

1 **Calculation and meaning of "excess mortality": A comparison of Covid- and pre-Covid mortality**
2 **data in 31 Eurostat countries from 1965 to 2021.**

3

4 Bernhard Gill, Theresa Kehler, Michael Schneider, Institute for Sociology, Ludwig-Maximilians-
5 Universitaet Muenchen, Germany, December 2022

6

7 **0) Abstract**

8 Determining "excess mortality" makes it possible to compare the burden of disasters between
9 countries and over time, and thus also to evaluate the success of mitigation measures. However,
10 the debate on Covid-19 has exposed that calculations of excess mortalities vary considerably
11 depending on the method and its specification. Moreover, it is often unclear what exactly is meant
12 by "excess mortality". We define excess mortality as the excess over the number of deaths that
13 would have been expected counter-factually, i.e. without the catastrophic event in question. That
14 is, we include all normally occurring flu and heat waves, which are excluded by some authors with
15 the consequence that they almost always record low expected values and correspondingly high
16 excess mortality rates. Based on this definition, we use a very parsimonious calculation method
17 that is easy to understand even for laypersons, namely the linear extrapolation of death figures
18 from previous years to determine the excess mortality during the Covid-19 pandemic. But unlike
19 other literature on this topic, we first evaluated and optimised the specification of our method
20 using a larger historical data set in order to identify and minimise estimation errors and biases. The
21 result shows that the excess mortality rates continuously published by international statistical
22 offices – OECD and Eurostat – are often inflated and would have exhibited considerable excess.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

23 mortalities in many countries and periods before Covid-19, if this value had already been of public
24 interest at that time. It also reveals that mortality rates already fluctuated strongly in the past and
25 that in a third of the countries studied, individual values from the past exceed the current
26 fluctuations due to the Covid-19 pandemic. Three conclusions can be drawn from this study and its
27 findings: 1) All calculation methods for current figures should first be evaluated against past
28 figures. 2) The definition of excess mortality used should be made explicit. 3) Statistical offices
29 should provide more realistic estimates.

30

31 **1) Introduction**

32 Covid-19 was undoubtedly associated with increased mortality in many countries. But how high
33 was this mortality compared to temporary excess mortality waves in the recent past? "Excess
34 mortality" is a concept that has been known in expert circles since the mid-19th century
35 (Honigsbaum 2020), but Covid-19 has brought it to wider attention even in the general public. It
36 has recently become a means of public alarm, which makes it advisable to take a closer look at the
37 semantic and statistical construction of this term.

38 Excess mortality is generally defined and determined quite simply: The current mortality is
39 compared with the average mortality figures of the past for the territorial unit and season in
40 question. If the present number of deaths exceeds the average, one speaks of "excess mortality".
41 This simple demographic method has many advantages: The results are promptly available since
42 there is no need to rely on often uncertain declarations concerning the cause of death. This
43 method has already been used for some time to analyse, for example, the extent of flu epidemics
44 in winter or heat waves in summer without resorting to medical diagnoses (e.g. Canoui-Poitaine et
45 al. 2006; Nunes et al. 2011). Since death itself is a fact that requires little interpretation, excess

46 mortality can also be compared across long periods of time and different cultural groups
47 exclusively on the basis of reasonably reliable demographic statistics.

48 In the case of Covid-19, tests for detecting an infection were developed very early on.
49 Nevertheless, doubts concerning the daily numbers of Covid-associated deaths persisted: Were
50 sufficient tests available everywhere? How many people died indirectly due to the epidemic, for
51 example due to overcrowded hospitals? To what extent did all those who died with a positive test
52 actually die from the disease or rather because of pre-existing co-morbidities? Since these
53 questions are often difficult to answer, epidemiologists such as Thomas Beaney and colleagues
54 (2020) refer to the excess mortality method as the "gold standard" to assess the extent of the
55 Covid-19 pandemic.

