

# Predicting total weekly death occurrences in England and Wales methodology

Outlining a way in which we can estimate the total number of deaths that occurred within the latest week in England and Wales.

Contact:  
Neil Hopper  
health.data@ons.gov.uk  
+44 (0) 1633 580080

Release date:  
29 May 2020

Next release:  
To be announced

## Table of contents

1. [Overview](#)
2. [Introduction to death registration and occurrences data](#)
3. [Exploring the relationship between death occurrence and registration](#)
4. [Predicting death occurrences from up to three weeks registration data](#)
5. [Caveats of the approach](#)
6. [Appendix 1: Relationship between death registrations and death occurrences](#)
7. [Appendix 2: Modelling registration delay](#)
8. [Appendix 3: Measuring performance against 2019 data](#)
9. [Related links](#)

# 1 . Overview

## Registration delays

Deaths are not generally registered on the actual date of death, so there is a delay between occurrence and registration. We regularly publish analysis of these [registration delays](#). These registration delays make it difficult for us to report timely and complete statistics on death occurrences as the total number of deaths that occurred in a given week can only be known some time later. For this reason, we have historically published weekly counts of death registrations rather than death occurrences.

In exceptional situations there can be excess deaths brought about by environmental factors such as the weather or disease, as seen in the coronavirus (COVID-19) pandemic. A method to estimate the total number of deaths occurring based on the limited registration data available is therefore of value.

We publish this method with the intention of using it in the future to help inform analyses needing time-sensitive monitoring of numbers of deaths. The most appropriate way of publishing such estimates is still being considered. This method is experimental and no estimates relating to deaths in 2020 are included in the paper.

## Weekly data on deaths

To have clear information on trends in deaths in the short-term, it is useful to estimate the number of deaths in the most recent weeks.

In our [weekly deaths publications](#) we report the number of deaths that were registered within a particular week in England and Wales. We report each Tuesday on the deaths registered in the week ending on the Friday 11 days before.

From week 13 2020 (week ending 27 March, published 7 April 2020), because of the increased need for data on deaths during the coronavirus (COVID-19) pandemic, we started reporting the number of deaths involving COVID-19 by date of occurrence as well as date of registration, including deaths occurring in the reference week that are registered as close to the publication date as possible (usually 10 days after the end of the reporting week). However, figures reported by date of occurrence relatively soon after the date of death are incomplete because of the known delays.

## Relationship between death occurrences and registrations

In this article, the relationship between weekly total death occurrences and registrations of those deaths occurring in the same week and those immediately following is investigated. The method outlined uses patterns of registration delay from previous years to estimate the number of deaths likely to have occurred in a time period but not yet been registered and produces an estimated total number of deaths that occurred within a specific week with confidence intervals. The further in time from the week of interest, the more complete the data and the more accurate the estimated total number.

This method also takes account of weeks that include bank holidays, when there is greater delay because of closed services.

Based on data covering years 2015 to 2018, it is found that means of around:

- 46% of deaths are registered in the same week as occurrence
- 85% are registered by the end of the subsequent week
- 91% are registered by the end of the subsequent second week

Simply speaking the reciprocal of these values can be applied to current registration data to estimate the true numbers of occurrences. Here the role of seasonal factors such as holiday periods on the variation of these registration rates is explored.

## Results and recommendations

It is found that bank holidays have a strong effect on reducing the rate of registration in the same week and immediately adjacent weeks. It is also found that there is a seasonal effect on registration rates that correlates with the seasonal increase in death occurrences. This compromises the approach to the extent that increased deaths could result in a reduction in the rate of registration and an under-estimate of total deaths. In addition, the method assumes that current registration patterns follow historical norms, an assumption that may not hold in all circumstances.

Relatively simple models of registration delay can be used to estimate the total number of occurrences based on early registration data. Application of these models including seasonal and bank holiday effects based on 2015 to 2018 data to predict total death data in 2019 showed strong performance and utility. A predictive accuracy of within 3% is observed when estimating total deaths from registrations occurring in the same week, improving to 0.7% when estimating total deaths from cumulative registrations by end of the subsequent week (based on mean absolute error). There is evidence that excess deaths resulting from time specific events such as a heatwave can be predicted from cumulative registrations by end of the subsequent week.

## 2 . Introduction to death registration and occurrences data

Registration of deaths is not instantaneous. To register a death in England and Wales either a medical certificate of death or permission from a coroner is required. According to the Births and Deaths Registration Act 1953, unless the death is referred to a coroner, it should be registered within five calendar days. There is no time restriction placed on coroners to provide permission to register a death once it has been referred, and because of the time that can be needed to hold an inquest, registration may be weeks, months or even years after the death has occurred. The time between death occurring and being registered is referred to as registration delay. In 2018, the [median registration delay in England and Wales](#) was five days.

Once a death is registered at the local registration office, the data are sent overnight to the Office for National Statistics (ONS) and are available for coding within two working days and usually for analysis on the third working day. The ONS publishes [weekly provisional death registrations](#) every Tuesday. These data are for the week ending on the Friday 11 days earlier and are extracted from ONS systems on the following Thursday.

Weekly counts of deaths are published using date of registration of the death because all registrations for the week are usually captured; the data do not usually need to be revised and so can be released in a timely fashion. However, generally only about 46% of registered deaths occurred in the same registration week, with a further 39% occurring in the week before (based on mean values for 2018). Most of the remaining registrations are for deaths that have been referred to a coroner and occurred more than four weeks earlier. So weekly registered deaths cover a large period of when the death occurred.

This article investigates the practicality of publishing weekly death occurrences, bearing in mind that the ONS does not know about a death until it is registered. As a proportion of all occurrences, the number of deaths registered over time becomes ever closer to the true value. The question investigated is how many weeks of registration data are required following an occurrence week for us to precisely estimate the total number of deaths that occurred in that week? This requires understanding of the variation of registration rates by week delay so that the amount of unexplained variation in rates is small and concomitantly our confidence intervals are narrow. This question will be particularly useful in estimating excess deaths that occur because of time-specific events.

The current reporting of weekly death registrations has limited use for time-sensitive surveillance but can be produced at speed. We show that it is possible to estimate weekly death occurrences with high accuracy and precision (narrow confidence intervals) using the limited registration data that are available by the end of the week following the week of occurrence.

#### **Looking for information on the coronavirus?**

- Our [weekly deaths dataset](#) has the most up-to-date figures on deaths involving the coronavirus (COVID-19).
- More detailed commentary on deaths involving COVID-19 is available in the [weekly deaths bulletin](#).
- Find the latest on [coronavirus in the UK](#).

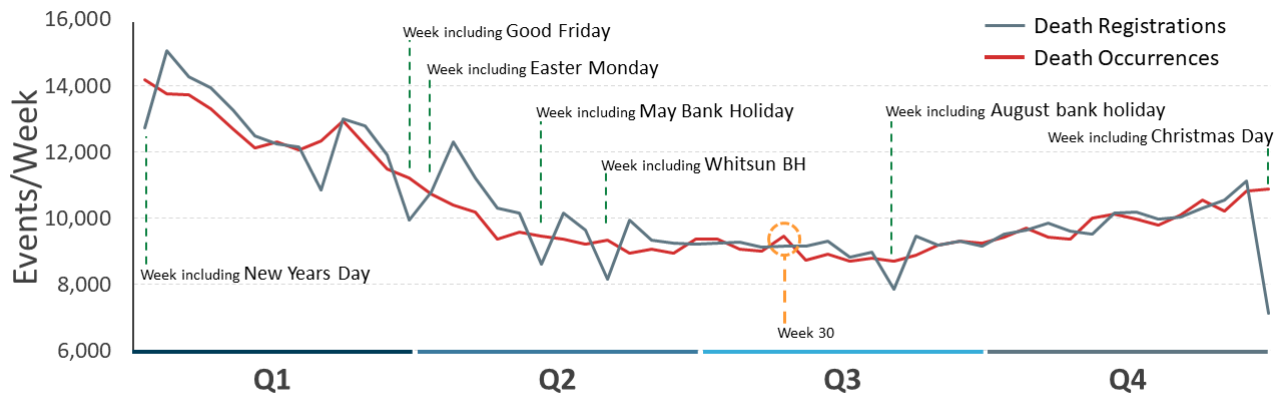
### **3 . Exploring the relationship between death occurrence and registration**

The weekly death registration data that the Office for National Statistics (ONS) publishes define a week as running from Saturday to Friday. The weekly account of death occurrences and registrations for 2018 in England and Wales is shown in Figure 1.

The number of death registrations during a week containing bank holidays is typically reduced by around 10% and there is a bounce of increased registrations most obviously in the subsequent week.

