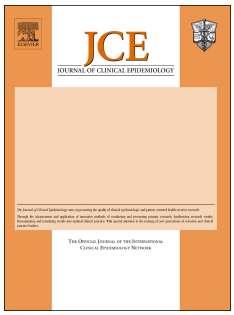


Estimates of years of life lost depended on the method used: tutorial and comparative investigation

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

PII: S0895-4356(22)00163-9

DOI: https://doi.org/10.1016/j.jclinepi.2022.06.012

Reference: JCE 10862

To appear in: *Journal of Clinical Epidemiology*

Received Date: 7 February 2022

Revised Date: 28 April 2022

Accepted Date: 20 June 2022

Please cite this article as: Chudasama YV, Khunti K, Gillies CL, Dhalwani NN, Davies MJ, Yates T, Zaccardi F, Estimates of years of life lost depended on the method used: tutorial and comparative investigation, *Journal of Clinical Epidemiology* (2022), doi: https://doi.org/10.1016/j.jclinepi.2022.06.012.

AUTHOR STATEMENT

Yogini V Chudasama: Conceptualization, Methodology, Formal analysis, Investigation, Methodology, Writing – Original Draft, Project administration

Kamlesh Khunti: Conceptualization, Writing – Review & Editing

Clare L Gillies: Writing – Review & Editing

Nafeesa N Dhalwani: Conceptualization, Data curation, Writing – Review & Editing

Melanie J Davies: Writing – Review & Editing

Thomas Yates: Writing – Review & Editing

Francesco Zaccardi: Supervision, Conceptualization, Methodology, Writing – Review & Editing

Estimates of years of life lost depended on the method used: tutorial and comparative investigation

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 200; Manuscript 4,322; References 48; Tables 1; Figures 2

Appendix: Methods 7; Table 1; Figures 1; STROBE Checklist

Declaration 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.

ABSTRACT

Objective. This review aims to summarise key methods for estimating years of life lost (YLL), highlighting their differences and how they can be implemented in current software, and applies them in a real-world example.

Study Design and Setting. We investigated the common YLL methods: (1) Years of potential life lost (YPLL); (2) Global Burden of Disease (GBD) approach; (3) Life tables; (4) Poisson regression; and (5) Flexible parametric Royston-Parmar regression. We used data from UK Biobank and multimorbidity as our example.

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 other methods both the remaining life expectancy within each group and the YLL could be estimated. At 65 years, the YLL in those with vs without multimorbidity was 1.8, 1.2, and 2.7 years using the life tables approach and the Poisson, and Royston-Parmar regression, respectively; corresponding values were -0.73 and -0.05 years for YPLL and using the GBD approach.

Conclusion. While deciding among different methods to estimate YLL, researchers should consider the purpose of the research, the type of available data, and the flexibility of the model.

Key word. Life expectancy, years of life lost, methods, statistics, epidemiology, multimorbidity.

Running title. Estimating life expectancy and years of life lost

HIGHTLIGHTS

- Years of life lost (YLL) is easier to understand compared to traditional estimates from survival analysis, but very few studies report it.
- A range of methods for 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 among 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
FP	Flexible parametric
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] Yet, the majority of studies using time-to-event analysis still presents 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,[2, 3] and requires fewer assumptions compared to some methods for estimating YLL. However, HRs can be difficult to understand and frequently misinterpreted;[4] furthermore, large HRs may translate into negligible absolute differences.[2] Presenting a variety of measures (relative and absolute) provides therefore a much wider perspective in terms of public health relevance.[5]

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);[1] (iii) the population used for the estimates (based only on individuals who have died, or on extrapolation beyond the last observed follow-up as most studies do not follow all the patients to the end of life).[10] To our knowledge, there has not been an overall review summarising the most common YLL methods, their advantages and disadvantages, and available statistical software.

In this context, our aim is to provide a tutorial style review summarising YLL methods and their differences, and to apply them to a clinically-relevant example of multimorbidity (i.e. the presence of two or more chronic conditions) using the UK Biobank.

