Estimating life expectancy and years of life lost in epidemiological studies: a review of methods using an example from multimorbidity

Authors

Yogini V Chudasama, PhD ¹

Kamlesh Khunti, FMedSci ¹

Clare L Gillies, PhD ¹

Nafeesa N Dhalwani, PhD

Melanie J Davies, MD ²

Thomas Yates, PhD ²

Francesco Zaccardi, PhD ¹

Author affiliations

¹ Leicester Real World Evidence Unit, Diabetes Research Centre, Leicester General Hospital, University of Leicester, Leicester, UK

² NIHR Leicester Biomedical Research Centre, Leicester Diabetes Centre, Leicester, UK

Please add affiliations

Corresponding author: Yogini V Chudasama; Tel 0116 258 8572; yc244@leicester.ac.uk

Word count

Abstract 362; Manuscript 4,706; References 60; Tables 2; Figures 2

Supplementary Material: Methods 6; Figures 2; STROBE Checklist

ABSTRACT

Background and objective: There has been an increasing interest in using life expectancy metrics, such as years of life lost (YLL), to explore epidemiological associations. YLL is easier to understand for both healthcare professionals and the lay people and has become a common measure for evaluating public health priorities. As the literature presents a range of approaches to estimate it, this review aims to: (1) summarise the key methods; (2) show how to implement them using current software; (3) apply them in a real-world example.

Methods: We investigated simpler nonparametric as well as parametric, model-based methods to estimate of YLL, including: (1) Years of potential life lost (YPLL); (2) Global Burden of Disease (GBD) approach; (3) Chiang's life tables; (4) Epi-demographic approach; and (5) Flexible Royston-Parmar parametric survival model. We used data from the UK Biobank with baseline measures collected in 2006-2010 and linkage to mortality records. We selected 36 chronic conditions: participants with two or more conditions were categorised as having multimorbidity.

Results: For the YPLL and GBD method, the analytical procedures allow only to quantify the average YLL within each group (with and without multimorbidity) and, from them, their difference. Conversely, for the Chiang's life tables, the epi-demographic approach, and the Royston-Parmar survival model, both the remaining life expectancy within each group and the YLL could be estimated. In 499,992 UK Biobank participants (white ethnicity, 94%; women, 55%) with a median (IQR) age of 58 (50-63) years, 98,605 (20%) had multimorbidity and 11,871 deaths occurred during the follow-up. The YLLs comparing subjects with vs without multimorbidity varied significantly according to the technique and the modelling approach used: from a longer life expectancy in subjects with multimorbidity using the YPLL and the GBD method to a shorter one using the other three methods (i.e., at 65 years, the

YLL were 1.8, 1.3, and 4.6 years using Chiang's, epi-demographic, and Royston-Parmar approach, respectively).

Conclusions: When comparing the burden of a disease on life expectancy across studies caution is needed as methods may estimate different quantities. While deciding among different methods to estimate YLL, researchers should consider such differences in relation to the purpose of the research and the type of available data.

Key words: Life expectancy, years of life lost, methodology, statistical code, epidemiology, multimorbidity.

SUMMARY BOX

- The concept of years of life lost (YLL) is easier to understand compared to traditional estimates from survival analysis, such as hazard ratios, but very few studies report it.
- A range of different methods of estimating YLL are reviewed: from basic methods – such as life tables – to most recent and advanced methods using statistical modelling.
- Using the example of multimorbidity, the estimated numbers of YLL differs between methods, as each method focused on estimating different quantities.
- This review will help promote a better understanding and use of life expectancy and YLL metrics in a wide range of studies in health care research.

LIST OF ABBREVIATIONS

CI	Confidence intervals
GBD	Global Burden of Disease
HRs	Hazard ratios
ONS	Office for National Statistics
QoF	Quality and outcomes framework
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
UK	United Kingdom
WHO	World Health Organization
YLL	Years of life lost
YPLL	Years of potential life lost

INTRODUCTION

Recently, there has been an increasing interest in using life expectancy measures, such as years of life lost (YLL), as metrics to explore epidemiological associations. Life expectancy estimates are indeed easier to understand and can be simply translated to deliver public health priorities and messages.^{1,2} Yet, the majority of epidemiological studies using time-to-event analysis still present results using the most common metric of hazard ratios (HRs). This is a relative measure, with the potential advantage of being more transportable, i.e. can be applied to populations different from those where they have been estimated,^{3,4} and requires fewer assumptions compared to some methods for estimating YLL. However, HRs can be difficult to understand and frequently misinterpreted;⁵ furthermore, large (or small) HRs may translate into negligible absolute differences.³

A better understanding of the impact of an exposure can be achieved by considering also absolute differences. The term “YLL” broadly encompasses a range of different definitions, names, methods, and quantities, including loss in expectation of life, loss of life expectancy, life lost, lost-lifetime, excess years of life lost, and difference in remaining or residual life expectancy.⁶⁻¹³ Differences among these metrics depend on a number of factors, such as: (i) the reference group for the comparisons with those having the condition of interest (either the standard life expectancy from the general population – assuming they are free from the condition – or those without the condition in the same (observed) population); (ii) the starting age (from birth, commonly used in demography; or from a specific age – at disease diagnosis or at death among those with the disease);² (iii) the population used for the estimates (based only on those individuals who have died, or based on extrapolation beyond the last observed follow-up as most studies do not follow all the patients to the end of life).¹⁰

One approach to estimate YLL is based on quantifying the mean survival time, obtained by constructing the survival curve: the area under the curve, with respect to age, corresponds to the mean age at death; for two populations – one including individuals with a certain

characteristic or condition (i.e., cancer, diabetes, multimorbidity) and the other including subjects without such condition – two survival curves can be obtained: the difference between the areas of the two survival curves indicates the modelled difference in the age at death associated with the presence of the condition (i.e., YLL).

To our knowledge, there has not been an overall review of YLL methods most commonly used in epidemiological studies, their relative advantages and disadvantages, and the software available to compute relevant metrics. In this context, our study aims to summarise and illustrate the differences and interpretations between the most common methods of estimating YLL and apply them to a clinically-relevant example of multimorbidity (i.e. the presence of two or more chronic conditions) using the UK Biobank, a large contemporary cohort; examples of different methods are complemented by statistical codes.

YEARS OF LIFE LOST: DEFINITIONS and METHODS

From reviewing the literature, we identified the most common methods to estimate life expectancy/YLL in epidemiological investigations, from simpler nonparametric to more advanced parametric, model-based methods, reflecting in part the progressively larger access to individual-level rather than aggregate data. We give a brief summary of each method before comparing and contrasting them using multimorbidity as an example.

Although these methods allow to quantify the difference in the years of life in relation to the presence of a condition, we note that the underlying estimands are different, thus we do not *a-priori* expect consistent findings across all the explored methods. Further details about the methods, the analytical procedures, and the statistical codes are reported in the

Supplementary Material.

1. Years of potential life lost

Years of potential life lost (YPLL) is calculated as the difference between the age at death and age from a selected cut-off, e.g. 65, 75, or 85 years. This concept dates back to the 1940s – when it was first introduced to compare deaths due to tuberculosis with heart disease and cancer – and, for each death, the number of years of life remaining up until premature mortality was calculated.¹⁴ For example, if premature mortality is taken to be at 65 years, an individual who has died at 55 years has lost 10 years of potential life. Deaths after the cut-off age are ignored, giving more value and weight to deaths occurring at younger ages.⁶ The data for this method is restricted to only those who have died. A disadvantage of using YPLL is that the selected cut-off age is arbitrary, making it difficult to compare results across studies. However, there have been further adaptations of standardising YPLL to overcome this issue.⁶ Another disadvantage is that accurate information on the cause of death is required.¹⁰ YPLL is usually based on ranking data by the leading causes of death for a given period of time: as such, the results emphasise the causes of death most commonly found in the younger population.⁶

Software

YPLL is simple to compute. This method can be applied using a calculator or an Excel spreadsheet.

2. Global Burden of Disease approach

The Global Burden of Disease (GBD) approach is an extension of the YPLL measure and, although restricted to those who have died, this method includes deaths beyond an arbitrary cut-off. It is based on comparing the age of death to an external standard life expectancy curve, and can further incorporate time discounting and age weighting.^{7,15} Since the YPLL method can be overly sensitive to deaths at younger ages, the GBD approach uses therefore a discounting strategy based on the idea that the productive years of life are more valuable than the very young or the very old. From the introduction of the GBD study in 1993,¹⁶ the burden of disease concept has expanded to numerous countries and is used by major health organisations, such as the World Health Organization (WHO), as well as an outcome measure in many cost-effectiveness analysis.¹⁶ The general equation for the GBD YLL approach is developed on key assumptions of standardised weightings for age, discount rate, and constants. For each age group, it requires details on the number of participants, number of deaths, and the standard life expectancy. The benefit of using the GBD approach is that the age-adjusted rates allow comparisons of causes of deaths in rank order across groups, conditions, and over time,¹⁵ which is important from a public health perspective because it identifies patterns of preventable YLL.

