Excess mortality in England

Methodology for the weekly reports
About Public Health England

Public Health England exists to protect and improve the nation’s health and wellbeing and reduce health inequalities. We do this through world-leading science, research, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

Public Health England
Wellington House, 133-155 Waterloo Road, London, SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

Prepared by:
Sharmani Barnard, Senior Statistical Advisor; Sebastian Fox, Principal Data Scientist; Zachary Waller, Senior Public Health Data Scientist; Leigh Dowd, Senior Public Health Intelligence Analyst; Allan Baker, Deputy Head of Population Health Analysis; Paul Fryers, Deputy Head of Public Health Data Science; Paul Burton, Professor of Data Science for Health at Newcastle University and Honorary Consultant in Public Health (Epidemiology and Statistics) to PHE; Peter Goldblatt, Senior Advisor University College London Institute of Health Equity and Statistical Advisor to PHE; Justine Fitzpatrick, Head of Population Health Analysis

© Crown copyright 2021
You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit OGL. Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Published January 2021
PHE publications gateway number: GW-1880

PHE supports the UN Sustainability Development Goals
## Contents

About Public Health England  2  
Summary  4  
Introduction  4  
Methods  5  
Setting  5  
Analyses  5  
Generating expected deaths  5  
Baseline model  8  
Deaths registered during the pandemic  9  
Excess deaths  11  
COVID-19 deaths  11  
Strengths and limitations  11  
Interpretation of data and comparing groups  13  
What this means when interpreting difference between groups  13  
References  15
Summary
This report describes the methods used by Public Health England (PHE) in the Excess mortality in England: Weekly reports, it details how weekly and total excess deaths since 21 March 2020 in England have been estimated. Excess deaths are estimated by comparing the number of observed deaths during the pandemic with the number of expected deaths. Expected deaths are estimated using models based on 5 years of data from 2015 to 2019.

Introduction
Excess mortality is the number of observed deaths relative to the number of deaths that would have been expected based on previous data. Weekly monitoring of excess mortality from all causes throughout the COVID-19 pandemic provides an objective and comparable measure of the scale of the pandemic [1]. Measuring excess mortality from all causes, instead of focusing solely on mortality from COVID-19, overcomes the issues of variation in testing and differential coding of cause of death between individuals and over time [1]. Because of this, weekly monitoring of excess mortality has become increasingly important [1].

Due to the increased importance of this measure, it is essential that the most accurate, timely and reliable measure of excess is used. This report describes the methods used by PHE to estimate excess mortality during the COVID-19 pandemic from all causes of death and specific causes, by place of death and among subgroups of the population in England.
Methods

Setting

The first reported death due to COVID-19 in England was on 6 March 2020. By 20 March, 109 deaths due to COVID-19 had been registered. PHE began receiving a daily feed of deaths registered from the Office of National Statistics (ONS) on 20 March 2020. The analyses presented in the Excess Mortality in England report include deaths occurring in England, registered between 21 March 2020 and the date for the final week of each report. This period captures all the registered deaths recorded as mentioning COVID-19 in England (to usual residents of England), except for the 109 COVID-19 deaths that were registered before 21 March.

Analyses

Analyses were carried out for deaths from all causes, deaths from specific causes and deaths by place of death. Geographical analyses were carried out for all causes of deaths for England and 9 English regions (East Midlands, East of England, London, North East, North West, South East, South West, West Midlands and Yorkshire and The Humber). Regions were constructed from 2019 Upper Tier Local Authorities (UTLAs).

Analysis by cause of death was carried out using the underlying cause of death (UCOD) of the deceased and, for selected causes, any mention of the cause on the death certificate. Deaths were selected using International Classification of Diseases, 10th revision (ICD-10) [2] codes assigned by Office for National Statistics (ONS).

Place of death analyses were carried out solely for all causes of death. Place of death was determined using the methodology described in Table 2 of the Classification of Place of Death Bulletin [3]. All deaths were disaggregated into those occurring: at home; in care homes (nursing or residential); in hospitals (acute or community, excluding psychiatric); in hospices; or elsewhere (other places) [3].

