Options Paper: Estimating the number of expected deaths (baseline) for calculating excess deaths

1. EXECUTIVE SUMMARY

This paper reports on an option appraisal on possible approaches to defining the expected number of deaths (baseline) for the calculation of excess mortality in official statistics such as the Deaths registered weekly releases produced by ONS, NRS and NISRA. Defining the baseline is a key requirement within the review of excess mortality statistics.

We ask the Methodological Assurance Review Panel (MARP) to review and comment on:

- The process and methods of the option appraisal
- The appropriateness of the short-listed options for the stated purpose
- Which factors point to any one of the short-listed options as the preferred choice

MARP's advice will inform the National Statistician's decision on methods and the next steps of the review project.

The option appraisal by a cross-GSS (Government Statistical Service) working group (plus experts from the actuarial profession) considered a long-list of 14 methods of defining the baseline, some of which themselves included several possible variants. The majority were rejected on one or more decision criteria (see Section 4 and Annex 4) leaving a short-list of two methods (Section 5) shown below along with two alternative options (Table 1).

TABLE 1. SHORT-LISTED OPTIONS AND RECOMMENDATIONS

Option	Name	Recommendation	Pros	Cons
0	No change	Reject	Continuity, no cost of changes	No benefit – project aims not met
1	Five-year trend (five-year average of ASMRs adjusted for trend)	First preference	Easy to explain. Takes into account recent trends in mortality. Fairly unbiased. Less sensitive to large annual changes. Transparent and easy to replicate. Good fit to historical data. Suitable for UK and the four nations Easily scalable by subnational area or population characteristic. Preferred methodology by the majority of the working group.	Novel methodology. Fairly volatile. Requires bank holiday adjustment or caveat. Some concern from the group in regard to robustness of method, including relative influence of past years.
2	ARIMA	Second preference	Established time-series methodology. Takes into account recent mortality trends. Fairly unbiased. Easy to apply bank holiday adjustments. Suitable for UK and the four nations.	Less transparent - harder to explain. Needs more statistical expertise to use. More complex on scalability, needs further investigation – applications at different scales may be incoherent with each other. Fairly volatile, sensitive to large changes in past years.
3	No decision now - further testing	Reject	Opportunity to consider more methods	Decision needed to progress

Note that the above recommendation represents the majority of views among the working group - there was no unanimity on the preferred final choice, or on whether the drawbacks of certain long-listed methods were enough to clearly exclude them without further testing.

There are elements of methodology that the group have not addressed in this paper but will be investigating further. These can be found in Annex 3.

2. BACKGROUND

2.1 INTRODUCTION

The background was previously outlined in a paper to the Methodological Assurance Review Panel's (MARP) April 2023 meeting, which is attached below as Annex 1. The remit and membership of the cross-GSS (Government Statistical Service) working group are also attached for reference as Annex 2.

Excess mortality is defined as the number or rate of deaths above what we would expect in a given period based on historical data. The number or rate of deaths we would expect, or baseline, is an essential part of the calculation and clearly has a major influence on the resulting excess mortality figure. As highlighted in an ONS blog in February 2023, various organisations report on excess mortality using different calculations, including different approaches to the baseline. This is likely to be confusing for users who want a definitive answer to the extent of excess mortality and it is important we work towards a coherent approach across government.

The baseline is often described as the mortality of an 'average year', that is, what we would most commonly experience or expect in the absence of a pandemic or other exceptional cause of mortality. We are not including the probability of a future pandemic or the probability of a high or low flu year in the baseline. That is, we want to estimate the expected number or rate of deaths based on no pandemic, an average flu year, and no other exceptional circumstances affecting health. These are events which we would expect to contribute to the calculated excess mortality as deviations from the 'average' or 'usual'. Therefore, from a statistical perspective, where we say average in this context we actually mean something closer to 'Mode' rather than 'Mean', as we don't want outliers such as a pandemic to influence the baseline.

2.2 USER NEEDS AND STAKEHOLDER ENGAGEMENT

Reporting of excess mortality on a regular basis is of high public interest and helps to guide government policy and the management of NHS services. ONS weekly, monthly and annual mortality publications are used widely and comparable publications released by other UK statistical producers also have a high profile. It is important for data on excess mortality to be timely, accurate, and understandable to the range of likely users. The method should be relatively straightforward for most users to replicate and verify.

There was especially strong public interest in these figures during the pandemic periods of March 2020 onwards, highlighting the need for the methods used to be transparent, easily replicable, and clearly communicated. The use of different methods of calculation by different statistics providers is liable to cause confusion. The importance of excess mortality figures in the pandemic also highlighted some weaknesses in the method historically used by ONS, which this project aims to solve.

The February 2023 ONS blog asked for user input on the issues and resulted in twelve responses, some of which had suggestions regarding further information on methodology or support for methods suggested. Suggestions which contributed to the working group's discussions mentioned particular time-series methods, length of the baseline time period, relative weighting of long-term and short-term trends, handling of the pandemic period, and mortality displacement. Comments on the accuracy of death certification were considered out of scope. An additional methodological approach suggested was ensemble forecasts as used in some weather applications, however this was not considered relevant to the excess mortality reporting scenario as it works by creating multiple versions of a forecast which we considered overly complex and less reproducible.

2.3 Scope of the project and of this paper

For the purposes of the project, the following overall definition has been applied:

> *The primary purpose of excess mortality reporting is to monitor short-term deviation from the expected level, so as to warn of emerging health emergencies and inform response to health-related hazards. The purpose is not to measure long-term trends or patterns in mortality, for example, the slow-down in mortality improvements seen across Europe in the early 2010s. The expected number of deaths reflects a 'usual' level for the UK and constituent countries, in a contemporary time period during which there are no unusual factors or substantial increasing or decreasing mortality.*

The scope of this project covers excess mortality methods as a whole, including issues around the handling of bank holidays and weekends (where we see few registrations), future treatment of the 2020 pandemic mortality 'spikes', and the delay between date of death and registration. These questions will be further addressed in the light of the choice of method for calculation of the baseline (the subject of this paper) and will form the next stage of the project. Brief background information on the further issues mentioned above can be found in Annex 3 for reference.

The final stage of the project will be communication and implementation of the agreed methods. Once the methodology has been agreed, colleagues from the group will create a plan on how to communicate the change to users and how this will be implemented in a coherent way across organisations.

3. Decision Criteria

The first set of decision criteria was used to narrow the long-list of methods we considered to a smaller number. The criteria we applied were that the method must:

- Be suitable at the UK, England, Wales, Scotland and Northern Ireland levels. Thus, the data used must only include sources which are available and comparable for all four constituent countries.
- Be sufficiently transparent and easy to calculate for public understanding and replication by a range of interested stakeholders.
- Take into account long-term trends in mortality, demographic change (population size and age distribution) and expected seasonal variation in mortality.
- Be appropriate for reporting on all standard lengths of periods – annual, and more importantly, weekly and monthly.
- Allow disaggregation by geography, at least down to lower tier local authority level.
- Allow disaggregation by sex and preferably other demographic characteristics for which data may be available.
- Allow the calculation of an interval estimate/measure of uncertainty as well as a point estimate.

The second set of criteria was used to choose which of the remaining, short-listed methods was the most preferred. We measured both approaches against the following criteria:

- Volatility – not overly susceptible to large changes in mortality from one year to the next (e.g the known increased mortality in 2015 and decreased mortality in 2019)
- Bias – not estimating deaths consistently either above or below average but broadly representative of the historical data
- Ease of use – a method that can be easily applied using a range of common software
- Transparency – a method that is transparent is preferable, in line with the code of practice
- Ease of replication – able to be implemented by different organisations and individuals for various breakdowns, geographies, etc
- Ease of explanation – a method that is easy to explain to users is preferable to one that has complex methodology
- Scalable - appropriate to run at the UK and four nation level, and for lower geographies and flexible areas.

For testing we used the expected number of deaths annually from 2006 and monthly from 2007. Data from 2001 onwards was used for both methods.

4. LONG-LISTED OPTIONS

4.1 OVERVIEW OF PROCESS

The project started with a list of 14 different methods that had the potential to provide an expected number of deaths to be used in calculating excess mortality. More information on each different method can be found in Annex 4.

Five methods were discounted prior to running analysis. These were:

- CMI mortality projections and CMI pandemic monitor, which were disregarded because the methodology is only available for paid subscribers, which clearly does not meet the requirements for transparency and reproducibility.
- EuroMoMo as it was designed to compare mortality to a ‘best-case’ scenario baseline, rather than a ‘most-likely’ baseline and because of other issues, particularly its volatility.
- OHID’s excess deaths methodology as it was specifically designed for the pandemic, uses variables not available at all UK geographies and is more complex than some of the other methodologies. However, there was disagreement on this method which some members of the working group considered feasible with development (see Annex 4, A4.5 below).
- The WHO pandemic mortality model as the methodology was deemed too complex to be transparent for users and not entirely relevant to the UK.

Following early testing we decided to stop testing 9 other methods:

- Five-year average and SARIMA based on numbers were not taken forward as they do not account for population change.
- Segmented regression analysis did not account for seasonality or provide feasible expected deaths at the weekly level.
- Holt-Winters provided estimates of expected mortality, however the results were very sensitive to recent trends and were incorrectly estimating high flu years.
- Various neural networks were tested, however the results were less robust than some of the other methods and neural networks are more difficult to explain to users, so these were also ruled out at this stage.
- The five-year average (on rates) and rASMRs were ruled out as they did not account for mortality trends.

However, we did introduce a new method based on the issues discussed up to that point, which was the five-year average with a trend element which we are currently calling the trend-adjusted five-year average. We initially ruled out ARIMA due to the fact it does not account for seasonality, however we then discussed performing the calculations at an annual level and applying this trend to the monthly and weekly data. Therefore, the trend-adjusted five-year average and ARIMA continued to be tested.