Software

An open source programming code is available in R that allows to estimate GBD YLL:⁷ all that is required is a value of the number of deaths, average age of death, and the standard life expectancy for that age.¹⁷ Alternatively, an Excel template created by the WHO National tools allows the calculation to be carried out for all ages at the same time.¹⁸

3. Chiang's life tables

Life tables provide a description of the mortality rates and are one of the most important tools used in demography and actuarial science. The two forms of life tables are: cohort or period/current life tables. The cohort life table identifies the actual observed mortality experience of individuals born at the same time followed up throughout time, and is similar to national life table data. The period/current life table describes the mortality experience of a population subject to the age-specific mortality rates currently observed and applied to a hypothetical cohort, and any future changes to mortality rates would not be taken into account, which is used in epidemiological studies. There are a number methods to construct life tables;¹⁹⁻²¹ the most common is the Chiang's life table method,⁸ as this approach is widely used for population estimates (i.e., by the Office for National Statistics (ONS)) in the UK^{17,21} and has been applied to large epidemiological studies.²² Chiang's life table method usually groups the data by five year age intervals, yet this depend on the size of the study. The life table is then constructed with the number of participants and deaths in each age interval, and is based on conditional probability of surviving to the start of the age group. The survival probabilities are applied to a hypothetical cohort and using cumulative distribution the expected number of years is estimated. The population can be stratified into groups where two life tables are constructed separately and the expected remaining years of life is subtracted to find YLL. The results from this method have been found to be less stable in smaller populations, and the chosen value of the ending age interval is arbitrary, therefore making it difficult to produce reliable estimates and compare results between studies.^{19,20} Nevertheless, life tables are one of the oldest and more traditional methods of calculating remaining life expectancy, and works well in large national studies and applied to YLL analyses.

Software

Templates are freely available online from the ONS,¹⁷ or the Association of Public Health Epidemiologists in Ontario.²³ They can also be calculated using R,²⁴ SAS,²⁵ or Stata.²⁶

4. Epi-demographic approach: Epi package

By linking a demographic and survival analysis approach within a parametric modelling framework, YLL can be estimated using the Epi package in R.²⁷ Firstly, the follow-up time is split in small intervals: this results in multiple observations for the same subject, with age, time into the study, and calendar time increasing by the same amount: for example, a subject starts the study/observation at age 45.5 years in year 2000.5 with time into study split in intervals of six months (0.5 years): thus, at the end of the next interval, age will be 46 years, calendar time 2001, and the time into the study 0.5 years.²⁸ Age is then modelled with a Poisson regression (with a natural spline) using each small time interval into the study as offset: the estimated coefficients allow the prediction of age-specific rates and, from them, the survival functions/curves; the area under the curve allows to estimate the residual life. This method has been used for large national registries,^{9,29} and it is straightforward to use.

Software

The Epi package has the functions to calculate estimated residual life (*erl*) and the YLL (*yll*); it also allows modelling the possible transition from a state without to a state with the disease/condition of interest across which the YLL is computed.³⁰ This method has also been used in Stata.³¹

5. Flexible Royston-Parmar parametric survival model

The Royston-Parmar flexible parametric survival model allows for greater flexibility to accurately capture data and extrapolate for future predictions.³² This is done by using restricted cubic splines thus obtaining smooth parametric survival curve.³² It was first introduced by Royston and Parmar in 2002 as an alternative to the Cox proportional model

and implemented in Stata.^{32,33} The most recent user-written Stata command *stpm2* is a quick and efficient method to fit flexible parametric models.³⁴ Previous studies have used this method mainly in cancer epidemiology.³⁵⁻³⁷ There are a number of possible outputs available from these models: this review focuses on calculating YLL. Firstly, *stpm2* is used to fit the regression model using age as time scale and including the condition of interest as covariate.³⁸ Then, the postestimation *predictnl* command enables the estimation of the survival curves across age; calculating the area under the curve is then straightforward, and YLL is calculated as the area between the two survival curves of interest. The advantages of using this method is that it performs better for smaller samples compared to some of the previously reported methods;³⁹ it enables 95% confidence intervals estimation (with delta method) around YLL; and potential confounders can be accounted for (both conditional and average/standardised estimation).⁴⁰ The disadvantage of this method is the computationally intensive process, particularly for large datasets and when confidence intervals are required.

Software

This method is very broad and there are a number of functions that can be applied.³¹ As well as having commands available in Stata, it can also be carried out in SAS⁴¹ and R.⁴² In this review, we will consider its implementation in Stata.

EXAMPLE FROM MULTIMORBIDITY

This real-world example focuses on estimating the number of YLL in a population of individuals with multimorbidity compared to those without multimorbidity. This study is reported in line with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline (**Supplementary Checklist S1**).

Data

We used UK Biobank data (Application Number 14146), one of the world's largest Biobank cohorts, designed to improve the prevention, diagnosis, and treatment of chronic diseases in middle-aged adults, recruited from 22 sites across England, Wales and Scotland. This longitudinal study included 502,629 participants with baseline measures collected between 2006 and 2010 and linkage to mortality records.⁴³ Written informed consent was obtained prior to data collection and ethics was approved by the North-West Research Ethics Committee.⁴⁴ Overall, 91 participants withdrew from the study. The deaths occurring within the first two years at baseline were removed (n=2,516) to reduce the possible impact of reverse causation, i.e. participants with undiagnosed or sub-clinical disease(s) leading to mortality. There was a small number of participants before the age of 45 years at follow-up (n=30, of which 29 died) who were excluded from analyses to avoid uncertainty of deaths at younger age (**Supplementary Figure S1**).

Mortality

All-cause mortality data were obtained from the National Health Service (NHS) Information Centre for participants from England and Wales and the NHS Central Register for participants from Scotland. Time for survivors was censored on 31st January 2016 for England and Wales and 30th November 2015 for Scotland.

Definition of multimorbidity

The UK Biobank collected self-reported medical information based on physician diagnosis at the baseline assessment. Three sources of selecting long-term cardiovascular, non-cardiovascular, or mental health conditions was used. The first included conditions from the Quality and Outcomes Framework which reports the most common diseases in the UK;⁴⁵ the second from a large UK based study, containing 40 of the recommended core disorders for any multimorbidity measure;⁴⁶ and the last from a systematic review on multimorbidity indices that included 17 conditions.⁴⁷ Based on these sources, we selected a total of 36 chronic conditions. Participants with two or more of these 36 chronic conditions were categorised as having multimorbidity (**Supplementary Methods S1**).

Statistical methods

Baseline characteristics were stratified by multimorbidity status and summarised. To model the association between multimorbidity and all-cause mortality, unadjusted Cox proportional hazards regression was used with age as time scale. HRs and corresponding 95% confidence intervals (95% CI) were calculated. Participants without multimorbidity constituted the reference category. For each YLL method described above, we provide detailed steps and example statistical code in the **Supplementary Material Methods S2-S6** to demonstrate their implementation. Stata version 17.0, R studio version 1.4, and Microsoft Excel were used to perform the analyses.

RESULTS

Baseline characteristics

A total of 499,992 participants were included in this study, with most participants being white (94%) and women (55%); the median (IQR) age was 58 (50-63) years. Subjects with multimorbidity (n=98,605; 19.7%) were older compared to participants without (>60 years: 52% vs. 35%, respectively; **Table 1**).

Multimorbidity and all-cause mortality

During a mean study follow-up of 7 (range 2-10) years and 3.5 million person-years at risk, 11,871 deaths were recorded: 4,587 (4.7%) in subjects with multimorbidity and 7,284 (1.8%) in those without. The median [IQR] age at death was 61 [54-65] and 57 [49-63] years in subjects with and without multimorbidity, respectively (**Supplementary Figure S2**). Using age as the time scale, the Cox regression model showed that the mortality rates were twice as high for those with multimorbidity compared to those without, with an unadjusted hazard ratio of 2.00 (95% CI: 1.93, 2.08).

Years of life lost estimates across methods

The results for the five methods are presented in **Table 2**. Given the analytical procedures underpinning the estimations, for the YPLL and GBD method only the average YLL could be estimated within each group (multimorbidity and no multimorbidity): then, from them, the difference between the two groups could be estimated. Conversely, for the Chiang's life tables, the epi-demographic approach with the *Epi* package, and the flexible Royston-Parmar parametric survival approach with *stpm2*, both the remaining life expectancy within each group and the YLL could be estimated (**Figure 1**).

