# Office for National Statistics

# Estimating good health prevalence for use in healthy life expectancy outputs

Application of a new interim modelling approach which estimates the probability of general health status for use in health state life expectancy publications.

## Table of contents

# 1 . Main changes

- Healthy life expectancy (HLE) is a summary measure of population health that measures the average number of years someone is expected to live in good health across their life course, and is calculated by applying age-specific good health prevalence to age-specific person years lived.

- The decline in the sample size of the Annual Population Survey (APS) has limited our ability to continue using the previous method to estimate age-specific proportions of people reporting their health as "good" by local authority.

- We used logistic regression applied to APS microdata to estimate age-specific probabilities of reporting good health by sex and local authorities, instead of using the APS's observed percentage prevalence.

- The method models the odds of reporting good health as a function of age, sex and local authority of residence; age is interacted with sex and region or country of residence, but not local authority, because of low cell sizes.

- Using the model, we estimated probabilities for all combinations of sex, single year of age and local authorities and then calculate the probabilities for the relevant age groupings required for the Sullivan life table; further aggregation up to larger geographical areas such as combined authorities, integrated care boards, regions, constituent countries and a UK measure is also possible.

- We have applied the model to the existing time series, starting in 2011 to 2013, rebasing the time series on the new method.

# 2 . Overview of method change

Our previous method calculated the prevalence of good health by age, sex and local area directly from survey responses, before applying a regression model with a quadratic feature for age and good health prevalence from the census to adjust for potential bias. More information is available in our Proposed method changes to UK health state life expectancies article. In addition, we used imputation based on the proportional difference observed in the census between age groups. Using the APS's observed prevalence, we imputed from the age groups 16 to 19 years and 80 to 84 years to the age groups:

- less than 1 year

- between 1 and 4 years

- between 5 and 9 years

- between 10 and 14 years

- between 85 and 89 years

- over 90 years

The resultant fitted values are used in an abridged Sullivan life table (PDF, 928KB) to calculate healthy life expectancy (HLE).

Since our last Health state life expectancies, UK bulletin, APS sample sizes have declined. This limits our ability to use our previous method to estimate age-specific good health prevalence in a growing number of local areas and leads to instances of implausibly low HLE estimates in some areas. This affects measurement of spatial gaps and volatility over time.

We have developed and implemented an interim methodological solution to enable continuity in local area reporting with a viable, comparable time series. Our new method applies logistic regression modelling to APS microdata in the age range 16 to 84 years, excluding those age groups for which imputation is applied. The use of microdata and single year of age enables us to borrow strength from the larger sample available.

The model estimates the log odds of reporting good health using a combination of parameters entered as main and interaction terms, specifically:

- age

- sex

- local area

- the full interactions between age, sex and region or country of residence.

However, because of low cell sizes, we did not interact age with local area.

The log odds are transformed back to odds facilitating the estimation of predicted probabilities, which can be aggregated to the age groupings, sex, local areas and higher geographical breakdowns required for HLE estimation. The result is a model-based solution, similar to the previous method, but containing fully interacted features.

The modelled estimates replace observed survey prevalence in the imputation step. The remainder of the imputation and smoothing steps from the previous method are unchanged; details can be found in our Health state life expectancies quality and methodology information (QMI) report.

# 3 . Previous method

Since 2012, the Annual Population Survey (APS) has been the primary source of health data for estimating healthy life expectancy (HLE). This source allows classification into states of "good" and "not good" health using the APS question that asks about a person's health in general.

Our APS guidance, which advises using a sample of 25 or above to measure characteristics such as health status, caused us to re-assess the appropriateness of our methodology. For example, in England, the percentage of local areas with a sample base below 25 for males in the influential age group 16 to 19 years grew from 2.7% of local areas in 2011 to 2013 to 36.7% in 2020 to 2022.

Smaller samples increase the scale of random error, leading to instances of extreme outliers (such as an implausibly low or high percentage prevalence of "good" health). The previous method was unable to correct for these errors. Examples of this are illustrated in Figure 1, alongside the more regular distributions observed at national level.

