Biology Methods and Protocols, 2024, bpae031
https://doi.org/10.1093/biomethods/bpae031
Advance Access Publication Date: 17 May 2024
Review

Meaning and prediction of 'excess mortality': a comparison of Covid-19 and pre-Covid-19 mortality data in 31 Eurostat countries from 1965 to 2021

Bernhard Gill [1,*] Theresa Kehler [1] and Michael Schneider [1]

[1]Institute for Sociology, Ludwig-Maximilians-Universitaet Muenchen, Munich, Germany

*Correspondence address. Institute for Sociology, Ludwig-Maximilians-Universitaet Muenchen, Munich, Germany. Tel: +49(89)2180-3222; E-mail: Bernhard.Gill@LMU.de

Abstract

Determining 'excess mortality' makes it possible to compare the burden of disasters between countries and over time, and thus also to evaluate the success of mitigation measures. However, the debate on coronavirus disease 2019 (Covid-19) has exposed that calculations of excess mortalities vary considerably depending on the method and its specification. Moreover, it is often unclear what exactly is meant by 'excess mortality'. We define excess mortality as the excess over the number of deaths that would have been expected counter-factually, that is without the catastrophic event in question. Based on this definition, we use a very parsimonious calculation method, namely the linear extrapolation of death figures from previous years to determine the excess mortality during the Covid-19 pandemic. But unlike most other literature on this topic, we first evaluated and optimized the specification of our method using a larger historical data set in order to identify and minimize estimation errors and biases. The result shows that excess mortality rates in the literature are often inflated. Moreover, they would have exhibited considerable excess mortalities in the period before Covid-19, if this value had already been of public interest at that time. Three conclusions can be drawn from this study and its findings: (i) All calculation methods for current figures should first be evaluated against past figures. (ii) To avoid alarm fatigue, thresholds should be introduced which would differentiate between 'usual fluctuations' and 'remarkable excess'. (iii) Statistical offices could provide more realistic estimates.

Keywords: excess mortality; Covid-19; prediction accuracy; alarm fatigue; introducing thresholds

Introduction

The coronavirus disease 2019 (Covid-19) was undoubtedly associated with increased mortality in many countries. But how high was this mortality compared to temporary excess mortality waves in the recent past? 'Excess mortality' is a concept that has been known in expert circles since the mid-19th century [1], but Covid-19 has brought it to wider attention even in the general public. It has recently become a means of public alarm, which makes it advisable to take a closer look at the semantic and statistical construction of this term.

Excess mortality is generally defined and determined quite simply: The current mortality is compared with the average mortality figures of the past for the territorial unit and season in question. If the present number of deaths exceeds the average, one speaks of 'excess mortality'. This simple demographic method has many advantages: The results are promptly available since there is no need to rely on often uncertain declarations concerning the cause of death. This method has already been used for some time to analyse, for example the extent of flu epidemics in winter or heat waves in summer without resorting to medical diagnoses, see for example [2, 3]. Since death itself is a fact that requires little interpretation, excess mortality can also be compared across long periods of time and different cultural groups exclusively on the basis of reasonably reliable demographic statistics.

In the case of Covid-19, tests for detecting an infection were developed very early on. Nevertheless, doubts concerning the daily numbers of Covid-19-associated deaths persisted: Were sufficient tests available everywhere? How many people died indirectly due to the epidemic, for example due to overcrowded hospitals? To what extent did all those who died with a positive test actually die from the disease or rather because of pre-existing co-morbidities? Since these questions are often difficult to answer, epidemiologists such as Thomas Beaney and colleagues refer to the excess mortality method as the 'gold standard' to assess the extent of the Covid-19 pandemic [4].

In the context of Covid-19, however, demographic excess mortality was not only treated as a scientific topos in comparatively small specialist circles, but was also received by the general public. Since the onset of the pandemic, national and international statistical agencies such as the OECD and Eurostat have begun publishing excess mortality figures regularly and as quickly as possible. These figures are then frequently disseminated in the mass media. Thus, 'excess mortality' has developed from a *terminus technicus* in limited expert circles to an object of public alarm, which is why increased methodological care is required to avoid misunderstandings. As with other politically relevant thresholds

Received: 28 December 2023. **Revised:** 01 May 2024. **Editorial decision:** 09 May 2024. **Accepted:** 16 May 2024

[5], more thought and discussion is needed to shape the concept of excess mortality.

First of all, it is necessary to clarify what is meant by excess mortality. In the literature, there are two meanings that must be distinguished from each other. In the first meaning, years with particularly low mortality are taken as the starting point for determining excess mortality. For example, the excess mortality of a current influenza wave is then calculated as the difference between the currently increased death figures and the particularly low mortality count in years in which influenza is hardly noticeable [6]. The second meaning, and the one most commonly used in the case of Covid-19, was already briefly outlined above: One takes the average mortality of the past (including influenza and all other causes of death) as a basis for estimating an expected value (Y) and calculates the excess mortality count (EM) as a difference between the expected value (Y) and the number of deaths currently observed (OD):

$$EM = OD - Y \tag{1}$$

Excess mortality, expressed as a percentage (EM%) of the expected number of deaths, is then calculated:

$$EM\% = (OD - Y)/Y \cdot 100 \tag{2}$$

The next question is how to determine the value of the expected mortality. The simplest method has already been indicated above: one takes an average value over a defined number of past years. The OECD's statistical office, for example, takes the average death counts of the last 5 years [7], while Eurostat uses the average of the last 4 years before Covid-19 as the expected value (for Eurostat, see https://ec.europa.eu/eurostat/cache/metadata/en/demo_mexrt_esms.htm). However, there is criticism of this method. While averages from the past are suitable for smoothing out random fluctuations between years, they are not able to capture falling or rising trends in mortality figures.