1. Years of potential life lost (**Methods S2**)

Using a selected cut-off age of 65 years, there was a total of 25,504 potential years of life lost. The average YPLL lost for a subject with multimorbidity was 6.15 (SD, 4.81) years and for a subject without 6.88 (SD 4.95), giving a difference of 0.73 years. Using a selected cut-off point of 75 years, corresponding figures were: 103,552, 8.52 (SD, 6.05), 9.67 (SD, 6.64), and 1.15 years, respectively. Since there was a higher number of deaths in subjects without vs with multimorbidity (61.4% vs. 38.6%), the results showed larger YPLL values in those without multimorbidity. In addition, the baseline age for the majority of those with multimorbidity was over 60 years (52%) compared to those without multimorbidity (35%); therefore, when followed-up until the age of death, participants with multimorbidity were less likely to lose more years of potential life than those without multimorbidity.

2. Global Burden of Disease approach (**Methods S3**)

The GBD approach estimate of the total YLL was 62,538 and 103,595 years in subjects with and without multimorbidity, respectively. For participants who died between 45-49 years, the average YLL was 21.61 years in those with multimorbidity and 21.63 years in those without: the difference in YLL was therefore 0.01 years. For participants who died between 65-69 years, corresponding figures were 13.80, 13.85, and 0.05 years.

3. Chiang's life tables approach (**Methods S4**)

Using Chiang's life tables, the results showed that, at the age of 45-49 years, the remaining life expectancy was 19.99 (95% CI: 19.30, 20.67) years in participants with multimorbidity and 27.54 (27.36, 27.73) in those without: the YLL was therefore 7.56 years. At the age of 65-69 years, corresponding figures were: 9.72 (9.60, 9.84), 11.55 (11.49, 11.61), and 1.83 years.

4. Epi-demographic approach: Epi package (**Methods S5**)

The results using the Epi package showed that, at the age of 45 years, the remaining life expectancy in subjects with multimorbidity was 46.25 years and in those without 51.16

years, resulting in YLL of 4.91 years. At the age of 65 years, corresponding estimates were 31.54, 32.82, and 1.27 years.

5. Flexible Royston-Parmar parametric survival model (**Methods S6**)

The flexible Royston-Parmar parametric survival model showed that, at the age of 45 years, participants with multimorbidity had 39.52 (95% CI: 38.85, 40.19) remaining years of life while those without 44.71 (44.03, 45.39) years, resulting in YLL of 5.19 (4.81, 5.58) years. Corresponding figures at the age of 65 years were: 21.05 (20.38, 21.71), 25.63 (24.95, 28.89), and 4.59 (4.24, 4.93) years. When the effect of multimorbidity was assumed not constant over time (i.e., non-proportionality of hazards), the results were: at 45 years, 40.34 (39.35, 41.33) remaining years of life in subjects with multimorbidity and 43.97 (95% CI: 43.05, 44.88) those without, resulting in YLL of 3.63 (2.27, 4.99) years; at 65 years, 22.19 (21.21, 23.17) in and 24.82 (23.89, 28.56) subjects with and without multimorbidity, respectively, resulting in YLL of 2.62 (1.27, 3.98) years.

DISCUSSION

The YLL measures are becoming a key health indicator in understanding a population's health. In this article, a range of different methods for estimating YLL were reviewed, from the oldest – such as YPLL and life tables – to the most recent and advanced method using statistical modelling. The example demonstrated that the total numbers of YLL due to multimorbidity varied according to the technique used, as each method focuses on estimating different quantities, emphasising the need to understand the different concepts and use the most suitable method for the database and the specific research question. In turn, our results also underline the need to consider such differences when comparing (and summarising) studies reporting “years of life lost” related to a certain condition.

An overview of the YLL methods described in our manuscript is illustrated in **Figure 2**. The simplest method is the YPLL, which is especially useful in a clinical setting requiring minimal data input. Though, YPLL is usually found in large studies that help quantify the total disease burden and identify the most common causes of premature deaths. When applied to our multimorbidity example, the results showed a gain in YLL, since the baseline age for the majority of those with multimorbidity was much older than those without multimorbidity; therefore, participants with multimorbidity were less likely to lose more YPLL than those without multimorbidity at age of death. The GDB approach, which additionally accounted for time discounting and age weighting, showed a small difference in the average YLL. These two methods are based on only participants who have died and do not consider people who are alive during the time period; it is thus most useful, probably, in extremely large national databases.

The results from Chiang's life tables showed a higher number of YLL only at the youngest age – at 45-49 years, YLL was 7.56 years, compared to the epi-demographic approach (at 45 years, the YLL was 4.91 years) or the Royston-Parmar model (at 45 years, the YLL was 5.19 years) – and then decreases. This is because: (1) the results were based on

aggregated data where abridged life tables with five year age intervals were used, as the sample size was not large enough for individual year data (unabridged or complete life table),²⁰ and the mortality rate in each age interval is assumed to be constant; (2) the chosen value for the last age interval was 7, which was the difference between 75 (inclusive) and the last observed age, 81 years. This value is arbitrary and varies in the literature, as any age interval value could be applied, or it could be left open-ended.^{19,20,48} Leaving this open-ended interval assumes a constant hazard of death for the remaining ages, which may lead to erroneous conclusions about mortality measures.²⁰ Therefore, in relation to the type of life table (abridged or unabridged) and the last age interval value, the results for the remaining life expectancies and YLL from the Chiang's life table method will vary.

The epi-demographic approach using the *Epi* package is based on survival curves obtained from mortality rates, following a Poisson modelling on split data: this fully parametric approach gives more flexibility (i.e., may allow multiple time scales and/or other covariates for conditional estimates). The results from the example showed higher remaining life expectancies than other methods, as this was extrapolated to 100 years; of note, the smaller the extrapolation, the smaller the remaining life expectancy (i.e., the area under the curve). Additionally, the estimated survival curve depends also on the functional form of age in the Poisson regression (i.e., continuous vs natural cubic spline – with possible effects in relation to the number and the position of the knots). An advantage of the *Epi* package is the possibility to easily model the transition from one group to the other exposure group across which YLL is estimated, in our case from no multimorbidity to multimorbidity.

The final method is based on the Royston-Parmar model implemented in the *stpm2* command in Stata, which enables to estimate the survival functions (or differences among them) and their uncertainties. Similar to the *Epi* package, the results of this fully parametric approach may reflect decisions around the maximum age used for the survival prediction (in this analysis we used 100 years as for the Poisson modelling approach); the

number/position of the knots for the cubic spline; or interactions between the covariate of interest and age. Besides conditional estimates, i.e. for specific values of the covariates included in the model, recent further development also enables marginal (standardised) average effect estimates, including YLL.⁴⁰ This method also allowed the estimate of the 95% CI, giving a measure of uncertainty. Notably, running the full code was time consuming.

Modelling approaches have advantages in that they allow for greater flexibility for both individual-level and grouped data.⁴⁹ As most studies do not follow participants to the end of life, the risk of extrapolating (i.e. to 100 years) needs to be taken into consideration. This means that factors such as the time-scale, length of follow-up, number of participants and deaths, as well as the age distribution of the population and among those who died, can influence the results. To date, there is no standard method for estimating life expectancy and YLL, and the most appropriate method may vary depending on the research question being addressed and the available data.

From the example presented here, the mortality rates were twice as high for those with multimorbidity compared to those without (HR, 2). Using the Royston-Parmar model, the YLL at the age of 45 years in participants with multimorbidity was on average 5.2 years lower than in participants without; one advantage of presenting HRs is that, as a relative measure, the estimate is more transportable and can be used to compare and combine estimates from different populations with different baseline risk.^{3,4} In this perspective, a recent study confirmed that the HR estimates from the UK Biobank are similar to those obtained from studies based on the general population.⁵⁰ On the other hand, presenting the absolute risk in terms of 'years of your life' may be far more influential from a public health perspective and policy makers, as the concept is easier to grasp.¹ Therefore, presenting both relative and absolute measures strengthens the reporting of associations from an epidemiological, clinical, and public perspective, and aligns with current guidelines for conducting and reporting epidemiological studies (STROBE; item #16).⁵¹

This review has some limitations. We focused only on the most commonly used YLL methods; others methods (or further extension of those mentioned in this study) are available and may also enable the estimation of the YLL by different causes of death.⁵²⁻⁵⁴ UK Biobank participants were volunteers, with slightly higher representation from affluent groups; therefore, participants may not be completely representative of the UK population.⁵⁵ Participants who died within the first two years of follow-up were excluded to reduce the risk of reverse causation, though it is still possible that participants with multimorbidity may generally be less well, which could result in a higher mortality rate. Additionally, data on chronic diseases was collected at baseline, yet participants may have subsequently developed other chronic conditions throughout follow-up. As most studies predict survival beyond the duration of the study data, further investigations with longer follow-up may explore the impact of the model structural uncertainty on the YLL estimates.⁵⁶ Lastly, the minimum sample size of participants and the number of events required to estimate life expectancy and YLL for each method is still yet to be clarified, and further research is needed, possibly through simulation studies.²⁰