**Figure 1: Death registrations are affected by bank holidays**

**Number of deaths occurring and registered by week in 2018, England and Wales**



**Source: Office for National Statistics**

**Notes:**

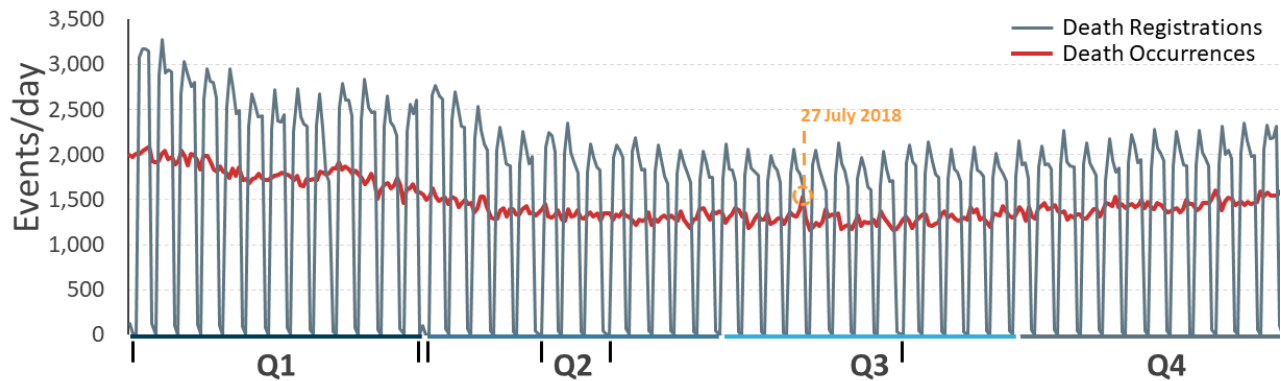
1. Week 1 begins Saturday 30 December 2017.
2. The hottest day of 2018 was recorded in week 30.

The daily account of death occurrences and registrations for 2018 in England and Wales using the same data is shown in Figure 2. Daily registrations show a weekly pattern, with a peak on a Tuesday and few or no registrations at the weekend or on bank holidays (Figure 2; note that the weekly peak of registrations is pushed to a Wednesday following a bank holiday Monday). So death registrations are most obviously affected by available working days, though it is also clear that there are more death registrations in winter than in summer months.

In contrast, there is little evidence that being a working day, or a specific weekday, affects the number of deaths that occur. However, as for registrations, the number of deaths occurring is higher in winter than in the summer months.

**Figure 2: Number of death registrations varies by day of the week.**

Number of deaths occurring and registered by day in 2018, England and Wales.



Source: Office for National Statistics

Notes:

1. Aligned to Figure 1; week 1 begins Saturday 30 December 2017.
2. The hottest day of 2018 was recorded on 27 July.
3. Vertical bars represent bank holidays.

The hottest day of 2018 was [Friday 27 July](#), with 35.6 degrees Celsius recorded in Felsham, Suffolk. This event does not produce an observable effect upon daily or weekly death registrations, but it does coincide with a peak in both the daily and weekly number of death occurrences (day peak occurred on 27 July 2018 (Figure 2, highlighted), weekly peak for week beginning 21 July 2018 (Figure 1, highlighted)). This suggests that to assess the public health impact of rare, time-specific, events it would be more useful to prioritise examination of death occurrence over death registration data.

## Measuring variability in the registration delay

The time between the death occurrence and registration is the registration delay. Historic patterns of registration delay can potentially be applied to current registration data to predict the total of death occurrences.

So for a given time  $X$  after an occurrence period, historically the mean proportion of those occurrences are registered by that time is  $p_x$ , then the current number of occurrences registered by the same time delay divided by  $p_x$  gives an estimate of the true number of death occurrences in the period of interest. The success of this method depends upon an assumption that the current year will have a similar pattern and variation in registration delay as the historic data. We need to measure the size of this historic variation to ascribe an uncertainty to our estimate. A general caveat to the approach is that registration delay may be correlated with the number of deaths ([Appendix 1: Relationship between death registrations and death occurrences](#)).

The problem under investigation is framed as what proportion of deaths occurring within a certain time period are expected to be registered by time  $X$  after that period? This analysis is aligned to the publication of weekly provisional death registrations. These weeks begin each Saturday and end on the following Friday. The first week of any given year ends on the first Friday of that year. Deaths can be identified with two weeks: the week in which the death occurred (occurrence period) and the week in which the death was registered (registration delay). The time in weeks between week of occurrence and the week of registration is the delay period in weeks. The analysis covers a delay period of up to five weeks and assumes that all deaths that occurred in any given week in 2015 to 2018 are registered by the end of 2019.

A fuller account of the methodology is given in [Appendix 2: Modelling registration delay](#). In brief, a given week of death occurrences is treated as a set of  $n$  trials (deaths) with each trial sharing a probability of  $p_x$  that each death will be registered at time  $X$ . The proportion of deaths ( $p$ ) that are registered at time  $X$  is an estimator of  $p_x$ , with a variance of  $p(1-p)/n$ . As the number of deaths is large (more than 8,000 per week; Figure 1),  $p_x$  is a precise estimator allowing us to ignore the variance of the estimator. The mean and variance of the population of proportions ( $p_x$ ) are under investigation here.

This population of proportions is made up of the 52 (or 53) weeks of death occurrences in a year. For the period of analysis, 2015 to 2018, there are 209 weeks. The proportion of deaths that are registered in the same week that they occurred is referred to as  $p_0$ . The proportion of deaths that are registered by the end of the subsequent week that they occurred in is referred to as  $p_1$  and so on to  $p_5$ , the limit of this study.

As this investigation is a study of a population of proportions, which may not be normally distributed, values undergo the logit transformation ( $\text{logit } p_x = \log_e(p_x/(1-p_x))$ ). Following this transformation, normality is assumed for analysis and variance calculations. In most cases, values are presented back in the probability scale using the following conversion:

$$p_x = 1/(1+\exp(-\text{logit}(\text{value})))$$

## Overall registration delay between years 2015 and 2018

The mean proportion of deaths registered by weeks delay by year is shown in Table 1 and are calculated through the logit transformation with 95% confidence intervals as shown. For comparison, means calculated without logit transformation are shown in square brackets.

Null models are simply these means (expected values), which can be used to make predictions; the reciprocal of these mean proportions can be used to estimate total occurrences based on registrations by delay week 0 to 5 (Table 1; reciprocal of the lower confidence interval of proportion gives an estimate of the upper confidence of predicted values).

As [reported in Impact of registration delays on mortality statistics in England and Wales: 2018](#), the registration delay appears to lengthen from 2015, suggesting a trend that using historic data to predict current occurrences may systematically under-estimate values (Table 1).

Table 1: The mean proportion of all weekly deaths registered by delay and year, England and Wales, 2015 to 2018

	2015	2016	2017	2018
p	48.6% (40.6% - 56.6%) [48.6%]	45.7% (37.3% - 54.3%) [45.7%]	45.5% (37.6% - 53.6%) [45.5%]	45.5% (36.3% - 55.0%) [45.6%]
p	86.7% (82.9% - 89.8%) [86.6%]	84.9% (80.5% - 88.3%) [84.7%]	84.9% (80.3% - 88.6%) [84.8%]	84.5% (80.2% - 88.0%) [84.4%]
p	91.4% (90.0% - 92.6%) [91.3%]	90.5% (89.2% - 91.7%) [90.5%]	90.6% (89.3% - 91.8%) [90.6%]	90.5% (89.2% - 91.6%) [90.5%]
p	92.3% (91.0% - 93.5%) [92.3%]	91.8% (90.9% - 92.7%) [91.8%]	92.0% (91.2% - 92.7%) [92.0%]	91.9% (90.8% - 92.9%) [91.9%]
p	92.8% (91.6% - 93.9%) [92.8%]	92.4% (91.4% - 93.2%) [92.4%]	92.5% (91.7% - 93.3%) [92.5%]	92.5% (91.4% - 93.5%) [92.5%]
p	93.8% (92.0% - 94.2%) [93.1%]	92.8% (91.8% - 93.6%) [92.7%]	92.9% (92.0% - 93.7%) [92.9%]	92.9% (91.8% - 93.8%) [92.8%]

Source: Office for National Statistics

#### Notes

1. Registration delay is measured in weeks and the proportion of occurrences registered by weeks delay given by p [Back to table](#)
2. Means with 95% confidence intervals are calculated using through the logit transformation before transforming back into the probability scale. [Back to table](#)
3. Means calculated directly in the probability scale are shown in square brackets. [Back to table](#)

## Largest cause of variability in registration delay of deaths is bank holidays

In this section, the rates of registration delay by weeks delay are examined. In Figure 3, the week of occurrence is plotted on the x axis and the proportion of deaths registered by weeks delay shown for death registrations in England and Wales in 2018.

The mean proportion of weekly deaths that were registered in the same Saturday to Friday week of occurrence was 45.5% in 2018 (Table 1; p<sub>0</sub>). There appears to be a trend of lower rates in the winter months compared with the summer, but with a general range of 40% to 50% (visual inspection, Figure 3). This winter effect may be correlated with the number of deaths as there is correlation between the number of occurrences and proportion of weekly deaths that were registered in the same week ([Appendix 1: Relationship between death registrations and death occurrences](#)).

Irrespective of this winter effect, almost all troughs correspond with weeks containing bank holidays (Figure 3). The largest exception is the 37% seen for week 9 of occurrence beginning 24 February 2018. The reason for this is unclear; though this week was bitterly cold, the 2018 mid-late quarter one peak of occurrences is for the following week (Figure 1).