4.2 Summary of Findings

An overview of the methods considered is shown in Table 2, this table is completed on the basis of using age-standardised mortality rates.

Table 2 - Table of methods and suitability to each criteria element

				Take into account			Suitable for			
Corresponding section	Method	Suitable for UK and four nations	Transparent and easy to replicate	long-term trends in mortality	Demographic changes	Seasonal variation	Weekly, monthly, annual data	Geography breakdowns to Local Authority	Demographic characteristics	Interval estimate
A4.2	CMI Mortality Projections	✓	✓	✓	✓	✓	✓	✓	✓	✓
A4.2	CMI pandemic Monitor	✓	✓	✓	✓	✓	✓	✓	✓	✓
A4.3	EuroMoMo	✓	✓	✓	✓	✓	×	×	×	✓
A4.4	Five-year average	✓	✓	×	✓	✓	✓	✓	✓	✓
A4.5	OHID pandemic excess deaths measure	(✓)	(✓)	✓	✓	✓	✓	✓	(✓)	✓
A4.6	UKHSA daily excess mortality method	⃠	✓	✓	✓	✓	×	⃠	×	✓
A4.7	World Health Organisation (WHO) model	⃠	×	✓	✓	⃠	✓	⃠	⃠	⃠
A4.8	Farrington model	⃠	(✓)	✓	✓	✓	⃠	⃠	⃠	⃠
A4.9	Holt-Winters model	⃠	(✓)	✓	✓	✓	×	⃠	⃠	✓

				Take into account			Suitable for			
Corresponding section	**Method**	**Suitable for UK and four nations**	**Transparent and easy to replicate**	**long-term trends in mortality**	**Demographic changes**	**Seasonal variation**	**Weekly, monthly, annual data**	**Geography breakdowns to Local Authority**	**Demographic characteristics**	**Interval estimate**
A4.10	Neural networks	⃠	×	✓	✓	✓	⃠	⃠	⃠	✓
A4.11	SARIMA	✓	✓	✓	✓	✓	✓	(✓)⃠	⃠	✓
A4.12	Segmented regression analysis	×	(✓)	×	✓	×	×	✓⃠	(✓)	✓
5.3	ARIMA	✓	✓	✓	✓	×	✓	(✓)⃠	✓ ⃠	✓
5.4	Trend-adjusted five-year average	✓	✓	✓	✓	✓	✓	✓ ⃠	✓ ⃠	✓

Key: ✓ Meets Criteria (✓) Possibly meets criteria × Does not meet criteria ⃠ Not tested

5. SHORT-LISTED OPTIONS

5.1 OVERVIEW

The two methods short-listed were ARIMA and a trend-adjusted five-year average rate, which were both based on ASMRs. (The latter could also be called a 'five-year trend' for simplicity, the acronym .) For these methods we used an annual model to determine the trend and applied this to monthly and weekly time periods. This avoids the volatility found at smaller breakdowns of mortality data. However, we need to do further testing to make sure we are using the most robust option.

5.2 SUMMARY – ARIMA

ARIMA is an autoregressive integrated moving average model without seasonality, SARIMA is the same with a seasonality element. More details on SARIMA can be found in Annex 4, A4.11.

An ARIMA model uses past mortality data in order to predict a future value, giving more weight to recent years. It is a well-known time series methodology that works on any data that exhibits a pattern and is available in constant, regular intervals. Rather than using the observed data to predict future data, an ARIMA model uses the difference between the observed and a fitted value in order to make a prediction of the future value. As we are running the models at an annual period we do not need to take into account any seasonality, for this reason we have chosen ARIMA over SARIMA. Clarification is needed on how well this accounts for the longer-term trends.

5.3 SUMMARY – TREND-ADJUSTED FIVE-YEAR AVERAGE RATES

The working group recognised the benefits of the five-year average currently used by ONS, which is easy to understand and smooths random fluctuations, however it does not take into account the long-term trends seen in mortality data. There are benefits to a trend calculation as in OHID's method, to take into account the mortality improvements seen across years. However this method tends to be reactive to any large fluctuations year on year. By combining the two methodologies we intended to combine the strengths while reducing the limitations.

We tested a number of combinations of trend and average, the preferred version was to use the linear trend of the last five years to predict the number of deaths in the current and next two periods. The average is then taken of the two previous years and the predicted values for the three further periods. For example, for reference year X, we use a least squares regression trend between X-5 to X-1 to predict X, X+1 and X+2. We then take an average of X-2, X-1, X, X+1 and X+2.

5.4 Volatility

To test for volatility we performed a sensitivity analysis by artificially increasing or decreasing the ASMR in one year by 10%. We did this for each year and direction of change independently. We were then able to record the average change a year made and how close the year was to the reference period. Table 3 shows the average, minimum and maximum percentage changes in expected mortality rates when altering an ASMR used by 10%. For example, when decreasing the year before the reference period's ASMR (X-1) by 10% then on average the expected ASMR for the reference period, on average, increases by 0.1% in SARIMA and decreases by 8.1% using the trend-adjusted five-year average.

As we do not want the methodology we choose to be susceptible to large changes (e.g a particularly good or bad flu season or mortality year) we would hope that this wouldn't have a large impact on the mortality baseline.

For this measure the trend-adjusted five-year average rate performs better as the impact is within 10% and only lasts for five years with less of an impact if it is in the middle years. ARIMA shows a large change in expected ASMRs when a 10% change is introduced and this has an impact for the whole time series.

Table 3 - Impact of changes to past year ASMRs on the expected current year ASMR

		ARIMA			**Trend-adjusted five-year average rate**		
Direction of change	**Year changed**	**Average**	**Min**	**Max**	**Average**	**Min**	**Max**
10% Lower	**X-1**	0.1%	-6.6%	5.3%	-8.1%	-8.4%	-8.0%
	X-2	13.0%	-0.9%	25.4%	-5.8%	-6.0%	-5.5%
	X-3	7.4%	0.6%	11.7%	-1.3%	-1.3%	-1.2%
	X-4	10.7%	0.0%	19.9%	1.3%	1.2%	1.4%
	X-5	9.9%	2.7%	14.7%	3.9%	3.5%	4.2%
	X-6	11.2%	4.6%	17.7%	-	-	-
	X-7	10.9%	6.5%	15.7%	-	-	-
	X-8	11.7%	4.3%	16.6%	-	-	-
	X-9	10.4%	6.1%	15.9%	-	-	-
	X-10	9.3%	4.8%	12.0%	-	-	-
	X-11	9.9%	6.9%	13.5%	-	-	-
	X-12	10.4%	4.0%	16.2%	-	-	-
	X-13	9.5%	5.0%	13.3%	-	-	-
	X-14	8.7%	5.0%	11.4%	-	-	-
	X-15	8.8%	2.3%	14.5%	-	-	-
	X-16	10.2%	9.3%	11.4%	-	-	-
	X-17	10.4%	7.5%	13.3%	-	-	-
	X-18	9.3%	9.3%	9.3%	-	-	-
10% Higher	**X-1**	19.2%	9.5%	27.1%	8.1%	8.0%	8.4%

	X-2	10.5%	3.5%	19.0%	5.8%	5.5%	6.0%
	X-3	13.7%	3.8%	22.2%	1.3%	1.2%	1.3%
	X-4	9.6%	3.8%	15.2%	-1.3%	-1.4%	-1.2%
	X-5	11.3%	2.2%	18.9%	-3.9%	-4.2%	-3.5%
	X-6	9.2%	1.4%	13.1%			
	X-7	10.5%	1.8%	17.8%			
	X-8	9.9%	3.0%	16.8%			
	X-9	11.6%	7.0%	16.8%			
	X-10	12.8%	11.7%	14.5%			
	X-11	11.5%	1.7%	14.5%			
	X-12	9.6%	2.4%	16.4%			
	X-13	11.3%	6.9%	17.3%			
	X-14	11.9%	7.4%	15.0%			
	X-15	15.4%	11.6%	17.9%			
	X-16	13.2%	11.6%	15.0%			
	X-17	13.6%	10.8%	16.3%			
	X-18	7.4%	7.4%	7.4%			

5.5 Bias

One of the problems with the current five year average methodology is that it is biased to produce a high expected number of deaths, mainly due to the fact we do not include any long-term mortality improvements in the method.

We have looked into two ways to check for bias. We compared the percentage difference between observed and expected mortality, with a percentage difference between 5% above and 5% below expected being classed as a small difference and a difference larger than 5% above or below expected being classed as large. The second test looks at the percentage of periods where the ASMR is significantly higher or lower than expected. Significance has been determined using the 95% confidence intervals, where instances of non-overlapping confidence intervals between estimates indicate the difference is unlikely to have arisen from random fluctuation.