Strengths of this study include the utilisation and illustration of different methods to estimate life expectancy and YLL, with statistical codes reported for the available software. Furthermore, we applied the methods to an emerging clinical and research area, as one of the four recommendations for research from the National Institute for Health and Care Excellence (NICE) guidance for multimorbidity focuses on predicting the life expectancy.⁵⁷ Finally, the methodologies used to estimate the number of YLL could be also be used to enhance the interpretation and potentially the promotion of beneficial interventions, such as gain in disease-free life expectancy,⁵⁸ or lifestyle modifications.^{59,60} For instance, patients with multimorbidity may be motivated to exercise more if their doctors communicated a simple health care message, such as 10 minutes of brisk walking a day could lead you to having up to 4 years of additional life.⁶⁰

RECOMMENDATIONS

YLL metrics are simple summary measures that can enhance the interpretation of the findings from epidemiological studies for health care professionals and the public. When possible, we therefore recommend that, alongside a relative risk (i.e. rate or hazard ratio), also the absolute risk (i.e. number of events, rates, YLL) should be reported, thus providing a more complete and actionable information. As the available methods to estimate YLL metrics could relate to different quantities, comparing the spatiotemporal burden of diseases on the life expectancy may be difficult. Furthermore, investigators should not only choose the YLL method in view of the purpose of the research, the type of available data, and the required flexibility but – when a modelling strategy is considered – they should also assess the consistency of the findings across different models.

DECLARATIONS

Funding: YC acknowledges funding from the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: KK, FZ and ND conceived the idea of the study; ND acquired the data; YC carried out the review and statistical analysis; FZ supervised the analysis. YC, FZ, KK, and CG interpreted the findings; and YC drafted the manuscript. YC had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors critically reviewed the manuscript and YC revised the manuscript for final submission.

Acknowledgements: This research has been conducted using the UK Biobank Resource (Reference 14614). We would like to acknowledge Dr Sarwar Islam Mozumder for his comments on an initial draft of the manuscript. We acknowledge the support from the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM), and the NIHR Leicester Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Ethical approval and informed consent: All participants gave written informed consent prior data collection. UK Biobank has full ethical approval from the NHS National Research Ethics Service (16/NW/0274). UK Biobank Reference 14614.

Data access and responsibility: The data that support the findings of this study are available from UK Biobank project site, subject to registration and application process. Further details can be found at <https://www.ukbiobank.ac.uk>.

Conflict of interest: KK is the National Lead for multimorbidity for National Institute for Health Research Applied Research Collaboration. MJD is Convenor of the NIHR Diet and Activity Research Translation Collaboration. All other authors have no conflict of interest.

REFERENCES

1. Fraser GE, Shavlik DJ. Ten years of life: Is it a matter of choice? *Arch Intern Med* 2001;161(13):1645-52. [published Online First: 2001/07/24]
2. Andersen PK. Life years lost among patients with a given disease. *Statistics in Medicine* 2017
3. Kent DM, Nelson J, Dahabreh IJ, et al. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *International journal of epidemiology* 2016;45(6):2075-88.
4. Schwartz LM, Woloshin S, Dvorin EL, et al. Ratio measures in leading medical journals. *BMJ* 2006 doi: doi:10.1136/bmj.38985.564317.7C
5. Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Ann Epidemiol* 2018;28(1):54-57. doi: 10.1016/j.annepidem.2017.10.014 [published Online First: 2017/12/15]
6. Gardner JW, Sanborn JS. Years of Potential Life Lost (YPLL)—What Does it Measure? *Epidemiology* 1990;1(4):322-29.
7. Aragón TJ, Lichtensztajn DY, Katcher BS, et al. Calculating expected years of life lost for assessing local ethnic disparities in causes of premature death. *BMC Public Health* 2008;8:116-16. doi: 10.1186/1471-2458-8-116
8. Chiang CL. The life table and its applications: R.E. Krieger Publishing Company 1984.
9. Carstensen B, Kristensen JK, Ottosen P, et al. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. *Diabetologia* 2008;51(12):2187-96.
10. Andersson TML, Dickman PW, Eloranta S, et al. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in medicine* 2013;32(30):5286-300.
11. Martinez R, Soliz P, Caixeta R, et al. Reflection on modern methods: years of life lost due to premature mortality—a versatile and comprehensive measure for monitoring non-communicable disease mortality. *International journal of epidemiology* 2019;48(4):1367-76.
12. Andersen PK. Decomposition of number of life years lost according to causes of death. *Statistics in Medicine* 2013;32(30):5278-85. doi: <https://doi.org/10.1002/sim.5903>
13. Bonneux L. How to measure the burden of mortality? *Journal of Epidemiology and Community Health* 2002;56(2):128-31. doi: 10.1136/jech.56.2.128
14. Dempsey M. Decline in tuberculosis: the death rate fails to tell the entire story. *American review of tuberculosis* 1947;56(2):157-64.
15. Murray CJL, Lopez A D. The global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: Harvard School of Public Health on behalf of the World Health Organization and the World Bank 1996:990.
16. Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health policy and planning* 2001;16(3):326-31. [published Online First: 2001/08/31]
17. Office for National Statistics. Dataset National life tables: UK 2021 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lif>

- [eexpectancies/datasets/nationallifetablesunitedkingdomreferencetables](#) accessed 18-01 2022.
18. The Global Burden of Disease concept. 3. The Global Burden of Disease concept 2021 [Available from: https://www.who.int/quantifying_ehimpacts/publications/en/9241546204chap3.pdf accessed 18-01 2022.
19. Silcocks PBS, Jenner DA, Reza R. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. *Journal of Epidemiology and Community Health* 2001;55(1):38-43. doi: 10.1136/jech.55.1.38
20. Eayres D, Williams ES. Evaluation of methodologies for small area life expectancy estimation. *Journal of Epidemiology and Community Health* 2004;58(3):243-49. doi: 10.1136/jech.2003.009654
21. Office for National Statistics. Guide to calculating national life tables: Explanation of the methodology used to create the national life tables 2019 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/methodologies/guidetocalculatingnationallifetables> accessed 18-01 2022.
22. Wen CP, Wai JP, Tsai MK, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *Lancet* 2011;378(9798):1244-53. doi: 10.1016/s0140-6736(11)60749-6 [published Online First: 2011/08/19]
23. Association of Public Health Epidemiologists in Ontario. Life Table Template V1.2 2006 [Available from: <https://core.apheo.ca/index.php?pid=223> accessed 18-01 2022.
24. RDocumentation. lifetable: Construct lifetables from mortality rates [Available from: <https://www.rdocumentation.org/packages/demography/versions/1.22/topics/lifetable> accessed 18-01 2022.
25. Vincent A, Zheng S. Life Expectancy Tables Getting SAS® to Run the Hard Math. *SAS Global Forum Proceedings* 2019;paper 3064-2019.
26. Ramalheira C. Global and multiple causes-of-death life tables from complete or aggregated vital data. *Stata Technical Bulletin* 2001;10(59)
27. B C, Plummer M, Laara E, et al. Epi: Statistical Analysis in Epidemiology 2021 [Available from: <https://cran.r-project.org/web/packages/Epi/index.html> accessed 18-01 2022.
28. Carstensen B. Analysis of follow-up data using the Lexis functions in Epi 2021 [Available from: <http://bendixcarstensen.com/Epi/flup.pdf> accessed 18-01 2022.
29. Hansen M, Jensen M, Carstensen B. Causes of death among diabetic patients in Denmark. *Diabetologia* 2012;55(2):294-302.
30. Carstensen B. Years of Life Lost - theory and practice 2017 [Available from: <http://bendixcarstensen.com/Epi/Courses/EDEG2017/> accessed 18-01 2022.
31. Dickman PW, Lambert PC, Eloranta S, et al. Statistical methods for population-based cancer survival analysis Computing notes and exercises 2013 [Available from: <http://www.pauldickman.com/survival/labs.pdf> accessed 18-01 2022.
32. Royston P, Lambert PC. Flexible parametric survival analysis using Stata : beyond the Cox model. College Station: Stata Press 2011.
33. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and