YEARS OF LIFE LOST: DEFINITIONS and METHODS

In a systematic search, we identified five common YLL methods (**Appendix Methods S1**). The modelling methods share similar concepts but the theoretical properties differ. We give a brief summary of each method before comparing them using our example. Although these methods allow to quantify the difference in the years of life in relation to the presence of a condition (i.e., cancer, diabetes, multimorbidity), we note that the underlying estimands are different, thus we do not *a-priori* expect consistent findings across all the explored methods.

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 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. The benefit of using the GBD approach is that the age-adjusted rates allow comparisons of causes of death 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 in each age group for 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. 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 frequently used is the life tables approach based on methods from Chiang,[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] Life tables method usually groups the data by five year age intervals, yet this depends 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 work well in large national studies and have been applied in YLL estimations.

Software

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

4. Rates and survival using Poisson regression

By linking a demographic and survival analysis approach within a parametric modelling framework, YLL can be estimated using a Poisson regression approach. 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.[27] 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] and it is straightforward to use.

Software

The Epi package has the functions to calculate estimated residual life (*erl*) and the YLL *(yll*) using Poisson regression; 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.[28] This method has also been used in Stata.[29]

5. Survival modelling using flexible parametric Royston-Parmar regression

The flexible parametric (FP) Royston-Parmar survival model allows for greater flexibility to accurately capture data and extrapolate for future predictions.[30] This is done by using

restricted cubic splines thus obtaining smooth parametric survival curve.[30] It was first introduced by Royston and Parmar in 2002 as an alternative to the Cox proportional model and implemented in Stata.[30, 31] The most recent user-written Stata command *stpm2* is a quick and efficient method to fit FP Royston-Parmar models.[32] Previous studies have used this method mainly in cancer epidemiology.[33] 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. 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;[34] it enables 95% confidence intervals estimation (with delta method) around YLL difference; and potential confounders can be accounted for (both conditional and average/standardised estimation).[35] 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.[29] As well as having commands available in Stata which we have considered, it can also be carried out in SAS[36] and R.[37]

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, using the UK Biobank data (Application Number 14146). [38] Full details are provided in **Appendix Methods S2**. This study is reported in line with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline (**Appendix Checklist S1**).

Statistical analyses

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 group. For each YLL method described above, we provide detailed steps and example statistical code in **Appendix Methods S3-S7**. For each method, the main analysis stratified the calculations (or regression model) for each exposure category, i.e. using data in those with and without multimorbidity separately. To compare the results of the parametric modelling methods (Poisson and FP Royston-Parmar regression), we used the same number and position of knots (ranging from 3 to 7 knots) and extrapolated the survival curves within the range of observed data (age at last observed follow-up: 81 years) and then to 90 and 100 years. Stata version 17.0, R studio version 1.4, and Microsoft Excel were used for analyses.

RESULTS

Multimorbidity and all-cause mortality

499,992 participants were included (white ethnicity, 94.1%; women, 54.5%; with multimorbidity, 19.7%; **Appendix Methods S2**). 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 (**Appendix Figure S1**). 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 are summarised in **Table 1**. 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 these values, the difference between the two groups could be estimated. Conversely, for the life tables, the Poisson regression, and the FP Royston-Parmar regression, both the remaining life expectancy within each group and the YLL could be estimated (**Figure 1**).

The results for the YPLL using a selected cut-off age of 65 years found a total of 25,504 potential years of life lost. The average YPLL lost for a subject with multimorbidity was 6.15 (standard deviation (SD), 4.81) years and for a subject without 6.88 (SD 4.95), giving a difference of -0.73 years (with vs without multimorbidity). 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 (**Table 1**). 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 of those with multimorbidity was over 60

years in 52% compared to 35% in those without; 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.

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 -0.02 years. For participants who died between 65-69 years, corresponding figures were 13.80, 13.85, and -0.05 years.

Using 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 7.55 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.

