 2002 used IMD 2004 2003 to 2006 used IMD 2007 2007 to 2009 used IMD 2010 2010 to 2013 used IMD 2015 2014 to current used IMD 2019

3.2. How we created smoothed life tables

We smooth the mortality data by age to overcome three issues:

- the high volatility of mortality at the youngest ages

- the high risk of disclosure for mortality at the youngest ages
- the need to extrapolate mortality rates for elderly.

Smoothing mortality rates allows us to present life tables in more detail. This means that we can produce life tables by a region and IMD quintile level, which are needed for cancer survival estimates.

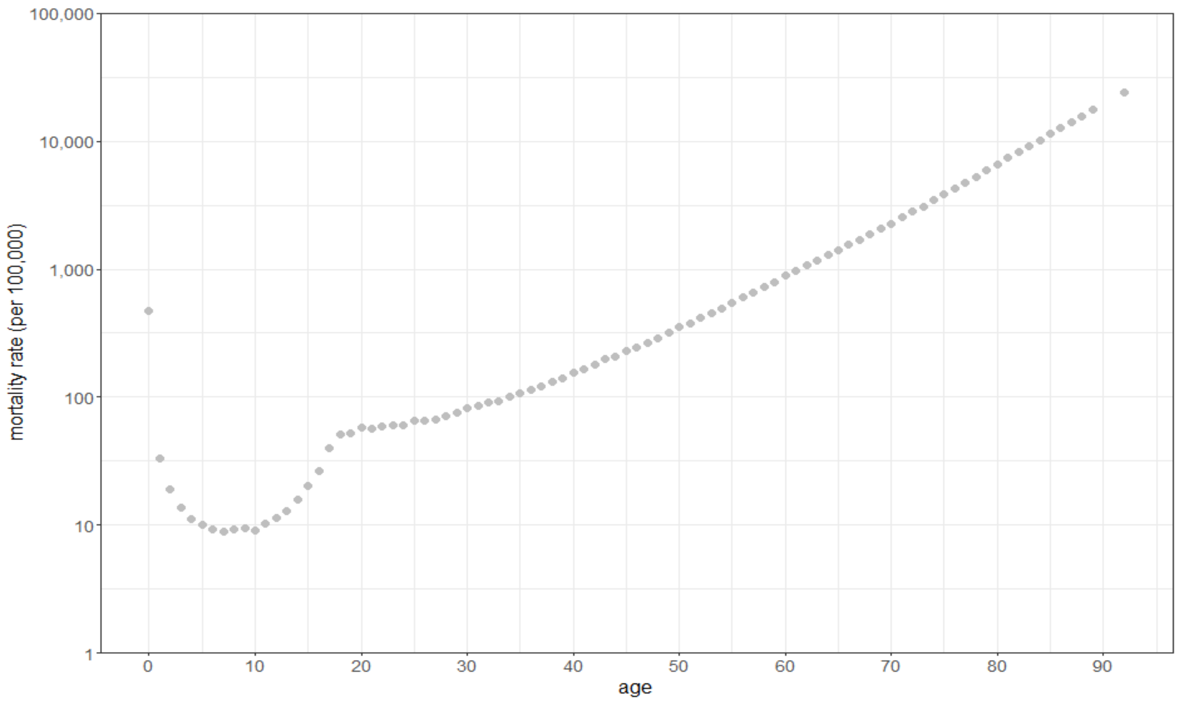
The method for creating smoothed life tables uses a flexible Poisson model with restricted cubic splines. This is the recommended method of the Cancer Survival Group, London School of Hygiene and Tropical Medicine (LSHTM). This recommendation comes the paper Multivariable flexible modelling for estimating complete, smoothed life tables for sub-national populations.

Restricted cubic splines are used when the data has a non-linear trend that does not fit standard distributions such as quadratic, cubic, exponential or logarithmic distributions. Instead of trying to fit one line to all the data, the data is broken into sections at break points (also known as knots). A trend line (spline function) is then fit to each of these sections within the series.