Generating expected deaths

Data sources

Models to develop baseline estimates of the expected number of death registrations on a given day of the year were constructed using a combination of deaths and population denominator data from 2015 to 2019.
Mortality data

Deaths for the years 2015 to 2019 were drawn from fully coded and cleaned annual extracts supplied to PHE by ONS. For analyses of all-cause mortality, deaths were disaggregated by age, sex, region (former Government Office Regions), ethnic group, and deprivation quintile. For analyses of all-cause mortality by UTLA, deaths were disaggregated by age, sex and UTLA. For analyses of cause of death and place of death, deaths were disaggregated by age, sex and UTLA.

Adjusting for changes to cause coding in 2020

ONS code causes of death using specialist software which automatically translates the text written on death certificates into ICD-10 codes. For the majority of deaths, the software also selects which of the causes on the death certificate was the underlying cause of death. In January 2020, ONS changed the software used to a version which reflects the latest medical and epidemiological thinking. This means that for some deaths a different underlying cause would be assigned using the old and new software. ONS have documented these changes [4]. To ensure that data between years are comparable, numbers of deaths for causes which were affected by the coding changes have been adjusted in 2015 to 2019, using comparability ratios based on bridge coded data released by ONS. This ensures that numbers of deaths by underlying cause in 2015 to 2019 and 2020 onwards are consistent.

Assigning ethnicity

To undertake the ethnicity analysis, data linkage to hospital records was required to obtain the necessary information on ethnicity, as this is not recorded at death registration. To obtain ethnicity information for the baseline data, an existing file that links ONS mortality records to Hospital Episode Statistics records (HES-ONS) was used [5].

Deaths from the HES-ONS mortality file were matched to HES records to obtain ethnicity codes. The most recent usable ethnicity code from 3 HES databases (Admitted Patient Care (APC), Outpatient (OP) and Accident and Emergency (A&E)) in 2003 to 2004 up until year of death were assigned. For remaining records where ethnicity could not be assigned, historic records from APC from 1997 to 2003 were used to assign ethnicity.

The method used results in a linkage rate of 98-99% (varying year on year). Of all linked records, 2.7% were linked to a ‘not stated’ or ‘not known’ ethnic group and less than 1% had differing codes for sex in each file. Among those not linked, not stated, not known
and mismatched records it was viewed as reasonable to assume that data were missing completely at random (MCAR). For not stated, not known and not linked data, ethnic group was assigned proportionally based on 2011 ONS Census populations, stratified for age, sex and region.

The HES-ONS file only contains death records for individuals that have attended hospital in the past. Because some people die without ever having attended a hospital there are fewer deaths in this file than the annual mortality files [5]. This results in roughly 5,000 fewer deaths per year. To adjust for this, first the distribution of counts of deaths for each day (taken as an average of counts over the 17 preceding days and the 17 subsequent days) for ethnic group, by age, sex, and region was derived using the HES-ONS file. These distributions were then applied to the annual mortality file. Using this approach, the ethnicity findings were consistent with other analyses.

**Denominator data**

ONS mid-year population estimates for England (2015 to 2019) and projections (2020 onwards) were used to estimate population denominators within each group on each day [6,7].

For all-cause mortality analyses that included ethnicity, ONS data that provide population denominators by ethnic group from 2015 to 2018 were used to obtain distributions of ethnic groups by age, sex and region [8]. These analyses were developed at regional level because the computational cost was too great at lower geographies. Using an approach analogous to that used for ethnicity deaths data, the distributions of the ethnic groups by age, sex and region were then applied to ONS midyear population estimates for England for 2015 to 2019 and projections for 2020 onwards [5,6]. Using this approach, the ethnicity analyses use population denominators consistent with other analyses.

A similar approach was taken to ensure that the denominators used for the deprivation model were consistent with the other analyses. Deprivation was measured using the 2019 Index of Multiple Deprivation (IMD). In England it is available at various geographic levels, including lower-layer super output areas (LSOAs) [9]. 2019 IMD scores for LSOA of usual residence were used to allocate national deprivation quintiles [9]. LSOA populations were available for 2015 to 2019 but population projections for 2020 onwards were only available for higher geographies. Therefore, to estimate denominators for LSOAs for 2020/21, the distributions of age, sex and LSOA populations from 2019 were applied to the UTLA population projections in which each LSOA was located.