For the two parametric modelling methods, when extrapolated within the range of observed data, we found similar remaining life expectancy and YLL estimates; however, when extrapolated to 90 or 100 years, the results slightly differed between the two methods, at the age of 45 years and slightly more at the age of 65 years (**Appendix Table S1**). Using the Poisson regression, extrapolating to 100 years the results at the age of 45 years showed that the remaining life expectancy in subjects with multimorbidity was 47.88 years and in those without 51.44 years (**Appendix Table S1**), resulting in YLL of 3.56 years (**Table 1** and **Figure 1**). At the age of 65 years, corresponding estimates were 31.49, 32.71 and 1.22 years. Using the FP Royston-Parmar regression, at the age of 45 years participants with multimorbidity had 40.45 (95% CI 39.37, 41.53) remaining years of life, while those without 44.18 (43.40, 44.95) years (**Appendix Table S1**), resulting in 3.73 YLL (**Table 1** and **Figure 1**). Corresponding figures at the age of 65 years were: 22.32 (21.24, 23.40), 25.03 (24.25, 25.81) and 2.71 years.

Using the proportional hazards model (instead of a stratified model which allows non-proportionality of hazards), 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 (**Table 1** and **Figure 1**). 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.

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”.

A graphical overview of the YLL methods described in our manuscript is shown in **Figure 2**. The simplest method is the YPLL, which is especially useful in a 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 the life tables approach showed a higher number of YLL only at the youngest age – at 45-49 years, YLL was 7.55 years, compared to the Poisson regression

approach (3.56 years) or the Royston-Parmar model (3.73 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, 39] 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 life tables method will vary.

The Poisson regression approach 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 our example showed higher remaining life expectancies than other methods; of note, the smaller the age used for the extrapolation, the smaller the remaining life expectancy (i.e., the area under the curve), as demonstrated in **Appendix Table S1**. While noting that not only the life expectancies within each group but also the YLL comparing the two groups were larger for longer extrapolation, in our database the impact of the number/position of knots was negligible (regardless of the extrapolated age or the conditional age used in the models). An advantage of this method in its implementation in the “Epi” package in R 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 last method we considered is based on the stratified FP Royston-Parmar model, which enables to estimate the survival functions (and the difference among them) and their uncertainties. Similar to the Poisson regression, the results of this fully parametric approach may reflect decisions around the maximum age used for the survival prediction (within the range of observed data or extrapolated beyond); the number/position of the knots for the cubic spline; or interactions between the covariate of interest and age (non-proportional hazard in a single model instead of stratified by presence/absence of multimorbidity). In our example, also for the Royston-Parmar models the impact of the number/position of knots was negligible (across the extrapolated and conditional age) and the values of YLL were greater for longer extrapolations; notably, YLL estimates at 45 years were similar between the two parametric models but, at 65 years, values were greater in the Royston-Parmar model. Overall, these results indicate that the age used for extrapolation has a relevant effect on the differences in YLL estimates within and between the two models and, in turn, between the methods explored in our review. Besides conditional estimates, i.e. for specific values of the covariates included in the model, recent further software development of the Royston-Parmar model also enables marginal (standardised) average effect estimates, including YLL.[35] Lastly, running the full code in our example was time consuming.

Parametric modelling approaches have advantages in that they allow for greater flexibility for both individual-level and grouped data.[40] 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 from Cox PH model, 2). Using the PH Royston-Parmar model, the YLL at the age of 45 years in participants with multimorbidity was on average 5.2 (95% CI 4.8, 5.6) 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.[2, 3] 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.[41] On the other hand, presenting the absolute risk in terms of `years of your life' may be far more influential from a public heath perspective and policy makers, as the concept is easier to grasp. 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).[42] It should be also noted, however, that measures of life expectancy/years of life lost should be complemented by metrics which are not based exclusively on the mortality of a population/cohort, such as years lived without disability – which are relevant for both the individual and the healthcare system (i.e., costs associated with multimorbidity).

This review has some limitations. We focused only on commonly used YLL methods; others (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.[43, 44] 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. 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 illustration of different YLL methods with available statistical codes. We applied the methods

to an emerging clinical 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.[45] 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,[46] or lifestyle modifications.[47, 48] 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.[48]

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 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. Furthermore, researcher should choose the YLL method in view of the purpose of the research, the type of available data, and the required flexibility when a modelling strategy is considered.