**Figure 1: Highly irregular distributions of female good health prevalence by age group were observed in Barnsley, Merton, Wandsworth and Carmarthenshire between 2020 and 2022**

**Female good health prevalence by age, using the previous method, for selected areas and countries, 2020 to 2022**



Figure 1: Highly irregular distributions of female good health prevalence by age group were observed in Barnsley, Merton, Wandsworth and Carmarthenshire between 2020 and 2022

Female good health prevalence by age, using the previous method, for selected areas and countries, 2020 to 2022
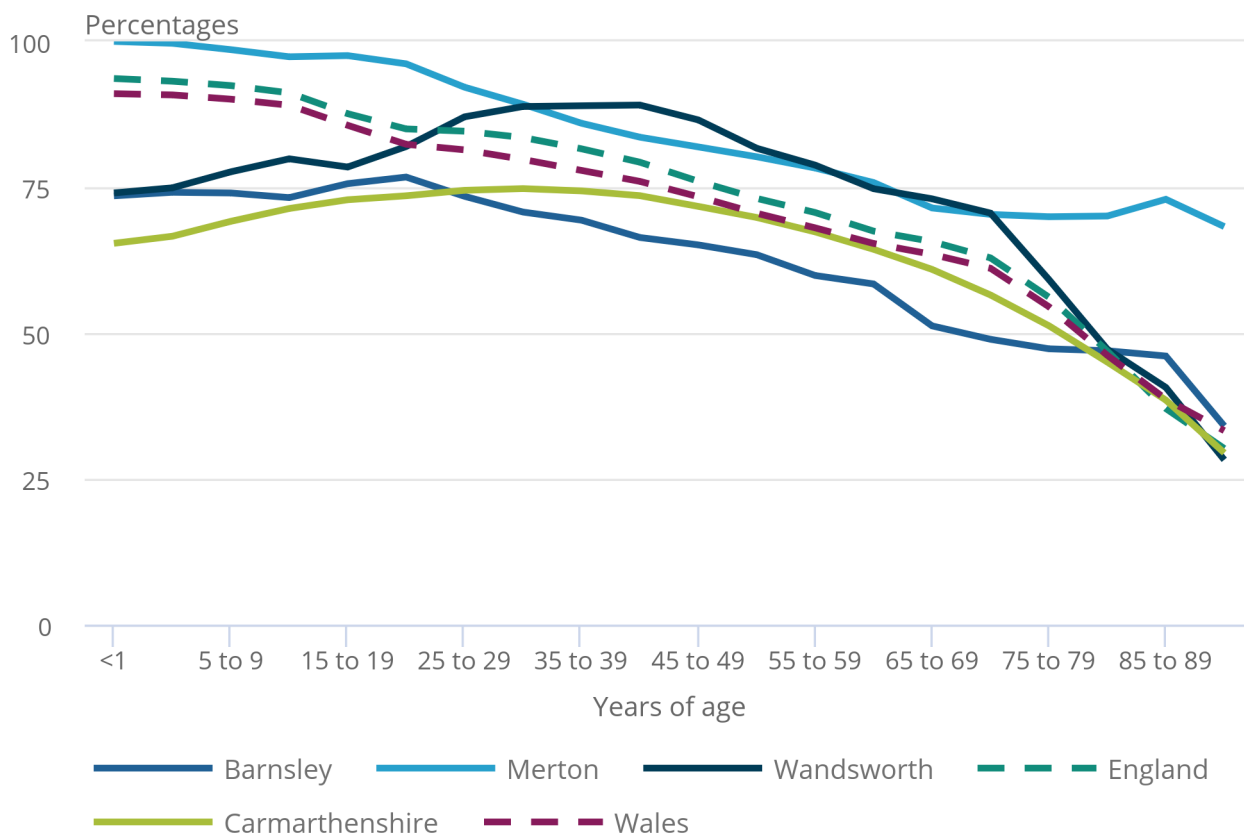
**Source: Office for National Statistics**

For example, good health prevalence in Barnsley and Carmarthenshire is implausibly low among young females. In Merton, it remains very high at most ages, and implausibly high at very old ages. At a national level, the association between age and general health has a familiar distribution with steady declines, which intensify after the age of 75 years.

Estimating HLE using the distributions observed in places such as Barnsley and Merton becomes challenging when using the previous method. For example, there are notably different estimates of female HLE at birth between these areas in the pre-coronavirus (COVID-19) pandemic period (2017 to 2019) and the pandemic period of 2020 to 2022. In the pre-pandemic period, the gap was 2.3 years between these areas; however, in 2020 to 2022 the estimated gap was 19.5 years. This is a consequence of the large contrast in distributions of health seen in Figure 1.

# 4 . Change in estimating good health prevalence

## Model composition

Our new method aims to reduce the impact of increasing sampling error caused by diminishing samples over time by using a statistical model. Our requirement is to estimate an outcome of interest (in our case, good health prevalence) using a set of covariates. The outcome variable and covariates of our model are:

- self-reported health (outcome)

- age

- sex

- local area

- region or country

As health varies by age, sex and local area, including these as covariates in our model was not controversial. However, we also tested whether differences observed in local area health patterns could be represented more validly using interactions. Interactions between covariates allow a covariate's main effect with an outcome to be different depending on the value of another covariate. In our scenario, we want the model to have flexibility in allowing any relationship between age and health to be different depending on its relationship in a larger denomination, such as region or country. The model comprises two categories.