FIGURE 1 - BOX-WHISKERS CHARTS OF PERCENTAGE OF EXCESS MORTALITY BY METHODOLOGY, ANNUAL AND MONTHLY, UK

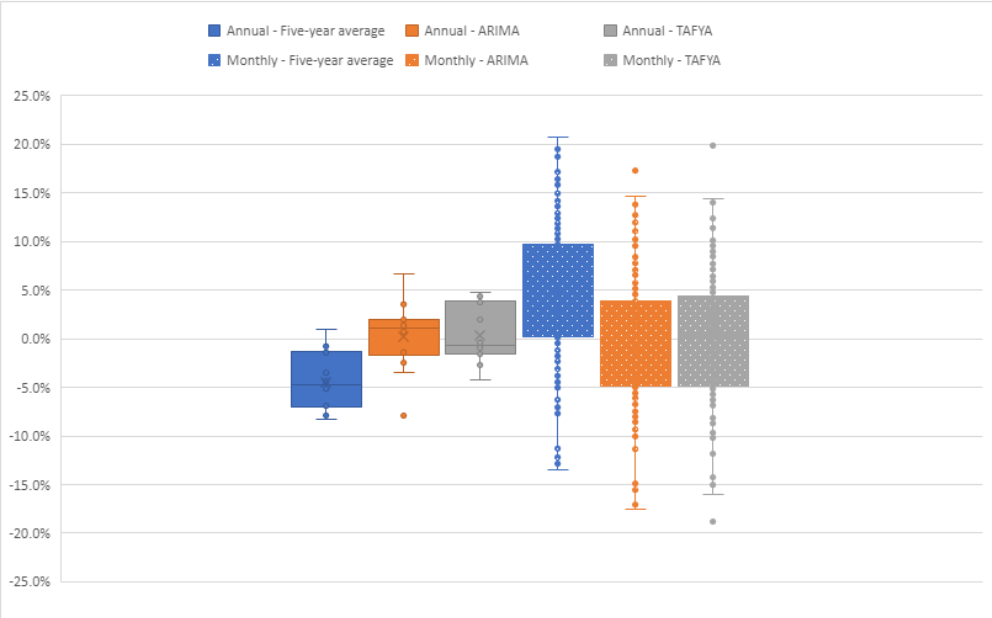


TABLE 4 - MEASURING BIAS BETWEEN METHODOLOGIES

		Annual (14 periods)			Monthly (156 periods)		
		Five-year average (current method)	ARIMA	Trend-adjusted five-year average rate	Five-year average (current method)	ARIMA	Trend-adjusted five-year average rate
Percentage compared to expected	**Percentage of years over 5% above expected**	0%	7%	0%	47%	20%	22%
	Percentage of years between -5% and 5% excess	57%	86%	100%	47%	57%	54%
	Percentage of years over 5% below expected	43%	7%	0%	5%	23%	24%
Significant difference	**ASMR significantly higher than expected**	7%	57%	36%	20%	46%	42%
	No significant difference	0%	0%	7%	13%	13%	14%
	ASMR significantly lower than expected	93%	43%	57%	67%	40%	44%

5.6 QUALITATIVE EVALUATION

TABLE 5 – DESCRIPTIVE CRITERIA OF METHODOLOGIES

	ARIMA	Trend-adjusted five-year average rate
Scalability	Potentially applicable to lower geographies but needs further testing. Practicable small	Easily applied to any geography or population group in a uniform way, subject to small numbers at

	number limits not known. Use at different scales may introduce incoherence.	lowest geographies. Any one ASMR needs 20+cases to calculate reliably.
Ease of use	End result easy to use but applying method requires expertise and/or precise reliance on ‘black box’ code and instructions.	End result easy to use, method could be applied either using prepared code or from scratch by a reasonably numerate user.
Transparency	Automatic values are taken for ARIMA, though we can likely make it more transparent. Reliable use requires considerable expertise.	Transparent – uses published ASMRs and formula can be published.
Ease of replication	If code is made available and there is access to a statistical software package then it can be replicated by someone with sufficient expertise. Parameters may have to be fixed to ensure replicability, potentially lowering model fit.	Replicable by someone with basic statistical knowledge. Could be reproduced in a spreadsheet, does not require statistical software.
Ease of explanation	The estimated expected age-standardised mortality rates uses past forecast errors in order to predict future data. Exact method might be difficult to explain to lay users.	The linear trend between year X-5 and X-1 is used to estimate years X, X+1 and X+2. The average of X-2 to X+2 is then taken as the expected number of deaths for year X.
Ease of implementation	More computationally intensive and requires some expertise, may be difficult for some users to apply.	Should be relatively simple to implement as it’s an extended version of current methodology
Ease of producing uncertainty measures	Need to look into more detail in terms of the best way to calculate uncertainty but standard methods are available.	We can adapt our current ASMR confidence interval methodology but we need to look into more detail.

5.7 FIT TO HISTORICAL DATA

Looking across UK annual ASMRs from 2001 to 2019, there are certain years which saw a slightly larger year-on-year percentage change than others and we have used these to determine how well the expected values fit historical data. Notably, there was a dip in mortality in 2009, 2014 and 2019 but an increase in mortality in 2015. Both ARIMA and trend-adjusted five-year average show the 2009 decrease and 2015 increase clearly, while only the latter method shows the 2014 and 2019 decreases.

FIGURE 2 - OBSERVED AND EXPECTED AGE-STANDARDISED MORTALITY RATES, 2006 TO 2019, UK

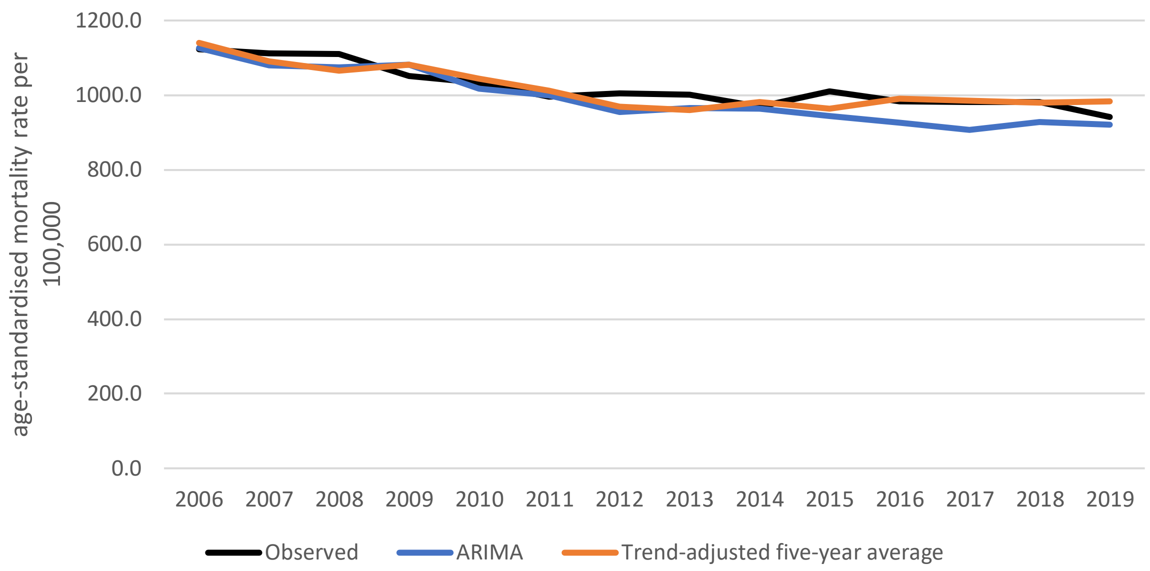
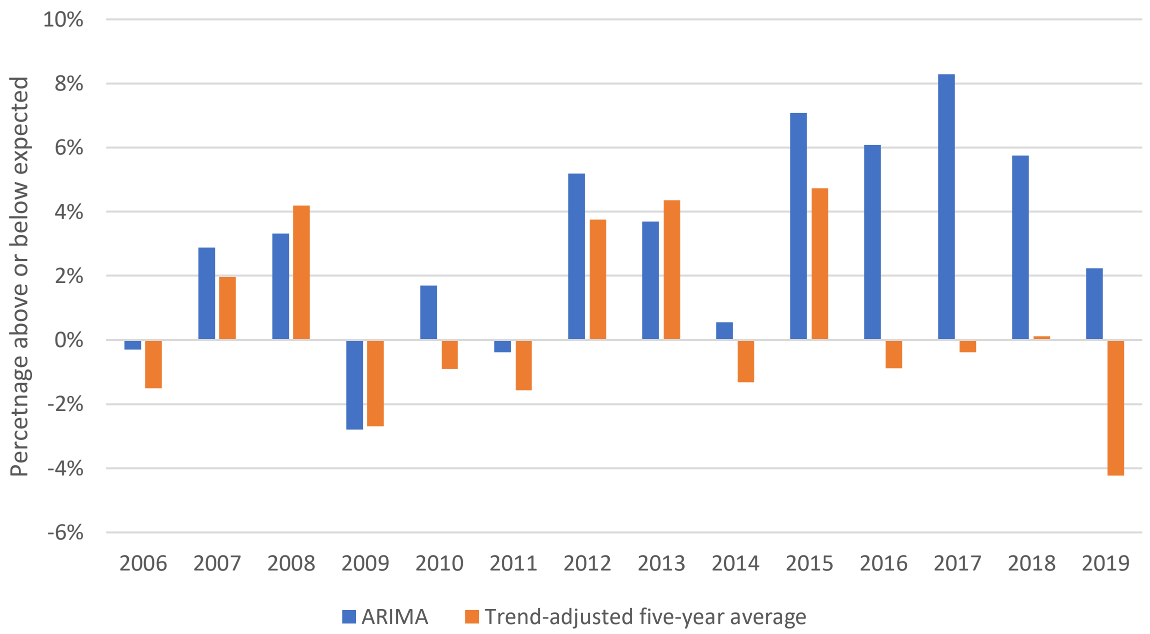


FIGURE 3 - PERCENTAGE ABOVE OR BELOW EXPECTED AGE-STANDARDISED MORTALITY RATES, 2006 TO 2019, UK



5.8 CONCLUSIONS

TABLE 6 - PROS AND CONS OF SHORT-LISTED OPTIONS

ARIMA	Trend-adjusted five-year average rate
Pro	

Established time-series methodology. Takes into account recent mortality trends. Fairly unbiased. Easy to apply bank holiday adjustments. Suitable for UK and the four nations.	Easy to explain. Takes into account recent trends in mortality. Fairly unbiased. Less sensitive to large annual changes. Transparent and easy to replicate. Good fit to historical data. Suitable for UK and the four nations Easily scalable by subnational area or population characteristic. Preferred methodology by the majority of the working group.
Con	
Less transparent - harder to explain. Needs more statistical expertise to use. More complex on scalability, needs further investigation – applications at different scales may be incoherent with each other. Fairly volatile, sensitive to large changes in past years.	Novel methodology. Fairly volatile. Requires bank holiday adjustment. Some concern from the group in regard to robustness of method, including relative influence of past years.