Because anybody in the population may die at home or in a hospital on any given day there is no defined subpopulation of denominators for place of death. This is also true
for cause of death. These are simply risks faced by the general population. Therefore, for place of death and cause of death analyses, denominators were derived from the age-sex distribution of the general population within a UTLA. To avoid introducing a competing risk within the model, each place of death and cause of death model is fitted separately. This means that any shortfall or excess of deaths in one place may be mirrored (in part) by a corresponding excess or shortfall of deaths in one or more of the other places of death.

**Baseline model**

**Outcome**

Counts of excess deaths are presented by week of registration at national and subnational level, and for subgroups of the population. Weeks run from Saturdays through to Friday, in-line with the weeks used by ONS. Week of registration, rather than week of occurrence, was used because the latter requires either a delay of at least 3 weeks before publication or making an estimate of the delay between week of death and week of registration. This adds uncertainty to the estimate, particularly among subnational groups.

**Data structure and covariates**

In-line with the ‘rising activity, multi-level mixed effects, indicator emphasis’ (RAMMIE) model [10], independent variables included whether the week contained a public holiday, and time of year, allowing for seasonal effects. Specific adjustments were made for registration fluctuations around public holidays. Covariates were included, allowing for the effect of age, gender, deprivation, ethnicity and geographical area. A linear trend was also included in the model to take in to account any systematic changes in the rate of death that are not reflected in the changing age structure of the population. The trend was constructed by giving each week a numerical value relating to the mean number of days from 31 December 2016 and dividing the value by 52.

For analysis of all-cause mortality, data were broken down by age groups (0-14, 15-44, 45-64, 65-74, 75-84, 85+) derived from age in years at the time of death. Data are presented by sex (male/female) based on sex reported in the death record. Deaths were allocated to UTLAs based on April 2019 UTLA boundaries [11]. Ethnic group was derived from the linkage process described above. Deprivation was measured using the 2019 IMD by LSOA. 2019 IMD scores for LSOA of usual residence were used to allocate each death to a national deprivation quintile [9].

The structure of the models used was hierarchical with population denominators and counts of death each being fully disaggregated by age, sex, geographic area, ethnicity and deprivation.
Statistical modelling

Quasi-Poisson regression models were fitted on the logarithmic scale [12]. Quasi-Poisson models were used because when counts of daily deaths are independent of one another they theoretically follow a "Poisson" distribution. This has the characteristic property that as its mean (the expected number of deaths) increases, the variability of the observed count of deaths (its variance) rises in parallel such that the variance always equals the mean. However, in the real world, the underlying risk of death varies between different population subgroups and as this cannot usually be modelled perfectly, observed counts of deaths are not completely independent. In consequence, the variance then increases faster than the mean and this is referred to as "overdispersion". Because Quasi-Poisson models allow the linear relationship between variance and mean to have a slope other than unity, they appropriately analyse rates of death when overdispersion exists [12].

The models contained the set of covariates outlined in the section 'Data Structure and Covariates' above. To allow for effects to vary between groups, interaction terms were added between age and sex, age and deprivation, age and time of year, and age and ethnicity. Finally, to model the underlying rate of death, an offset was included in every model. This reflected the logarithm of the estimate of the size of the population in each population subset.

Prediction intervals were calculated to quantify the uncertainty around baseline predictions. Prediction intervals account for both the uncertainty around estimate and the random variation in the individual values. The width of prediction intervals was estimated by generating random samples from a Quasi-Poisson distribution with the appropriate modelled rates of death and the estimated dispersion parameter. Upper and lower prediction intervals were then obtained by estimating the 0.001 and 0.999 quantiles (ie ±3 standard deviations) of this distribution.

All data were analysed using the generalised linear modelling function in the statistical package R 4.0.3 [13].

Deaths registered during the pandemic

Data source

Results of the analyses are presented on a weekly basis. ONS provided PHE with a daily feed of registered deaths data that began on 20 March 2020. The first full week of analysis begins on 21 March. These data are provisional and subject to change. As
each weekly report is run on the latest version of the data, small changes may occur to previous weekly numbers, reflecting improvements to cause of death coding or the addition of registrations not previously received.