Falling trends may be the result of increasing life expectancy, while rising trends are caused by increasing population numbers, especially in older birth cohorts. In many European countries, we have been observing upward trends in death counts for some time now, as is in fact predicted by the theory of the second demographic transition [8]. In our sample of 31 European countries, the total annual mortality increased from 3.67 million deaths in 1960 to 4.78 million in 2019, or by about 30%. If this trend towards rising death figures is not taken into account, excess mortality is regularly overestimated because the average of the past years is lower than the upward trend.

To address this problem, some authors suggest taking into account the varying age cohort composition and their specific mortality rates, thus capturing the trend towards higher mortality rates in ageing populations, for example, [9, 10]. Conversely, however, this creates the problem that rising life expectancy is disregarded; the expected values may therefore be too high and excess mortality would be underestimated accordingly [11].

As long as the excess mortality is very significant, as was the case with Covid-19 particularly concerning outbreaks in some local hotspots, a lack of precision in the expected values presents no problem as the estimation error tends to be irrelevant when the number of excess deaths is high. This changes, however, when entire countries are compared on the basis of longer periods—for example over 1 or 2 years—because here the excess mortalities sometimes do not turn out very clearly and even lower than expected mortalities may be recorded [12]. Often, these country comparisons, which are now increasingly emerging in the literature, implicitly or explicitly raise the question: How well did each country cope with the pandemic [13–16]? This also involves the question of how successful the bundles of measures chosen by the respective governments—restrictive or less restrictive—were [17]. Because of the relatively small differences between countries, a higher degree of precision is required here than in the study of individual outbreaks.

It is all the more surprising that to our knowledge only few authors have evaluated their own estimation method on the basis of past mortality data [18–20]: How well would the mortality figures have been predicted by the method employed for a period before Covid-19, that is before the state of emergency? For only when a systematic assessment of the measurement error of the respective method has been carried out on the basis of comparative data can one actually validate the resulting statements on excess mortality during the Covid-19 pandemic. Indeed, several recently published papers show that the figures on excess mortality vary greatly depending on the method and its specification [9, 11, 17, 21–23]. The over- and underestimates are also quite divergent for different countries, so that not only does the overall level rise or fall, but also the ranking of countries changes with the method—that is individual countries may do 'well' with one method and 'poorly' with another.

Levitt *et al.* [9] compare four prominent methods as well as their own age-stratified model with the finding that in the 33 countries studied, the excess mortality of all these countries together is between 1.6 and 2.8 million deaths, in the highest case almost twice the lowest estimate. For individual countries like Germany, the deviation is even more pronounced—the lowest estimate here results in 55 000, the highest in 203 000 additional deaths. The relevant studies also emphasize the enormous sensitivity of most methods to the time frame chosen to calculate the expected values [17, 21–23]. Since most studies for determining the expected value for the Covid-19 pandemic refer to the five preceding years, that is 2015–2019, the high number of deaths in 2015 and the low number of deaths in 2019 result in a trend that is relatively flat or—in the original and later corrected spline model of the WHO for Germany—even declining due to these coincidences. In consequence, the expected value for 2020 and 2021 is estimated significantly too low (cf. figure 1 in [23]).

In contrast to other alarm values [5], it is also noticeable that threshold values are rarely given. These could serve to distinguish usual mortality fluctuations from extraordinary and therefore politically really noteworthy excess mortalities. First of all, it should be noted that the term 'excess mortality' in everyday language automatically implies the idea of an exceptional event, although according to the definition used by experts a certain excess mortality is completely usual, because the fluctuations of excess mortality and lower than expected mortality balance out to zero over the years, provided that the expected value was estimated correctly. So, in order to determine whether excess mortality is exceptional for a particular country in a particular year, one has to study the figures from the past: Taking Germany as an example, is an excess mortality of 2.5% in the Covid-19 year 2020 'exceptional' compared to a longer period before?

In conclusion, methods for determining excess mortality have rarely been evaluated and compared on the basis of actual and longer-term mortality figures. The present article attempts to fill this gap. A computation method for the expected value is proposed that takes into account both fluctuations and trends while being very simple to calculate, correspondingly easy to understand and parsimonious in its data requirements. The method described is then tested for accuracy using mortality data from 31 Eurostat countries for the period from 1965 to 2019. It is shown that it can compete in accuracy with more sophisticated methods and performs significantly better than the calculation methods used by the OECD and Eurostat. Figures for excess mortality for 2020 and 2021 are then derived and discussed with respect to historic excess mortality between 1965 and 2019. This reveals that the excess mortality of the Covid-19 years is within the range of the 95%-prediction interval in 8 countries while it exceeds it in 23 countries. The discussion concludes with two policy recommendations: First, statistical offices could use better estimation methods, while remaining consistent with their plausible goal of 'following a simple and easily interpretable calculation' (see [7], p. 1). Secondly, as with other warnings, appropriate thresholds should be set. According to our findings, alert for excess mortality in the past would have been quite common with the calculation methods currently used—that is, if excess mortality had already been reported in the period before Covid-19.

Method development

As the central basis of our calculations, we use the annual death figures of all 31 countries in Eurostat which provide complete data for the period from 1960 to 2021 (code: demo_magec). To compensate for distortions due to leap years, the death figures in these years were multiplied by a correction factor (1/366*365) and thus all years were normalized to 365 days. Data for Germany before 1990 includes both the Federal Republic and the Democratic Republic, that is West- and East-Germany combined. France is indicated there partly with and partly without its non-European territories. Based on overlaps, a conversion factor was estimated and thus the missing figures—2013 to 2021—were imputed for metropolitan France. The year 1960 is missing for Cyprus.

