

# Coronavirus (COVID-19) related deaths by hearing and vision impairment status, England methodology

Technical appendix to accompany Estimates of coronavirus (COVID-19) related deaths by hearing and vision impairment status in England for deaths occurring up to 20 July 2022.

Contact:  
Perrine Machuel, Seth Thomas,  
Josephine Foubert  
health.data@ons.gov.uk  
+44 1633 651602

Release date:  
18 November 2022

Next release:  
To be announced

## Table of contents

1. [Overview](#)
2. [Data sources](#)
3. [Coronavirus \(COVID-19\) related deaths](#)
4. [Hearing and vision impairment definition](#)
5. [Hospital variables](#)
6. [Primary care variables](#)
7. [Vaccination variables](#)
8. [Crude and age-specific death rates](#)
9. [Modelling analysis](#)
10. [Related links](#)
11. [Cite this methodology](#)

# 1 . Overview

This article provides details of the data and methods used in the article [Estimates of Coronavirus \(COVID-19\) related deaths by hearing and vision impairment status, England: 24 January 2020 to 20 July 2022](#).

## 2 . Data sources

These analyses use data from the Office for National Statistics' (ONS) Public Health Data Asset (PHDA). The PHDA is a unique linked dataset that encompasses 2011 Census records, death registrations, [Hospital Episode Statistics \(HES\)](#) and primary care records retrieved from the [General Practice Extraction Service \(GPES\) Data for Pandemic Planning and Research \(GDPPR\)](#). The PHDA covers England only and was created by:

- using deterministic and probabilistic linkages, NHS numbers were obtained for individuals present in the 2011 Census and in the NHS Patient Register (PR) records between 2011 and 2013
- using NHS number, death registrations data were linked to the 2011 Census records
- using NHS number, HES records from April 2017 and GPES records from January 2015 were linked onto the Census deaths linked data

We linked vaccination data from the [National Immunisation Management Service \(NIMS\)](#) to the PHDA based on NHS number to adjust for vaccination status.

The study population comprises 19.3 million respondents to the 2011 Census, that:

- were aged between 30 and 100 years in 2020
- had not died before 24 January 2020
- had at least one hospital record between 1 January 2011 and 23 January 2020 found in the [Hospital Episode Statistics \(HES\) database](#)
- could be linked to the 2011 to 2013 Patient Registers and [GDPPR dataset](#) (which comprises active NHS patients at the start of the coronavirus (COVID-19) pandemic and are unlikely to have emigrated between 2011 and 2020)

The study population is not currently refreshed with immigrations. Some deaths involving COVID-19 will therefore have occurred to immigrants entering the country since 2011.

## 3 . Coronavirus (COVID-19) related deaths

Causes of death were defined using the International Classification of Diseases, 10th Revision (ICD-10). Deaths involving COVID-19 include those with an underlying cause, or any mention, of ICD-10 codes U07.1 (COVID-19, virus identified), U07.2 (COVID-19, virus not identified) or U09.9 (Post-COVID condition, also known as long COVID).

## 4 . Hearing and vision impairment definition

In this publication, we defined our exposure as the presence of a hearing, vision and dual-sensory impairment in people with an electronic hospital record found in the [Hospital Episodes Statistics \(HES\)](#) database.

HES records comprise ICD-10 codes recorded upon diagnosis during a singular hospital episode. ICD-10 codes are medical codes used for the [International Statistical Classification of Diseases and Related Health Problems](#). Specific ICD-10 codes and descriptions were used as proxies to deduce hearing or visual impairments from HES records, in collaboration with medical and academic experts. We assessed ICD-10 codes in both primary and secondary diagnosis fields in HES records taken from 1 January 2011 to 23 January 2020.