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 statement:
Yogini V Chudasama: Conceptualization, Methodology, Formal analysis, Investigation, Methodology, Writing – Original Draft, Project administration
Kamlesh Khunti: Conceptualization, Writing – Review & Editing
Clare L Gillies: Writing – Review & Editing
Nafeesa N Dhalwani: Conceptualization, Data curation, Writing – Review & Editing
Melanie J Davies: Writing – Review & Editing
Thomas Yates: Writing – Review & Editing
Francesco Zaccardi: Supervision, Conceptualization, Methodology, Writing – Review & Editing

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 Reference14614.

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.

REFERENCES

[1] Andersen PK. Life years lost among patients with a given disease. Statistics in Medicine. 2017.

[2] Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. International journal of epidemiology. 2016;45:2075-88.

[3] Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals. BMJ. 2006.

[4] Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Ann Epidemiol. 2018;28:54-7.

[5] Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS medicine. 2007;4:e297.

[6] Gardner JW, Sanborn JS. Years of Potential Life Lost (YPLL)—What Does it Measure? Epidemiology. 1990;1:322-9.

[7] Aragón TJ, Lichtensztajn DY, Katcher BS, Reiter R, Katz MH. Calculating expected years of life lost for assessing local ethnic disparities in causes of premature death. BMC Public Health. 2008;8:116-.

[8] Chiang CL. The life table and its applications: R.E. Krieger Publishing Company; 1984.

[9] Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K, Register SGotND. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. Diabetologia. 2008;51:2187-96.

[10] Andersson TML, Dickman PW, Eloranta S, Lambe M, Lambert PC. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. Statistics in medicine. 2013;32:5286-300.

[11] Martinez R, Soliz P, Caixeta R, Ordunez P. 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;48:1367-76.

[12] Andersen PK. Decomposition of number of life years lost according to causes of death. Statistics in Medicine. 2013;32:5278-85.

[13] Bonneux L. How to measure the burden of mortality? Journal of Epidemiology and Community Health. 2002;56:128-31.

[14] Dempsey M. Decline in tuberculosis: the death rate fails to tell the entire story. American review of tuberculosis. 1947;56:157-64.

[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.

[16] Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. Health policy and planning. 2001;16:326-31.

[17] Office for National Statistics. Dataset National life tables: UK. 2021.

[18] World Health Organization. WHO methods and data sources for global burden of disease estimates 2000-2019. 2020.

[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;55:38-43.

[20] Eayres D, Williams ES. Evaluation of methodologies for small area life expectancy estimation. Journal of Epidemiology and Community Health. 2004;58:243-9.

[21] Office for National Statistics. Guide to calculating national life tables: Explanation of the methodology used to create the national life tables. 2019.

[22] Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, Lee MC, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet. 2011;378:1244-53.

[23] Association of Public Health Epidemiologists in Ontario. Life Table Template V1.2. 2006.

[24] RDocumentation. lifetable: Construct lifetables from mortality rates.

[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.

[27] Carstensen B. Analysis of follow-up data using the Lexis functions in Epi. 2021.

[28] Carstensen B. Years of Life Lost - theory and practice. 2017.

[29] Dickman PW, Lambert PC, Eloranta S, Andersson T, Rutherford MJ, Johansson A, et al. Statistical methods for population-based cancer survival analysis Computing notes and exercises. 2013.

[30] Royston P, Lambert PC. Flexible parametric survival analysis using Stata : beyond the Cox model. College Station: Stata Press; 2011.

[31] 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;21:2175-97.

[32] Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. Stata Journal. 2009;9:265.

[33] Syriopoulou E, Bower H, Andersson TM, Lambert PC, Rutherford MJ. 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;117:1419.

[34] 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.

[35] Lambert P. standsurv. 2019.

[36] Dewar R, Khan I. A new SAS macro for flexible parametric survival modeling: applications to clinical trials and surveillance data. Clin Invest. 2015;5:855-66.

[37] Clements M, Liu X, Christoffersen B, Lambert P, Jakobsen LH, Gasparini A, et al. rstpm2: Smooth Survival Models, Including Generalized Survival Models. 2021.

[38] UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource Protocol No: UKBB-PROT-09-06 (Main Phase) 2007.

[39] 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;372:293-9.

[40] Rutherford MJ, Dickman PW, Lambert PC. Comparison of methods for calculating relative survival in population-based studies. Cancer Epidemiology. 2012;36:16-21.

[41] Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. 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.

[42] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344-9.