## Main Effects

- Age (single year, between 16 and 84)

- Sex (dummy variable with female coded 1 and male 0)

- Local Area (upper tier local authorities in England, unitary authorities in Wales, local government districts in Northern Ireland, council areas in Scotland)

## Interaction Effects

- Age multiplied by Sex

- Age multiplied by Region or Country

- Sex multiplied by Region or Country

- Age multiplied by Sex multiplied by Region or Country

In this model, England's local area estimates are modified through age and sex interactions with their region membership; for Wales, Northern Ireland and Scotland, the country was used as the higher geographical unit in the interaction terms.

Logistic regression is appropriate for modelling binary choice outcomes for classification. It applies the method of maximum likelihood to determine the values of parameters included in a model. The model we have implemented in our latest [Healthy life expectancy in England and Wales bulletin](#) uses the following equation:

$$LN\left(\frac{P}{1-P}\right) = \hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 sex + \sum_{a=1}^{215}\hat{\beta}_{a+2} local\ area + \hat{\beta}_{218} age*sex + \sum_{b=1}^{11}\hat{\beta}_{b+218} age*region\ (country) +$$

Where:

LN(p/1-p)

is the log odds of good health,

hat0

is the constant and the

hat1.......hatn

are parameters to be estimated.

The exponential function "e" converts log odds back to odds, which then enables the probability of good health to be estimated from the model's parameter values using the following ratio:

Predicted probability of good general health for each age, sex and local area =

$$\frac{e^{(\hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 sex + \sum_{a=1}^{215}\hat{\beta}_{a+2} local\ area + \hat{\beta}_{218} age*sex + \sum_{b=1}^{11}\hat{\beta}_{b+218} age*region(country) + \sum_{c=1}^{11}\hat{\beta}_{c+229} sex*region(country) + \sum_{i=1}^{11}\hat{\beta}_{i+240}\ age}}{1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 sex + \sum_{a=1}^{215}\hat{\beta}_{a+2} local\ area + \hat{\beta}_{218} age*sex + \sum_{b=1}^{11}\hat{\beta}_{b+218} age*region(country) + \sum_{c=1}^{11}\hat{\beta}_{c+229} sex*region(country) + \sum_{i=1}^{11}\hat{\beta}_{i+240}\ a}}$$

Although logistic regression is a generalised linear model, the resultant predicted probabilities are a nonlinear transformation of the log odds, and therefore capable of representing the curvilinear relationship that is found between health status and age. It is an established tool for classification.

We tested eight model variants in our explorations and the details will be made available on GitHub in due course, including annotated R code. Using a combination of model fit statistics and visual inspection of resultant estimated good health prevalence, this model was our preferred choice. The chosen model smooths distributions across the entire range of areas, and is particularly effective in those local areas with notably spurious distributions.

We apply the model to estimate good health prevalence through aggregation using a weighted mean from survey weights to cover the age groups, sex and geographical units required for healthy life expectancy (HLE) estimation. The following age groups are then imputed from these modelled estimates:

- less than 1 year

- between 1 and 4 years

- between 5 and 9 years

- between 10 and 14 years

- between 85 and 89 years

- over 90 years

The imputation method itself and the further smoothing step used in the previous method have not changed. More information can be found in our Health state life expectancies quality and methodology information (QMI) report.

# Results

The method change produces a more plausible distribution of good health prevalence by age across local areas in general. We discuss the effect of the new method on prevalence in Barnsley, Carmarthenshire and Merton in Figures 2 to 4, respectively. The model has very little additional effect on the distributions at national level, when compared with the previous method.

**Figure 2: The new method corrects for the previous method's implausibly low good health prevalence at younger ages in Barnsley**

**Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Barnsley, females, 2020 to 2022**

## Figure 2: The new method corrects for the previous method's implausibly low good health prevalence at younger ages in Barnsley

Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Barnsley, females, 2020 to 2022
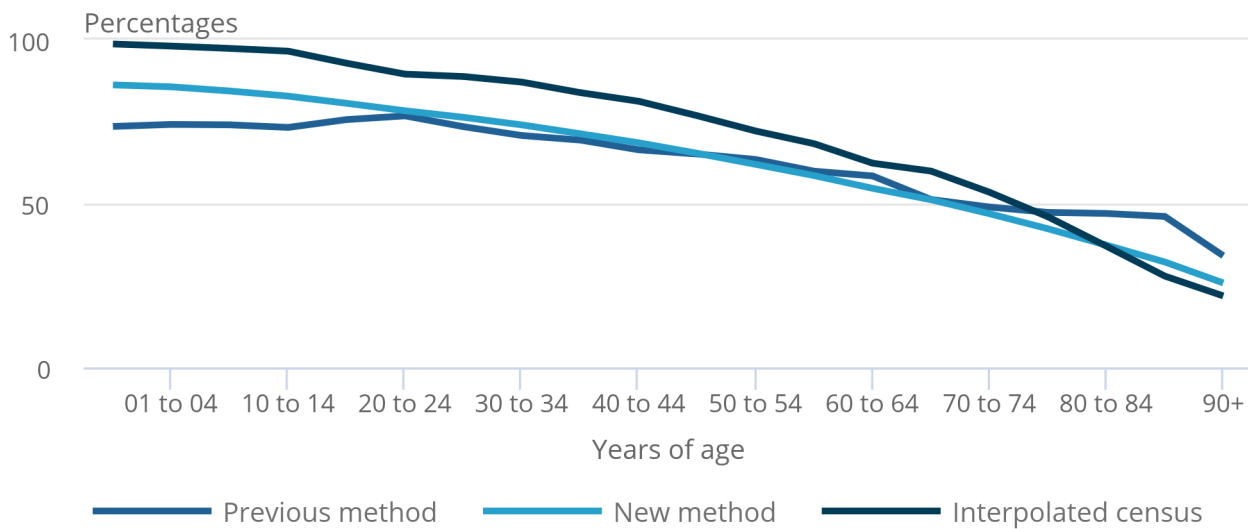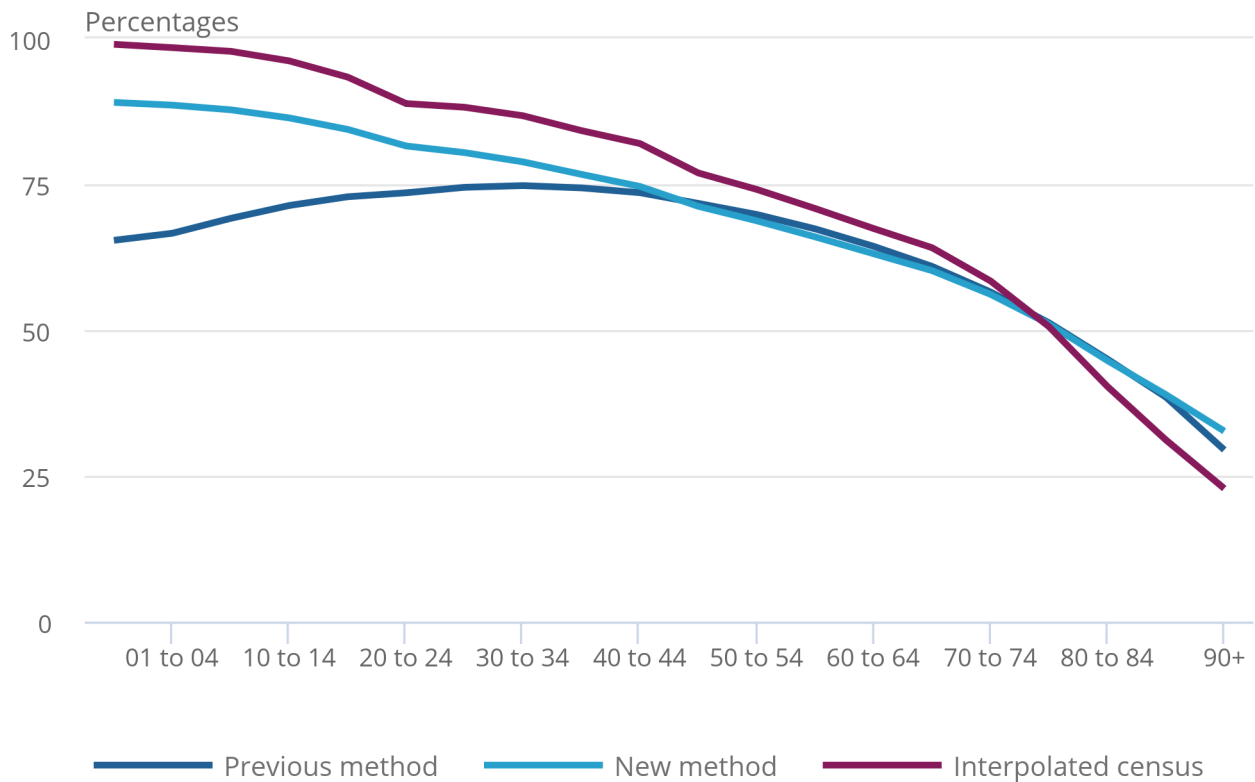