- estimation of treatment effects. *Statistics in Medicine* 2002;21(15):2175-97. doi: 10.1002/sim.1203
34. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata Journal* 2009;9(2):265.
35. Syriopoulou E, Bower H, Andersson TM, et al. Estimating the impact of a cancer diagnosis on life expectancy by socio-economic group for a range of cancer types in England. *British journal of cancer* 2017;117(9):1419.
36. Syriopoulou E, Morris E, Finan PJ, et al. Understanding the impact of socioeconomic differences in colorectal cancer survival: potential gain in life-years. *British Journal of Cancer* 2019;120(11):1052-58. doi: 10.1038/s41416-019-0455-0
37. Rutherford MJ, Andersson TM, Björkholm M, et al. Loss in life expectancy and gain in life years as measures of cancer impact. *Cancer epidemiology* 2019;60:168-73.
38. Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2015;85(4):777-93. doi: 10.1080/00949655.2013.845890
39. Dehbi H-M, Royston P, Hackshaw A. Life expectancy difference and life expectancy ratio: two measures of treatment effects in randomised trials with non-proportional hazards. *Bmj* 2017;357:j2250.
40. Lambert P. standsurv 2019 [Available from: <https://pclambert.net/software/standsurv/> accessed 18-01 2022.
41. Dewar R, Khan I. A new SAS macro for flexible parametric survival modeling: applications to clinical trials and surveillance data. *Clin Invest* 2015;5(12):855-66.
42. Clements M, Liu X, Christoffersen B, et al. rstpm2: Smooth Survival Models, Including Generalized Survival Models 2021 [Available from: <https://cran.r-project.org/web/packages/rstpm2/index.html> accessed 18-01 2022.
43. UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource 2007 [updated 21 March 2007. Available from: <https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf> accessed 18-01 2022.
44. UK Biobank. UK Biobank ethics and governance framework 2007 [Available from: <https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf> accessed 18-01 2022.
45. Doran T, Kontopantelis E, Reeves D, et al. Setting performance targets in pay for performance programmes: what can we learn from QOF? *BMJ: British Medical Journal (Online)* 2014;348
46. Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 2012;380(9836):37-43.
47. Diederichs C, Berger K, Bartels DB. The measurement of multiple chronic diseases—a systematic review on existing multimorbidity indices. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 2011;66(3):301-11.
48. Antiretroviral Therapy Cohort Collaboration. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet* 2008;372(9635):293-9. doi: 10.1016/s0140-6736(08)61113-7 [published Online First: 2008/07/29]

49. Rutherford MJ, Dickman PW, Lambert PC. Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology* 2012;36(1):16-21.
50. Batty GD, Gale CR, Kivimäki M, et al. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. 2020 doi: 10.1136/bmj.m131
51. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61(4):344-9. doi: 10.1016/j.jclinepi.2007.11.008 [published Online First: 2008/03/04]
52. Overgaard M, Andersen PK, Parner ET. Regression analysis of censored data using pseudo-observations: an update. *The Stata Journal* 2015;15(3):809-21.
53. Plana-Ripoll O, Canudas-Romo V, Weyer N, et al. lillies: An R package for the estimation of excess Life Years Lost among patients with a given disease or condition. *PloS one* 2020;15(3):e0228073.
54. Mozumder SI, Rutherford MJ, Lambert PC. Estimating restricted mean survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models. *BMC Medical Research Methodology* 2021;21(1):52. doi: 10.1186/s12874-021-01213-0
55. Swanson JM. The UK Biobank and selection bias. *The Lancet* 2012;380(9837):110. doi: 10.1016/S0140-6736(12)61179-9
56. Kearns B, Stevens J, Ren S, et al. How Uncertain is the Survival Extrapolation? A Study of the Impact of Different Parametric Survival Models on Extrapolated Uncertainty About Hazard Functions, Lifetime Mean Survival and Cost Effectiveness. *Pharmacoeconomics* 2020;38(2):193-204. doi: 10.1007/s40273-019-00853-x [published Online First: 2019/11/26]
57. National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. NICE in development [GID-CGWAVE0704]: NICE; 2016 [Available from: <https://www.nice.org.uk/guidance/NG56> accessed 18-01 2022.
58. Dorresteyn JAN, Kaasenbrood L, Cook NR, et al. How to translate clinical trial results into gain in healthy life expectancy for individual patients. *BMJ* 2016;352:i1548. doi: 10.1136/bmj.i1548
59. Chudasama YV, Khunti K, Gillies CL, et al. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: A longitudinal cohort study. *PLoS medicine* 2020;17(9):e1003332.
60. Chudasama YV, Khunti KK, Zaccardi F, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. *BMC Medicine* 2019;17(1):108. doi: 10.1186/s12916-019-1339-0

Table 1. Participant characteristics and outcomes by multimorbidity status

Characteristics	All participants (N = 499,992)	Multimorbidity status	
		With multimorbidity (n = 98,605)	Without multimorbidity (n = 401,387)
Men	227,507 (45.5)	45,655 (46.3)	181,852 (45.3)
Women	272,485 (54.5)	52,950 (53.7)	219,535 (54.7)
Age, y (median [IQR])	58 [50-63]	61 [54-65]	57 [49-63]
≤ 50	131,885 (26.4)	15,041 (15.3)	116,844 (29.1)
51-60	176,540 (35.3)	32,327 (32.8)	144,213 (35.9)
> 60	191,567 (38.3)	51,237 (52.0)	140,330 (35.0)
White	470,292 (94.1)	92,748 (94.1)	377,544 (94.1)
Non-white	26,952 (5.4)	5,352 (5.4)	21,600 (5.4)
Missing	2,748 (0.6)	505 (0.5)	2,243 (0.6)
Deaths	11,871 (2.4)	4,587 (4.7)	7,284 (1.8)

Shown are numbers (%) unless stated otherwise. Y=years; IQR=interquartile range.

Table 2. Estimating years of life lost in subjects with multimorbidity compared to those without multimorbidity using different methods

Method	Age (years)	Average years of life lost (SD)		Difference in average years of life lost, without vs with
		With multimorbidity	Without multimorbidity	
Years of potential life lost	65 cut-off	6.15 (4.81)	6.88 (4.95)	0.73
	75 cut-off	8.52 (6.05)	9.67 (6.64)	1.15
Global burden of disease approach	45-49	21.61	21.63	0.01
	65-69	13.80	13.85	0.05
Method	Age (years)	Remaining life expectancy (95% CI)		Years of life lost (95% CI), without vs with
		With multimorbidity	Without multimorbidity	
Chiang's life tables	45-49	19.99 (19.30, 20.67)	27.54 (27.36, 27.73)	7.56
	65-69	9.72 (9.60, 9.84)	11.55 (11.49, 11.61)	1.83
Epi-demographic approach	45	46.25	51.16	4.91
	65	31.54	32.82	1.27
Royston-Parmar PH	45	39.52 (38.85, 40.19)	44.71 (44.03, 45.39)	5.19 (4.81, 5.58)
	65	21.05 (20.38, 21.71)	25.63 (24.95, 28.89)	4.59 (4.24, 4.93)
Royston-Parmar non PH	45	40.34 (39.35, 41.33)	43.97 (43.05, 44.88)	3.63 (2.27, 4.99)
	65	22.19 (21.21, 23.17)	24.82 (23.89, 28.56)	2.62 (1.27, 3.98)

Years of life lost = without multimorbidity – with multimorbidity.

SD=standard deviation; CI=confidence interval; PH=proportional hazards

FIGURES

Figure 1. Number of years of life lost in subjects with multimorbidity compared to those without in the UK Biobank

Legend

PH=proportional hazards; Y=years.

Marker estimates on the Chiang's life table method indicates the years of life lost for the five-year age interval (i.e. 45 to 50 years). The last age interval was 75 to 81 years, as the last age at follow-up was 81 years.

Shaded area indicates 95% confidence interval.

Figure 2. Summary of methods to estimate years of life lost

Legend

YLL=years of life lost; YPLL=years of potential life lost; GBD=Global Burden of Disease.

Figure 1. Number of years of life lost in subjects with multimorbidity compared to those without multimorbidity in the UK Biobank

PH=proportional hazards; Y=years.

Marker estimates on the Chiang's life tables method indicates the years of life lost for the five-year age interval (i.e. 45 to 50 years). The last age interval was 75 to 81 years, as the last age at follow-up was 81 years.

Shaded area indicates 95% confidence interval.