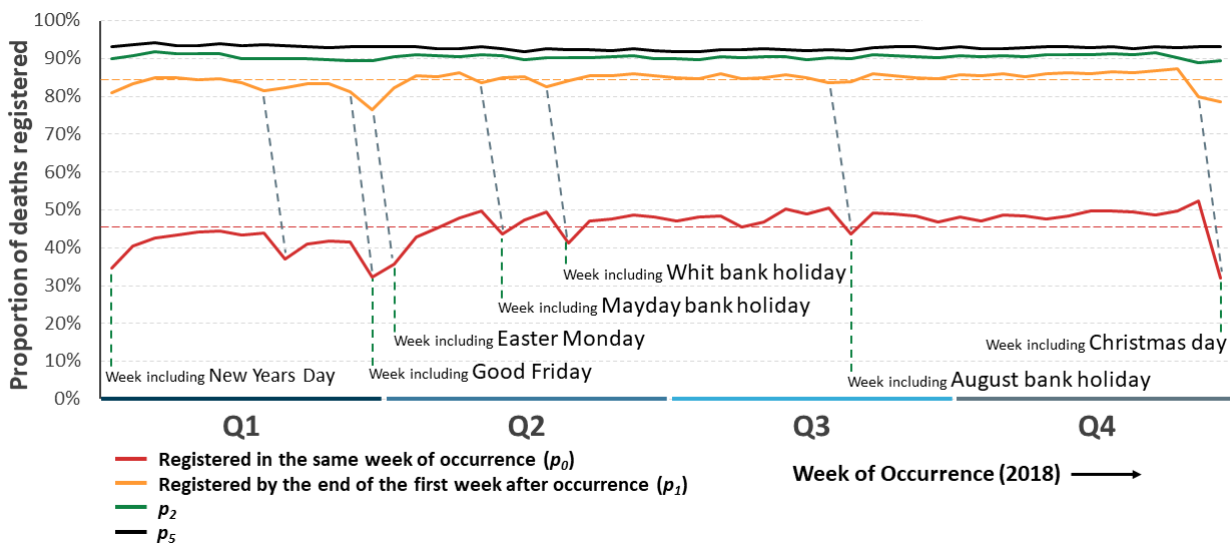


A mean proportion of 84.4% deaths weekly were registered by the end of the first week after occurrence in 2018 (Table 1;  $p_1$ ). There is less variation around this mean; the winter effect has lessened. However, bank holidays in the subsequent registration week to occurrence do seem to add to a delay (Figure 3; troughs for one week's delay proportions ( $p_1$ ) are generally a week earlier than for week 0 ( $p_0$ ) and are linked by dashed lines in the figure).

The most obvious troughs in the proportion of deaths that were registered by the end of the first week after occurrence is for those with a public holiday in both the week of occurrence and the subsequent week (Easter and Christmas holiday periods).

**Figure 3: The variation in the proportion of deaths that are registered decreases over time since occurrence**

Proportion of occurred deaths registered in the same week ( $p_0$ ), by the subsequent week ( $p_1$ ), by two ( $p_2$ ) and five ( $p_5$ ) weeks, England and Wales, 2015 to 2018



Source: Office for National Statistics

Notes:

1. Data for England and Wales.
2. Mean proportions for  $p_0$  and  $p_1$  are included (dashed lines)

By the end of the second week after occurrence, a mean proportion of 90.5% deaths were registered in 2018 (Table 1;  $p_2$ ). The impact of winter is no longer apparent, and the effects of bank holidays have become minor.

Taken together, this suggests that we should account for the effect of bank holidays upon registration delay when modelling registration delay close to occurrence. As bank holidays can occur in different reporting weeks, with Easter being the most volatile holiday, the method used must be able to handle this. The winter effect is defined as quarter one, weeks 1 to 13 in any given year. Our method allows interaction between these effects, such that Easter, for example, can occur either on top of, or independently to, the winter effect (quarter one).

## 4 . Predicting death occurrences from up to three weeks registration data

Based on data from years 2015 to 2018, a series of expected values and their confidence intervals are generated based on the modelling described in [Appendix 2: Modelling registration delay](#). These regression models aim to explain the variance of the population of proportions ( $p_x$ ) following their logit transformation. This is calculated as follows:

$$\text{logit } p_x = \log_e(p_x/(1-p_x))$$

This transformation aims to overcome the lack of normality in the distribution of proportions ( $p_x$ ) and allows linear models to be built on the transformed data. Models are built at weekly time points, based on registrations that are recorded by week of, and following, the week of occurrence.

Two sets of explanatory factors are considered: a binary variable of whether the week of occurrence is in quarter one (winter effect) and a categorical variable describing the presence of bank holidays ([Appendix 2: Modelling registration delay](#)). Models are based on two datasets: a single year's data for 2018 and the combined data from 2015 to 2018. The models created from the combined data are the most robust and are presented in the following sub-sections.

The performance of thmber of registrate models is assessed against 2019 data and based on the expected proportion registered by time  $x$ ,  $p_x$ . The nuions observed by time  $x$  ( $R_x^{\text{obs}}$ ) is multiplied by the reciprocal of  $p_x$ , such that:

$$\text{Predicted} = (R_x^{\text{obs}})/p_x$$

As the data in this study are limited to deaths that were registered by the end of week 16 2020, we may not know the true number of occurrences in 2019 and this problem becomes more acute as the occurrence data analysed get closer to the end of 2019. After 26 weeks following occurrence, typically, 98% of occurrences are registered (2015, 98.1%; 2016 to 2018, 97.8% for each year).

Assuming this holds for 2019, to ensure that at least 98% of all occurrences to be captured for 2019, death occurrences through to, and including, week 42 are analysed. This week ran from Saturday 12 October to Friday 18 October 2019. This allows at least 26 weeks for registration to occur for all 2019 deaths in the analysis; though the assumption for analysis is that all deaths occurring in 2019 up to Friday 18 October 2019 were registered and are included in this analysis, as the week analysed approaches 42, it is likely that around 98% of deaths are captured. Some consideration of this will be required when interpreting the assessment.

Finally, the [hottest day on record for the UK](#) was recorded at Cambridge Botanic Gardens on 25 July 2019, with a temperature of 38.7 degrees Celsius . In October 2019, we reported in [Do summer heatwaves lead to an increase in deaths?](#) that the number of deaths increased around the same time as this hottest date on record, without estimating an excess of deaths, citing the delay in registration of deaths resulting in insufficient (or provisional) data as the reason. In November 2019, Public Health England published an estimate of there being [572 excess deaths in those aged 65 years and over between 21 and 28 July 2019](#). This event is used as an exemplar of this method to test how soon could we estimate excess deaths caused by this heat wave.

## Expected proportion of deaths registered in the same week as occurrence ( $p_0$ )

The proportion of deaths registered in the same week as occurrence is significantly lower in Quarter 1 (Jan to Mar) than in other quarters ([Appendix 1: Relationship between death registrations and death occurrences](#)). Likewise, the expected proportion of deaths registered in the same week as occurrence is severely reduced by bank holidays in the week of occurrence ([Appendix 2: Modelling registration delay](#)).

This effect is enhanced when there is also a bank holiday in the subsequent week, presumably because Easter and Christmas are significant holiday periods. These two factors are combined into a single model (called  $p0\_bhw0q1$ ) that is highly significant; approximately 71% of the weekly variation in the proportion of occurrences that are registered in the same week as occurrence (logit scale) is explained by the  $p0\_bhw0q1$  model ([Appendix 2: Modelling registration delay](#)). The expected values derived from this  $p0\_bhw0q1$  model are shown in Table 2.

Table 2: Deaths occurring in Quarter 1 (Jan to Mar) and/or weeks containing bank holidays are expected to have lower rates of registration in that same week

Expected proportions of deaths registered in the same week as occurrence based on the  $p0\_bhw0q1$  model

Quarter (q1)	Bank holiday week category (bhw0)	Expected proportion of deaths registered
Q2-4	No bank holidays	48.6% (43.8% - 53.5%)
Q1	No bank holidays	44.5% (39.7% - 49.3%)
Q2-4	Bank holiday in occurrence and following week	35.2% (30.9% - 39.8%)
Q1	Bank holiday in occurrence and following week	31.5% (27.4% - 35.8%)
Q2-4	Bank holiday in week of occurrence	40.8% (36.2% - 45.6%)
Q1	Bank holiday in week of occurrence	36.9% (32.4% - 41.5%)

Source: Office for National Statistics

### Notes

1. Quarter (q1); Is the week of occurrence in Quarter 1 (Jan to Mar). [Back to table](#)
2.  $bhw0$  is a categorical variable that takes one of three values: there is a bank holiday in the week of occurrence and the following week; there is a bank holiday in the week of occurrence only; and there are no bank holidays in the week of occurrence or the following week. [Back to table](#)

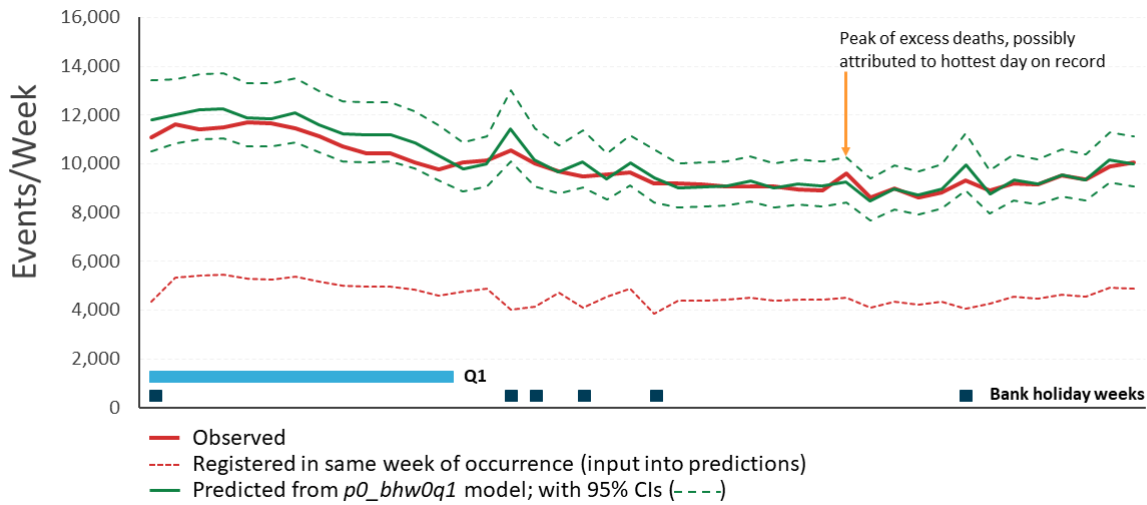
To test whether 2019 showed a broadly similar delay profile as 2015 to 2018, the mean proportion of deaths occurred by weeks registration delay was examined for all data in the analysis (209 weeks combined for 2015 to 2018; 42 weeks combined for 2019). The mean proportion of deaths registered in the week of occurrence for the combined 2015 to 2018 years was 46.0%, compared with 47.0% for 2019. This may be because the period for registering 2019 deaths is shorter. Consequently, the model tends to over-estimate death occurrences (Figure 4).