**Source: Office for National Statistics**

**Figure 3: The new method has corrected the parabola shape of the age-health distribution in Carmarthenshire**

**Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Caramarthenshire, females, 2020 to 2022**

## Figure 3: The new method has corrected the parabola shape of the age-health distribution in Carmarthenshire

Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Caramarthenshire, females, 2020 to 2022
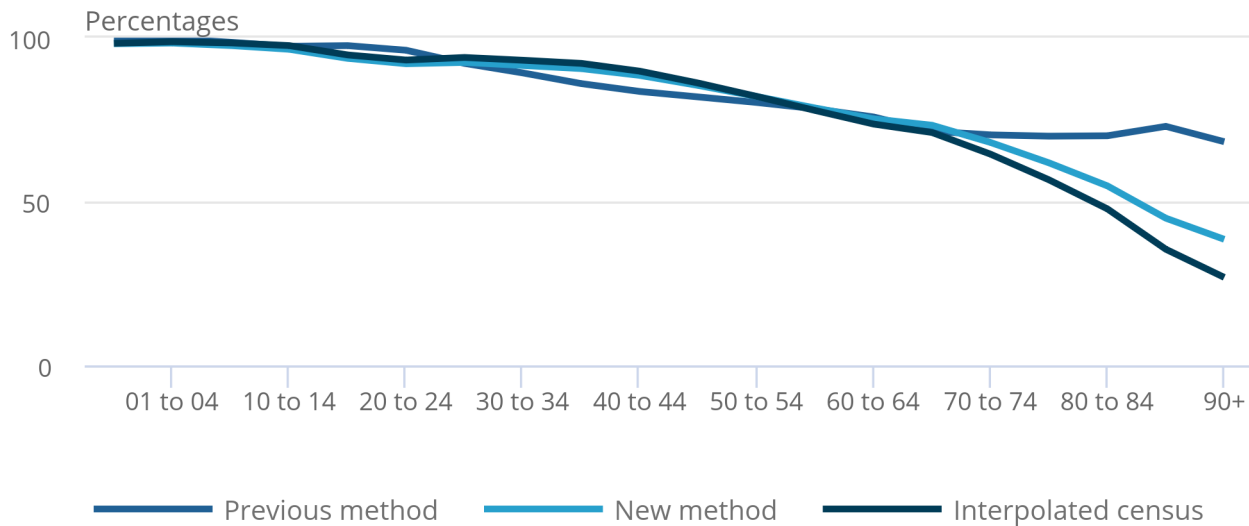
Source: Office for National Statistics

**Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Merton, females, 2020 to 2022**

Figure 4: The new method has corrected the plateauing of good health prevalence at older ages in Merton

Good health prevalence by age using the new method, previous method, and interpolated census prevalence, Merton, females, 2020 to 2022



**Source: Office for National Statistics**

The effect of these changed distributions on female HLE at birth in 2020 to 2022 in these areas, when compared with the previous method, is shown in Figure 5. The new method moderates estimates of HLE in areas where it is very high (Merton) or very low (Barnsley). The areas affected by uncharacteristically very low good health prevalence at young ages (Carmarthenshire and Wandsworth) have seen their HLE increase to a more plausible value across their time series.

In general, the effect of applying the new method is to reduce HLE in those areas with very high estimates and to increase HLE in areas with very low estimates.

**Figure 5: The very low estimate of female healthy life expectancy at birth for Barnsley and very high estimate for Merton are moderated using the new method**

**Female healthy life expectancy at birth, comparing the new method with the previous method for selected areas, 2020 to 2022**



Figure 5: The very low estimate of female healthy life expectancy at birth for Barnsley and very high estimate for Merton are moderated using the new method

Female healthy life expectancy at birth, comparing the new method with the previous method for selected areas, 2020 to 2022
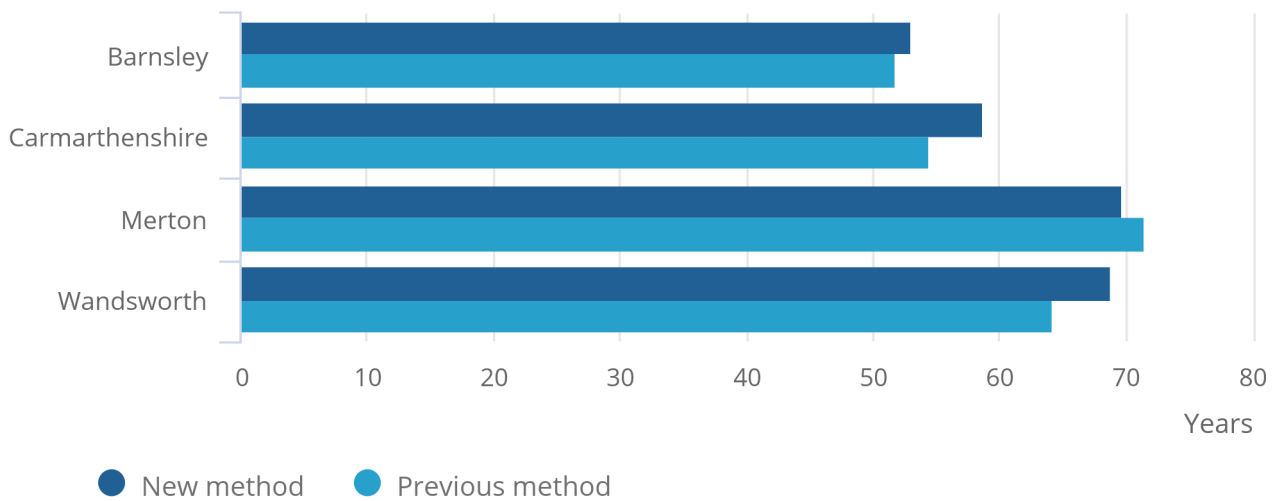
Source: Office for National Statistics