[43] Overgaard M, Andersen PK, Parner ET. Regression analysis of censored data using pseudo-observations: an update. The Stata Journal. 2015;15:809-21.

[44] 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;21:52.

[45] National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. NICE in development [GID-CGWAVE0704]. Guidance and guidelines NICE; 2016.

[46] Dorresteijn JAN, Kaasenbrood L, Cook NR, van Kruijsdijk RCM, van der Graaf Y, Visseren FLJ, et al. How to translate clinical trial results into gain in healthy life expectancy for individual patients. BMJ. 2016;352:i1548.

[47] Chudasama YV, Khunti K, Gillies CL, Dhalwani NN, Davies MJ, Yates T, et al. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: A longitudinal cohort study. PLoS medicine. 2020;17:e1003332.

[48] Chudasama YV, Khunti KK, Zaccardi F, Rowlands AV, Yates T, Gillies CL, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. BMC Medicine. 2019;17:108.

Table 1. Estimating years of life lost in subjects with multimorbidity compared to those without multimorbidity using different methods

Methods	Age (years)	Average years of life lost (SD)		Difference in average years of life lost
		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	**-0.02**
	65-69	13.80	13.85	**-0.05**
Methods	**Age (years)**	**Remaining life expectancy (95% CI)**		**Years of life lost (95% CI)**
		With multimorbidity	Without multimorbidity	
Life tables #	45-49	19.99 (19.30, 20.67)	27.54 (27.36, 27.73)	**7.55**
	65-69	9.72 (9.60, 9.84)	11.55 (11.49, 11.61)	**1.83**
Poisson regression *	45	47.88	51.44	**3.56**
	65	31.49	32.71	**1.22**
FP Royston-Parmar stratified *	45	40.45 (39.37, 41.53)	44.18 (43.40, 44.95)	**3.73**
	65	22.32 (21.24, 23.40)	25.03 (24.25, 25.81)	**2.71**
FP 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)**

Years of life lost = without multimorbidity – with multimorbidity.

SD=standard deviation; CI=confidence interval; FP=flexible parametric; PH=proportional hazards

Life tables method indicate the years of life lost for the five-year age interval (i.e. 45 to 49 years). Note that these estimates are based on the maximum age of 81 years, which was the last age at follow-up (last age interval 75 to 81 years). Comparisons with Royston-Parmar and Poisson estimates using the same extrapolation until 81 years are reported in Appendix Table S1.

* Poisson and stratified Royston-Parmar models stratified by exposure (multimorbidity), with the same number and position of knots (3 internal knots and two boundaries knots: centile position 0, 25, 50, 75, 100), and predictions extrapolated for age up to 100 years.

^ Proportional hazards Royston-Parmar models with 3 internal knots and two boundaries knots (centile position 0, 25, 50, 75, 100) and predictions extrapolated for age up to 100 years.

FIGURES

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

Legend

PH=proportional hazards; y=years.

Poisson and stratified Royston-Parmar models stratified by exposure (multimorbidity), with the same number and position of knots (3 internal knots and two boundaries knots: centile position 0, 25, 50, 75, 100), and predictions extrapolated for age up to 100 years.

Proportional hazards Royston-Parmar models with 3 internal knots and two boundaries knots (centile position 0, 25, 50, 75, 100) and predictions extrapolated for age up to 100 years.

Marker estimates on the life tables method indicate the years of life lost for the five-year age interval (i.e. 45 to 49 years). Note that these estimates are based on the maximum age of 81 years, which was the last age at follow-up (last age interval 75 to 81 years). Comparisons with Royston-Parmar and Poisson estimates using the same extrapolation until 81 years are reported in Appendix Table S1.