The ICD-10 codes and descriptions used for deduction of hearing impairment from HES records were:

- H903 – sensorineural hearing loss, bilateral
- H904 – sensorineural hearing loss, unilateral with unrestricted hearing on the contralateral side
- H905 – sensorineural hearing loss, unspecified
- H906 – mixed conductive and sensorineural hearing loss, bilateral
- H907 – mixed conductive and sensorineural hearing loss, unilateral with unrestricted hearing on the contralateral side
- H908 – mixed conductive and sensorineural hearing loss, unspecified
- H918 – other specified hearing loss
- H919 – hearing loss, unspecified

The ICD-10 codes and descriptions used for deduction of vision impairment from HES records were:

- H541 – severe visual impairment, binocular
- H542 – moderate visual impairment, binocular
- H544 – blindness, monocular
- H545 – severe visual impairment, monocular
- H546 – moderate visual impairment, monocular
- H549 – unspecified visual impairment (binocular)

Since our method of identification uses hospital records to identify presence of the exposure, we defined our comparison group as unexposed individuals with a hospital record over the same period.

We derived an indicator for impairment status composed of four groups:

- people with a hearing impairment – individuals whose hospital record(s) mentioned at least once a presence of a hearing impairment, between 2011 and 2020
- people with a vision impairment – Individuals whose hospital record(s) mentioned at least once a presence of a vision impairment between 2011 and 2020 as proxied by a list of ICD-10 codes
- people with a dual-sensory impairment – individuals whose hospital record(s) mentioned a vision and a hearing impairment between 2011 and 2020 as proxied by a list of ICD-10 codes; these mentions can be found within the same hospital record or via separate records across different years
- comparison group – individuals whose hospital record(s) did not feature any mention of a hearing or vision impairment between 2011 and 2020 as proxied by a list of ICD-10 codes

This approach to defining impairment relies on electronic health records and the diagnosis of a clinician. It differs from the understanding of disability under [the Equality Act \(2010\)](#) and from social model understandings of disability.

This approach yields a selective exposed population and should not be used to provide information on the prevalence of hearing, vision or dual-sensory impairments in the general population. This approach is relevant when recording the association between a health outcome of interest, here COVID-19 related death; and the presence or absence of an exposure, here a hearing, vision or dual-sensory impairment, as identified through hospital records.

We use this definition as a solution to the scarcity of current indicators available to identify individuals with specific impairment types. This solution was developed by the Office for National Statistics (ONS) in 2022. Please find more details in our [Improving disability data to understand the effects of coronavirus \(COVID-19\) on people with different impairment types article from June 2022](#).

For more details about the terminology, strengths and limitations of this approach, please refer to the glossary and strengths and limitations sections of [our main article, Estimates of coronavirus \(COVID-19\) related death by hearing and vision impairment status](#).

## 5 . Hospital variables

This analysis uses additional hospital data to get information on the number and duration of hospital admissions in the study population, in the years before the start of the coronavirus (COVID-19) pandemic.

To get this information, we used [Hospital Episode Statistics \(HES\)](#) data from April 2017 to January 2020 sourced from Admitted Patient Care (APC) records. The information within this dataset is at episode level (each finished period of care under a consultant). We created a person-level dataset from the episode HES data to preserve all information when linking to the 2011 Census and deaths data.

The analytical variables derived from HES records between 2017 and 2020 were:

- the number of first admission episode flags in the APC dataset to derive the number of admissions per person
- the number of days spent in admitted patient care from the APC dataset

These were then aggregated up to the person level by stacking and deduplicating all datasets on the NHS number and date of birth, to create one row per individual. Records with blank or invalid NHS numbers and/or dates of birth were dropped, as these could not be linked to the 2011 Census.

HES data were linked to the 2011 Census and deaths data by NHS number. Of people with at least one HES record between 2017 and 2019, 83.4% could be linked to the 2011 Census and the 2011 to 2013 Patient Register via NHS number. The remaining unlinked 16.6% are likely to have not been registered on the 2011 Census. This could be because they were born after 27 March 2011, migrated to England after that date, or were not counted at the 2011 Census despite being a resident.

In addition, some individuals in the unlinked group may not have been able to have an NHS number assigned to their 2011 Census record. This could be because of conflicting addresses or name changes, and so the deterministic and probabilistic linkage methods would have failed. However, this is only in a small number of cases.

## 6 . Primary care variables

Primary care records were extracted from the [General Practice Extraction Service \(GPES\) Data for Pandemic Planning and Research \(GDPPR\) dataset](#), which contains 55,199 [SNOMED codes, as shown on the NHS website](#) . Of these codes, 28,561 concern dispensary information, prescriptions, and medications, and 20,306 describe diagnoses and findings (including resolved and remission).