The previous method also caused some instances of extreme volatility in estimates for periods since the APS sample size decline. Figure 6 shows selected area trajectories between 2017 to 2019 and 2020 to 2022 using the previous method and the new method.

Using the previous method resulted in a very high level of volatility. For example, an implausibly large fall of 9.8 years in female HLE at birth in Barnsley, estimated using the previous method, is moderated to 5.3 years using the new method. Much more plausible trajectories are also observed in Merton, Wandsworth and Carmarthenshire, when using the new method.

**Figure 6: The extremely large fall in female healthy life expectancy at birth in Barnsley between 2017 to 2019 and 2020 to 2022 reduces when using the new method**

**Change in female healthy life expectancy at birth between 2017 to 2019 and 2020 to 2022 by method for selected areas**

Figure 6: The extremely large fall in female healthy life expectancy at birth in Barnsley between 2017 to 2019 and 2020 to 2022 reduces when using the new method

Change in female healthy life expectancy at birth between 2017 to 2019 and 2020 to 2022 by method for selected areas
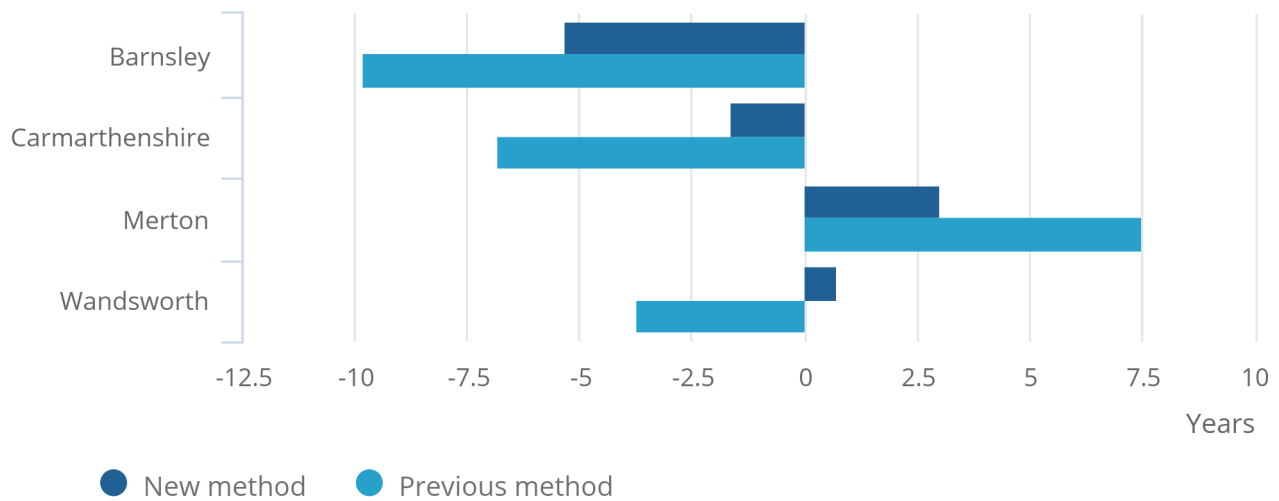


**Source: Office for National Statistics**

The new method retains a plausible rank order of authorities by level of HLE, conforming with an expected alignment to other health outcomes and influences, such as area deprivation. It also closely aligns with the rank order observed using the previous method.

The following sections list the top and bottom local areas in England and Wales for female healthy life expectancy at birth for the 2020 to 2022 period. These lists show how the rankings differ between the new and previous methods for calculating our estimates. A number in brackets denotes an area that was in the top 20 or bottom 20 using the previous method but not using the new method. The number in brackets refers to its rank using the new method. Areas marked with the word "stable" in brackets indicate those that remained in the top or bottom twenty across both the new and old methods.