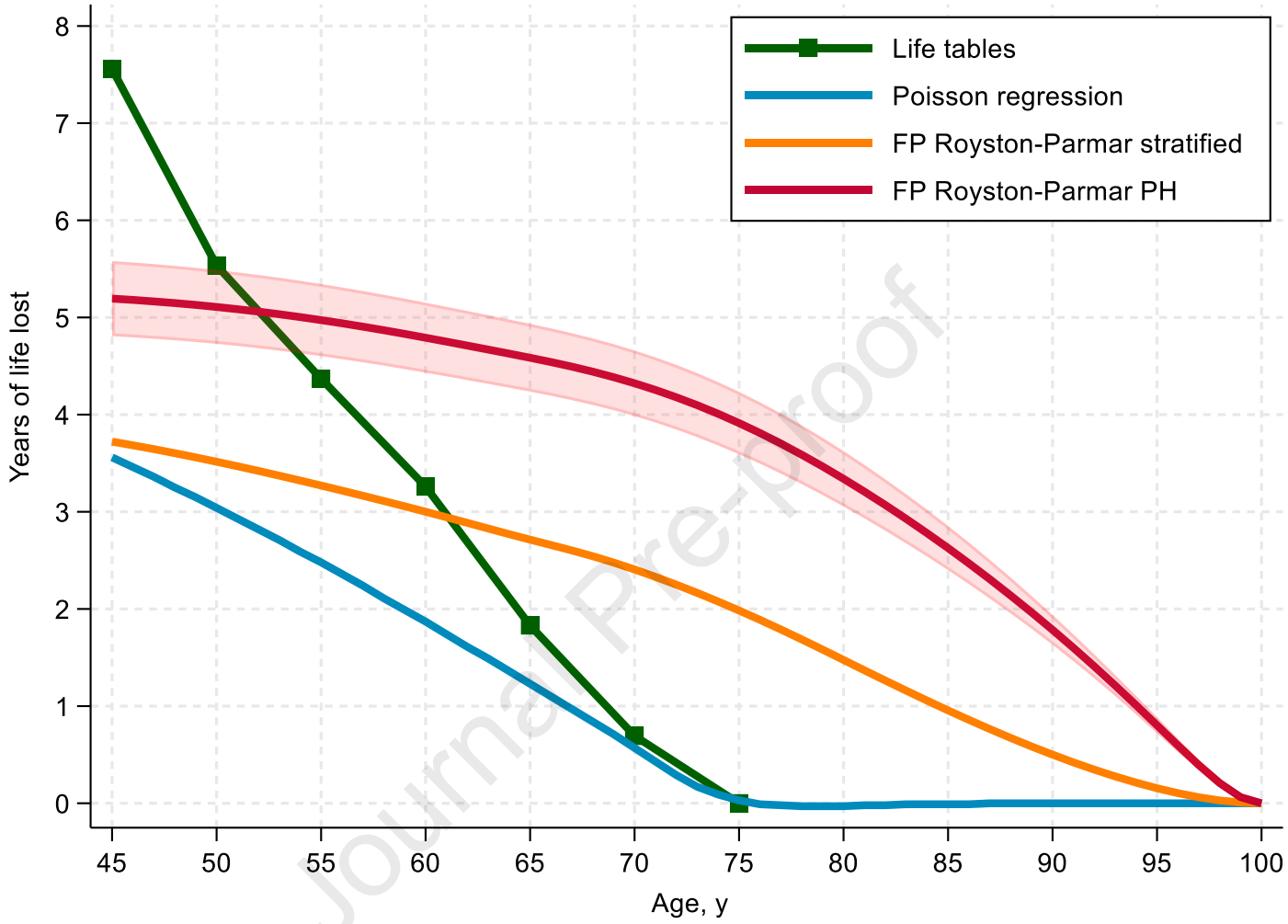
Shaded area indicates 95% confidence interval

Figure 2. Summary of common 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 using different methods



PH=proportional hazards; y=years.