Both of these offer improvements over the existing method and are both suitable methodologically speaking. While there are pros and cons of each method, there was not a unanimous decision with the majority of the working group preferring the trend-adjusted five-year average over the ARIMA methodology. ARIMA models do show promise, but require further work to ensure that in particular we fully develop the model and understand their volatility – and whether they are a framework that can be extended and improved with experience of applying them in this application.

We recommend adopting the trend-adjusted five-year average of rates (‘five-year trend’) as the excess mortality baseline, although the working group was not unanimous in its views. The benefits of the five-year trend are that it takes account of both population change and mortality trends while still being relatively easy to explain and apply. It shows a good fit to historical data when tested. Although there are some aspects where an ARIMA model performs better, we consider ARIMA less transparent, i.e. more difficult for users to understand or replicate without considerable expertise. Both methods require some further testing and development, in particular the five-year trend needs decisions about adjustment for bank holidays. ARIMA needs further investigation of its scalability and ability to ensure coherence when reporting at geographies if it is taken forward.

For any methodology taken forward we will need to investigate ways to overcome the atypical mortality in the pandemic years, the group will look at this next. More information can be found in Annex 3.

ANNEX 1 – CALCULATING THE EXCESS MORTALITY BASELINE (PAPER FOR MARP APRIL 2023 MEETING)

Purpose

This paper discusses the issues to consider when estimating the number of expected deaths (or 'baseline')) in a given period, and possible methods being explored by a cross-UK and cross-organisation baseline working group. Estimating the number of expected deaths is a key step in calculating the number of excess deaths. The outcome of the working group will be to provide a recommendation paper on the most appropriate method and bring it to MARP for further assurance. Therefore, this paper is intended to provide MARP with the background of this project and a chance to provide advice at an early stage.

Key asks of MARP

We would welcome feedback on the methods discussed and the proposed criteria for determining which methodology to take forward.

- Are there other methods we should be investigating?
- Are there other elements we should consider when looking at results?

1. Background information

Excess mortality is the difference between the actual number of deaths observed and the number that were expected. Across the UK we are aware of the number of deaths observed (although there is more than one measure of this – registrations, occurrences and notifications), however there are different methods available for estimating the expected number of deaths.

The impact of the pandemic means that the question we are seeking to answer isn't as clear as it was prior to the pandemic and has now widened to 'What do we mean by expected number of deaths?'. Are we trying to calculate the number of deaths had there not been a pandemic, i.e. the impact of the pandemic on mortality? Or are we trying to calculate the number of deaths given there is a pandemic, i.e. to inform us about what is happening now? Or could there be a different question we need to answer? The baseline working group have decided the question we should be answering is "*Do the latest reported data on deaths, in isolation or cumulatively, indicate a rise, or fall, in mortality compared with previously observed mortality?*"

The Office for Statistics Regulation have been involved in discussions around this work. In February 2023 the OSR released a report with recommendations around ONS' excess deaths reporting. We continue to liaise with OSR and use their recommendations as part of the working group. There were six recommendations that came from this report:

- Be open and proactively engage with a wide range of users about the development work it is undertaking and its decision making throughout this process.
- Improve how its data are presented and how any uncertainty in the data is explained to users. While there isn't one definitive right answer on what the expected number of deaths should be, users should be supported in understanding this.
- Review its understanding of users' needs, as different users may have different uses of these data and therefore different methodologies may be suitable in different circumstances. This means reviewing, considering and explaining what it means by 'expected deaths' when calculating excess mortality.
- Provide users with clear, accessible information on:
    - the methods and the reasons why the group has chosen to apply those methods

- the differences in approach taken to measuring excess mortality including signposting to the different sources of data
- how to interpret differences in the resultant published data

As well as this review, the OSR has recently carried out a data compliance check on mortality statistics (because of the increased interest in the topic) and the Methods and Quality Directorate in ONS are also doing a review. We will take the recommendations from the three areas and make sure we incorporate them in this project and the way we work moving forward.

The cross-UK excess mortality baseline group was established in February 2023 and benefits from expertise spanning many organisations and fields. As well as representatives from different areas of ONS (mortality experts, methodologists and demographers), the group has representation from Cabinet Office, Office for Health Improvements and Disparities (OHID) within the Department of Health and Social Care (DHSC), UK Health Security Agency (UKHSA), Public Health Wales, Welsh Government, National Records of Scotland, Northern Ireland Statistics and Research Agency, Department of Health Northern Ireland, and the actuarial profession (Lane Clark & Peacock LLP and the Continuous Mortality Investigation (CMI)).

The group will bring its recommendations to MARP at a later date for consideration before presenting to the National Statistician to make the final decision. This methodology will then be adopted across organisations as the standard methodology to calculate excess deaths with only a few exceptions to this rule. A paper will be published at the end of the project looking at the results of each option and why the final method was chosen. As part of this paper we will discuss the exceptions to this and when they should apply.

2. Terms and outcomes of the group

The resulting methodology will need to provide measures of excess deaths across a range of geographies (UK, England, Wales, Scotland, Northern Ireland, Local Authorities) and time periods (Weekly, Monthly, Annual). The group will also take in to account the possibility of providing further disaggregation such as by cause of death and socio-demographic characteristics such as age, sex and possibly ethnicity.

The group outcomes are:

1. To provide clarity on the question(s) we are seeking to answer (e.g. ‘How many excess deaths are there?’)
2. To evidence the most appropriate method to be used to answer this/these question(s) and calculate excess deaths
3. Make sure that user needs are at the heart of evidencing 1 and 2, and that the resulting publication clarifies the excess mortality landscape

An options paper with a recommendation will be sent to the National Statistician who will make the final decision on the most appropriate method.

The fundamental question that the group will focus on is “*Do the latest reported data on deaths, in isolation or cumulatively, indicate a rise, or fall, in mortality compared with previously observed mortality?*”

We will prioritise and test available methods against data from at least 2015 using the following criteria when calculating the expected number of deaths. The method needs to:

- account for the context that the pandemic has occurred
- account for demographic change, mainly age distribution
- account for the seasonal variation seen in mortality

- be appropriate for all lengths of periods – Weekly, monthly and annual
- be suitable to run at the UK, England, Wales, Scotland and Northern Ireland levels when broken down by sex
- provide an interval estimate around the point estimate (confidence intervals)
- be understandable to users. A simpler method is preferable to a more complex method if the methods work as well as each other
- be relatively unbiased across a range of scenarios

The analysis will be carried out by the ONS mortality team and other member organisations, dependent upon available resources.

As well as meeting the above criteria we will test using statistics such as the mean absolute percentage error (MAPE) across the non-pandemic years to see how well the methods would have predicted previously observed death rates. We will also look to see if we see deviation from predicted values where expected. For example, we would expect to see excess in 2015 and negative excess in 2019 due to the differences in flu prevalence in those years particularly.

3. Methodology already available and methodology being investigated

There are multiple measures for the number of expected deaths used to measure excess deaths. These different measures produce different excess deaths figures. They include:

- five-year averages (number of deaths and age-standardised mortality rates)
- relative age-standardised mortality rates
- segmented regression analysis
- the Continuous Mortality Investigation (CMI) mortality projections
- CMI pandemic monitor
- EuroMOMO
- UK Health Security Agency (UKHSA) daily mortality method
- Office for Health Improvement and Disparity's (OHID) excess death model

Some of the commonly used methods are summarised below.

ONS uses the average of the previous five years (excluding 2020) to calculate the expected number of deaths in a given period. We use this method as it ensures comparison with a recent period which was similar in life expectancy, advances in healthcare, population size and shape. Using multiple years removes the fluctuations that can be seen year-on-year when looking at mortality. Some of the benefits of this method are that it is easy to understand and does not rely on different data breakdowns. For instance, it requires only the number of deaths for the period and the five years prior. Some of the limitations are that it does not take into account any trend in the data, nor any other variables than deaths (when looking at numbers of deaths), population, and age-structure (when looking at rates).

OHID's estimates of excess deaths are based on a method developed specifically to measure excess deaths during the coronavirus (COVID-19) pandemic. It uses a (Quasi-Poisson regression) model to estimate expected deaths each week, based on the trend in mortality rates from 2015 to 2019. OHID's model accounts for changes to the population, including ageing. The model accounts for ethnicity, sex, and levels of deprivation, so excess deaths for these factors are reported separately. OHID's complex method assumes that the trend in mortality rates before the pandemic (from 2015 to 2019) would have continued had there been no pandemic. This trend is not completely clear, however. The long-term downward trend in mortality rates slowed in the 2010s, though there was more improvement in 2019 than in other recent years.

The Continuous Mortality Investigation (CMI) Pandemic Mortality Monitor was developed specifically to measure excess deaths during the pandemic and uses analysis based on Standardised Mortality

Rates (SMRs). Provisional weekly deaths data published by the ONS is adjusted to control for changes in the size, age, and gender distribution of the population over time. Calculated SMRs are compared with SMRs from 2019, the last pre-pandemic year.

The UKHSA mortality baseline used for EuroMOMO is based on modelling week of death with a Serfling wave function and linear trend or for younger ages with no seasonality, just a linear trend. Only spring and autumn weeks are used in this modelling. This is to have a function that fits well to years, with no large excesses due to flu in the winter and heatwaves in the summer. For their daily model, the baseline until November 2020 was from the same day of the year in the previous five years, plus or minus seven days, with an extrapolated time trend. The baseline from December 2020 to March 2021 only uses the same days, plus or minus seven days, from the past three low flu years with no trend. The baseline from April 2021 onwards is set to be the same as the previous year’s baseline.