The mean absolute model error for the 42 predictions based on 2019 data is 3.0% (on average the difference between predicted and observed deaths is 3.0% of the predicted value). The observed value was within the model predicted 95% confidence intervals for all 42 estimates.

The observed occurrences show there to be around 600 to 800 excess deaths in week 30, compared with the trend. This week includes the hottest day on record, 25 July 2019. As there is no spike in registrations of deaths that occurred in week 30 in registration week 30, it is impossible to predict excess deaths from the registration data available in this same week (Figure 4).

**Figure 4: The model predicted total deaths from registrations in the same week of occurrence is accurate but with low precision**

Predicted occurrences based on numbers registered in the same week as occurrence from the  $p0\_bhw0$  model plotted against observed 2019 data for England and Wales (weeks 1 to 42). Wide 95% confidence intervals indicate low precision



Source: Office for National Statistics

Notes:

1. Wide 95% confidence intervals indicate low precision.

## Expected proportion of deaths registered by the end of the week following occurrence ( $p_1$ )

The effect of bank holidays on the proportion of deaths registered by the end of the week following occurrence is much stronger than the effect of Quarter 1 (Jan to Mar) ([Tables 8 and 10, Appendix 2: Modelling registration delay](#)). Consecutive bank holidays in the week of occurrence and the following week have the strongest effect (Table 3), followed by consecutive bank holidays in the two weeks following occurrence. In the absence of consecutive bank holidays, a bank holiday in either the week of occurrence or the following week has a significant effect.

The bank holiday and Quarter 1 effect are combined into a single model (called  $p1\_bhw1q1$ ) that is highly significant; approximately 61% of the weekly variation in the proportion of occurrences that are registered by the end of the week following occurrence (logit scale) is explained by the  $p1\_bhw1q1$  model ([Table 11, Appendix 2: Modelling registration delay](#)). The expected values derived from the  $p1\_bhw1q1$  model are shown in Table 3.

Table 3: Deaths occurring in Quarter 1 (Jan to Mar) and weeks containing, or followed by weeks containing, bank holidays are expected to have lower rates of registration by the end of the following week  
 Expected proportions of deaths registered by the end of the week following occurrence based on the p1\_bhw1q1 model

Quarter (q1)	Bank holiday week category (bhw1)	Expected proportion of deaths registered
Q2-4	No bank holidays	86.2% (83.5% - 88.5%)
Q1	No bank holidays	85.1% (82.3% - 87.6%)
Q2-4	Bank holiday in occurrence and following week	77.9% (74.1% - 81.3%)
Q1	Bank holiday in occurrence and following week	76.4% (72.4% - 80.0%)
Q2-4	Bank holiday in occurrence or the following week	84.2% (81.2% - 86.8%)
Q1	Bank holiday in occurrence or the following week	83.0% (79.9% - 85.8%)
Q2-4	Consecutive bank holidays in the two week following occurrence	82.4% (77.7% - 84.1%)
Q1	Consecutive bank holidays in the two weeks following occurrence	81.1% (82.3% - 87.6%)

Source: Office for National Statistics

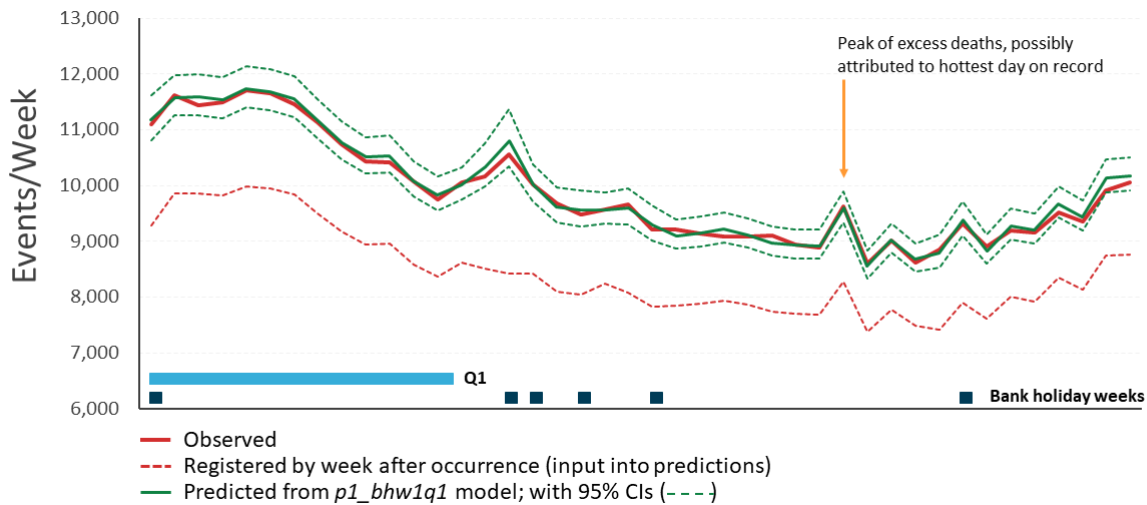
Notes

1. Quarter (q1); Is the week of occurrence in Quarter 1 (Jan to Mar). [Back to table](#)
2. bhw1 is a categorical variable which takes one of four values: there are bank holidays in the week of occurrence and the following week; there is a bank holiday in the week of occurrence or the following week only; there are bank holidays in the two consecutive weeks following occurrence; and there are no bank holidays in the week of occurrence or the following week. [Back to table](#)

The mean proportion of deaths registered by the end of the week following occurrence for the combined 2015 to 2018 years was 88.0%, compared with 88.5% for 2019. Again, this leads to a tendency for the model to over-estimate death occurrences (Figure 5). The mean absolute model error for the 42 predictions based on 2019 data is 0.7% and the observed value was within the model predicted 95% confidence intervals for all 42 estimates.

**Figure 5: The model predicted total deaths from registrations by the end of the week following occurrence is both accurate and precise**

Predicted occurrences based on numbers registered by the end of the week following occurrence from the p1\_bhw1 model plotted against observed 2019 data for England and Wales (weeks 1 to 42)



Source: Office for National Statistics

## Expected proportion of deaths registered by the end of the second week following occurrence ( $p_2$ )

There is no significant difference between the proportion of deaths that are registered by the end of the second week after occurrence for Quarter 1 compared with other quarters ( $p=0.09$ ). However, the bank holiday effect remains significant, though much less than in earlier models ([Table 12, Appendix 2: Modelling registration delay](#)).

The bank holiday effect has three categories, relating to whether there are zero, one or two weeks containing bank holidays in the week of occurrence and subsequent two weeks. The interaction of bank holidays with Quarter 1 is not significant and was dropped.

The final model called p2\_bhw2 is significant and explains approximately 23% of the weekly variation in the proportion of occurrences that are registered by the end of the second week following occurrence (logit scale; [Table 12, Appendix 2: Modelling registration delay](#)). The expected values derived from the p2\_bhw2 model are shown in [Table 4](#).

Table 4: Deaths occurring in weeks containing, or followed by weeks containing, bank holidays are expected to have moderately lower rates of registration by the end of the second week  
 Expected proportions of deaths registered by the end of the second week following occurrence based on the p2\_bhw2 model

Bank Holiday week category (bhw2)	Expected proportion of deaths registered
No Bank Holidays	91.0% (89.6% - 92.1%)
One Bank Holiday week in the occurrence or following two weeks	90.5% (89.2% - 91.8%)
Two Bank Holiday weeks in the occurrence and following two weeks	89.6% (88.2% - 91.0%)

Source: Office for National Statistics

Notes

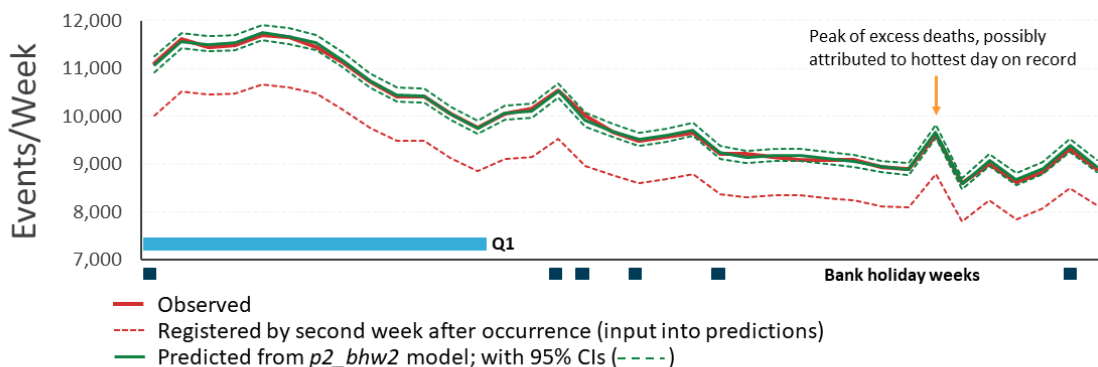
1. bhw2 is a categorical variable that takes one of three values: there is a single bank holiday week in the week of occurrence or the following two weeks; there are two bank holiday weeks in the week of occurrence and the following two weeks; and there are no bank holidays in the week of occurrence or the following two weeks. [Back to table](#)

The mean proportion of deaths registered by the end of the second week following occurrence for the combined 2015 to 2018 years was 90.7%, compared with 91.2% for 2019. Again, this leads to a tendency for the model to over-estimate death occurrences (Figure 6).