The method for estimating the expected values is based on the consideration that there are annual fluctuations and longer-term upward and downward movements, as already indicated above. Figure 1 shows the situation before the pandemic in the years 1960–2019, using Germany as an example. It can be seen that the number of deaths tended to rise until the mid-1970s, then fell until the mid-2000s and has since been rising again. It is also worth noting the significant jumps that can occur between individual years.

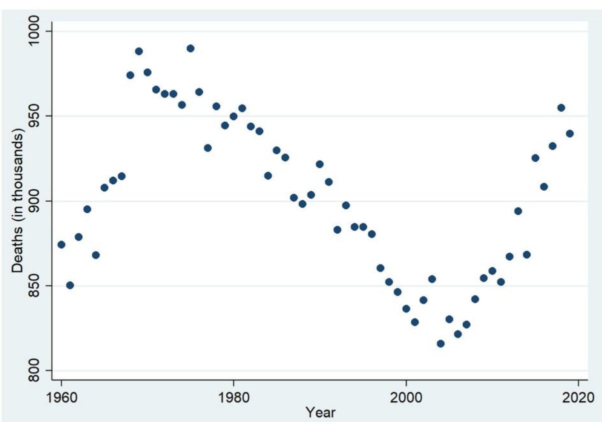

Figure 1. Fluctuating death counts in Germany before Covid-19

To capture longer term tendencies and smooth out year-to-year fluctuations, a linear trend is estimated using the death count in a specified number (*n*) of previous years:

$$Y(t) = c + \beta \cdot t + \varepsilon \tag{3}$$

where Y(*t*) is the expected value in focus year (*t*), *c* is the constant, and β is the slope factor. The residual (ε), that is the difference between the trend (*Y*) and the number of deaths actually observed (OM), is at the same time the excess mortality count (EM) for the respective year [see equation (1)].

The constant and slope are determined using the most common method, ordinary least squares (OLS) regression. Demographers often use Poisson regression instead since the response variable, that is the number of deaths, is 'count data' and therefore not continuous [24]. However, this is not necessary here, as the number of deaths over whole countries and whole years is large enough to be far from zero and the residuals, which are the EM count at the same time, are normally distributed (see Fig. 5). Calculations were made with LibreOffice (Calc) Version 7.3.6.2 (x64) and uploaded to an Open Science Repository (OSF, see below). Graphs were edited with STATA Version 13.1.

How many previous years are included in the estimation is theoretically a question of balance: the more years are included, the better random fluctuations between the years are evened out, but the slower the expected value reacts to changes in the trend. Hence, in the following, it is empirically determined which number (*n*) of previous years provides the most accurate estimate and is therefore recommended.

But how is the quality of the expected value to be assessed? Here we roughly follow the concept of [11]. First, the Mean Absolute Percentage Error (MAPE) of the estimation procedure must be determined. To do this, the absolute value of the respective deviations of the observed value from the expected value is added up and divided by the number (*N*) of corresponding years (e.g. 1965–2019 = 55 years):

$$MAPE = 1/N \cdot \sum_{t=1}^{t=N} |EM\%(t)| \tag{4}$$

This first parameter should be relatively low, but it will be one to two percentage points above zero, as the annual variation, mainly due to the presence or absence of more severe influenza waves in winter and heat waves in summer, cannot be predicted. Accordingly, a value of 'zero' for MAPE should not be expected, as this would imply an 'overfitting model' (see [25], 29ff.). Such an overfitting model would follow the short-term fluctuations of past mortality exactly and would therefore fail to capture the long-term mortality trend of a population as determined by its size, age structure and life expectancy. Or, to put it another way, if the regression line were identical to the actual number of deaths, the excess (and lower than expected) mortality of the past would no longer exist. This means that MAPE is the indistinguishable sum of 'excess' and 'error'. Thus, although a zero MAPE should not be aimed for, the relative performance of an estimation method can be identified: The lower the MAPE on a larger set of comparable data, the better the method.

Secondly, the predominant direction of the estimation error, its bias, must be determined. The calculation is similar; but instead of the absolute value, the actual values are summed in the conventional way so that positive and negative deviations balance each other out. However, only the last 5 years before the focussed event are considered (for Covid-19: 2015–2019), because this shorter period (n) is more relevant for the present than the longer period, in which the bias levels out if the trend is reversed:

$$Bias = 1/n \cdot \sum_{t=1}^{t=n} EM\%(t) \quad (5)$$

Ideally, this second parameter should be as close to zero as possible. A positive bias means that the expected value tends to underestimate the observed value and consequently, on average, shows inflated excess mortality rates. The reverse is true for a negative bias; in this case, excess mortality is underestimated.

Figure 2 compares two methods for forming the expected value based on the death counts in Germany. It can be seen that the five-year mean, as used by the OECD, always lags behind the actual development and estimates the expected values too low when the numbers are rising and, conversely, too high when the mortality is declining. In contrast, the 5-year trend is often much closer to the actual development. The difference becomes most pronounced when looking at either rising or falling phases. On the other hand, at turning points, the trend method tends to overshoot (as is most evident around 1970), while the mean-based method reacts more smoothly to the change in direction.

In addition, in order to judge about alarm thresholds, we should introduce prediction intervals (sometimes synonymously called 'forecast intervals', but not to be confused with the more commonly used 'confidence intervals', see [26], 205ff). Figure 3 shows how the width of the 95%-prediction intervals for the 5-year trend fluctuates considerably, depending on whether the figures for the previous 5 years follow a straight line or suddenly jump. It can be observed that even in the pre-Covid-19 period, the upper threshold was exceeded twice, first in 1975 and then again in 2002. Yet, the historical records of these years do not recall any epidemics or other elevated life threats in Germany. See https://www.hdg.de/lemo/jahreschronik/1975.html and https://www.hdg.de/lemo/jahreschronik/2002.html. Thus, even with a rather high threshold of 95%, false alarms are not completely excluded.