The GDPPR dataset was first used to identify individuals in the study population in 2020. Of people with at least one GPES record between 2015 and 2019, 68.8% could be linked to the 2011 Census and 2011 to 2013 Patient Register via NHS number.

Secondly, as with the hospital episode statistics (HES) data, episode data for relevant conditions (listed in this section) were converted to binary (except for body mass index, chronic kidney disease and type 1 and type 2 diabetes), person-level variables by grouping by NHS number.

The GDPPR dataset was used to identify individuals who had primary care contact over 5 years between 1 January 2015 to 31 December 2019 for a range of conditions. These comorbidities were chosen because they were previously implicated in raising the risk of death from coronavirus (COVID-19) by the [QCOVID algorithm for predicting hospital admission and mortality from COVID-19 in adults](#). The list of conditions we adjust for is updated to align with the [update of the COVID-19 risk prediction model known as QCovid2](#) used by the NHS.

Use of certain health variables in the QCOVID algorithm was precluded by either:

- an insufficient number of cases for analysis (bone marrow transplant, congenital heart disease, rare neurological conditions, venous thromboembolism, sickle cell disease and severe combined immunodeficiency)
- a lack of permissions to use these data (chemotherapy or radiotherapy treatment)
- or omission of indicator in the public health data asset (PHDA) (osteoporotic fracture, solid organ transplant) or the requisite clinical codes from the GDPPR (HIV/AIDS and inflammatory bowel disease).

The full list of health variables included is:

- body mass index
- chronic kidney disease (CKD)
- diabetes type 1
- diabetes type 2
- respiratory cancer
- blood cancer
- chronic obstructive pulmonary disease (COPD)
- cerebral palsy
- asthma
- rare pulmonary diseases
- pulmonary hypertension or pulmonary fibrosis
- coronary heart disease
- stroke
- atrial fibrillation
- heart failure
- peripheral vascular disease
- dementia
- Parkinson's disease
- epilepsy
- rheumatoid arthritis or systemic lupus erythematosus
- cirrhosis of the liver
- severe mental illness (schizophrenia or bipolar disorder)

## 7 . Vaccination variables

We used vaccination data from the [National Immunisation Management Service \(NIMS\)](#) for the period 8 December 2020 (the day of the first vaccination in England) to 20 July 2022.

Our analysis of the second wave of the coronavirus (COVID-19) pandemic includes first and second vaccination doses, and includes first, second and third doses for the third wave. The analysis does not differentiate between booster doses and third doses provided for other reasons.

Vaccination status was included in the model as a time-varying covariate, and we considered a person vaccinated once 14 days had passed since the dose was administered. More information can be found in the [UK Health Security Agency's blog post COVID-19: analysing first vaccine effectiveness in the UK](#). Of people aged 30 years and over who received at least one dose of a vaccine, 82.3% were linked to the Office for National Statistics (ONS) public health data asset (PHDA).

## 8 . Crude and age-specific death rates

### Crude death rate

Crude death rates (per 100,000 person-years at-risk) is defined as total deaths per 100,000 person-year, or:

$$\left( \frac{\text{Total deaths}}{\text{Total person - years at risk}} \right) \times 100,000$$

To calculate total person-years at risk, we first divide the number of days each individual spends in the study (from 24 January 2020 to date of death or 20 July 2022) by 365.25 to convert to years, then sum this quantity across all individuals in the study population.

### Age-specific death rates

Age-specific death rates may be calculated for each age group. These are defined as the number of deaths in the age group per 100,000 person-years at-risk in the same age group or:

$$M_k = \left( \frac{d_k}{p_k} \right) \times 100,000$$

Where:

M = age-specific death rate for age group k

d = deaths in age group k

p = person-years at risk in age group k

= age

Age-specific rates may be calculated separately for males and females or for both sexes combined.

## Confidence intervals

To help assess the variability in the rates, they have been presented alongside 95% [confidence intervals](#).

The choice of the method used in calculating confidence intervals for rates will, in part, depend on the assumptions made about the distribution of the deaths data on which these rates are based.