The analysis presented in [Appendix 3](#) shows that the model performs poorly when assessed against week 37 and later for 2019 data. This is probably because the true number of occurrences is likely to be under-estimated by up to 2% (based on 98% of total deaths being registered by 26 weeks following occurrence). To week 36, there are no breaches of the model predicted total deaths 95% confidence intervals, but every week thereafter breaches. The mean absolute model error for the predictions based on 2019 data to week 36 is 0.4%.

**Figure 6: The model predicted total deaths from registrations by the end of the second week following occurrence is accurate and very precise**

Predicted occurrences based on numbers registered by the end of the second week following occurrence from the p2\_bhw2 model plotted against observed 2019 data for England and Wales (weeks 1 to 36)



Source: Office for National Statistics



The model accurately predicts excess deaths in week 30; a predicted 9,678 total deaths (confidence interval: 9,553 to 9,824) compared with the observed 9,625. The precision of the estimates is greater at the end of the second week following occurrence than a week earlier, but the accuracy shows little improvement.

The variation in the proportion of deaths registered at subsequent weeks following occurrence becomes less and though some effects remain significant, the expected proportions are all within 0.5% of each other. Should estimation beyond the second week after occurrence be required, it is proposed that the null model is used for predictions after a delay of two or more weeks (Table 1).

## 5 . Caveats of the approach

There are two large caveats with this approach. The first is that there is evidence that an increased number of deaths can itself add to registration delay ([Appendix 1: Relationship between death registrations and death occurrences](#)). If the primary purpose of this method is to estimate excess deaths, then this is compromised; the method is likely to under-estimate excess deaths (beyond what is captured by the winter, or Quarter 1 (Jan to Mar), effect).

Secondly, the method relies on the past being a reliable predictor of the future. Recent events following the coronavirus (COVID-19) pandemic have shown that extraordinary events can change registration behaviour. Registrations are currently happening over weekends and bank holidays at a rate they have not previously done and the level of control over the death registration process has relaxed. It is not clear to what extent these factors affect the modelling shown.

The models presented are based on weekly data and so can be limited by the day of week any bank holiday falls. An alternative approach is to model delay by day of occurrence and delay period in days or working days. Preliminary trials of such methods found that though the models fitted historic data closely, when they were used to predict total deaths it was found that there were many breaches of the confidence intervals produced by the model predictions by the observed data. We interpret this as over-fitting to the original data and moved to the methods described in this article.

Though less precise, we are satisfied that the confidence intervals produced by the weekly method are a greater reflection of the predictive power of the models compared with those built on daily data. We have shown that a week after the occurrence we can reliably predict total deaths from partial registration data with high accuracy and precision, using 2019 registration data.

## 6 . Appendix 1: Relationship between death registrations and death occurrences

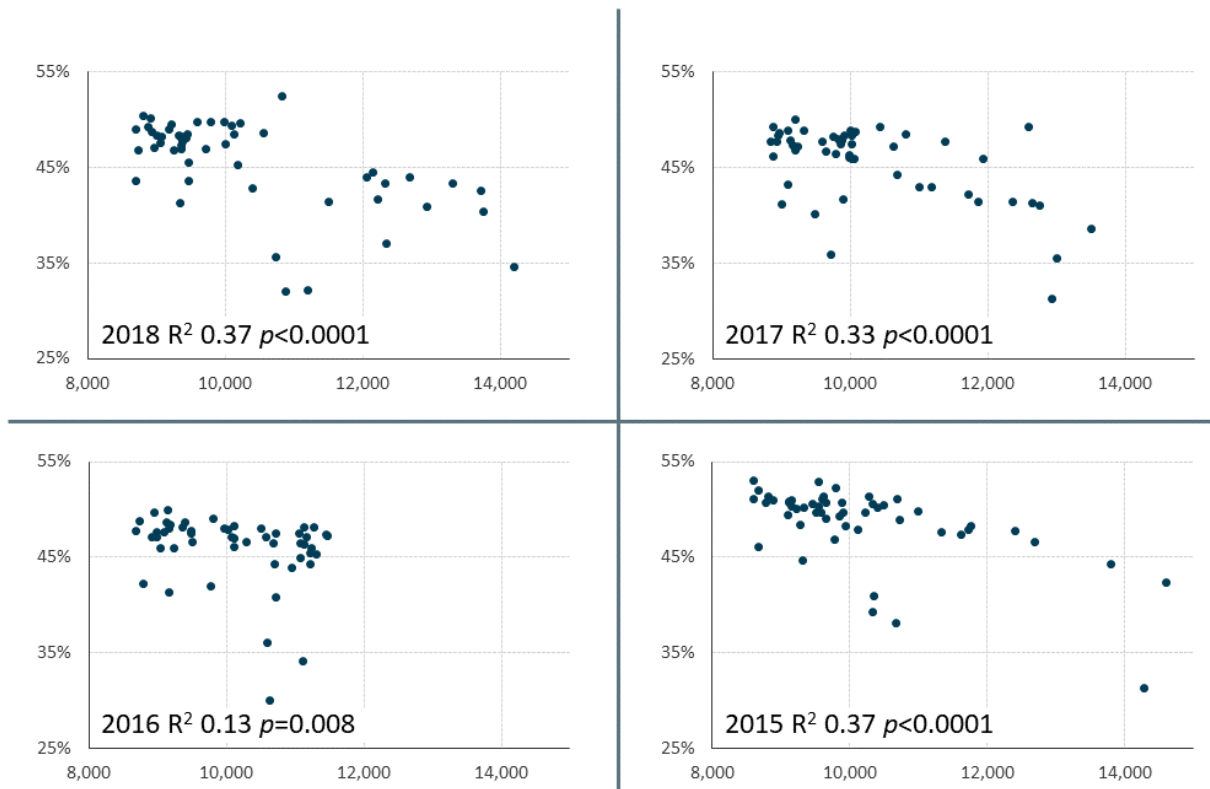
**The higher the number of occurrences, the lower the proportion of deaths that are registered in the same week as occurrence.**

There are lower proportions of deaths registered in the same week as occurrence in Quarter 1 (Jan to Mar) compared with the summer (Figure 3). There is strong evidence that the number of weekly deaths strongly associates with Quarter 1 in 2018 (t test,  $p < 0.001$ ; Mean weekly deaths Q2-4 = 9,529 ( $\pm 1,195$ ), Q1 (1) = 12,641 ( $\pm 1,760$ ); see also Figure 7). This raises the question whether the Quarter 1 effect is a consequence of seasonal excess deaths. The Quarter 1 effect is highly significant in  $p_0$  models built with 2018 and with combined 2015 to 2018 data, but not significant using 2016 data (see Table 5). Quarter 1 of 2016 was mild and there were on average 1,277 extra deaths per week in Quarter 1 of 2016 compared with the rest of the year. In 2018, there were 3,111 extra deaths per week in Quarter 1 compared with the rest of the year.



The number of deaths correlates with the proportion of weekly deaths that were registered in the same week (Figure 7). The year 2016 had no weeks with over 12,000 deaths and the correlation between the number of occurrences and proportion of weekly deaths that were registered in the same week is weaker (Figure 7).

**Figure 7: The number of death occurrences reduces the proportion that are registered within the same week**



**Notes:**

1. Data for England and Wales.

## 7 . Appendix 2: Modelling registration delay

The following equation is used to explore the relationship between explanatory factors and the proportion of deaths registered by time X:

$$\text{logit } p_{xi} = x_0 + x_1 X_{1i} + \dots + x_n X_{ni}$$

where:

- $p_{xi}$  is the proportion of deaths registered by time x for occurrence week i
- $X_{ni}$  is the value of the explanatory variable,  $X_n$ , for occurrence week i and  $x_n$  is the estimated size of the effect of  $X_n$  on  $\text{logit } p_x$
- $x_0$  is the intercept

Null models are expressed as follows:

$$\text{logit } p_{xi} = x_0$$

where the intercept  $x_0$  is calculated as the mean logit  $p_{xi}$ . These values were calculated and for  $p_0$  to  $p_5$  and presented in Table 1 (where  $p_0$  are the proportion of deaths that are registered in the same week that they occurred and  $p_5$  are the proportion of deaths that are registered by the end of the fifth subsequent week that they occurred).

An advantage of this formulation is that the variance is calculated as a single term. Models are built from two data sources: the single year's data for 2018 and the combined data from 2015 to 2018. It is expected that models created from the combined data will be most robust, as single year effects are smoothed out.