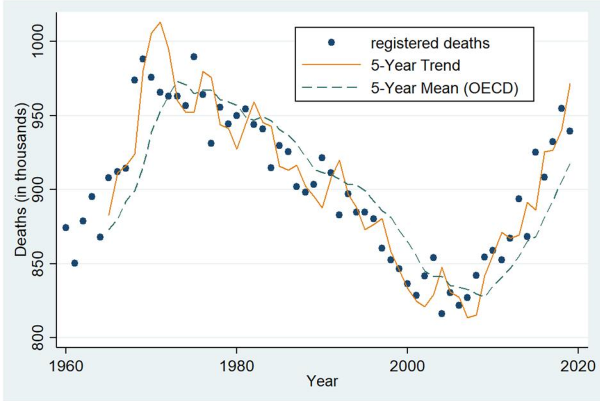

Figure 2. Two different prediction methods try to follow the registered number of deaths. The lines shown are the results of 55 forecasts, each based on the values of the previous 5 years. Thus, the first year with predicted values is 1965 (as calculated from the years 1960 to 1964)

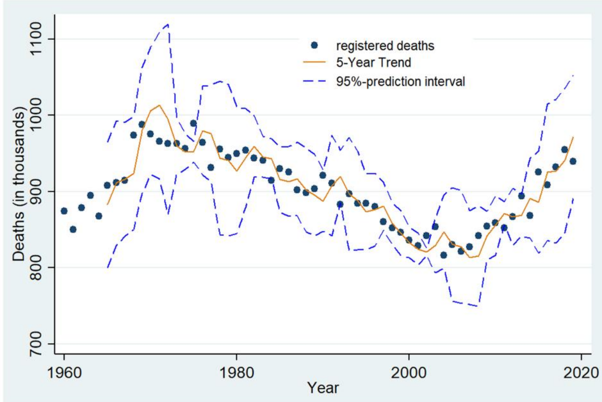

Figure 3. Fluctuating death counts, the 5-year trend, and its respectively floating 95%-prediction intervals.

Results

First, we tried to find out which trend has the most favourable performance indicators. Table 1 shows that, calculated over all countries and years (1960–2019), the estimation error (MAPE) is lowest for the 7-year trend. In contrast, the 3-year trend has the lowest bias. In this respect, the 5-year trend represents a good compromise between the two performance indicators—accordingly, our further calculations are based on this choice. If we compare the 5-year trend with the standard method of statistical offices, the 5-year or 4-year mean, we see a particularly marked difference in terms of bias; this is 9 times higher for the 5-year mean, that is the OECD method, and 7 times higher for the 4-year mean, the Eurostat method, than for our 5-year trend. The difference in the estimation errors (MAPE) is not as drastic and amounts to 22% and 11%.

In the following, we examine how the use of the 5-year trend, that is the method proposed here, affects the estimate of excess mortality in the Covid-19 period, that is for the years 2020 and 2021 together. For simplicity, only the OECD method (the 5-year mean) is now used for comparison, as the Eurostat method (the 4-year mean) would yield very similar results. For the second pandemic year (2021), we have used the trend or mean from 2015 to 2019 to estimate the expected value so that the calculations are not biased by the increased number of deaths in the first pandemic year (2020). Figure 4 shows the average annual excess mortality over the 2-year period 2020–2021, ordered from left to right using the excess mortality values by country derived from the 5-year trend. The differences between the two methods are very pronounced for some countries. They result from the fact that the mean method, as already mentioned, does not take into account underlying gradients: If the death figures show an increasing tendency, as is the case in Northern and Western Europe, the mean method overestimates excess mortality; if they show a decreasing tendency, as is partly the case in Eastern Europe, this leads to an underestimation (cf. also the changing gradients in Fig. 2).

It can also be seen in Fig. 4 that excess mortality during the pandemic in Europe shows a high variance—ranging from 0.7% in Iceland to 27% in Bulgaria. These are the currently much-noticed differences between countries, which are often spontaneously attributed to the measures against the pandemic and their more or less strict compliance. In general, it becomes obvious that the lowest excess mortality is predominantly found in Northern Europe, followed by Western Europe, Southern Europe, and Eastern Europe, thus clearly following a negative socio-economic gradient. The correlation coefficient between excess mortality and wealth is indeed quite strong (in the negative direction) and highly significant. (Wealth is measured as GDP per capita in 2020 in logarithmic form. The correlation coefficient is −0.82, with $P<0.001$).

Table 1. Comparison of estimation errors: MAPE and Bias of different methods for 31 nations

	MAPE		Bias	
Method	**1960–2019 mean (%)**	**95% CI (±)**	**1960–2019 mean (%)**	**95% CI (±)**
2Y-Trend	4.54	0.23	0.23	0.98
3Y-Trend	3.33	0.16	**0.20**	0.64
4Y-Trend	2.90	0.14	0.29	0.48
5Y-Trend	2.77	0.13	0.32	0.48
6Y-Trend	2.71	0.13	0.56	0.45
7Y-Trend	**2.69**	0.13	0.78	0.44
8Y-Trend	2.70	0.13	1.01	0.44
9Y-Trend	2.72	0.13	1.24	0.43
4Y-Mean	3.07	0.13	2.43	0.45
5Y-Mean	3.38	0.14	2.86	0.47

The most favourable values are highlighted in bold. The 5-year trend thus represents the best compromise (marked yellow).