# New method

## Top twenty areas for female HLE at birth

1. Rutland (stable)

2. Wokingham (stable)

3. Windsor and Maidenhead (stable)

4. Bromley (stable)

5. Kingston upon Thames (stable)

6. Merton (stable)

7. Richmond upon Thames

8. Wandsworth

9. Kensington and Chelsea

10. Buckinghamshire (stable)

11. Oxfordshire (stable)

12. Sutton

13. West Berkshire

14. Hertfordshire

15. Bath and North East Somerset (stable)

16. North Somerset (stable)

17. Hammersmith and Fulham (stable)

18. Cheshire East (stable)

19. North Yorkshire

20. Surrey (stable)

**Bottom twenty areas for female HLE at birth**

154. Sunderland

155. Redcar and Cleveland

156. Stoke-on-Trent

157. Blackburn with Darwen

158. Caerphilly (stable)

159. Middlesbrough (stable)

160. Rotherham

161. Rhondda Cynon Taf (stable)

162. Nottingham (stable)

163. Gateshead

164. Derby (stable)

165. Hartlepool (stable)

166. North East Lincolnshire (stable)

167. Knowsley (stable)

168. Torfaen (stable)

169. Blaenau Gwent (stable)

170. Kingston upon Hull, City of (stable)

171. Barnsley (stable)

172. Blackpool (stable)

173. Merthyr Tydfil (stable)

# Previous method

## Top twenty areas for female HLE at birth

1. Kingston upon Thames (stable)

2. Rutland (stable)

3. Merton (stable)

4. Wokingham (stable)

5. Bromley (stable)

6. Oxfordshire (stable)

7. Ealing (27)

8. Buckinghamshire (stable)

9. Hammersmith and Fulham (stable)

10. Windsor and Maidenhead (stable)

11. Cheshire East (stable)

12. Bath and North East Somerset (stable)

13. Bexley (51)

14. Cheshire West and Chester (32)

15. Shropshire (21)

16. North Somerset (stable)

17. Surrey (stable)

18. Monmouthshire (23)

19. Reading (34)

20. Gwynedd (31)

**Bottom twenty areas for female HLE at birth**

154. Derby (stable)

155. Leicester (150)

156. Kingston upon Hull, City of (stable)

157. Southwark (60)

158. North East Lincolnshire (stable)

159. Stockton-on-Tees (147)

160. Rhondda Cynon Taf (stable)

161. Knowsley (stable)

162. Hartlepool (stable)

163. Merthyr Tydfil (stable)

164. Nottingham (stable)

165. Salford (149)

166. Caerphilly (stable)

167. Manchester (151)

168. Carmarthenshire (131)

169. Middlesbrough (stable)

170. Blaenau Gwent (stable)

171. Torfaen (stable)

172. Barnsley (stable)

173. Blackpool (stable)

Thirteen areas featuring in the top 20 using the new method are also present in the top 20 using the previous method. Fourteen areas featuring in the bottom 20 using the new method are also present in the bottom 20 using the previous method.

An outlier in rank order using the new method, when compared with the previous method, was Southwark, which ranked 157th for HLE at birth using the previous method, compared with 60th using the new method. This arose because of the anomalous age health distribution in observed APS prevalence in 2020 to 2022. The previous method was unable to correct for an implausibly large fall in good health between age groups 35 to 39 years and 16 to 19 years. In 2017 to 2019, the previous method and the new method had a similar distribution in this area, suggesting underlying data issues in 2020 to 2022.
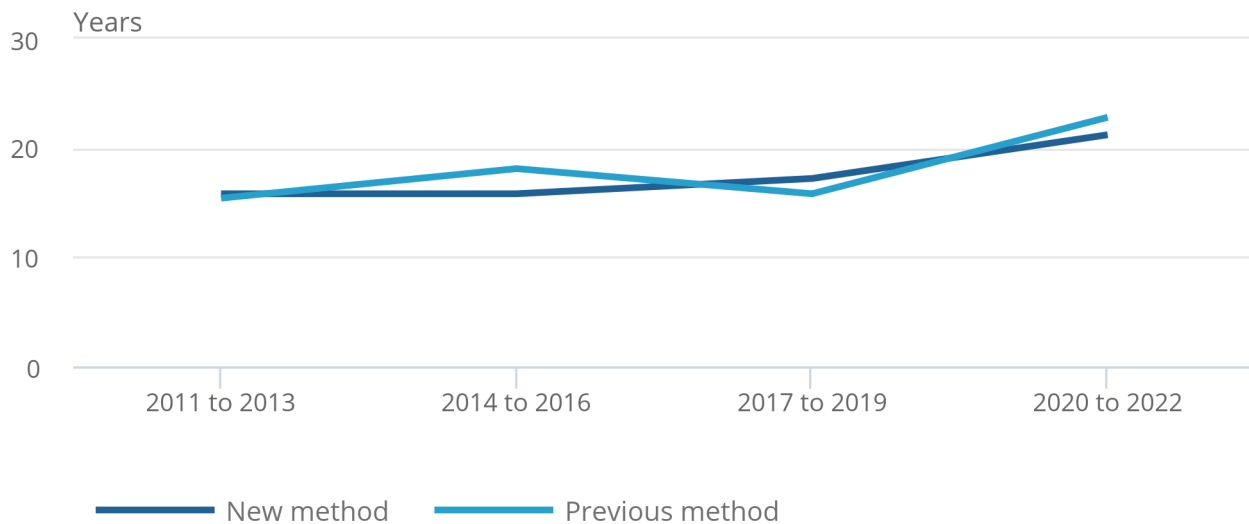
The spatial gap in HLE at birth between areas stood at 22.8 years using the previous method by the 2020 to 2022 period. However, applying the new method produced a gap of 21.2 years (Figure 7).