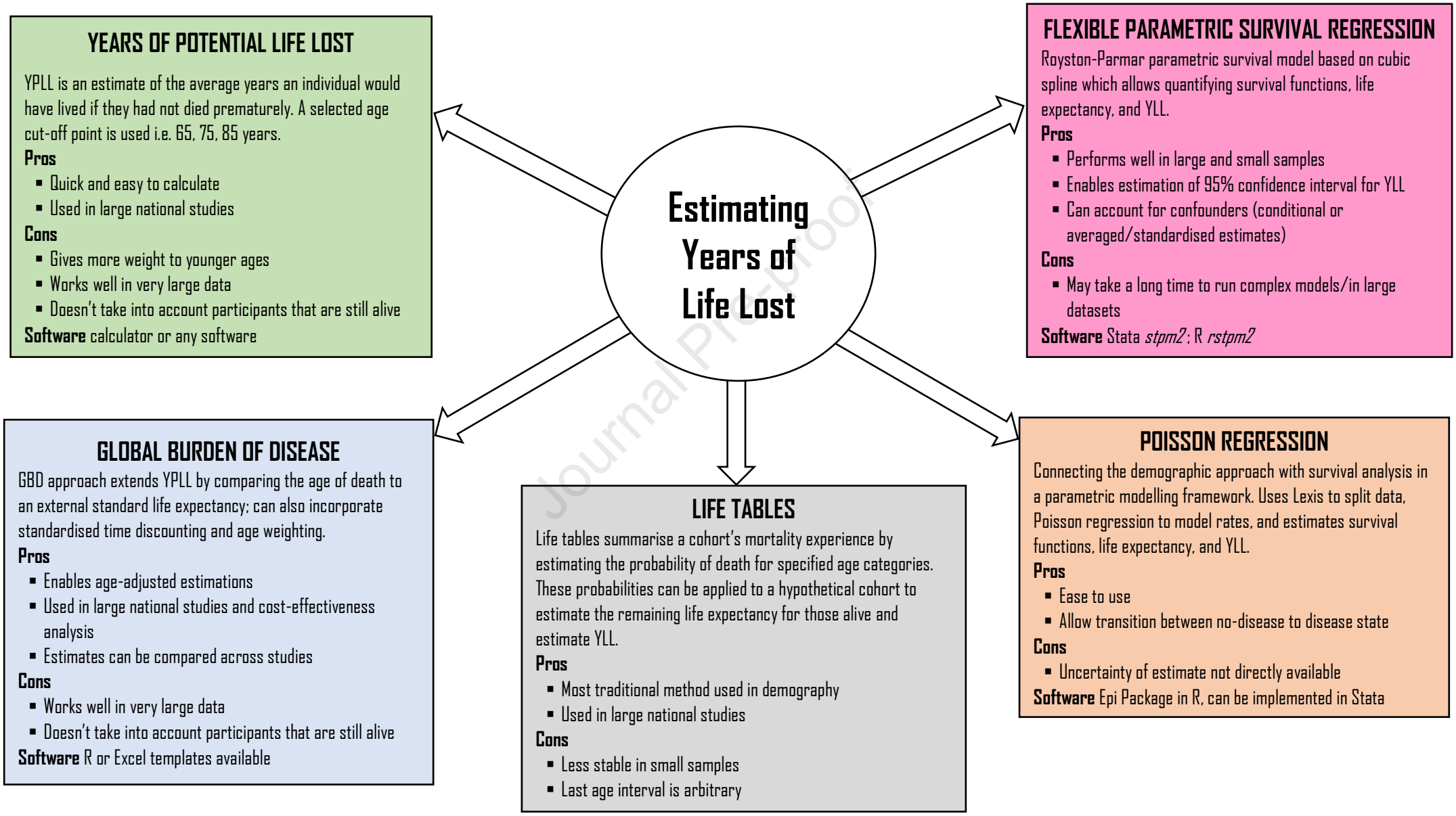
Poisson and stratified Royston-Parmar models stratified by exposure (multimorbidity), with the same number and position of knots (3 internal knots and two boundaries knots: centile position 0, 25, 50, 75, 100), and predictions extrapolated for age up to 100 years.

Proportional hazards Royston-Parmar models with 3 internal knots and two boundaries knots (centile position 0, 25, 50, 75, 100) and predictions extrapolated for age up to 100 years.

Marker estimates on the life tables method indicate the years of life lost for the five-year age interval (i.e. 45 to 49 years). Note that these estimates are based on the maximum age of 81 years, which was the last age at follow-up (last age interval 75 to 81 years). Comparisons with Royston-Parmar and Poisson estimates using the same extrapolation until 81 years are reported in Appendix Table S1.

Shaded area indicates 95% confidence interval.

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



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

What is new?

- Years of life lost (YLL) is easier to understand compared to traditional estimates from survival analysis, but very few studies report it.
- A range of methods for 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 among 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.

Declaration 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.