We will be exploring alternative methods such as Holt-Winters (a method that uses exponential smoothing to predict future values), ARIMA/SARIMA models (autoregressive integrated moving average model with or without seasonality included), Neural networks (a simplified model based on the way the human brain processes information), Farrington models (a quasi-Poisson regression-based model), and investigating the models used by WHO (more information can be found here).

ANNEX 2 – TERMS OF REFERENCE OF THE CROSS-GSS WORKING GROUP ON EXCESS MORTALITY

1. Purpose

The group has been set up in response to the UK Statistics Authority Board to clarify the landscape of excess mortality for users. This work is to be done across government and cross-UK, with ideally one measure to calculate excess mortality across geographies and time periods.

The aim of this group is to clarify the statistical need, complete the analysis, produce a recommendation, and resulting publication by the end of April. The work will report to the UKSA board in May 2023.

The group will be led and chaired by ONS, but the decisions and assumptions made by the group will be decided collaboratively. The group will meet fortnightly, which may be more frequent during particular times and for particular members if it would benefit delivery.

2. Scope of requirements

The resulting methodology will need to provide measures of excess deaths across geographies (UK, England, Wales, Scotland, Northern Ireland, Local Authorities) and time periods (Weekly, Monthly, Annual). The group will also take in to account the possibility of providing further disaggregation such as by cause of death.

3. Outcomes

The group outcomes are:

- To provide clarity on the question(s) we are seeking to answer (e.g. 'How many excess deaths are there?')
- To evidence the most appropriate method to be used to answer this/these question(s) and calculate excess deaths
- Make sure that user needs are at the heart of evidencing 1 and 2, and that the resulting publication clarifies the excess mortality landscape

An options paper will be sent to the National Statistician who will make the final decision on the most appropriate method

4. Approach

The fundamental question that we are trying to address is "Do the latest reported data on deaths, in isolation or cumulatively, indicate a rise, or fall, in mortality compared with previously observed mortality?"

We will prioritise and test available methods against data from at least 2015 using the following criteria when calculating the expected number of deaths. The method needs to:

- account for the context that the pandemic has occurred
- account for demographic change, mainly age distribution
- account for the seasonal variation seen in mortality
- be appropriate for all lengths of periods – Weekly, monthly and annual
- be suitable to run at the UK, England, Wales, Scotland and Northern Ireland levels when broken down by sex

- provide an interval estimate as well as a point estimate

The analysis will be carried out by the ONS mortality team and other member organisations, dependent upon available resources.

5. Decision making process

The chair will ensure that all contributions are heard, either during meetings, via correspondence or exercises such as short surveys to acquire views.

The views of the majority (two-thirds of respondents agreeing) will carry wherever possible. If there is no clear majority view, then the different views will be noted and further work identified. If a difference of views significantly risks delivery, a summary of the options (including advantages and disadvantages) will be produced and presented to the National Statistician to provide a steer.

6. Membership

The technical working group on excess deaths has representation from:

- Office for National Statistics (ONS)
- Cabinet Office
- Office for Health Improvement and Disparities (OHID)
- UK Health Security Agency (UKHSA)
- Public Health Wales
- Welsh Government
- National Records of Scotland (NRS)
- Northern Ireland Statistics and Research Agency (NSIRA)
- Department of Health Northern Ireland
- Lane Clark & Peacock LLP (LCP; Actuary)
- Continuous Mortality Investigation (CMI; Actuary)

ANNEX 3 – FURTHER ISSUES NOT COVERED IN THIS PAPER

This annex outlines further issues around the calculation of excess mortality, which are not part of the option appraisal presented in this paper but will be addressed in the next stage of the project.

A3.1 Bank holiday and weekend effects

Local register offices are closed on weekends and bank holidays, which means there are no (or very few) death registrations. During weeks with bank holidays, we tend to see a decrease in registered deaths in the week the bank holiday occurred but increases in the week prior and post bank holidays. Bank holidays do not always occur in the same weeks each year, especially Easter which moves between a number of weeks.

To address this, various bank holiday adjustments have been tested to make data comparable. These methods involved applying regressors to the week with a bank holiday and sometimes the weeks prior and after. There was not a large difference between the various bank holiday methods.

A bank holiday adjustment is likely to be used at the weekly deaths level, but not in our monthly or annual breakdowns. For monthly figures, the number of weekends in the month can have an impact on the number of excess deaths. The more weekends in a month, the less days available to register a death which then produces a lower number of deaths registered. For this project, we have not yet looked into adjustments for weekends.

A3.2 Treatment of the pandemic period

The years 2015 to 2019 have been initially used as the testing period for excess mortality methods as these were recent years that included both high and low flu years, but were not affected by the COVID-19 pandemic. It is difficult to test methods of expected deaths in 2021 onwards as we are currently still experiencing excess mortality (by current measures) and 2020 will have a large impact on whatever methodology is chosen. In particular, the spring 2020 ‘spike’ and other periods of unusually high mortality caused by the pandemic, such as the high mortality at the start of 2021, should not be counted towards the definition of a ‘usual’ level of deaths.

There were four approaches investigated to date for the handling of the peak pandemic period when calculating excess mortality for subsequent years, and consideration of how to handle 2020 and 2021 will continue in the next stage of the project.

Approach	Pros	Cons
Use deaths observed in 2020 (i.e. make no special arrangement)	No adjustment means 2020 is treated like other years. Available for all breakdowns.	Large increase in deaths from the pandemic would artificially increase the expected number of deaths for future years. Would impact breakdowns to different extents.
Use expected deaths in 2020 as 2020 figure when calculating excess	Removes the direct and indirect effects of the coronavirus pandemic (in 2020) on mortality statistics. Impacts each breakdown consistently.	Does not take into account mortality displacement. Does not use any 2020 data.
2020 data adjusted by OHID’s mortality displacement method	Takes into account mortality displacement brought on by the pandemic (in 2020 only).	Only available for England and all-cause mortality, proportion change

	Uses 2020 data.	can then be applied to other breakdowns. Relies on COVID-19 testing data, which means that some excess deaths (that did not have COVID-19 on the death certificate) remain.
Treat specific periods in 2020 as outliers (to be excluded or capped at a specified level)	Uses 2020 data with specific time periods noted as outliers. Can be applied to all breakdowns.	Does not take into account mortality displacement. Would need to be applied to different breakdowns independently.

A3.3 Registration delays and death occurrences

In the UK, we are only aware that a death has occurred and can include it in our figures when it is registered, and there can be a variable and sometimes substantial delay between death occurrence and registration. Some deaths may not be registered for weeks, months or occasionally years – the longer delays are where a coroner's inquest and even a criminal trial are involved. The issues in England and Wales and Northern Ireland are similar, whereas in Scotland the registration system works differently and delays do not occur to the same extent.

The majority of deaths are registered within 7 days of occurring, with 67.2% of deaths being registered within 7 days in 2021 in England and Wales. Although the median registration delay (the time between a death occurring and being registered) for all-cause mortality is 5 days, this can differ by cause with the largest delay being seen in assaults with 447 days.

Analysing deaths by date of registration date allows us to produce timely statistics that are stable over time, and comparable across locations. An alternative is to analyse by date of death (occurrence date) which is important to look at seasonal patterns, which is why organisations usually release statistics on both death registrations and occurrences. However, to have robust information on deaths by date of occurrence it is necessary either to wait until the majority of deaths are likely to be registered or to create an adjustment or nowcasting methodology to estimate the number of deaths likely to have occurred based on previously observed registration delays. These methods are widely used by actuaries.

There are also alternative sources of data on deaths which might have some value in increasing the timeliness or completeness of reporting, notably the NHS central database. However, previous research has shown this is unlikely to improve on death registration data because of inaccuracies in NHS administrative data and other issues. Because of this and the likely cost of setting up a new data flow/linkage, ONS does not intend to follow up this possibility at present.

ANNEX 4 – FURTHER INFORMATION ON THE LONG-LISTED OPTIONS

A4.1 CMI MORTALITY PROJECTIONS

The Continuous Mortality Investigation (CMI) mortality projections is a core CMI output – made available to subscribers for the purposes of projecting future mortality (a best estimate and alternative scenarios). It is a model refreshed annually for predicting future mortality rates (akin to the mortality forecasting in the ONS population projections). CMI fit an Age-Period-Cohort model to historical mortality improvements (i.e. year on year change in mortality rates) to determine the recent/current amount of mortality improvement. This then blends towards a long-term age structure for mortality improvements which is set by the user of the model. This has not been taken forward as it is designed to look at long-term estimates of annual mortality, whereas we are interested in short-term mortality estimates from a weekly to annual level. There would also be an issue with transparency as this is a proprietary method only available to subscribers.

A4.2 CMI PANDEMIC MONITOR

CMI pandemic monitor was established during the pandemic and compares the age-standardised mortality rate in a given week with the mortality rate seen in 2019. For expected and hence excess deaths during the pandemic the CMI was aware that users would not be experts and hence ease of explanation was as important as statistical robustness. CMI noted that mortality rates in the first quarter of 2020 were very similar to the first quarter of 2019 (there was a two-year period of consistently low mortality from around April 2018 to March 2020). This is a simple pragmatic approach which does not allow for longer-term trends. Because it has a fixed baseline year it is not suitable for future monitoring, despite meeting all criteria in table 1.

A4.3 EUROMOMO

EuroMOMO is specifically designed to look at excess mortality with a 'best case' baseline, for example times with low flu, and is designed for international comparisons. The expected count is based on modelling week of death with a Serfling wave function and linear trend (or for younger ages with only a linear trend. Only spring and autumn weeks are used in this modelling to aim to have a function that fits well to years with no large excesses due to flu in the winter and heatwaves in the summer. Estimation is used to allow for registration delays to obtain a correction to the observed death count. This baseline methodology did not meet our criteria of measuring an 'average' year, however this methodology will still be used to provide figures consistent with other countries and to measure mortality against the 'best case' baseline. This methodology is also very volatile for smaller areas and results for the current period are heavily influenced by volatility in past periods.