Cubic splines use cubic polynomial functions that have smooth joins at the knot placements. Restricted cubic splines force the cubic splines to be linear beyond the first and last knots (boundary knots). The position of the knots is used to partition the mortality into different segments. This allows the mortality to fluctuate between age groups, without too many restrictions. Knots do not need to be spaced evenly. For example, if the variation in mortality is less volatile between the ages of 15 to 45, further knots between these ages are not needed.

Figures 1 to 3 show how the mortality data can be fitted using restricted cubic splines. Figure 1 shows an average of the mortality rate by age in England between 2002 and 2020.

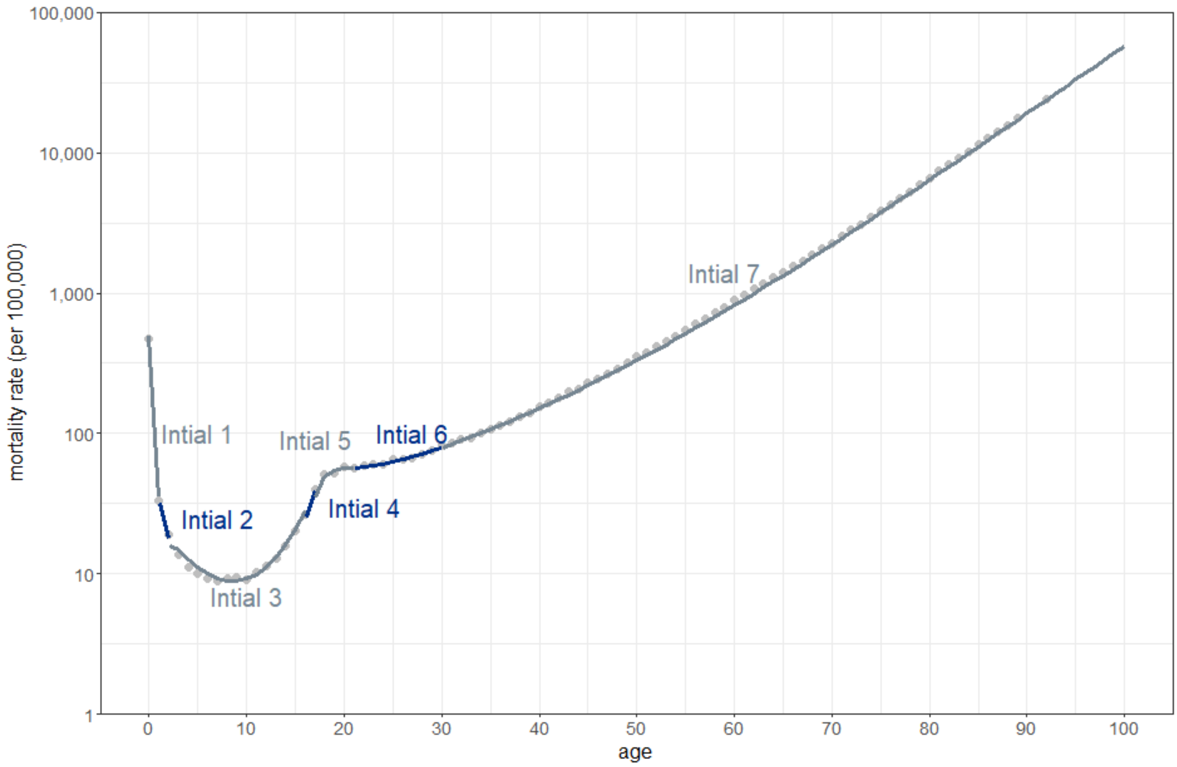
Figure 1: Average mortality rates (per 100,000 people) on the logarithmic scale



Source: National Cancer Registration and Analysis Service, NHS Digital

Figure 2 shows how restricted cubic splines first fit individual cubic polynomials between each pair of knots. Figure 2 demonstrates what the initial cubic polynomials would look like for these mortality rates. Between each pair of knots initial cubic polynomials 1 to 7 have been fitted to the data.