56 In the context of Covid-19, however, demographic excess mortality was not only treated as a
57 scientific topos in comparatively small specialist circles, but was also received by the general
58 public. Since the onset of the pandemic, national and international statistical agencies such as the
59 OECD and Eurostat have begun publishing excess mortality figures regularly and as quickly as
60 possible. These figures are then frequently disseminated in the mass media. Thus, "excess
61 mortality" has developed from a terminus technicus in limited expert circles to an object of public
62 alarm, which is why increased methodological care is required to avoid misunderstandings. As with
63 other politically relevant thresholds (Schmidt et al. 2008), more thought and discussion is needed
64 to shape the concept of excess mortality.

65 First of all, it is necessary to clarify what is meant by excess mortality. In the literature, there are
66 two meanings that must be distinguished from each other. In the first meaning, years with
67 particularly low mortality are taken as the starting point for determining excess mortality. For
68 example, the excess mortality of a current influenza wave is then calculated as the difference

69 between the currently increased death figures and the particularly low mortality count in years in
70 which influenza is hardly noticeable (Shkolnikov et al. 2022). The second meaning, and the one
71 most commonly used in the case of Covid-19, was already briefly outlined above: One takes the
72 average mortality of the past (including influenza and all other causes of death) as a basis for
73 estimating an expected value and calculates the excess mortality as a percentage of the
74 difference [1]:

$$75 \quad \text{Excess mortality in percent} = (\text{Observed mortality} / \text{Expected mortality} - 1) * 100$$

76 The next question is how to determine the value of the expected mortality. The simplest method
77 has already been indicated above: one takes an average value over a defined number of past years.
78 The OECD's statistical office, for example, takes the average death counts of the last five years,
79 while Eurostat uses the average of the last four years before Covid-19 as the expected value.
80 However, there is criticism of this method. While averages from the past are suitable for smoothing
81 out random fluctuations between years, they are not able to capture falling or rising trends in
82 mortality figures.

83 Falling trends may be the result of increasing life expectancy, while rising trends are caused by
84 increasing population numbers, especially in older birth cohorts. In many European countries, we
85 have been observing upward trends in death counts for some time now [2], as is in fact predicted
86 by the theory of the second demographic transition (Lesthaege 2014). If this trend towards rising
87 death rates is not taken into account, excess mortality is regularly overestimated because the
88 average of the past years is lower than the upward trend.

89 To address this problem, some authors suggest taking into account the varying age cohort
90 composition and their specific mortality rates, thus capturing the trend towards higher mortality
91 rates in ageing populations (e.g. Levitt et al. 2022a; Kowall et al. 2021). Conversely, however, this

91 creates the problem that rising life expectancy is disregarded; the expected values may therefore
92 be too high and excess mortality would be underestimated accordingly (Schöley 2021).

93 As long as the excess mortality is very significant, as was the case with Covid-19 particularly
94 concerning outbreaks in some local hotspots, a lack of precision in the expected values presents no
95 problem as the estimation error tends to be irrelevant when the number of excess deaths is high.
96 This changes, however, when entire countries are compared on the basis of longer periods – for
97 example, over one or two years – because here the excess mortalities sometimes do not turn out
98 very clearly and even lower than expected mortalities may be recorded (e.g. Sanmarchi et al.
99 2021). Often, these country comparisons, which are now increasingly emerging in the literature,
100 implicitly or explicitly raise the question: how well did each country cope with the pandemic
101 (Karlinsky/Kobak 2021; Wang et al. 2022; Aburto et al. 2022; Pifarré i Arolas et al. 2022)? This also
102 involves the question of how successful the bundles of measures chosen by the respective
103 governments – restrictive or less restrictive – were (Nepomuceno et al. 2022). Because of the
104 relatively small differences between countries, a higher degree of precision is required here than in
105 the study of individual outbreaks.

106 It is all the more astonishing that hardly anyone has evaluated their own estimation method so far.
107 The easiest way to do this is to use mortality figures from the past: How well would the mortality
108 figures have been predicted by the method employed for a period before Covid-19, i.e. before the
109 state of emergency? For only when a systematic assessment of the measurement error of the
110 respective method has been carried out on the basis of comparative data can one actually validate
111 the resulting statements on excess mortality during the Covid-19 pandemic. Indeed, several
112 recently published papers show that the figures on excess mortality vary greatly depending on the
113 method and its specification (Schöley 2021, Nepomuceno et al. 2022, Levitt et al. 2022a; Levitt et
114 al. 2022b; Mazzuco 2022; Ferenci 2022). The over- and underestimates are also quite divergent for

115 different countries, so that not only does the overall level rise or fall, but also the ranking of
116 countries changes with the method – i.e. individual countries may do "well" with one method and
117 "poorly" with another.