To ensure data for deaths in communal establishments were up to date, an individual’s place of residence was mapped to a list of communal establishments in England. This list is supplied by ONS every 2 months and reflected in the analysis accordingly.

Assigning ethnicity

To obtain ethnicity data for deaths during the pandemic, a method for linking to data held within PHE to match closely the method used for the HES-ONS mortality file was developed. A 20-year historic HES table that included NHS number, date of birth (DOB), home address, sex and ethnic group is used. HES tables included APC, OP and A&E. Additionally, a table of daily Secondary Uses Service (SUS) data was used because this contained more recent finished hospital episodes for patients covering all completed episodes of care from 1 December 2019. Duplicates were identified, and the most recent valid ethnicity was assigned to the death records. All linkage was deterministic, using NHS number, admission/attendance data, date of birth, address and sex.

Using this method 97% of records were linked to a valid ethnic code, 2% were linked to a not stated or not known record and less than 1% were not linked. As with the baseline data, the mechanism of missingness was assumed to be MCAR. The same method of assigning not linked, not stated or unknown ethnicity data was applied to these records as with the baseline data. This linkage is carried out weekly.

Causes of death

The causes of death included in the report were selected based on the leading causes of death in years before 2020 and on the causes of death with the highest numbers of deaths in April 2020. They are grouped according to standard groups of leading causes developed by ONS [14] but with some adjustments. All acute respiratory infections have been grouped together in this report (ICD-10 J00-J22) rather than reporting influenza and pneumonia separately (J09-J18), as in the ONS definition. At this stage in the pandemic it was also thought more appropriate to group all cancers together rather than reporting on specific cancer sites. Future analyses of excess deaths could examine the impact of the pandemic on specific cancers.

Although there is interest in the impact of the pandemic on external causes of death, such as transport accidents and suicides, it is not yet possible to look at this. Deaths from external causes are not generally registered until after a coroner’s inquest has taken place to examine the circumstances of the death. It may take months for an inquest to take place and for the death then to be registered. As the coronial system has
been disrupted during the pandemic, fewer deaths from external causes are being registered [15]. This aspect of the pandemic will be an important factor to examine when data do eventually become available.

**Excess deaths**

Total cumulative excess mortality was estimated by counting deaths since 21 March 2020 and subtracting the expected cumulative deaths since 21 March 2020 in the baseline model. Weekly excess mortality was calculated by taking the observed number of deaths registered in a week and subtracting the expected registered deaths for that week.

**COVID-19 deaths**

A separate colour is used on charts to represent the number of deaths where COVID-19 was mentioned on the death certificate. This includes any death record where COVID19 was the underlying cause of death or was mentioned anywhere (ICD10 code U07.1 or U07.2). This includes both confirmed and suspected COVID-19.

**Strengths and limitations**

This methodology has many strengths. It produces timely, reliable estimates of excess deaths by modelling week of registration rather than week of occurrence. This is especially important when the excess appears to be declining, because using week of occurrence too soon after the decline will overestimate how quickly the excess is declining. This method takes appropriate account of the age, sex, geographical distribution, ethnicity and deprivation quintile in each sub-group both in the baseline and pandemic period data. By including year on year trends in mortality and changing population denominators the method adjusts for trends in mortality and an ageing population. By modelling expected deaths for each week, based on the time of year and proximity to a public holiday, it accurately adjusts expected death registrations during the weeks around Easter (where the date varies year on year) and other public holidays. This is especially important for monitoring weekly excess over the Easter and Christmas periods in 2020, which both occurred at critical points in the outbreak. Finally, it provides a clear picture of the total excess attributable to the COVID-19 outbreak by only including excess mortality occurring from 21 March 2020.

This methodology has some limitations. Because baseline data are based on previous years’ registration practices, adjustments could not be made to the baseline model for changes that occurred to registration processes from 2020. One important difference between the pandemic period and previous years is that deaths could be registered on weekends and public holidays in April [16]. The increased capacity for registration over the Easter weekend may have resulted in registrations being brought forward. This may
slightly overestimate the excess [17]. This may be more pronounced in the week before Easter due to the reassignment of deaths from Friday and Saturday of the Easter weekend to the previous Thursday [17]. There may also be a greater delay in death registrations where the death has been referred to a coroner throughout the COVID-19 outbreak, resulting in an underestimate of the excess.