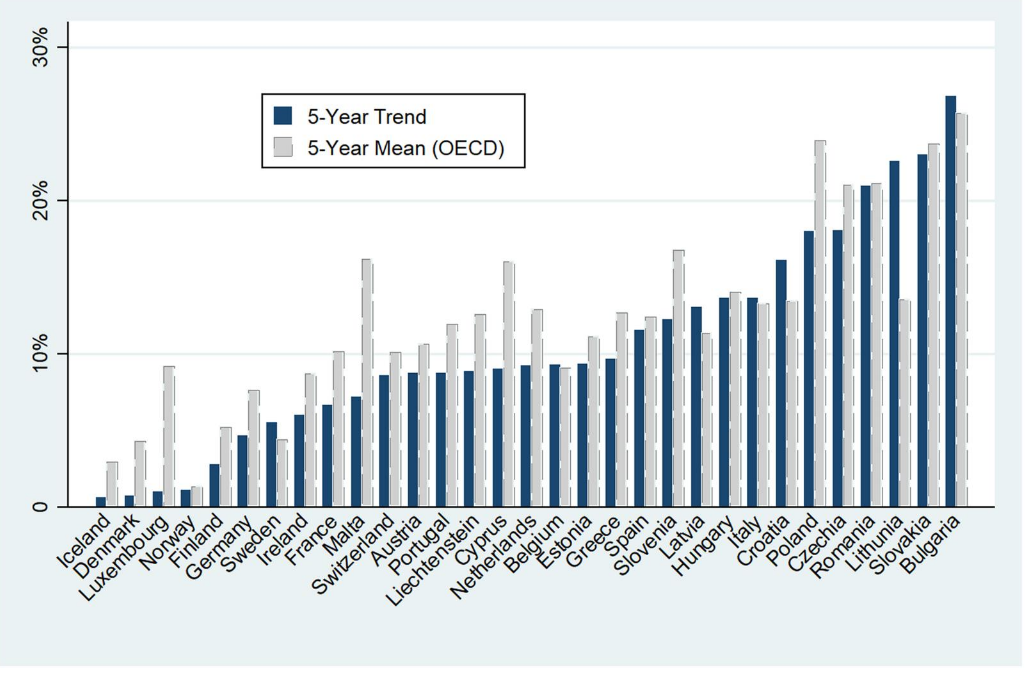

Figure 4. Average annual excess mortality in the 2-year period 2020 to 2021, that is during the Covid-19 pandemic. We compare our proposed method (5-year trend) with the estimates of the OECD statistical office (5-year mean).

However, we do not know how high the variations in excess mortality were in the past and how much they differed between the countries. In other words: Before we try to interpret death tolls and containment measures during the pandemic, we should analyse the usual mortality fluctuations from pre-pandemic years. Figure 5 attempts to answer this question by comparing the maximal annual excess mortality during the Covid-19 pandemic with the higher and the lower-than-expected mortality from 1965 to 2019. Here, it becomes visible that the mortality figures in countries with smaller populations—due to the law of small numbers, that is for statistical reasons—fluctuate strongly anyway. This applies, in particular, to Liechtenstein with only 38 000 inhabitants. Cyprus is a special case, since there are many breaks in the statistical time series.

However, even countries with comparatively large populations, such as Germany, France, Italy, and Spain, show rather wide fluctuations and accordingly rather wide prediction intervals. In respect to Covid-19, we see that in 23 out of 31 countries, the maximal annual excess mortalities in the period from 2020 to 2021 exceed the 95%-prediction interval threshold, while only eight countries remain in this range. However, even in the period before Covid-19, there were 49 other country–years with excess mortality above the respective 95%-prediction interval and 56 country–years below this threshold. These numbers are a bit higher than expected: With 31 countries times 55 years times 5% probability, only 85 extraordinary events should have occurred. Furthermore, it is noteworthy (from an analysis not shown here) that highly elevated excess mortality figures in the pre-Covid-19 period do not occur in or around any particular year—that is they cannot be explained by other pandemics, such as the Hong Kong flu, which spread worldwide from 1968 to 1970 [27].

Discussion and conclusions

Four main results should be noted and discussed: First, the clarification of different meanings of the term "excess mortality";