An initial categorical variable called `bhw_effect` was created. This variable takes one of six values based on the proximity of bank holidays to the week of occurrence. In the 209 weeks of the combined dataset:

- 142 weeks take the value "No\_BHW" as they do not occur within three weeks of a bank holiday
- 8 weeks take the value "OCCURRED\_BHW\_CONSEC" as the week of occurrence and the subsequent week both contain a bank holiday
- 20 weeks take the value "OCCURRED\_BHW" as the week of occurrence contains a bank holiday
- 12 weeks take the value "SUBSEQUENT\_BHW" as the week subsequent to occurrence contains a bank holiday
- 8 weeks take the value "CONSECUTIVE\_BHWs" as the week subsequent to occurrence and the following week both contain a bank holiday
- 19 weeks take the value "SECONDWEEK\_BHW" as the second subsequent week to occurrence contains a bank holiday

A second categorical value, called Q1, is created. This defines the winter (Quarter 1) effect and when the week of the year is 1 to 13 it takes a value of 1 otherwise 0.

## Estimating parameters of the expected proportion of occurrences that are registered in the same week ( $p_0$ )

Estimation of the intercept ( $\beta_0$ ) from the null model ( $\text{logit } p_{0i} = \beta_0$ ), based on data from 2018, is -0.1808 (Std Err, 0.0269) (Table 1; residual standard error 0.1941). Based on the combined data (2015 to 2018) the values are:  $\beta_0$  -0.1482 (Std Err, 0.0126); residual standard error 0.1825 [ $1/(1+\exp(\beta_0))= 46.3\%$ ].

The winter effect (Q1) is examined first (Table 5). The expected proportion of deaths registered in the same week of occurrence in Quarter 1, based on combined 2015 to 2018 data, is 43.3% and 47.3% outside of Quarter 1 (Table 5).

Table 5: Quantification of the winter effect, Q1, upon the expected proportion of deaths registered in the same week as occurrence (p), England and Wales, 2015 to 2018

	2016			2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	-0.1505	0.0275	<0.001	-0.1163	0.0256	<0.001	-0.1083	0.0135	<0.001
<b>Q1 (1)</b>	-0.0945	0.0556	0.0952	-0.258	0.0511	<0.001	-0.1601	0.0271	<0.001
<b>Residual</b>		0.1741	F, 2.9		0.1596	F, 25.5		0.1692	F, 35.0
<b>SSE/SST (R<sup>2</sup>)</b>	0.05			0.34			0.14		
<b>E[p   Q1 (0)]</b>	46.2% (37.9% - 54.8%)			47.1% (39.4% - 54.9%)			47.3% (39.2% - 55.6%)		
<b>E[p   Q1 (1)]</b>	43.9% (35.7% - 52.4%)			40.8% (33.5% - 48.5%)			43.3% (35.4% - 51.6%)		

Source: Office for National Statistics

#### Notes

1. The winter effect appears to depend on excess deaths as it is not significant in 2016 which saw a mild winter and fewer excess deaths than usual. [Back to table](#)

The effect of bank holidays is examined next, using the variable `bhw_effect` described previously. When examining the impact of this variable upon the proportion of deaths that are registered in the same week as occurrence it is found that two categories are not significant (“CONSECUTIVE\_BHWs” and “SECONDWEEK\_BHW”) and a third, “SUBSEQUENT\_BHW”, is only marginally significant (not shown). All these categories depend on bank holidays falling outside of the week of occurrence and not on those with bank holidays falling in the week of occurrence.

A second explanatory variable was constructed, `bhw0`, with just three categories:

- “OCCURRED\_BHW\_CONSEC” (Easter and Christmas, where the week of occurrence is followed by another bank holiday week)
- “OCCURRED\_BHW” (the remaining weeks of occurrence containing a bank holiday)
- “No\_BHW”, the reference category

The expected proportion of deaths registered in the same week as occurrence is strongly affected by bank holidays in the week of occurrence and the subsequent week (Table 6). The effect is still strong, but not as severe, when there is a bank holiday in the week of occurrence but not the subsequent week.

Table 6: Quantification of the bank holiday effect, bhw0, upon the expected proportion of deaths registered in the same week as occurrence ( $p_{0}$ ), England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	-0.1294	0.02	<0.001	-0.0977	0.0092	<0.001
<b>OCCURRED_BHW_CONSEC</b>	-0.6188	0.0969	<0.001	-0.5541	0.0445	<0.001
<b>OCCURRED_BHW</b>	-0.2876	0.0632	<0.001	-0.3061	0.0291	<0.001
<b>Residual</b>		0.134	F, 29.0		0.1233	F, 124.9
<b>SSE/SST (R<sup>2</sup>)</b>	0.54			0.55		
<b>E[p   No_BHW ]</b>	46.8%	(40.3% - 53.3%)		47.6%	(41.6% - 53.6%)	
<b>E[p   OCCURRED_BHW_CONSEC ]</b>	32.1%	(26.7% - 38.1%)		34.3%	(29.0% - 39.9%)	
<b>E[p   OCCURRED_BHW ]</b>	39.7%	(33.6% - 46.2%)		40.0%	(34.4% - 46.0%)	

Source: Office for National Statistics

The two effects were combined into a single model (called  $p_{0\_bhw0q1}$ ) that marks a significant improvement in model fit (Table 7). Between 71% and 82% of the weekly variation in the proportion of occurrences that are registered in the same week as occurrence is explained by the  $p_{0\_bhw0q1}$  model (Table 7; SSE/SST ( $R^2$ ), logit scale).

Table 7: Parameterisation of  $p_{0\_bhw0q1}$  models that utilise bhw0 and Q1 as explanatory factors to estimate the expected proportion of deaths registered in the same week as occurrence, England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	-0.0714	0.0142	<0.001	-0.0550	0.0084	<0.001
<b>OCCURRED_BHW_CONSEC</b>	-0.5582	0.0613	<0.001	-0.5548	0.0360	<0.001
<b>OCCURRED_BHW</b>	-0.2891	0.0397	<0.001	-0.3152	0.0235	<0.001
<b>Q1 (1)</b>	-0.2370	0.0272	<0.001			<0.001
<b>Residual</b>		0.0842	F, 74.3		0.0996	F, 164.5
<b>SSE/SST (R<sup>2</sup>)</b>	0.82			0.71		

<ons-table-v2

## Estimating parameters of the expected proportion of occurrences that are registered by the end of the subsequent week ( $p_1$ )

Estimation of the intercept ( $\alpha_0$ ) from the null model (logit  $p_{1i} = \alpha_0$ ), based on data from 2018, is 1.6949 (Std Err, 0.0210) (Table 1; residual standard error 0.1500). Based on the combined data (2015-2018) the values are:  $\alpha_0 = 1.7544$  (Std Err, 0.0118); residual standard error 0.1700 [ $1/(1+\exp(-\alpha_0)) = 85.3\%$ ].

The winter effect is much weaker in explaining variation in the proportion of occurrences that are registered by the end of the subsequent week ( $p_1$ ), compared with in the same week ( $p_0$ ). The winter effect is significant in 2018 but only marginally when combined 2015 to 2018 data are examined (Table 8).

Table 8: Quantification of the winter effect, Q1, upon the expected proportion of deaths registered by the end of the week following occurrence (p), England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	1.7366	0.0212	<0.001	1.7741	0.0133	<0.001
<b>Q1 (1)</b>	-0.1666	0.0424	<0.001	-0.0792	0.0266	0.003
<b>Residual</b>		0.1324	F, 15.4		0.1664	F, 8.8
<b>SSE/SST (R<sup>2</sup>)</b>	0.24			0.04		
<b>E[p   Q1 (0)]</b>	85.0% (81.4% - 88.0%)			85.5% (81.0% - 89.1%)		
<b>E[p   Q1 (1)]</b>	82.8% (78.8% - 86.2%)			84.5% (79.7% - 88.3%)		

Source: Office for National Statistics

The effect of bank holidays upon the proportion of deaths that are registered by the end of the week subsequent to occurrence was examined using `bhw_effect`. The “SECONDWEEK\_BHW” category, which relates to the presence of a bank holiday in the second week after occurrence, was not significant in models explaining the proportion of deaths that are registered by the end of the week subsequent to occurrence. A second explanatory variable was constructed, `bhw1_1`, that merges the “SECONDWEEK\_BHW” into the reference category “No\_BHW”.

This five-category variable, `bhw1_1`, explains about the same proportion of variation in the proportion of deaths that are registered by the end of the week subsequent to occurrence as the three-category variable, `bhw0`, did for variation in the proportion of deaths that are registered in the same week as occurrence (combined 2015 to 2018 data: Table 9, 56%; compared with Table 6, 55%).

Table 9: Quantification of the bank holiday effect, `bhw1_1`, upon the expected proportion of deaths registered by the end of the week following occurrence (p), England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	1.7491	0.0146	<0.001	1.8099	0.0089	<0.001
<b>OCCURRED_BHW_CONSEC</b>	-0.5103	0.0668	<0.001	-0.5707	0.0411	<0.001
<b>OCCURRED_BHW</b>	-0.1407	0.0438	0.002	-0.1356	0.0269	<0.001
<b>SUBSEQUENT_BHW</b>	-0.1438	0.0552	0.012	-0.1619	0.034	<0.001
<b>CONSECUTIVE_BHWs</b>	-0.3299	0.0668	<0.001	-0.2973	0.0411	<0.001
<b>Residual</b>		0.0923	F, 21.9		0.1135	F, 65.0
<b>SSE/SST (R<sup>2</sup>)</b>	0.65			0.56		
<b>E[p   No_BHW ]</b>	85.2% (82.8% - 87.3%)			85.9% (83.0% - 88.4%)		
<b>E[p   OCCURRED_BHW_CONSEC ]</b>	77.5% (74.2% - 80.5%)			77.5% (73.4% - 81.2%)		
<b>E[p   OCCURRED_BHW ]</b>	83.3% (80.7% - 85.7%)			84.2% (81.0% - 87.0%)		
<b>E[p   SUBSEQUENT_BHW ]</b>	83.3% (80.6% - 85.6%)			83.9% (80.6% - 86.7%)		
<b>E[p   CONSECUTIVE_BHWs ]</b>	80.5% (77.5% - 83.2%)			81.9% (78.4% - 85.0%)		

Source: Office for National Statistics

The categories "OCCURRED\_BHW" and "SUBSEQUENT\_BHW" relate to the week of occurrence containing a bank holiday or the week after, respectively. Based on visual examination of Figure 3, it was expected that the category "SUBSEQUENT\_BHW" would have a stronger effect than "OCCURRED\_BHW". However, they have an almost identical effect upon the proportion of occurrences registered by the end of the subsequent week (Table 9) and the categories are combined to create a new category, called "OCCURRED\_or\_SUBSEQUENT", in bhw1 (Table 10). The revised variable, bhw1, performs better than bhw1\_1, and is used going forwards (Table 10).