A4.4 FIVE-YEAR AVERAGE BASED ON COUNTS, ASMRS OR RASMRS

The five-year average based on counts of deaths has been used as the expected number of deaths by the Office of National Statistics (ONS), National Records of Scotland (NRS) and Northern Ireland Statistics and Research Agency (NISRA) for a number of years. This method is very transparent and can be easily applied to all geographical levels and disaggregations. However, it does not take account of changes in the population year on year or long-term trends in mortality. Large effects (such as the pandemic) can distort the number of expected deaths in future years. The figures have not been adjusted for bank holidays and weekends, which affect comparisons at weekly and monthly level.

The five-year average based on age-standardised mortality rates (ASMRs) ensures that we are comparing like for like in terms of population size and composition. It also smooths the volatility seen year-on-year in mortality statistics, especially at the weekly level. This method is relatively easy to explain and to apply to different levels and breakdowns. However, it does not take into account long-term trends in mortality.

Relative age-standardised mortality rates (rASMRs) have been used by ONS for international comparisons of excess mortality. For present purposes the method is effectively the same as the five-year average, however the five-year average takes an age-standardised mortality rate (ASMR) of the five-years aggregated together, whereas the rASMR averages the separate ASMRs for the five years. rASMRs have not been considered further as they produce similar results to the five-year average but are slightly more complex to explain. They share the limitation of ASMRs of not accounting for long-term trends.

A4.5 OHID PANDEMIC EXCESS DEATHS MEASURE

Office for Health Improvement and Disparity's (OHID) excess death model is an established indicator that was specifically designed to monitor excess mortality through the pandemic. It was not originally designed to take into account the pandemic as an element in calculations for future years. Expected deaths are calculated using a quasi-Poisson regression-based model (similar to Farrington) which incorporates trend information, seasonality, geography, age and sex, as well as added detail for ethnicity and deprivation. Due to the model specification reporting for different geographies, population breakdowns, place of death, and cause of death is possible. The OHID model currently takes into account bank holidays and demographic change. While the trend parameter is based on mortality rates, the model also produces counts of deaths as an output with corresponding uncertainty intervals.

A current limitation of the OHID model is that the trend parameter uses a 5-year average, although this parameter can be changed, with the most suitable range of years to be determined. As currently designed it is not suitable for cross-UK comparison because of differences in ethnicity and deprivation measures. A project that involves estimating mortality displacement has been undertaken by OHID which, at least partially, accounts for the pandemic deaths to create a new baseline for pandemic years (albeit this approach requiring some further evaluation). While some of the working group supported this as a feasible preferred method, the majority were concerned by the need for further development of the model to make it suitable for general excess mortality use in the future, and its relative complexity for implementation and explanation.

A4.6 UKHSA DAILY EXCESS MORTALITY METHOD

UK Health Security Agency (UKHSA) daily mortality method is run for heat alerts and pandemics with results going into heat health impact reports and into the UKHSA flu and covid surveillance report. The model uses daily data on deaths by date of death. The baseline until November 2020 was from the same day of the year in the previous 5 years plus or minus 7 days with an extrapolated time trend. The baseline from December 2020 to March 2021 only uses the same days plus or minus 7 days from the past 3 low flu years with no trend, and the baseline from April 2021 onwards is set to be the same as the previous year's baseline. Estimation is used to allow for registration delays to obtain a correction to the observed death count. For heat alerts the baseline is from periods just before and after the heat period. The method is designed for assessing mortality for specific health scenarios and as with EuroMOMO this method does not meet our criteria of an average year. Its application is complex in varying by year and difficult to justify to a wider audience. Therefore this was not considered suitable for ongoing excess mortality monitoring.

A4.7 WORLD HEALTH ORGANISATION (WHO) MODEL

The World Health Organisation (WHO) developed its own method for estimating excess mortality for the purpose of comparable international reporting on the pandemic. Full information on the model can be found here. The working group decided not to move forward with this method as it is specific to the pandemic and more complex than other methodologies, being designed for various different national situations some of which are not relevant to the UK. Because of its complexity it lacks transparency and ease of understanding or replication, while pandemic-specific features make it unsuitable for general excess mortality monitoring in the future.

A4.8 Farrington model

Farrington models are a quasi-Poisson regression-based model. It is very similar to the basis of the OHID method. The method described in Farrington et al (1996) is an algorithm for early detection of outbreaks of infectious disease that involves checking whether counts from a count series are outside of some threshold. Later developments of this approach are common in disease surveillance applications. Two different variants were investigated (see **Error! Reference source not found.** and **Error! Reference source not found.**). Although the 'Farrington flexible' model worked well for 2019 weekly data the model didn't work well around bank holidays or with years that had 53 weeks.

It should be noted that the calculation of a baseline is only one component of the overall Farrington approach which is not designed for that purpose. Further investigation is needed to assess its practical suitability for excess mortality monitoring, including the choice of various parameters within the model. The methods are probably not easy to explain or replicate. These factors suggest that the method should not be recommended at this time, but future research may provide more insight as to whether it might be a suitable method in the future.

Figure 4 - Numbers of observed deaths and expected deaths (Farrington models), England and Wales, Deaths registered weekly, 2015 to 2019

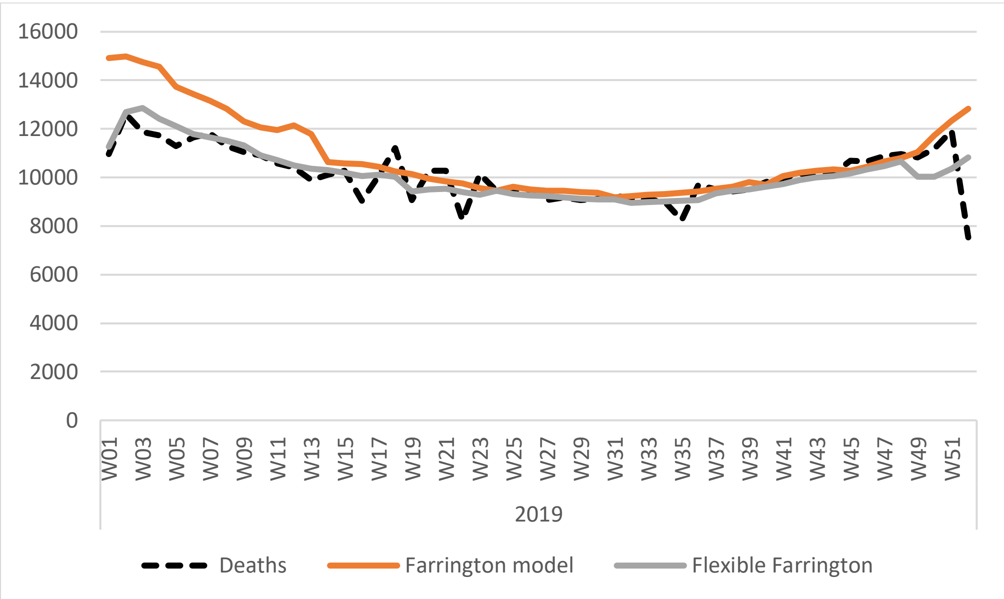
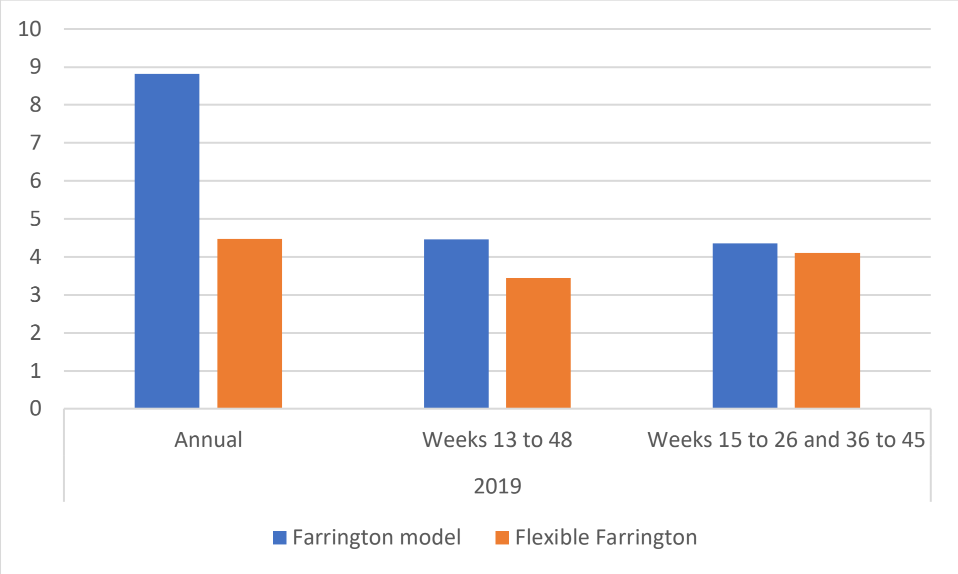


FIGURE 5 - MEAN ABSOLUTE PERCENTAGE ERROR, VARIOUS PERIODS, ENGLAND AND WALES



A4.9 HOLT-WINTERS MODEL

Holt-Winters is a method that uses exponential smoothing to predict future values and is sometimes used by actuaries. We tested two versions of Holt-Winters (see **Error! Reference source not found.** and **Error! Reference source not found.**) and although the results provided followed the historical pattern of mortality the differences were larger than some of the other methods we had investigated. The results were also very sensitive to the previous trends, with some years having a bad flu year predicted. It was agreed not to pursue this option further.