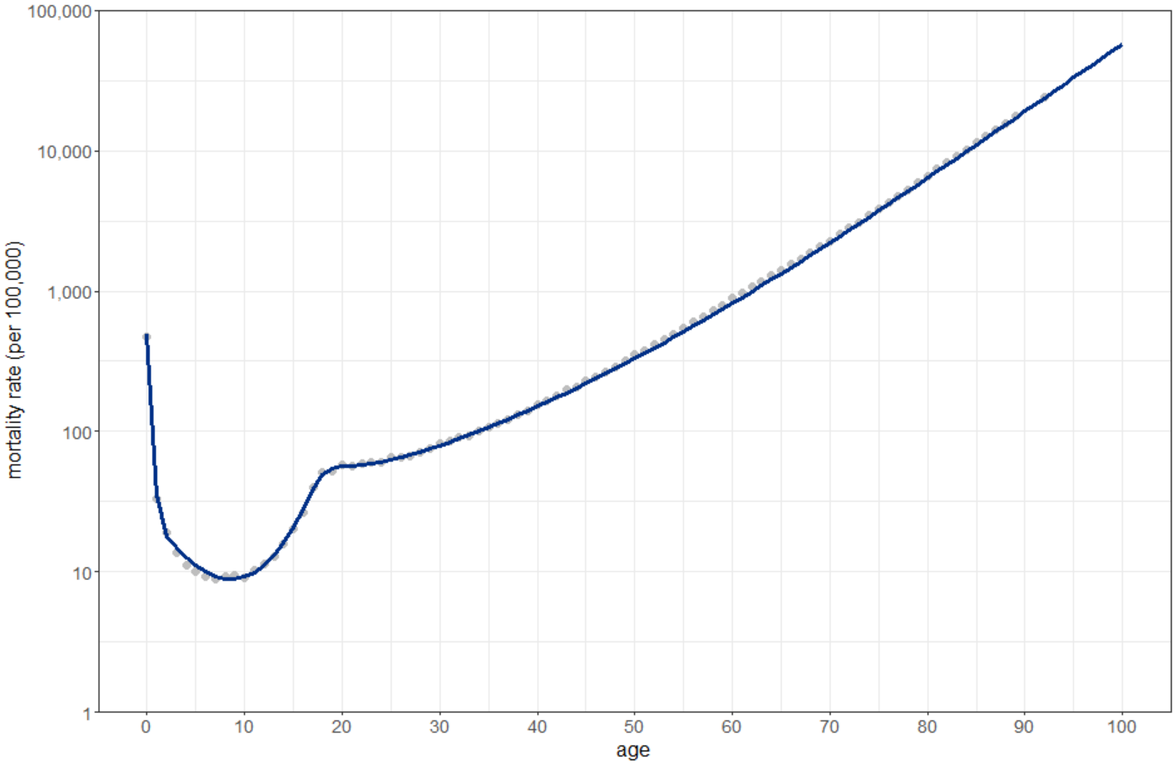
Figure 2: Average mortality rates (per 100,000 people) and initial cubic polynomials on the logarithmic scale



Source: National Cancer Registration and Analysis Service, NHS Digital

The piecewise polynomials are restricted to force each pair of adjoining cubic splines to meet at the knots. The final restricted cubic spline model is in Figure 3. All 7 cubic polynomials have now been joined to form one smooth function that is linear beyond age 92 years, the boundary knot.

Figure 3: Average mortality rates (per 100,000 people) and final cubic spline on the logarithmic scale



Source: National Cancer Registration and Analysis Service, NHS Digital

3.3. Knot selection

To create the smoothed life tables, we use the MVRS function (multivariable regression spline models) in Stata to fit flexible Poisson models with restricted cubic splines to mortality data. The models are stratified by each calendar period, gender, and region of England. Age and an age/deprivation interaction are included as flexible cubic splines in the models.

The number and location of knots is important to create a model that is the best possible fit to the data without over fitting. The process we use to choose the knot locations of these splines is described below.

Knots for age are selected based on prior knowledge and an exhaustive search for “best” fit. Prior knowledge means knots are placed at positions where we know that there is going to be significant change in the mortality rates. That is, if the curve continually decreases at age 1 year then decreases at a slower rate for ages 2 years, then we would select age 1 year as a position for a knot. We use different fixed knots for males and females due to the difference in mortality rates by age and gender.