Analysis of ethnicity was limited to 5 broad categories because of differences in the sources of death data and denominator data. Ethnicity was obtained by linking deaths to HES, while denominators for ethnicity are derived from 2011 Census data. While ethnicity is intended to be self-reported in both the Census and HES, it is possible that a proportion of HES ethnicity is recorded by health care staff or may be recorded by the head of household at Census. Compared with those coded as White, those coded as non-White are more likely to be misclassified in HES [18]. Health care staff may be more likely to record ethnicity as ‘other’ ethnic groups, (such as ‘Black Other’ and ‘Asian Other’), or to misclassify ethnicity if they are not sure of the ethnicity of the patient. This may have occurred more often during the COVID-19 outbreak, when fewer family members were present in hospitals. Therefore, there may be a greater number of excess deaths among ‘Black other’ and ‘Asian other’ ethnic groups and more misclassification of ethnicity during the COVID-19 outbreak that would not be reflected in the denominators. To reduce the impact of this potential bias, ethnic groups are presented as 5 broad categories. Future work will attempt to cross-validate ethnicity coding in HES and the Census.

A further potential sources of bias between deaths and populations may result from the fact that ethnicity was assigned using HES records, which include international migrants since 2011. The population estimates by ethnic group used in this analysis do not, however, reflect the effect of international migration since 2011 on the ethnic distribution of the population. This is likely to have led to underestimation of the population for some ethnic groups.

Deprivation is measured by LSOA of residence at time of death. Among individuals in care homes, the LSOA of the care home may not reflect the level of deprivation they experienced prior to entering the home.

The baseline is modelled using 5 years of historical data. These data include years of relatively high mortality and relatively low mortality. Although, prior to the COVID-19 outbreak, 2020 began as a year with relatively low mortality, this may have been due to the relatively mild flu season and would not necessarily have affected mortality compared with the baseline through the spring and summer months. Finally, there may be a small underestimation (including the 109 deaths coded to COVID-19 as well as early deaths caused by COVID-19, with no mention of COVID-19 on the certificate) of the cumulative excess due to the start date of 21 March 2020.
Interpretation of data and comparing groups

Among each subgroup (age, sex, region, ethnicity and deprivation) rates are calculated from the historic (2015 to 2019) data and then applied to the 2020/21 population for that same subgroup to calculate the expected number of deaths. When these are added up, they generate the total expected number of deaths that would be registered in a specified week, if the death rates based on the historic data were to apply in 2020/21. Observed deaths are compared with expected deaths to estimate both the excess number of deaths and the ratio of the observed to expected deaths. This can therefore be seen to be a form of indirect standardisation (equivalent to a Standardised Mortality Ratio).

What this means when interpreting difference between groups

In the comparative charts, showing total cumulative absolute numbers of excess and ratios of excess for particular groups (for example age groups, ethnic groups and deprivation quintiles), the excess deaths and ratios are generated for each group separately. Excess deaths can be interpreted as the additional deaths in the group over and above the number it would have experienced if the 5 years’ age-sex region, ethnic group and deprivation decile specific rates for that group had stayed the same as they were in the historic (2015 to 2019) period.

The ratio is the number of observed deaths divided by the number that would have been experienced if rates had stayed the same in that group. Therefore, both the excess and the ratio are standardised for changes in the age and sex distribution over time within that group. Any excess deaths or increase in the ratio are not due to the ageing of the population of that group.

Using the ratio of observed to expected deaths, comparisons of the excess can be made between groups of difference sizes. However, differences between the age, sex structure of the groups will still account for some of the difference seen in the ratio of observed to expected deaths between groups.

There are 2 reasons for this:

- the expected number of deaths are predicted, given the age, sex structure of that population
- the risk of death from COVID-19 (observed deaths) is differential between age groups and sex

Therefore, rather than removing the effect of age and sex between groups (as a directly standardised rate does), ratios of excess deaths take account of the age and sex structure of each group. Differences between groups (such as 2 ethnic groups), should
be interpreted as the relative difference from the expected, given the different age and sex structures of each group.
References