Figure 2. Summary of methods for estimating years of life lost

Supplementary material

Estimating life expectancy and years of life lost in epidemiological studies: a review of methods using an example from multimorbidity

Yogini V Chudasama, Kamlesh Khunti, Clare L Gillies, Nafeesa N Dhalwani, Melanie J Davies, Thomas Yates, Francesco Zaccardi

Table of Contents

Figure S1. Flow chart of participants included in the study	1
Figure S2. Distribution of age of death by multimorbidity status	2
Methods S1. List of the 36 chronic conditions included in the definition of multimorbidity	3
Methods S2. Years of potential life lost example using UK Biobank	4
Methods S3. Global Burden of Disease approach example using UK Biobank.....	5
Methods S4. Chiang's life tables example using UK Biobank.....	8
Methods S5. Epi package example using UK Biobank	12
Methods S6. Flexible Royston-Parmar parametric survival model example using UK Biobank	14
Checklist S1. Strengthening the reporting of observational studies in epidemiology (STROBE).....	17

Figure S1. Flow chart of participants included in the study

Figure S2. Distribution of age of death by multimorbidity status

Methods S1. List of the 36 chronic conditions included in the definition of multimorbidity

1. Anaemia
2. Angina
3. Anxiety or panic attacks
4. Asthma
5. Atrial fibrillation
6. Bronchiectasis
7. Cancer
8. Chronic kidney disease
9. Chronic obstructive pulmonary disease
10. Chronic sinusitis
11. Cirrhosis
12. Dementia
13. Depression
14. Diabetes
15. Eczema or dermatitis
16. Epilepsy
17. Glaucoma
18. Heart failure
19. Hepatitis
20. Hypertension
21. Inflammatory bowel disease
22. Irritable bowel syndrome
23. Meningitis
24. Migraine
25. Multiple sclerosis
26. Myocardial infarction
27. Osteoporosis
28. Parkinson's disease
29. Peripheral vascular disease
30. Prostate problem
31. Rheumatoid arthritis
32. Schizophrenia
33. Stroke
34. Thyroid problem
35. Tuberculosis
36. Vestibular disorder

Methods S2. Years of potential life lost example using UK Biobank

Step 1. Select an age cut-off point (i.e. 65, 75, 85 years).

Step 2. Exclude participants who have died or are still alive after the cut-off point.

Step 3. For each participant who died before the end point, calculate the YPLL by subtracting the age at death from the cut-off point.

$$\text{YPLL}_{\text{individual}} = \text{cut-off point} - \text{age at death}$$

Example

Selected cut-off = 65

Age at death = 55

YPLL individual = 65 - 55 = 10 years of potential life lost

Step 4. Total the individual YPLL.

Steps 5. To find the average YPLL, the total number of YPLL is divided by the total number of deaths.

Steps 6. The average YPLL for those with and without multimorbidity is subtracted to find the difference in YPLL.

Example (10 subjects with multimorbidity and 8 without)

Multimorbidity				No Multimorbidity			
Subject ID	Age at death	Cut-off	YPLL individual	Subject ID	Age at death	Cut-off	YPLL individual
1	55	65	10	11	59	65	6
2	53	65	12	12	59	65	6
3	45	65	20	13	62	65	3
4	44	65	21	14	62	65	3
5	59	65	6	15	63	65	2
6	61	65	4	16	64	65	1
7	61	65	4	17	58	65	7
8	63	65	2	18	57	65	8
9	49	65	16				
10	39	65	26				
Total			121	Total			36

The example above shows an average of $121/10 = 12.1$ YPLL in subjects with multimorbidity and $36/8 = 4.5$ YPLL in those without; the multimorbidity vs no multimorbidity difference is therefore $12.1 - 4.5 = 7.6$ YPLL.

Methods S3. Global Burden of Disease approach example using UK Biobank

Step 1. Calculate the number of participants and deaths in each age interval (i.e., participants and deaths occurring at age 50, 51, 52.....)

The number of deaths in each age interval was calculated for participants with and without multimorbidity, and by men and women. These numbers can be calculated using any software/calculator.

Step 2. Determine an external standard life expectancy, which is the average number of years that those aged x exact will live. This can be from the Coale and Demeny Model Life Tables West which is a set of standardised life expectancies, or from national registries such as the Office for National Statistics (ONS) in the UK.

In this study, the UK Biobank data was last censored on 31st January 2016 for England and Wales and 30th November 2015 for Scotland. We used the ONS National Life Tables by age and sex in 2013 to 2015 as the external standard life expectancy: this is based on population estimates from births and deaths by dates of registration for a period of 3 consecutive years.

National life tables: UK

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables>

(Accessed 18-01-2022)

Step 3. Substitute the values into the general YLL formula:

The crude expected years of life lost is:

$$YLL = (D_x)(e_x^s)$$

where D_x is the number of deaths and e_x^s is the standard age of death from the external life expectancy.

We could simply use the crude expected YLL to estimate the number of YLL by each age and then find the total (following table).

Example: female subjects with and without multimorbidity

With multimorbidity				Without multimorbidity			
Age at death	No. of deaths (D_x)	Standard life expectancy (e_x^s)	YLL	Age at death	No. of deaths (D_x)	Standard life expectancy (e_x^s)	YLL
50	17	34.2	581.40	50	8	34.20	273.60
51	19	33.28	632.32	51	10	33.28	332.80
52	18	32.35	582.30	52	12	32.35	388.20
53	15	31.43	471.45	53	18	31.43	565.74
54	34	30.52	1037.68	54	19	30.52	579.88
Total	103		3305.15		67		2140.22

The example above shows how to quantify the YLL for the age-group 50-54 years. The average of $3305.15/103 = 32.09$ YLL in female subjects with multimorbidity and $2140.22/67 = 31.94$ YLL in those without; the multimorbidity vs no multimorbidity difference is therefore $32.09 - 31.94 = 0.15$ YLL.

To further extend the basic formula and incorporate the discounting and age weighting requires a range of assumptions.

The following general YLL formula is used:

$$Y_x = d_x \left[\frac{K C e^{r(n^a x)}}{(r + \beta)^2} (e^z [-(r + \beta)(e_x^s + a_x) - 1] - e^{-(r+\beta)a_x} [-(r + \beta)a_x - 1]) + \frac{1 - K}{r} (1 - e^{r(e_x^s)}) \right]$$

where:

a = age of death (years)
 r = discount rate (usually 3%)
 β = age weighting constant (usually $\beta=0.04$)
 K = age-weighting modulation constant (usually $K=1$)
 C = adjustment constant for age-weights (usually $C=0.1658$)
 e = standard life expectancy at age of death (years)

The discount rate (r) is based on the concept that future mortality outcomes have less value than the present mortality outcomes. The discounting rate is a continuous and exponential function, therefore the higher the discount rate, the lower the impact is in the future. The usual time for the discounting rate is set at 3% per year. Age weighting is used because it enables the YLL equation to value differently at various ages, i.e. the value of years lived by a young adult could be noted to be worth more than that of a very young or an old individual. So, the parameters were estimated at $C=0.1658$ and $\beta=0.04$ to account for this, which was calibrated and chosen from the 1990 GBD study.^{1,2}

There is an open source programming code available in R that includes this formula. Once the function is loaded, all that is required is a value of the number of death, average age of death, and the standard life expectancy for that age. The individual YLL is estimated.

https://static-content.springer.com/esm/art%3A10.1186%2F1471-2458-8-116/MediaObjects/12889_2007_1086_MOESM3_ESM.pdf (Accessed 18-01-2022).

For an easier alternative to include all ages, an Excel calculation template can also be used from the World Health Organisation (WHO) National tools 2018, where the number of participants, deaths, and the standard life expectancy for that age can be entered into the template: https://www.who.int/quantifying_ehimpacts/publications/en/9241546204chap3.pdf (Accessed 18-01-2022).

The YLL for the UK Biobank example was calculated using the Excel template to get the full data. The YLL started at the age of 45 to 75+ in one year age intervals, for both men and women and with and without multimorbidity.

Step 4. The total YLL is summed, this can be aggregated.

Step 5. To find the average YLL, the total number of YLL is divided by the total number of deaths (in individual age groups if large enough, or in the aggregated age groups, i.e. 50-54 years).

Step 6. The average YLL for those with and without multimorbidity is subtracted to find the difference in YLL.

In our example, the average YLL was grouped into 5 year age group intervals to give an overall YLL, as the estimates for individual ages was too small. This method works best for large national data, as it is difficult to use when there are too few events and participants in the individual age groups.

References

1. Murray CJL, Lopez A D. The global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: Harvard School of Public Health on behalf of the World Health Organization and the World Bank 1996:990.
2. Aragón TJ, Lichtensztajn DY, Katcher BS, et al. Calculating expected years of life lost for assessing local ethnic disparities in causes of premature death. *BMC Public Health* 2008;8:116-16. doi: 10.1186/1471-2458-8-116

Methods S4. Chiang's life tables example using UK Biobank

Step 1. Identify which life table to use.^{1,2} The current life table was used for our example, as the UK Biobank included middle-aged participants, fewer deaths had occurred at younger ages than the older, and since the age structure is known to be affected by historical changes in fertility, mortality, and migration, which is unrelated to the current mortality. Therefore, to avoid historical changes, a hypothetical cohort that is based on the observed probabilities of death at corresponding ages was used to reflect the effect on only mortality.

Step 2. Determine whether the life tables will be unabridged (i.e., a complete) life table or abridged life table. Unabridged life tables display data for every single year of age until the last age category, whilst the abridge exhibit data by 5 or 10 year age intervals.

For our example, abridged life table with 5 year intervals was used starting from 45 years to the age of the oldest participant at follow-up, 81 years, since it has been recognised that large populations are needed to avoid systematic and random variations in mortality when constructing life tables.^{2,3}

Step 3. Determine the current age of participants as the age at last follow-up or the age at death and estimate the number of participants and deaths by age.