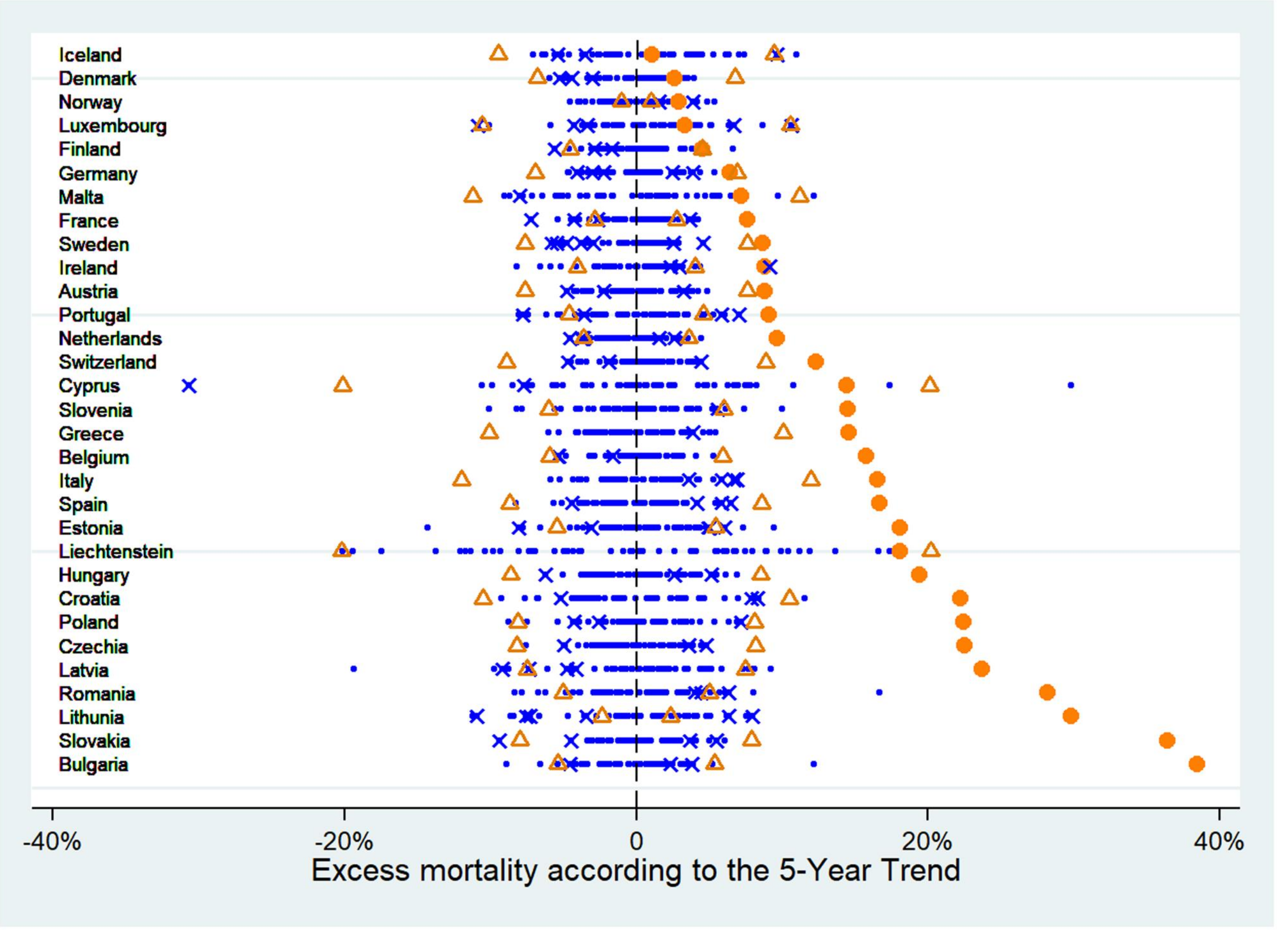

Figure 5. Annual excess mortality during the Covid-19 pandemic from 2020 to 2021 (large orange dots represent the higher value, either from 2020 or from 2021) compared with excess mortality values from 1965 to 2019 (small blue dots). Hollow orange triangles indicate the lower and upper 95%-threshold for the respective large orange dot (i.e. the Covid-19 excess). Large blue Xs indicate years before 2020 with mortality figures outside their respective 95%-prediction interval. The dashed line in the middle of the graph highlights the observation that 'excess mortality' and 'lower than expected mortality' events before 2020 are roughly symmetrically distributed around zero.

secondly, the reference of excess deaths to expected deaths and not to population numbers; thirdly, the need for an evaluation method; fourthly, the accuracy of our calculation method compared to other methods commonly used in the literature.

First, it has been shown that relatively high variations in death figures, and thus in excess mortality, already occurred in the past. Our estimation procedure also reveals years with remarkably lower-than-expected mortality. In the light of these fluctuations, the question arises as to which horizons of meaning should be invoked with the term 'excess mortality'. Three different notions are conceivable and should not be confused with each other:

i) In the epidemiological and demographic literature, it is often argued that all 'avoidable' deaths should be understood as excess mortality, that is also usual flu deaths, heat deaths, road traffic victims, etc. (e.g. [6, 28]). As a result, the baseline from which excess mortality is calculated is generally lowered and at the same time the normative reference point is shifted. The benchmark is not the socially accepted normality prior to the exceptional event under consideration, but ideas about which causes of death are preventable or treatable. For example, the OECD considers most deaths under the age of 75 years—that is more than 25% of all deaths in OECD countries—as avoidable [29]. While these ideas may be plausible, they are moving targets in scientific debates rather than fixed points of reference across time, space, and culture, that could be used to determine, for example, the magnitude of a pandemic on a historical scale [30].
ii) The notion established by William Farr [1], used in this text and still the most common among statistical offices today, is that of a deviation from a longer-term average or trend, which is understood as 'demographic normality'. Of course, ideas about normality are also normative, but they are less arbitrary than ideas about avoidability, since they are based on the judgement of the general public in a given society and a given time, that is on what is publicly perceived as acceptable in terms of the trade-offs between prevention, personal liberties and economic investments. However, a threshold value is not envisaged here, and if the calculation method avoids bias, lower-than-expected mortality can occur just as easily as higher-than-expected mortality.
iii) Finally, there is the possible, and in the public mind also observable misconception that 'excess mortality' would imply an alert threshold that is supposed to indicate an extraordinary event. Here, too, the point of reference is demographic normality, except that it is not the zero point of usual fluctuations between excess mortality and lower than expected mortality that would be taken as the baseline, but an upper

threshold value. In fact, as a result of our study, we would explicitly propose such a threshold value, for example beyond the 95%-prediction interval. Assuming this threshold, for Covid-19 no excess mortality alarm would be triggered in 8 out of the 31 countries in 2020 and 2021. Furthermore, in order to avoid 'alarm fatigue' [31], we would suggest to reconsider the use of the term 'excess mortality' in the mass media and in policy communication. In a public-understanding-of-science-perspective [32], for the usual fluctuations around the baseline we would recommend using the term 'higher than expected mortality' symmetrically to the term 'lower than expected mortality' while reserving the more alarmist-sounding term 'excess mortality' only for events above a clearly demarcated threshold. For the upper threshold of a 95%-prediction interval, this would mean that such an event would only be expected with a 2.5% probability, that is once in every 40 years. Of course, lower thresholds—for example 80% or 90%, corresponding to a probability of once in 10 or 20 years—would be possible, but would increase alarm frequency accordingly. (On the other hand, for statistical reasons, the higher frequency would allow for more accurate predictions.) Even with the rather high threshold of 95%, we have observed 49 events in the pre-Covid-19 period that exceeded the upper limit, which raises the question of whether an alarm would have been considered justified in each case.