FIGURE 6 - NUMBERS OF OBSERVED DEATH AND EXPECTED DEATHS (HOLT-WINTER), ENGLAND AND WALES, DEATHS REGISTERED WEEKLY, 2015 TO 2019

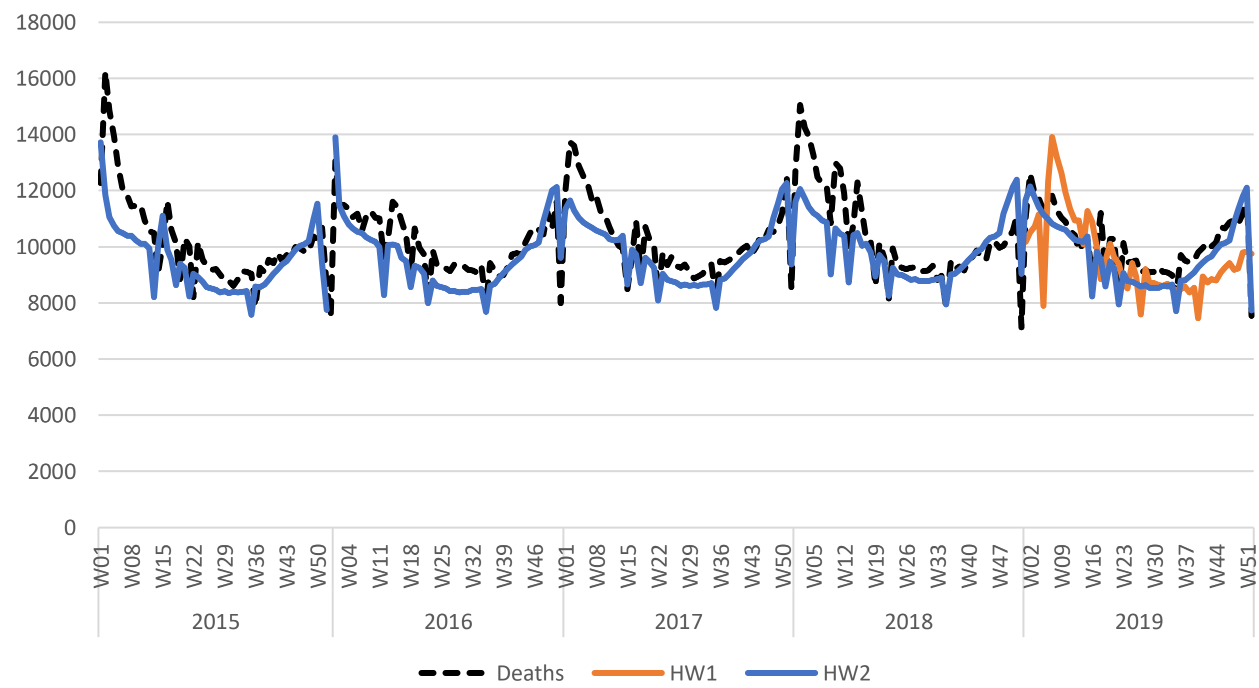
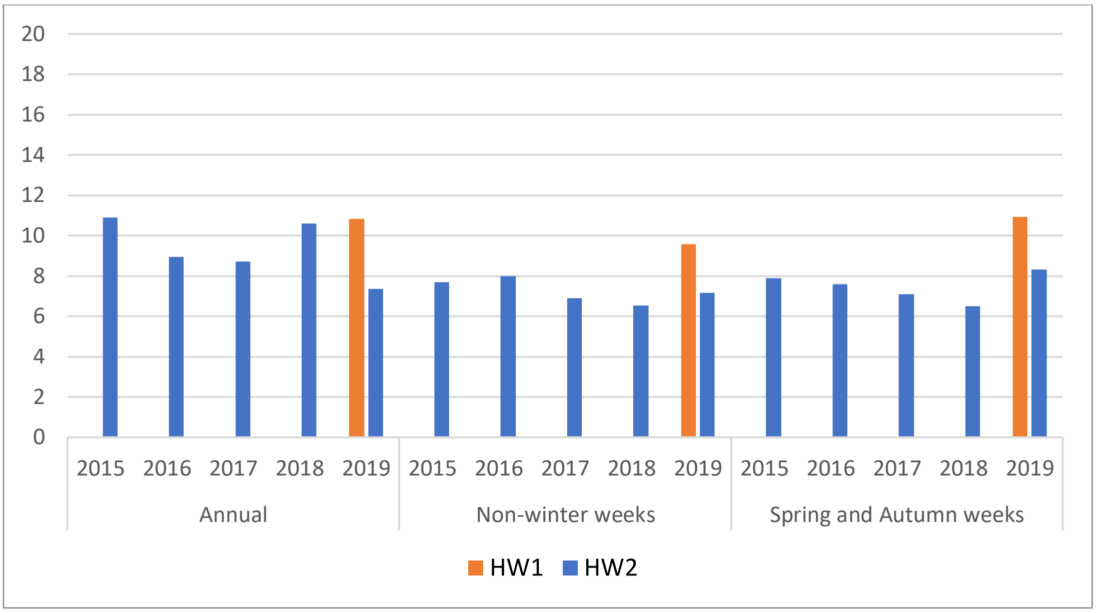


Figure 7 - Mean absolute percentage error, various periods, England and Wales



A4.10 Neural network time series models

Neural network methods are a variant of machine learning. These approaches have been developed over the last thirty years and have been used in various time series models. We tested four versions of neural networks ([Long Short-Term Memory [LSTM]](), [Stacked Recurrent Neural Network [Stacked RNN]](), [Convolutional Neural Network [CNN]](), [Gated Recurrent Unit]() [GRU]). Neural network methods are not transparent and would be difficult to explain to users, are not easily replicable, and can produce different figures for each 'run' based on the same data. These issues in combination mean that the method was considered unsuitable.

As LSTM and GRU showed similar results and Double RNN and CNN showed similar results we have only displayed one of each on the charts.

Figure 8 - Numbers of observed deaths and expected deaths (Neural networks), England and Wales, Deaths registered weekly, 2015 to 2019

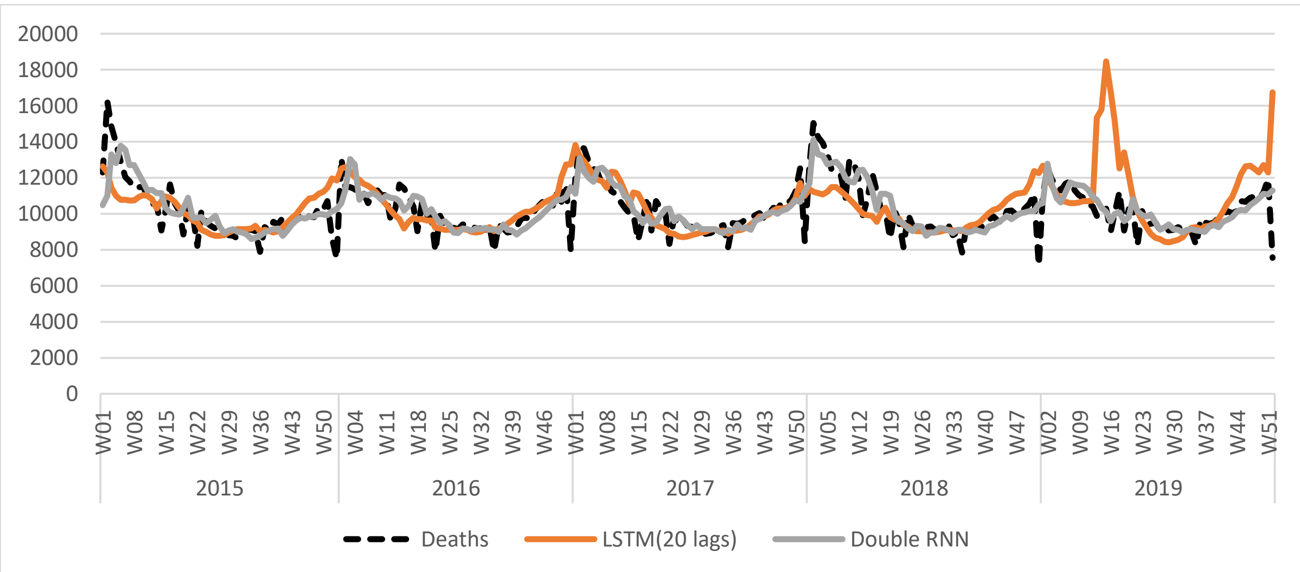
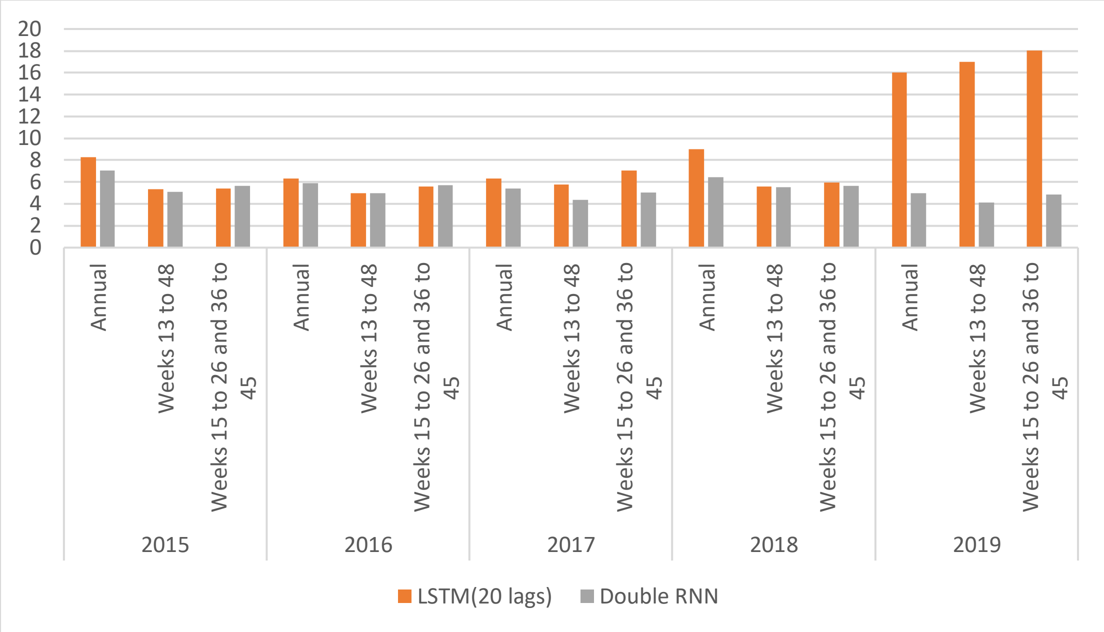


FIGURE 9 - MEAN ABSOLUTE PERCENTAGE ERROR, VARIOUS PERIODS, ENGLAND AND WALES



A4.11 SARIMA

SARIMA is an autoregressive integrated moving average model including seasonality, ARIMA (without seasonality) is described in section 4.2. These are both standard time-series models, used elsewhere in ONS. As mortality data is seasonal by nature, for example high mortality in the winter, the SARIMA model would be preferable over the ARIMA model for breakdowns by week and month. We tested the SARIMA model at a weekly and monthly level by applying an auto ARIMA that uses the Hyndman-Khandakar algorithm to automatically choose the trend elements and seasonal elements.