The number of deaths in each age interval was calculated for participants with and without multimorbidity.

Step 4. Construct the life table. The ONS provides further detailed information on the construction of National life tables.¹

<https://www.ons.gov.uk/ons/rel/subnational-health4/life-expec-at-birth-age-65/2004-06-to-2008-10/ref-life-table-template.xls> (Accessed 20-10-2021).

The following gives a brief overview of the values for the life table:

x = the starting age of the interval.

n = width of the age interval.

a_n = defined to be 0.5, as we assume the deaths are evenly distributed over the age interval.

m_x = the rate of mortality, defined as the number of deaths at age x (D_x) divided by the total population at that age (P_x).

$$m_x = \frac{D_x}{P_x}$$

q_x = the mortality rate between age x and $(x + 1)$, that is the conditional probability that a person aged x exact will die before reaching age $(x + 1)$, given survival to age x .

$$q_x = \frac{n * m_x}{1 + (1 - a_x) * n * m_x}$$

p_x = the survival rate between age x and $(x + 1)$ based on the conditional probability.

$$p_x = 1 - q_x$$

l_x = this is the number of survivors from a hypothetical cohort of new-born babies on which the life table is based. The initial value is arbitrary, i.e. 1, 100, 1000.

d_x = the number of deaths from a hypothetical cohort between exact age x and $(x + 1)$.

$$d_x = l_x - l_{x+1}$$

L_x = the number of person-years lived by a hypothetical cohort between exact age x and $(x + 1)$.

$$L_x = n * l_{x+1} * (a_x + d_x)$$

T_x = the number of person-years lived by a hypothetical cohort above exact age x , which is a cumulative distribution of the L_x column.

$$T_x = L_x + T_{x+n}$$

e_x = the expected number of years that those aged x exact will live thereafter in this hypothetical cohort.

$$e_x = \frac{T_x}{l_x}$$

By calculating the variance of the conditional probability of death, the standard errors and the 95% confidence intervals (CI) can be estimated.

For these calculations, we used the ONS Excel template:

<https://www.ons.gov.uk/ons/rel/subnational-health4/life-expec-at-birth-age-65/2004-06-to-2008-10/ref-life-table-template.xls> (Accessed 18-01-2022).

The number of participants and the number that have died by age groups and multimorbidity was inserted, and the hypothetical cohort started at 1000 (as shown in the Figure below).

Step 5. To find the difference in YLL between two groups, calculate two life tables separately (stratify) and then subtract the two expected number of years remaining.

The difference in YLL, which is age-group specific for the example, was calculated by subtracting the remaining life expectancy for those without multimorbidity and those with multimorbidity (for example, in the following Figure from the Excel template: $YLL_{45} = 27.54 - 19.99 = 7.55$; $YLL_{75} = 3.50 - 3.50 = 0$).

Excel template of example using Chiang's life tables in those with and without multimorbidity:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	With multimorbidity													
2	Age interv	x	n	a_x	pop	death	M_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
3	45-49	45	5	0.5	2574	82	0.031857	0.1475351	0.852465	1000	148	4631	19986	19.99
4	50-54	50	5	0.5	6945	165	0.023758	0.1121305	0.887870	852	96	4023	15355	18.01
5	55-59	55	5	0.5	10186	318	0.031219	0.1447956	0.855204	757	110	3510	11331	14.97
6	60-64	60	5	0.5	14858	718	0.048324	0.2155768	0.784423	647	140	2888	7821	12.08
7	65-69	65	5	0.5	25258	1,366	0.054082	0.2382032	0.761797	508	121	2236	4933	9.72
8	70-74	70	5	0.5	28087	1,639	0.058354	0.2546257	0.745374	387	98	1688	2697	6.97
9	75-81	75	7	0.5	10697	299	0.027952	0.1782263	0.821774	288	288	1009	1009	3.50
10					98605	4587								
11														
12														
13														
14	Without multimorbidity													
15	Age interv	x	n	a_x	pop	death	M_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
16	45-49	45	5	0.5	24459	192	0.007850	0.0384939	0.961506	1000	38	4904	27544	27.54
17	50-54	50	5	0.5	55623	412	0.007407	0.0363617	0.963638	962	35	4720	22640	23.55
18	55-59	55	5	0.5	61916	715	0.011548	0.0561194	0.943881	927	52	4503	17920	19.34
19	60-64	60	5	0.5	69129	1,239	0.017923	0.0857718	0.914228	875	75	4185	13417	15.34
20	65-69	65	5	0.5	91567	2,134	0.023305	0.1101112	0.889889	800	88	3778	9232	11.55
21	70-74	70	5	0.5	74988	2,239	0.029858	0.1389208	0.861079	711	99	3310	5455	7.67
22	75-81	75	7	0.5	23705	353	0.014891	0.0990758	0.900924	613	613	2144	2144	3.50
23														
24					401387	7284								
25														

References

1. Office for National Statistics. Dataset National life tables: UK 2021 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetable> accessed 18-01-2022.
2. Office for National Statistics. Guide to calculating national life tables: Explanation of the methodology used to create the national life tables 2019 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/methodologies/guidetocalculatingnationallifetables> accessed 18-01-2022.
3. Eayres D, Williams ES. Evaluation of methodologies for small area life expectancy estimation. *Journal of Epidemiology and Community Health* 2004;58(3):243-49. doi: 10.1136/jech.2003.009654
4. Silcocks PBS, Jenner DA, Reza R. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. *Journal of Epidemiology and Community Health* 2001;55(1):38-43. doi: 10.1136/jech.55.1.38

Methods S5. Epi package example using UK Biobank

Step 1. Using the R software, install the Epi package and load the data.

Step 2. Estimate the survival curves for the two groups (multimorbidity and no multimorbidity) to quantify years of life lost (YLL) associated with having multimorbidity.

There are two ways to calculate the survival curves. In the Epi package, the example uses two groups in relation to the presence of diabetes: “well” (i.e., without diabetes) and “diseased” (i.e., with diabetes). It is possible to assume either that the “Well” persons are immune to diabetes, so persons without diabetes will not develop diabetes over time; or to model the transition rates between the “well” and “diseased” status, i.e., including the possibility that people without diabetes may develop diabetes over time. In our example, we will consider the first case (“immune”), as we want to compare subjects with multimorbidity to those without multimorbidity at baseline.

The details on the calculations are reported at 1) <https://bendixcarstensen.com/>

2) <http://bendixcarstensen.com/Epi/Courses/EDEG2017/>

(Accessed 18-01-2022) and are briefly summarised here:

The survival function, for Well (W, no disease) and Disease (D) subjects, is derived from the instantaneous mortality rate (also known as the hazard function, hazard rate, μ) as follows:

$$S_W(a) = \exp\left(-\int_0^a \mu_W(u) du\right)$$

$$S_D(a) = \exp\left(-\int_0^a \mu_D(u) du\right)$$

These are the survival functions from age 0 to age a in the two groups of subjects, with and without disease. We want conditional survival functions, given survival to age A :

$$S_W(a) = \frac{S_W(a)}{S_W(A)}, \quad S_D(a) = \frac{S_D(a)}{S_D(A)}$$

The age-specific mortality rates can be calculated using a Poisson model, after splitting the follow-up time in small intervals of age (“splitLexis” in the code below) and using a spline transformation of age as covariate. Rates are then predicted separately for subjects with and

without multimorbidity, and subsequently used to obtain age-specific survival estimates (survival function) in each group.