Table 10: Quantification of the bank holiday effect, bhw1, upon the expected proportion of deaths registered by the end of the week following occurrence (p), England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	1.7491	0.0144	<0.001	1.8099	0.0089	<0.001
<b>OCCURRED_BHW_CONSEC</b>	-0.5103	0.0661	<0.001	-0.5707	0.0411	<0.001
<b>OCCURRED_or_SUBSEQUENT</b>	-0.1418	0.0354	<0.001	-0.1455	0.0219	<0.001
<b>CONSECUTIVE_BHWs</b>	-0.3299	0.0661	<0.001	-0.2973	0.0411	<0.001
<b>Residual</b>		0.0913	F, 29.9		0.1133	F, 86.8
<b>SSE/SST (R<sup>2</sup>)</b>	0.65			0.56		
<b>E[p   No_BHW ]</b>	85.2% (82.8% - 87.3%)			85.9% (83.0% - 88.4%)		
<b>E[p   OCCURRED_BHW_CONSEC]</b>	77.5% (74.3% - 80.5%)			77.5% (73.4% - 81.2%)		
<b>E[p   OCCURRED_or_SUBSEQUENT]</b>	83.3% (80.7% - 85.6%)			84.1% (80.9% - 86.8%)		
<b>E[p   CONSECUTIVE_BHWs]</b>	80.5% (77.6% - 83.2%)			81.9% (78.4% - 85.0%)		

Source: Office for National Statistics

The two effects were combined into a single model (called p1\_bhw1q1) that marks a significant improvement in model fit (Table 11). Between 61% and 82% of the weekly variation in the proportion of occurrences that are registered by the end of the week subsequent to occurrence is explained by the p1\_bhw1q1 model (Table 11; SSE/SST (R<sup>2</sup>), logit scale).

Table 11: Parameterisation of p1\_bhw1q1 models that use bhw1 and Q1 as explanatory factors to estimate the expected proportion of deaths registered by the end of the week subsequent to occurrence, England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	1.785	0.0118	<0.001	1.8333	0.0096	<0.001
<b>OCCURRED_BHW_CONSEC</b>	-0.4744	0.0484	<0.001	-0.5722	0.0388	<0.001
<b>OCCURRED_or_SUBSEQUENT</b>	-0.1598	0.0258	<0.001	-0.1579	0.0209	<0.001
<b>CONSECUTIVE_BHWs</b>	-0.294	0.0484	<0.001	-0.2878	0.0388	<0.001
<b>Q1 (1)</b>	-0.1436	0.0217	<0.001	-0.0876	0.0173	<0.001
<b>Residual</b>		0.0663	F, 53.4		0.1071	F, 79.4
<b>SSE/SST (R<sup>2</sup>)</b>	0.82			0.61		

Source: Office for National Statistics

## Estimating parameters of the expected proportion of occurrences that are registered by the end of the second week after occurrence ( $p_2$ )

Estimation of the intercept ( $\beta_0$ ) from the null model ( $\text{logit } p_{2i} = \beta_0$ ), based on data from 2018, is 2.2506 (Std Err, 0.0098) (Table 1; residual standard error 0.0703). Based on the combined data (2015 to 2018) the values are:  $\beta_0 = 2.2846$  (Std Err, 0.0061); residual standard error 0.0873 [ $1/(1+\exp(-\beta_0)) = 90.8\%$ ].

The total variation (SST) in the 2015 to 2018 derived null models has fallen from 6.9302 ( $p_0$ ) through 5.9798 ( $p_1$ ) to 1.5854 for the proportion (logit scale) of occurrences that are registered by the end of the second week after occurrence ( $p_2$ ). So by the end of the second week after occurrence there is much less variation to explain and the null model may be sufficient for our needs.

The bank holiday effect remains significant, though much less than in earlier models ( $F_{5,203} = 12.7$ ). The categories "OCCURRED\_BHW", "SUBSEQUENT\_BHW" and "SECONDWEEK\_BHW" all relate to containing a single week with bank holidays out of the three-week period. They had identical effects and are combined to create a new category, called "SINGLE\_BHW", in bhw2.

Similarly, the categories "OCCURRED\_BHW\_CONSECUTIVE\_BHWs" and "CONSECUTIVE\_BHWs" relate to containing two consecutive weeks with bank holidays out of the three-week period. These too had identical effects and are combined to create a new category, called "TWO\_BHWs", in bhw2. This variable has three categories, with "No\_BHW" being the reference category. The impact of bhw2 on explaining variation in the proportion of occurrences that are registered by the end of the second week after occurrence is investigated in Table 12.

Table 12: Quantification of the bank holiday effect, bhw2, upon the expected proportion of deaths registered by the end of the second week after occurrence ( $p$ ), England and Wales, 2015 to 2018

	2018			2015 to 2018		
	Est	Std Err	Pr	Est	Std Err	Pr
<b>Intercept</b>	2.2759	0.0096	<0.001	2.3081	0.0064	<0.001
<b>TWO_BHWs</b>	-0.1476	0.0299	<0.001	-0.1506	0.0203	<0.001
<b>SINGLE_BHW</b>	-0.0559	0.0184	0.004	-0.049	0.0125	<0.001
<b>Residual</b>		0.0567	F, 14.7		0.0768	F, 31.4
<b>SSE/SST (<math>R^2</math>)</b>	0.37			0.23		

Source: Office for National Statistics

## 8 . Appendix 3: Measuring performance against 2019 data

To test whether 2019 showed a broadly similar delay profile as 2015 to 2018, the mean proportion of deaths occurred by weeks registration delay was examined for all data in the analysis (209 weeks combined for 2015 to 2018; 42 weeks combined for 2019). The mean proportion of deaths registered by weeks since occurrence tended to be higher in 2019 than in the combined 2015 to 2018 data (Table 13). A caveat of this analysis is that part of this may be because the period for registering 2019 deaths is shorter.

Table 13: A greater proportion of deaths tended to be registered by each weeks' delay in 2019 than in 2015 to 2018, England and Wales

Proportion of deaths registered by weeks' delay for 2019 in comparison with 2015 to 2018 data

Year	Registration delay (weeks since occurrence)										
	0	1	2	3	4	5	10	15	20	25	
2019	<b>47.0%</b>	<b>85.5%</b>	<b>91.2%</b>	92.5%	93.1%	93.5%	94.8%	96.0%	96.9%	97.7%	
2015 to 2018	<b>46.0%</b>	<b>85.0%</b>	<b>90.7%</b>	92.0%	92.5%	92.9%	94.3%	95.6%	96.8%	97.7%	

Notes

1. Data for England and Wales. [Back to table](#)
2. The proportions registered after a short delay are the focus of this study (bold). [Back to table](#)

Model assessment is presented graphically and is based on analysis of variance of model predicted registration rates against those observed ( $SSE_{\text{model}}/SST$ ), the mean absolute error (proportion of the expected count) and the number of breaches of the 42 observed values over 95% confidence intervals.

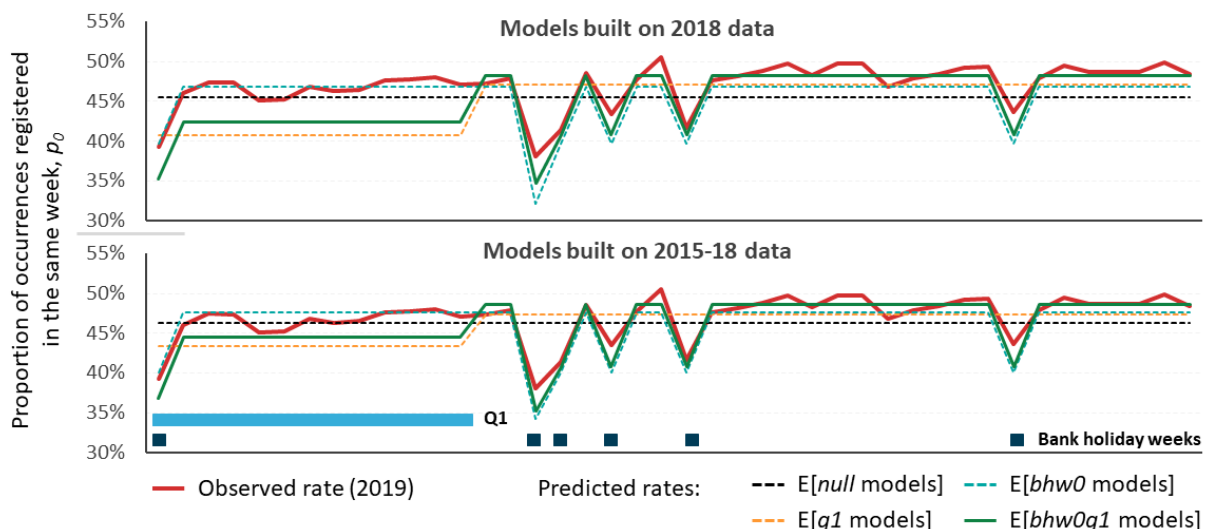
The predicted number of occurrences is based on the estimated proportion registered by time  $x$ ,  $p_x$ . The number of registrations observed by time  $x$  ( $R_x^{\text{obs}}$ ) is multiplied by the reciprocal of  $p_x$ , such that Predicted =  $(R_x^{\text{obs}})/p_x$ . Error in predictions is given as a proportion of expected values.