Secondly, we related the number of excess deaths to the number of expected deaths, rather than to population numbers, as demographers often do (e.g. [14]). Theoretically, this could make a big difference—not so much when comparing individual years in the same country, but considerably when comparing different countries in the same year. Depending mainly on the age distribution in a given country, usual mortality rates vary accordingly—for example in our sample from 640 deaths (Ireland) to 1532 deaths (Bulgaria) per 100 000 inhabitants per year during the pre-Covid-19 period from 2015 to 2019. Thus, countries with a low usual mortality rate would be better placed in a ranking with excess deaths in respect to the (relatively high) number of inhabitants than in a ranking with excess deaths in respect to the (relatively low) number of usual deaths. Empirically, however, the choice of denominator—usual death figures or number of inhabitants—is not so important in our case study: the correlation between excess mortality in terms of usual deaths and excess mortality in terms of 100 000 inhabitants is very high (0.96) and, accordingly, changing the reference would change the country rankings only minimally. We have therefore preferred to use the reference to usual death counts because it is more intuitive, better comparable to data from statistical offices and more parsimoniously, relying only on death counts rather than more error-prone population estimates.

Thirdly, it has been demonstrated that different calculation methods lead to quite different results. It is therefore necessary to calibrate the calculation method used against a past that is considered 'usual' before applying it to an exceptional event. The prevailing neglect of this necessity is all the more astonishing as the methodological reviews cited above show considerable differences in results due to the choice of method. However, considering that excess mortality rates are relevant to public controversies and thus politically sensitive, one should a priori choose a transparent and easily comprehensible method in order to avoid the suspicion of 'cherry-picking'—and corresponding alarmist or anti-alarmist motives. Most of the recently published reviews that refer to the problem of method selection remain rather vague in their recommendations for remedial action. Nepomuceno *et al.* [17], for example, continue to leave the decision on method to the discretion of the researcher: 'Moreover, the method used and the reference period (which may or may not include leap week years) should both be chosen carefully' (p. 21). Only Jonas Schöley [11] uses past data—similar to ours—to assess the accuracy of different methods, but unfortunately in a way that the results are not directly comparable to ours. (Schoeley divides the expected values by the observed values and not vice versa as is done here in the text. Especially in the case of high deviations, this leads to the problem that excess mortalities and lower-than-expected mortalities are not represented symmetrically.)

Fourthly, we have proposed and tested a parsimonious calculation method with comparatively good results, cf. [23]. We were able to demonstrate that it has a significantly lower bias and a slightly higher accuracy compared to the methods used by Eurostat and the OECD. Importantly, this is achieved without higher demands on data availability and mathematical understanding, unlike other methods described in the literature, for example [33]. We therefore believe that it represents a good compromise between ease of verification and statistical accuracy. In particular, the bias is very low and seems to be competitive with computationally much more sophisticated methods.

Concerning the deviation of expected from observed values, the accuracy may still be improved; the estimation error (MAPE) could possibly be diminished. Particularly for longer periods, for example the whole 5 years that a pandemic typically lasts [34], predictions of expected deaths should be based on a split procedure: According to the method established by [18], not the whole death counts, but only the mortality rates of different age groups should be extrapolated and then multiplied with the number of people in that respective groups (which may change over the course of longer periods). In terms of causality, it would also be worth considering how the remarkably lower-than-expected mortalities shown in Fig. 5 come about. In the epidemiological literature, mortality displacement effects are occasionally reported, that is in the case of heat or flu waves, severely pre-weakened persons die and therefore lower-than-expected mortalities can occur afterwards [35]. To the best of our knowledge, these effects have never been studied in larger demographic contexts over longer periods of time.

However, all methods for assessing excess mortality should be evaluated with regard to their estimation error and bias using standardized procedures, in order to be able to exploit their advantages over other measurement methods (such as direct disease diagnosis). This is particularly important since pandemics—throughout history [36]—have generated considerable controversy and therefore any appearance of non-transparency and political bias on the part of science should be avoided as far as possible. Excess mortality as a summary indicator of a pandemic crisis is too relevant for health policies and public health interventions to risk potentially misleading calculations and interpretations.

Author contributions

Bernhard Gill (Conceptualization [lead], Formal analysis [equal], Methodology [equal], Writing—original draft [lead]), Theresa Kehler (Data curation [lead], Formal analysis [equal], Visualization [equal], Writing—review & editing [equal]), and Michael Schneider (Methodology [equal], Project administration [lead], Writing—review & editing [equal])

Conflict of interest statement. None of the authors have any conflicts of interest to declare.

Funding

The study was partly supported by the Volkswagen Foundation (grant Az. 99 026).