Like LSHTM, we fixed three knots in advance at ages 0,1 and 2, due to age-specific mortality changing rapidly in the first few years of life. Due to inaccuracies in population counts above age 90 years, we calculated the cumulative number of deaths and population for both males and females aged 90 years and over and selected the median age at death in this group to be the upper boundary knot. Further knots, between the ages of 2 and 50 were required to capture changes in mortality rates.

We used the Stata command UVRS (univariate regression spline models) to fit flexible Poisson models with restricted cubic splines with knots fitted at 0,1,2 and median age (90 years and over) and all possible combinations of 1 to 4 knots between ages 2 and 50. We used the Akaike information criterion (AIC) from each model and selected the best possible fitting model for males and females separately.

Using the median age at death for people aged 90 years and over can result in a mortality rate greater than 1. Therefore, to ensure this does not occur we add additional knots beyond the median age at death for persons aged 90 and over. As we do not have data points beyond the median age at death, we take the predicted mean death rate from the UVRS models for people aged greater than the median age at death (for 90 years and over) and multiply this by the population estimate to generate the predicted number of deaths. We then use this information to add an additional 3 knots at equal intervals between the median age at death and 100 years (used as a final “boundary” knot).

We then use the MVRS function to fit the flexible Poisson model with splines for age (with knot selected described above) and the interaction between age and deprivation. We use the automated procedure within the user written command MVRS to select the number and position of the knots for the age and deprivation interaction.

By not pre-specifying the number and location of knots for the interaction, this allows for differences in the distribution of deaths by age for each deprivation quintile, e.g. in one quintile the relationship between mortality and age might be linear whereas in a different quintile the relationship might be cubic.

Once the life tables have been created, to expand the years of coverage for years where data does not yet exist, the final year of data is rolled forward as an approximation of the current experience of the general population.

If the mortality of the general population were to significantly change before the data is available to use in life tables, the impact in estimates of net survival are (broadly) as follows:

- If the gap between the mortality experienced by cancer patients and that experienced by the general population grows, the net survival estimates will be lower than if the current mortality data were available for creating life tables
- If the gap between the mortality experienced by cancer patients and that experienced by the general population shrinks, the net survival estimates will be higher than if the current mortality data were available for creating life tables

If a population experiences a sudden increase in mortality because of a pandemic, people with significant pre-existing health conditions like cancer have been found to have an elevated risk of mortality from the pandemic. If we use the actual mortality experience in cancer patients during the pandemic, but not the actual mortality experience of the general population, the gap between mortality of cancer patients and the rest of the population may increase. This may lead to lower net survival estimates being produced when using the final year of data rolled forward as an approximation of the current experience of the general population than if the actual current mortality experience were used for creating life tables.

This means net survival estimates for cancer survival will be conservative estimates for data covering the first year of the pandemic.

Further assessments of the impact of the pandemic on net survival estimates for cancer patients will be made for each year that includes the pandemic.

3.4. How we calculate life expectancy

Life expectancy calculations are the same as the National life tables, with the following exception. The mortality rates are based solely on the modelled mortality rates described above and, unlike the national life tables, make no adjustment for changes in infant mortality within the first year of life or use of populations estimates for people aged 90 to 110. These life tables use period life expectancy, which do not apply any potential future changes to mortality rates.

4. Strengths and weaknesses of smoothed life tables

4.1. Strengths:

- We can publish complete life tables at a sub-national level without encountering errors in small numbers
- By using a robust smoothing technique, we can produce reliable estimated mortality rates for all ages at a sub-national level.
- Smooth life tables are based on data for a period of three consecutive years to smooth fluctuations from exceptional events
- These are timely life tables designed for use in estimating net cancer survival
- These life tables can be used for estimating net survival for other diseases where there is a defined diagnosis date and the outcomes of a cohort can be traced

4.2. Weaknesses:

- Modelling the mortality rates reduces the amount that the rate representing the actual changes and variation in mortality experienced by the population. These smoothed rates are useful for input into specific calculations such as cancer survival. They may not be appropriate for the description of the individual geographical areas covered

- This method differs from those used to produce the unsmoothed life table products by the Office of National Statistics. The life expectancy figures from this method are not directly comparable with these other outputs because of the assumptions described above