Traditionally, a normal approximation method has been used to calculate confidence intervals on the assumption that deaths are normally distributed. However, if the number of deaths is relatively small (fewer than 100), it may be assumed to follow a Poisson probability distribution. In such cases, it is more appropriate to use the confidence limit factors from a Poisson distribution table to calculate the confidence intervals instead of a normal approximation method.

The method used in calculating confidence intervals for rates based on fewer than 100 deaths was proposed by [Dobson and others in Confidence intervals for weighted sums of poisson parameters \(1991\)](#). This is described in [the Association of Public Health Observatories' third technical briefing \(2008\) \(PDF, 2.088KB\)](#).

In this method, confidence intervals are obtained by scaling and shifting (weighting) the exact interval for the Poisson distributed counts (number of deaths in each year). The weight used is the ratio of the standard error of the crude rate to the standard error of the number of deaths.

The lower and upper 95% confidence intervals are denoted as crude rate lower and crude rate upper, respectively, and calculated as:

$$Crude\ rate_{lower} = Crude\ rate + (D_i - D) \cdot \sqrt{\frac{v(Crude\ rate)}{v(D)}}$$

$$Crude\ rate_{upper} = Crude\ rate + (D_u - D) \cdot \sqrt{\frac{v(Crude\ rate)}{v(D)}}$$

where:

- $D_i$  and  $D_u$  are the exact lower and upper confidence limits for the number of deaths, calculated using confidence limit factors from a Poisson probability distribution table
- $D$  is the number of deaths in each year
- $v(Crude\ rate)$  is the variance of the crude rate
- $v(D)$  is the variance of the number of deaths

Where there are 100 or more deaths in a year, the 95% confidence intervals for crude rates are calculated using the normal approximation method:

where:

$$Crude\ rate_{LL/UL} = Crude\ rate \pm 1.96 \times SE$$

$Crude\ rate_{LL/UL}$  represents the upper and lower 95% confidence limits, respectively, for the crude rate and SE is the standard error of the crude rate.



## 9 . Modelling analysis

We use Cox proportional hazard models to assess how the risk of death involving coronavirus (COVID-19) varies by hearing, vision and dual-sensory impairment status. This is done after adjusting for residence type (private household, care home, or other communal establishment) and a range of other characteristics. These characteristics include, location, measures of disadvantage, occupation, living arrangements, pre-coronavirus (COVID-19) pandemic health status and vaccination status.

Most individual characteristics were retrieved from the 2011 Census. This was except for hospital admissions, pre-existing health conditions and vaccination status, which were derived from pre-pandemic [Hospital Episode Statistics \(HES\)](#) records from April 2017 onwards, [General Practice Extraction Service \(GPES\) Data for Pandemic Planning and Research \(GDPPR\)](#) from 1 January 2015 to 31 December 2019, and the [National Immunisation Management System \(NIMS\)](#), respectively.

We model the hazard of death involving COVID-19 between 24 January 2020 and 20 July 2022. In our analytical dataset, we include all the people in our study population who died of any cause during this period. We then include all the exposed population (that is, people with a hearing, vision or dual-sensory impairment) who did not die. Finally, we include a 1% weighted random sample of the unexposed population who did not die.

The hazard function was modelled as follows:

$$h(t) = h_0(t) \times \exp(b_1 x_1 + b_2 x_2 \dots + b_i x_i)$$

where:

- $t$  is the survival time
- $h(t)$  is the hazard function at time  $t$
- $h_0(t)$  is the baseline hazard at time  $t$
- $b_i$  is the estimated coefficient for the  $i_{th}$  covariate
- $x_i$  is the value for the  $i_{th}$  covariate

The hazard ratio for the  $i_{th}$  term is calculated as:

$$\exp(b_i)$$

We present results from several models, adding different control variables step by step. This allows us to see how differences in the risk of death involving COVID-19 vary as we include further explanatory variables.

In our baseline model, we present hazard ratios adjusted for age. We include age as a second-order polynomial to account for the non-linear relationship between age and the hazard of death involving COVID-19. We then adjust for factors likely to affect the risk of infection, but also the risk of having a pre-existing condition and therefore prognosis.

Second, we adjust for residence type (private household, care home, other communal establishments). We use the 2019 NHS Patient Register to update place of residence for individuals recorded as living in a private household on the 2011 Census that had subsequently moved into a care home.