Data availability

The spreadsheet file with the respective data and computations is made available at: https://osf.io/7wxyt

References

1. Honigsbaum M. Revisiting the 1957 and 1968 influenza pandemics. *Lancet* 2020;**395**:1824–6.
2. Canouï-Poitrine F, Cadot E, Spira A *et al.* Excess deaths during the August 2003 heat wave in Paris, France. *Revue D'épidémiologie et de Santé Publique* 2006;**54**:127–35.
3. Nunes B, Viboud C, Machado A *et al.* Excess mortality associated with influenza epidemics in Portugal, 1980 to 2004. *PloS One* 2011;**6**:e20661.
4. Beaney T, Clarke JM, Jain V *et al.* Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? *J R Soc Med* 2020;**113**:329–34.
5. Schmidt M, Glasson J, Emmelin L *et al. Standards and Thresholds for Impact Assessment*. Berlin/Heidelberg: Springer, 2008.
6. Shkolnikov VM, Klimkin I, McKee M *et al.* What should be the baseline when calculating excess mortality? New approaches suggest that we have underestimated the impact of the COVID-19 pandemic and previous winter peaks. *SSM Popul Health* 2022; **18**:101118.
7. OECD. *Methodological Note—All-cause, Excess and COVID-19 Deaths*. https://stats.oecd.org/wbos/fileview2.aspx?IDFile=97aacc20-eac4-4e44-8f59-9fb5b7e25090 (23 February 2024, date last accessed).
8. Lesthaeghe R. The second demographic transition: a concise overview of its development. *Proc Natl Acad Sci USA* 2014; **111**:18112–5.
9. Levitt M, Zonta F, Ioannidis J. Comparison of pandemic excess mortality in 2020-2021 across different empirical calculations. *medRxiv*, 2022.
10. Kowall B, Standl F, Oesterling F *et al.* Excess mortality due to Covid-19? A comparison of total mortality in 2020 with total mortality in 2016 to 2019 in Germany, Sweden and Spain. *PloS One* 2021;**16**:e0255540.
11. Schöley J. Robustness and bias of European excess death estimates in 2020 under varying model specifications. *medRxiv*, 2021.
12. Sanmarchi F, Golinelli D, Lenzi J *et al.* Exploring the gap between excess mortality and COVID-19 deaths in 67 countries. *JAMA Netw Open* 2021;**4**:e2117359–e2117359.
13. Karlinsky A, Kobak D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife* 2021;**10**:e69336.
14. Wang H, Paulson KR, Pease SA *et al.* Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* 2022;**399**:1513–36.
15. Aburto JM, Schöley J, Kashnitsky I *et al.* Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int J Epidemiol* 2022; **51**:63–74.
16. Pifarré I Arolas H, Acosta E, López-Casasnovas G *et al.* Years of life lost to COVID-19 in 81 countries. *Sci Rep* 2021;**11**:3504–6.
17. Nepomuceno MR, Klimkin I, Jdanov DA *et al.* Sensitivity analysis of excess mortality due to the COVID-19 pandemic. *Popul Dev Rev* 2022;**48**:279–302.
18. De Nicola G, Kauermann G, Höhle M. On assessing excess mortality in Germany during the COVID-19 pandemic. *AStA Wirtschafts und Sozialstatistisches Archiv* 2022;**16**:5–20.
19. Duerst R, Schöley J, Bohk-Ewald C. *A Validation Workflow for Mortality Forecasting*. Max Planck Institute for Demographic Research, 2023. MPIDR Working Paper WP-2023-020, 27 pages.
20. Acosta RJ, Irizarry RA. A flexible statistical framework for estimating excess mortality. *Epidemiology* 2022;**33**:346–53.
21. Levitt M, Zonta F, Ioannidis J. Excess death estimates from multiverse analysis in 2009-2021. *medRxiv*, 2022.
22. Mazzuco S. Measuring mortality crises: a tool for studying global health. In: *Oxford Research Encyclopedia of Global Public Health*, Oxford University Press, 2022. https://doi.org/10.1093/acrefore/9780190632366.013.434.
23. Ferenci T. Comparing methods to predict baseline mortality for excess mortality calculations. *BMC Med Res Methodol* 2023; **23**:239.
24. Winkelmann R, Zimmermann KF. Count data models for demographic data. *Math Popul Stud* 1994;**4**:205–21, 223.
25. Burnham KP, Anderson DR (eds). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer New York, 2002.
26. Hamilton LC. *Statistics with Stata: Version 12*. Boston Mass: Cengage Learning, 2012.
27. Viboud C, Grais RF, Lafont BA, *et al.* Multinational impact of the 1968 Hong Kong influenza pandemic: evidence for a smoldering pandemic. *J Infect Dis* 2005;**192**:233–48.
28. Nielsen J, Vestergaard LS, Richter L *et al.* European all-cause excess and influenza-attributable mortality in the 2017/18 season: should the burden of influenza B be reconsidered? *Clin Microbiol Infect* 2019;**25**:1266–76.
29. OECD. Avoidable mortality (preventable and treatable). In: *Health at a Glance*. Paris: OECD, 2019, 72–74.
30. Gill B, Kehler T, Schneider M. Is Covid-19 a dread risk? The death toll of the pandemic year 2020 in long-term and transnational perspective. *Futures* 2022;**142**:103017.
31. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol* 2012;**46**:268–77.
32. Durant JR, Evans GA, Thomas GP. The public understanding of science. *Nature* 1989;**340**:11–14.
33. WHO. *Methods for Estimating the Excess Mortality Associated with the COVID-19 Pandemic*. https://www.who.int/publications/i/item/estimating-global-and-country-specific-excess-mortality-during-the-covid-19-pandemic (17 November 2022, date last accessed).
34. Kissler SM, Tedijanto C, Goldstein E *et al.* Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020;**368**:860–8.
35. Armstrong B, Bell ML, de Sousa Zanotti Stagliorio Coelho M *et al.* Longer-term impact of high and low temperature on mortality: an international study to clarify length of mortality displacement. *Environ Health Perspect* 2017;**125**:107009.
36. Cohn SK. Pandemics: waves of disease, waves of hate from the Plague of Athens to AIDS. *Hist J* 2012;**85**:535–55.
37. De Nicola G, Kauermann G. An update on excess mortality in the second year of the COVID-19 pandemic in Germany. *AStA Wirtschafts-Und Sozialstatistisches Archiv* 2022b;**16**:21–24.

Biology Methods and Protocols, 2024, 00, 1–8
https://doi.org/10.1093/biomethods/bpae031
Review