### Performance of models based on the number registered in the same week as occurrence ( $p_0$ )

All models tend to under-estimate the proportion registered in the same week of occurrence (Figure 8). In all cases, the 2015 to 2018 models perform better than the 2018 model (Table 14). The  $p0\_bhw0q1$  and  $p0\_bhw0$  models perform best (Table 14).



**Figure 8: Expected proportions of deaths that are registered in the same week as occurrence based on models built on 2018 and 2015-2018 data plotted against the observed 2019 data**



Source: Office for National Statistics

Table 14: Accuracy of the  $p_0$  models to predict total death occurrences from the number of registrations that occur in the same week as death occurrence in 2019

	2018			2015 to 2018		
	SSE/SST	Mean  error	Breaches	SSE /SST	Mean  error	Breaches
$p_0\_null$	0	6.0%	0	0	5.1%	0
$p_0\_q1$	0.043	7.2%	1	0.043	5.6%	1
$p_0\_bhw0$	0.766	3.7%	1	0.777	2.8%	0
$p_0\_bhw0q1$	0.733	4.5%	9	0.844	3.0%	0

Source: Office for National Statistics

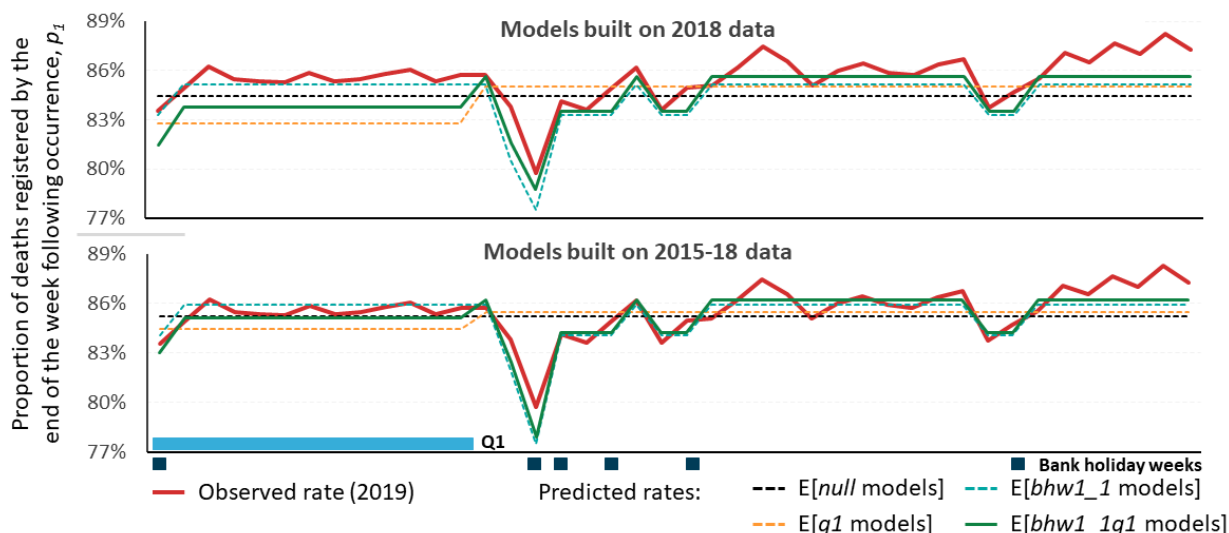
#### Notes

1. Proportion of sum of squares (rate measure) predicted by the models mean absolute error (proportion of E [count]) and breaches of 42 observed values over 95% confidence intervals are shown. [Back to table](#)

## Performance of models based on the number registered by the end of the week following occurrence ( $p_1$ )

All models tend to under-estimate the proportion registered in the same week of occurrence (Figure 9), but care should be taken interpreting data relating to late 2019 as the true number of occurrences is likely to be under-estimated by 2% (Table 13) leading to a higher than actual proportion of deaths registered by the end of the week following occurrence. In all cases, the 2015 to 2018 models perform better than the 2018 models and  $p_1\_bhw1q1$  and  $p_1\_bhw1$  models perform best (Table 15).

**Figure 9: Expected proportions of deaths that are registered by the end of the week following occurrence based on models built on 2018 and 2015-2018 data plotted against the observed 2019 data**



Source: Office for National Statistics

Table 15: Accuracy of the p1 models to predict total death occurrences from the number of deaths registered by the end of the week following occurrence in 2019.

	2018			2015 to 2018		
	SSE/SST	Mean  error	Breaches	SSE /SST	Mean  error	Breaches
<b>p1_null</b>	0	1.7%	2	0	1.2%	1
<b>p1_q1</b>	0.003	2.1%	3	0.003	1.4%	1
<b>p1_bhw1</b>	0.689	1.2%	4	0.701	0.8%	0
<b>p1_bhw1q1</b>	0.707	1.4%	13	0.776	0.7%	0

Source: Office for National Statistics

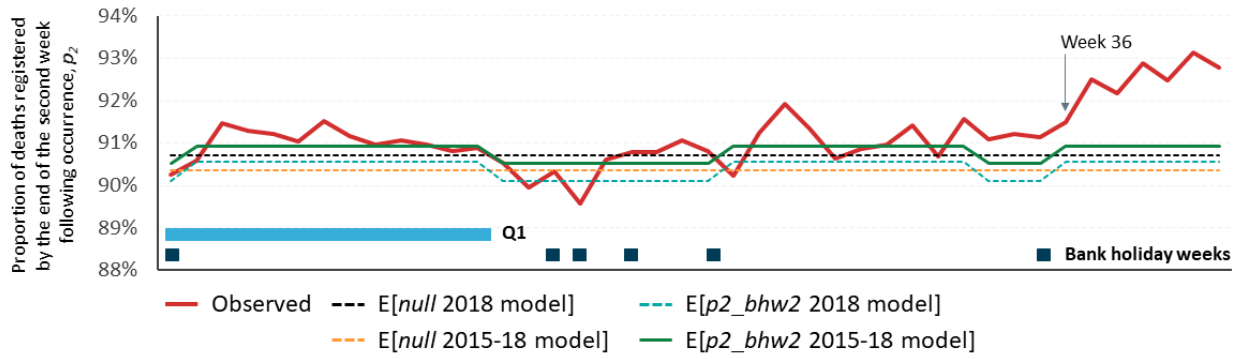
Notes

1. Proportion of sum of squares (rate measure) predicted by the models, mean absolute error (proportion of E [count]) and breaches of 42 observed values over 95% confidence intervals are shown. [Back to table](#)

**Performance of models based on the number registered by the end of the second week following occurrence (p<sub>2</sub>)**

From 37 weeks, the observed proportions registered are out of line with historical observations and their derived models (Figure 10). Evaluation is therefore based on the first 36 weeks of 2019 (Table 16). To week 36, there are no breaches of the model predicted total deaths 95% confidence intervals, but every week thereafter breaches. The mean absolute model error for the predictions based on 2019 data to week 36 is 0.4%.

**Figure 10: Expected proportions of deaths that are registered by the end of the second week following occurrence based on models built on 2018 and 2015-2018 data plotted against the observed 2019 data**



Source: Office for National Statistics

Table 16: Accuracy of the p2 models to predict total death occurrences from the number of deaths registered by the end of the second week following occurrence in 2019.

	2018			2015-18		
	SSE/SST	Mean  error	Breaches	SSE /SST	Mean  error	Breaches
<b>p2_null</b>	0	0.7%	4	0	0.7%	0
<b>p2_bhw2</b>	0.234	0.5%	9	0.234	0.4%	0

Source: Office for National Statistics

Notes

1. Proportion of sum of squares (rate measure) predicted by the models, mean absolute error (proportion of E [count]) and breaches of 36 observed values over 95% confidence intervals are shown. [Back to table](#)

## 9 . Related links

### [Deaths registered weekly in England and Wales, provisional](#)

Statistical bulletin | Released weekly

Provisional counts of the number of deaths registered in England and Wales, including deaths involving the coronavirus (COVID-19), by age, sex and region, in the latest weeks for which data are available.

### [Where to find statistics on UK deaths involving the coronavirus \(COVID-19\) and infection rates by country](#)

Web page | Updated regularly

Links to statistics on coronavirus (COVID-19) deaths and infection rates published by the different constituent countries of the UK.

### [Coronavirus \(COVID-19\) latest data and analysis](#)

Web page | Updated when new data are available

Brings together the latest data and analysis on the coronavirus (COVID-19) pandemic in the UK and its effect on the economy and society.

### [Coronavirus \(COVID-19\) roundup](#)

Article | Updated as and when data become available

Catch up on the latest data and analysis related to the coronavirus (COVID-19) pandemic and its impact on our economy and society.