The YLL is then calculated by integrating the survival functions with respect to age (t) of those well (no multimorbidity), $S_W(t)$, subtracted by those that have the disease (with multimorbidity), $S_D(t)$.

$$YLL = \int_0^{\infty} S_W(t) - S_D(t) dt$$

R STUDIO

```
# Input data (from Stata in this case)
UKbbdata <- read_dta("data.dta")
View(UKbbdata)

# Convert multimorbidity to a factor
UKbbdata$multimorbid <- factor(UKbbdata$multimorbid, labels=c("without", "with"))

# Lexis object - we can't have the same name as age as the list, so put as age_in (age at study start)
and age_out (age at study end)
mm_Lx <- Lexis(entry = list(age=age_in) ,
              exit = list(age=age_out),
              exit.status = factor(dead, labels=c("Alive","Dead")),
              data = UKbbdata)

# Split in small age intervals (1 year, from 38 to 81 years)
mm_Sc <- splitLexis(mm_Lx, breaks=seq(38,81,1))
summary(mm_sSc)

# Graphical description of Lexis object
boxes(mm_Sc, boxpos=TRUE, show.BE=TRUE, scale.R=100)

## Stratified model
# Knots for splines
a.kn <- 2:9*10

# Mortality among those without multimorbidity - Poisson model with natural spline (Ns)
mW <- glm( lex.Xst=="Dead" ~ Ns(age_out, knots=a.kn),
           offset = log(lex.dur),
           family = poisson,
           data = subset(mm_Sc,multimorbid=="without"))

# Mortality among with multimorbidity - Poisson model with natural spline (Ns)
mD <- glm( lex.Xst=="Dead" ~ Ns(age_out, knots=a.kn),
           offset = log(lex.dur),
           family = poisson,
           data = subset(mm_Sc,multimorbid=="with"))

# Models are used to derive fitted rates (in this case, intervals of 1 year)
int <- 1

# Frame to return predicted rates in units of cases per 1 year
nd_mm0 <- data.frame(age_out = seq(45+int, 100, int) - int/2, lex.dur = 1, multimorbid="without")
nd_mm1 <- data.frame(age_out = seq(45+int, 100, int) - int/2, lex.dur = 1, multimorbid="with")

muW <- predict(mW, newdata = nd_mm0, type = "response")
muD <- predict(mD, newdata = nd_mm1, type = "response")

# Estimated remaining life expectancy and YLL using the predicted rates (muW and muD) until 100 years
erl(int, muW, muD, age.in = 45.5, A =c(45.5:100), yll=T, note=F)

# Also possible to calculate using the YLL function
YLL <- yll(int, muW, muD, age.in = 45.5, A =c(45.5:100))
```

Methods S6. Flexible Royston-Parmar parametric survival model example using UK Biobank

Step 1. Set up the data using *stset*: age is used as the time scale.

Step 2. Use the *stpm2* command to model survival data. Detailed explanations of the *stpm2* command can be found at: <https://pclambert.net/software/stpm2/> (Accessed 18-01-2022).

Below is a brief explanation of the Royston-Parmar parametric survival model, implemented in the *stpm2* command.

A Cox proportional hazards model is defined as:

$$h_i(t|x_i) = h_0(t)\exp(x_i\beta)$$

In the Cox model, we don't estimate the baseline hazard $h_0(t)$ – this cancels out in the partial likelihood – and we only estimate relative effects (i.e. hazard ratios). In the Royston-Parmar model, an additional parameters is required to estimate the baseline. In the case of the log cumulative hazard, the linear predictor is therefore:

$$\ln[H(t|x_i)] = \eta_i(t) = s(\ln(t)|\gamma, k_0) + x_i\beta$$

with $s(\ln(t)|\gamma, k_0)$ being a restricted cubic spline function of log (time). Restricted cubic splines are used because they are forced to be linear before the first knot and after the final knot; this is where there is often less data and standard cubic splines tend to be sensitive to few extreme values.

Step 3. Use the *predictnl* command. This is used for postestimation predictions after fitting the model with *stpm2*. The command uses the delta method to calculate the variance (standard error, 95% confidence interval). The *surv* term (code below) indicates that the survival function is to be estimated for the group without (*at(multimorbid 0)*) and with (*at(multimorbid 1)*) multimorbidity. The *meansurv* is alternative to *surv* to obtain standardized (average) survival curves when including also covariates in the *stpm2* regression. Another command that does the two steps (*stpm2*, *meansurv*) together is *standsurv* (details are reported at: <https://pclambert.net/software/standsurv/>; Accessed 18-01-2022).

Step 4. Integrate the survival functions to find the remaining life expectancy for each group.

Step 5. YLL is estimated as the difference in the remaining life expectancy between the two groups.

STATA CODE

```
stset currentag, failure(dead == 1) enter(age)
stpm2 i.multimorbid, df(4) scale(hazard) eform nolog

** At age 45 years (until 100 years)
range tt45 45 100 100
gen t45 = 45 in 1/100

predictnl a45d0 = predict(surv timevar(tt45) at(multimorbid 0)) / predict(surv timevar(t45)
                        at(multimorbid 0)), ci(lci_m0 uci_m0) level(95)

predictnl a45d1 = predict(surv timevar(tt45) at(multimorbid 1)) / predict(surv timevar(t45)
                        at(multimorbid 1)), ci(lci_m1 uci_m1) level(95)

integ a45d0 tt45          // remaining life expectancy
integ lci_m0 tt45
integ uci_m0 tt45

integ a45d1 tt45          // subtract this to previous estimate to obtain the YLL at 45 years
integ lci_m1 tt45
integ uci_m1 tt45

** Or to estimate YLL and 95% CI directly
predictnl yll = [predict(surv timevar(tt45) at(multimorbid 0)) / predict(surv timevar(t45)
                        at(multimorbid 0))] -
                [predict(surv timevar(tt45) at(multimorbid 1)) / predict(surv timevar(t45)
                        at(multimorbid 1))], ci(lci_yll uci_yll) level(95)

integ yll tt45          // YLL at 45 years
integ lci_yll tt45
integ uci_yll tt45
```


Stata code to estimate years of life lost at different ages using flexible parametric Royston-Parmar survival model

Multimorbidity can be considered as comparing any two groups to calculate the YLL.

```
1  stset currentag, failure(dead == 1) enter(age)
2  stpm2 i.multimorbid , df(4) scale(hazard) eform nolog
3
4  preserve
5  clear
6  tempfile results1
7  save `results1', emptyok replace
8  restore
9
10 foreach i of num 45(1)100 {
11     preserve
12     gen t`i' = `i' in 1/100
13     range tt`i' `i' 100 100
14     predictnl mm`i' = [predict( surv timevar(tt`i') at( multimorbid 0) ) / predict( surv timevar(t`i') at(multimorbid 0))] ///
15                     - [predict( surv timevar(tt`i') at( multimorbid 1) ) / predict( surv timevar(t`i') at(multimorbid 1) ) ], ///
16                     ci(lci`i' uci`i') level(95)
17     integ mm`i' tt`i'
18     scalar area_mm`i' = r(integral)
19     integ lci`i' tt`i'
20     scalar area_lci`i' = r(integral)
21     integ uci`i' tt`i'
22     scalar area_uci`i' = r(integral)
23     clear
24     set obs 1
25     gen cond = `i'
26     gen yll_mm = area_mm`i'
27     gen yll_lci = area_lci`i'
28     gen yll_uci = area_uci`i'
29     append using `results1'
30     save `results1', replace
31     restore
32 }
33 use `results1', clear
34 sort cond
35 line yll_mm yll_lci yll_uci cond
```

1. Set up the time-to-event data, using *stset* and age as the time scale
2. *stpm2* regression model (option *tvc* can be added to allow a time-varying effect of multimorbidity)
3. Create an empty file to save the YLL results
4. Create the conditional age (45-100) and the time variable for the survival curve as a loop for each age
5. Use *predictnl* and *surv* to estimate the survival curves for group (multimorbidity) 0 and group 1
6. Integrate the area between the two survival curve to find YLL

The code can be further adjusted to:

- Different age intervals and upper limit, i.e. 5 year intervals until 95 years using *45(5)95*
- Account for confounders, where confounders can be added to the *sptm2* regression model and *meansurv* is used in the *predictnl* command instead of *surv* for standardised estimates
- Have more than two groups
- Can include a *timer* to determine how long the model takes to run

Checklist S1. Strengthening the reporting of observational studies in epidemiology (STROBE)

	Item No	Recommendation	Manuscript
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	Title and Abstract
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction paragraph 1-2
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction paragraph 4
Methods			
Study design	4	Present key elements of study design early in the paper	Methods
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods, Example from multimorbidity
Participants	6	(a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Methods, Example from multimorbidity
		(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Methods, Example from multimorbidity
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods, Example from multimorbidity
Bias	9	Describe any efforts to address potential sources of bias	Methods, Example from multimorbidity
Study size	10	Explain how the study size was arrived at	Supplementary Figure S1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Methods, Example from multimorbidity
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	Methods, Example from multimorbidity
		(b) Describe any methods used to examine subgroups and interactions	NA
		(c) Explain how missing data were addressed	Methods, Example from multimorbidity
		(d) Cohort study—If applicable, explain how loss to follow-up was addressed	NA
		(e) Describe any sensitivity analyses	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Methods, Example from multimorbidity
		(b) Give reasons for non-participation at each stage	Methods, Example from multimorbidity
		(c) Consider use of a flow diagram	Supplementary, Figure S1
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	Results, Baseline Characteristics
		(b) Indicate number of participants with missing data for each variable of interest	Supplementary, Figure S1
		(c) Cohort study—Summarise follow-up time (e.g., average and total amount)	NA
Outcome data	15*	Cohort study—Report numbers of outcome events or summary measures over time	Results
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Results
		(b) Report category boundaries when continuous variables were categorized	Results
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Results
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	Discussion paragraph 1
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion paragraph 8
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion paragraph 2-9
Generalisability	21	Discuss the generalisability (external validity) of the study results	Recommendations
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	End of the manuscript