We then adjust for geographical factors, derived from current postcodes held in GPES. The probability to be infected by COVID-19 is likely to vary by region of residence. We therefore allow the baseline mortality hazard to vary by local authority district. We also adjust for population density of the lower layer super output area (LSOA). To account for the non-linear relationship between population density and the hazard of death involving COVID-19, we include population density as a second-order polynomial. This allows for different slopes for the top 1% of the population density distribution to account for outliers.

We then account for deprivation and wider measures of socio-economic status. We adjust for neighbourhood deprivation by adding decile, based on the Index of Multiple Deprivation (IMD) 2019, to the model. The IMD is an overall measure of deprivation based on factors such as income, employment and health.

We also adjust for:

- the highest level of qualification of the individual (degree, A-level or equivalent, GCSE or equivalent, no qualification)
- the National Statistics Socio-Economic Classification (NS-SEC) of the household reference person (higher managerial, administrative and professional occupations, intermediate occupations, routine and manual occupations, never worked or long-term unemployed, not applicable)

We further adjust for household composition and circumstances. We include in our models:

- the number of people in the household
- the family type (not a family, couple with children, lone parent)
- household composition (single-adult household, two-adult household, multi-generational household (households with at least one person aged 65 years or over and someone at least 20 years younger), child aged 18 years or under in household)
- tenure of the household (owned outright, owned with mortgage, social rented, private rented, other)

We include an additional "not in a household" level for all household variables for people living in a care home or other communal establishment.

In addition, we adjust for a set of measures of occupational exposure. We include a variable indicating if the individual is a key worker, and if so, what type. These data are taken from occupation as recorded on the 2011 Census. We also include a binary variable indicating if anyone in the household is a key worker.

We account for exposure to disease and contact with others using scores ranging from 0 (no exposure) to 100 (maximum exposure). Exposure to disease and physical proximity scores were obtained using Occupational Information Network (O\*NET) data, based on US Standard Occupational Classification (SOC) codes, which were then mapped to UK SOC codes. The derivation of the scores is in line with the methodology previously used by the Office for National Statistics (ONS) in our article, [Which occupations have the highest potential exposure to the coronavirus \(COVID-19\)?](#) We include these scores for all individuals with a valid occupation and derive the maximum value among all household members.

Most of these characteristics were retrieved from the 2011 Census. We sought to increase the accuracy of the Census variables so that they more accurately reflect living circumstances in 2020. We did this by setting occupational exposure variables to 0 for people who were recorded as living in a private household on the 2011 Census but living in a care home on the 2019 Patient Register. In addition, people aged 10 to 17 years at the time of the 2011 Census were excluded from the calculation of household-level variables as they are likely to have left the household.

We adjust for the number of hospital admissions and number of days spent in admitted patient care over the past three years, derived from NHS HES records. We also adjust for the presence of pre-existing health conditions, derived from the GPES GDPPR. To allow for the effect of all these health-related factors to vary depending on the age of the individuals, we interact each of them with a binary variable indicating if the individual is aged 70 years or over.

Finally, we adjust for vaccination status as a time-varying covariate, and we consider a person vaccinated once 14 days had passed since the dose was administered. More information can be found in the [UK Health Security Agency's blog COVID-19: analysing first vaccine effectiveness in the UK.](#)

We report the hazard ratios for the exposure variables between 24 January 2020 and 20 July 2022, after adjusting for age, residence type, geographical factors, socio-economic and demographic factors, health-related variables and vaccination status. A hazard ratio greater than one indicates a greater rate of death involving COVID-19 than the reference group. A hazard ratio less than one indicates a lower rate of COVID-19 mortality than the reference group.

## 10 . Related links

[Estimates of coronavirus \(COVID-19\) related deaths by hearing or vision impairment status, England: 24 January 2020 to 20 July 2022](#)

Article | Released 18 November 2022

Estimates of differences in coronavirus (COVID-19) mortality risk by hearing or vision impairment status for deaths occurring up to 20 July 2022, using linked data from the Office for National Statistics Public Health Data Asset.

## 11 . Cite this methodology

Office for National Statistics (ONS), released 18 November 2022, ONS website, methodology, [Coronavirus \(COVID-19\) related deaths by hearing and vision impairment status, England: 18 November 2022](#)