SARIMA model is a form of regression analysis with additional seasonal terms. It gauges the strength of one dependent variable relative to other changing variables, applying the model to future patterns by examining the differences between values in the series instead of through actual values.

SARIMA can best be understood by examining its component parts

- **Autoregression:** a changing variable that regresses on its own lagged, or prior, values.
- **Integrated (I):** differencing of raw observations to allow the time series to become stationary.
- **Moving average (MA)**: incorporation of the dependency between an observation and a residual error from a moving average model applied to lagged observations.
- **Seasonal component**: The seasonal part of the SARIMA model captures repeating trends in the data. It is denoted by (P, D, Q)s, where P represents the number of seasonal autoregressive terms, D represents the number of seasonal differences, Q represents the number of seasonal moving average terms, and s represents the number of periods in a season.

A SARIMA model has the formula SARIMA(p,d,q)(P,D,Q)m where p,d,q are the non-seasonal elements and P,D,Q and m are the seasonal elements.

Trend elements:

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.

Seasonal elements:

- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

The group decided that applying an ARIMA at an annual level and applying that trend at further breakdowns would be a more appropriate method as the data is more robust at an annual level. The other option was to create a weighted ASMR based on monthly data for an annual period.

FIGURE 10 - NUMBERS OF WEEKLY OBSERVED DEATHS AND EXPECTED DEATHS (SARIMA), ENGLAND AND WALES, DEATHS REGISTERED WEEKLY, 2015 TO 2019

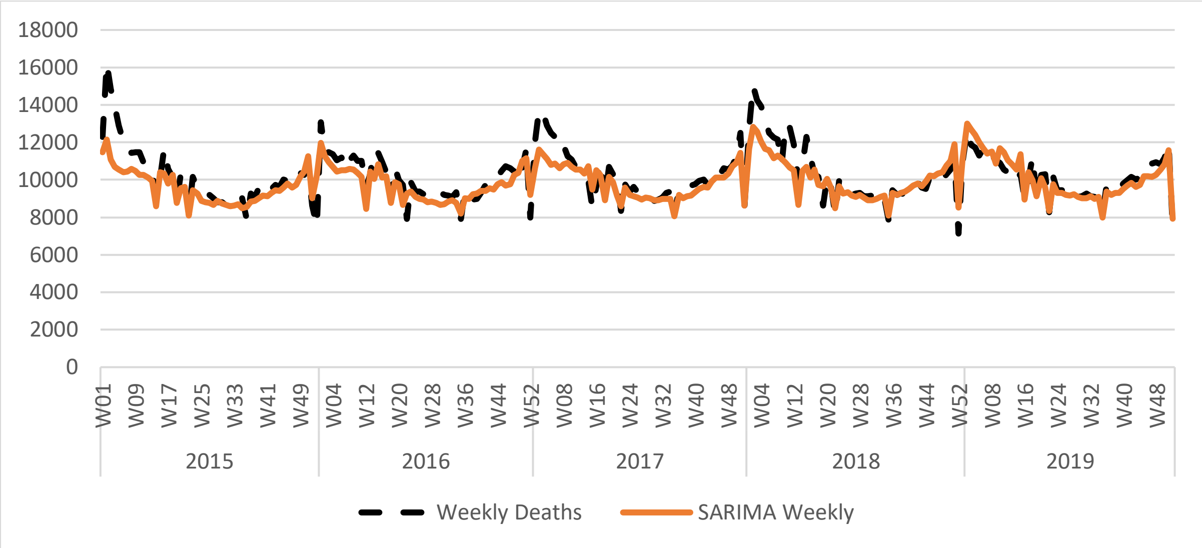


FIGURE 11 - MEAN ABSOLUTE PERCENTAGE ERROR, VARIOUS PERIODS, ENGLAND AND WALES

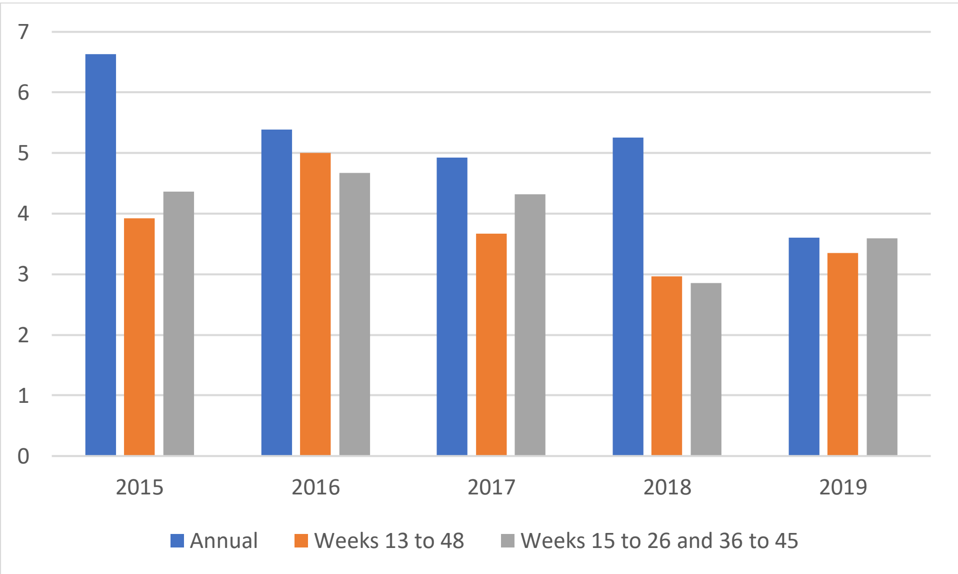
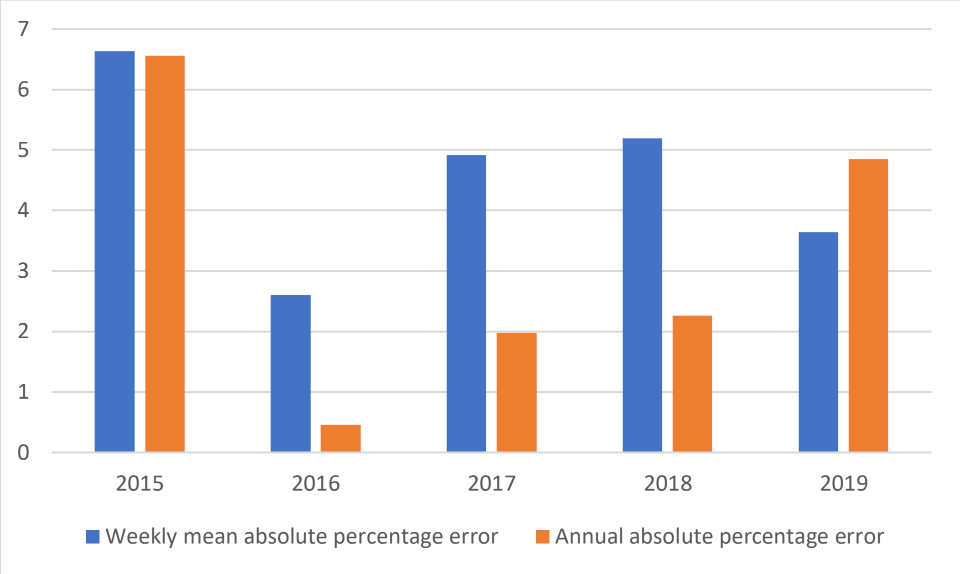


FIGURE 12 - MEAN ABSOLUTE PERCENTAGE ERROR OF WEEKLY SARIMA AND ANNUAL ARIMA, 2015 TO 2019



A4.12 SEGMENTED REGRESSION ANALYSIS

Segmented regression analysis has previously been used by the ONS to identify points in a time series where there has been a significant change in the trend. Although this works well at an annual level, it was not found suitable at the weekly deaths level and was not tested any further. As can be seen in **Error! Reference source not found.** below, this method was not able to detect seasonality when looking at numbers of deaths per week and assumed a linear trend to the number of weekly deaths registered. The mean absolute errors for this method were relatively large (some over 10%) for the periods where we would expect low excess deaths (e.g. spring and autumn). It was agreed that the technical requirements were not met.

FIGURE 13 - NUMBERS OF OBSERVED DEATHS AND EXPECTED DEATHS (SEGMENTED REGRESSION ANALYSIS), ENGLAND AND WALES, DEATHS REGISTERED WEEKLY, 2015 TO 2019