**Figure 7: The gap in female healthy life expectancy between the highest and lowest ranked areas increases at a slower rate using the new method**

**Trend in the gap in female healthy life expectancy between the highest and lowest ranked areas, comparing the new method with the previous method, selected periods between 2011 to 2013 and 2020 to 2022**

Figure 7: The gap in female healthy life expectancy between the highest and lowest ranked areas increases at a slower rate using the new method

Trend in the gap in female healthy life expectancy between the highest and lowest ranked areas, comparing the new method with the previous method, selected periods between 2011 to 2013 and 2020 to 2022

Years

30

20

10

0

2011 to 2013          2014 to 2016          2017 to 2019          2020 to 2022

— New method          — Previous method

**Source: Office for National Statistics**

The gap has widened sizably in 2020 to 2022 using both the new and previous methods, although the gap is smaller using the new method. The increase in the spatial gap observed during the 2020 to 2022 period is likely a result of:

- declining sample sizes

- differential effects on physical and mental health by locale caused by the coronavirus (COVID-19) pandemic

- change in modes of data capture

- selective non-response

We will monitor this further over future periods to judge how much the factors linked to the pandemic are responsible for the scale of this gap.

## Benefits of the new method

Using the new method to calculate HLE estimates has:

- improved the plausibility of good health prevalence patterns by age across the range of local areas included in our HLE release

- removed extreme outliers and unexpected patterns in prevalence, providing greater confidence in HLE metrics

- achieved a sensible rank order of local areas by level of HLE at birth; the only conflict with the previous method is likely to be a valid correction

- reduced volatility in estimates so they are more reliable for users over time

- narrowed spurious spatial gaps and lessened their volatility

# 5 . Future developments

The Office for National Statistics (ONS) is in the process of reviewing data sources and assessing their potential to provide robust, reliable and durable measures of health and disability status that are relevant for future health state life expectancy statistics (HSLE). The team responsible for delivering HSLE outputs at the ONS will work closely with internal and external partners over the coming months to determine which data sources to test for their potential. These will include data the ONS already publish and have started to review, as well as other sources that may need acquisition. During this process, we aim to harmonise the data used for health status measurement across the UK, by using agreed definitions for national and local reporting. We have strong links with our partner agencies in the constituent countries and will work collaboratively to deliver a data solution for future HSLE statistics.

Our interim method using Annual Population Survey (APS) data will continue alongside the review of data sources, definitions and methods. It will not change until that review has been completed and we have consulted on a proposal for future reporting.

# 6 . Glossary

## Binary logistic regression

Logistic regression is one of several generalised linear models. It can be used to predict a binary outcome, such as self-reports of good or not good health, based on characteristics such as age, sex and area of residence. The linear combination of variables measured using the log scale is then transformed back to odds to enable the model to predict the probability that someone with those characteristic combinations will report "good" or "not good" health.

## General Health

General health measures subjective health-related well-being. "Good health" is defined using the Annual Population Survey (APS) item: "How is your health in general? Is it 'Very Good'; 'Good'; (classified as 'Good' health); 'Fair'; 'Bad'; 'Very Bad'" (classified as 'Not Good' health)".

## Interpolated census prevalence

Age, sex and area-specific good health prevalence for intercensal years was estimated using linear interpolation between Census 2011 and Census 2021.

# 7 . Related links

[Healthy life expectancy in England and Wales: between 2011 to 2013 and 2021 to 2023](#)
Bulletin | Released 12 December 2024
The number of years people are expected to live in good general health in constituent countries and local areas of England and Wales.

# 8 . Cite this article

Office for National Statistics (ONS), released 12 December 2024, ONS website, article, [Estimating good health prevalence for use in healthy life expectancy outputs](#)