118 Levitt and colleagues (2022a) compare four prominent methods as well as their own age-stratified
119 model with the finding that in the 33 countries studied that have relatively reliable population
120 statistics, the excess mortality of all these countries together is between 1.6 and 2.8 million deaths,
121 in the highest case almost twice the lowest estimate. For individual countries like Germany, the
122 deviation is even more pronounced – the lowest estimate here results in 55,000, the highest in
123 203,000 additional deaths [3]. In general, the relevant studies also emphasise the enormous
124 sensitivity of most methods to the time frame chosen to calculate the expected values (Levitt et al.
125 2022b, Nepomuceno et al. 2022, Mazzuco 2022). In this context, Ferenci (2022) refers to the
126 significance of the strong influenza wave in 2015. Since most studies for determining the expected
127 value for the Covid 19 pandemic refer to the five preceding years, i.e. 2015 to 2019, the high
128 number of deaths in 2015 and the low number of deaths in 2019 result in a trend that is relatively
129 flat or – in the original and later corrected spline model of the WHO for Germany – even declining
130 due to these coincidences. In consequence, the expected value for 2020 and 2021 is estimated
131 significantly too low (Ferenci 2022, Figure 1).

132 In contrast to other alarm values (Schmidt et al. 2008), it is also noticeable that threshold values
133 are rarely given. These could serve to distinguish normal excess mortalities from extraordinary and
134 therefore politically really noteworthy excess mortalities. First of all, it should be noted that the
135 term "excess mortality" in everyday language automatically implies the idea of an exceptional
136 event, although according to the definition used by experts a certain excess mortality is completely
137 normal, because the fluctuations of excess mortality and lower than expected mortality would
138 balance out to zero over the years, provided that the expected value was estimated correctly. So,

138 in order to determine whether excess mortality is exceptional for a particular country in a
139 particular year, one has to study the figures from the past: For example, is an excess mortality of
140 2.5 percent in Germany in the Covid year 2020 exceptional, i.e. greater than most excess mortality
141 rates in the last 10 or 50 years?

142 In conclusion, methods for determining excess mortality have rarely been evaluated and compared
143 on the basis of actual and longer-term mortality figures. The present article attempts to fill this
144 gap. A computation method for the expected value is proposed that takes into account both
145 fluctuations and trends while being very simple to calculate, correspondingly easy to understand
146 and parsimonious in its data requirements. The method described is then tested for accuracy using
147 mortality data from 31 Eurostat countries for the period from 1965 to 2019. It is shown that it can
148 compete in accuracy with very complex methods and performs significantly better than the
149 calculation methods used by the OECD and Eurostat. Figures for excess mortality for 2020 and
150 2021 are then derived and discussed with respect to historic excess mortality between 1965 and
151 2019. This reveals that the excess mortality of the Covid-19 years is within normal fluctuations in
152 10 countries while it exceeds extreme values of the examined time frame since 1965 in 21
153 countries. The discussion concludes with two policy recommendations: First, statistical offices
154 could use better estimation methods without calculations becoming too complex and thus
155 suspected of being non-transparent. Second, as with other warnings, appropriate thresholds
156 should be set. According to our findings, alert for excess mortality in the past would have been
157 quite normal with the calculation methods currently used by statistical offices – that is, if excess
158 mortality had already been reported in the period before Covid-19.

159

160

161 2) Method development

162 As the central basis of our calculations, we use the annual death figures of all 31 countries in

163 Eurostat which provide complete data for the period from 1960 to 2020 (code: demo_magec) [4].

164 The values for 2021 were calculated from the more recent (and therefore provisional) data of

165 weekly death counts in Eurostat (code: demo_r_mwk_ts). To compensate for distortions due to

166 leap years, the death figures in these years were multiplied by a correction factor $(1/366 \cdot 365)$ and

167 thus all years were normalised to 365 days.

168 The method for estimating the expected values is based on the considerations that there are

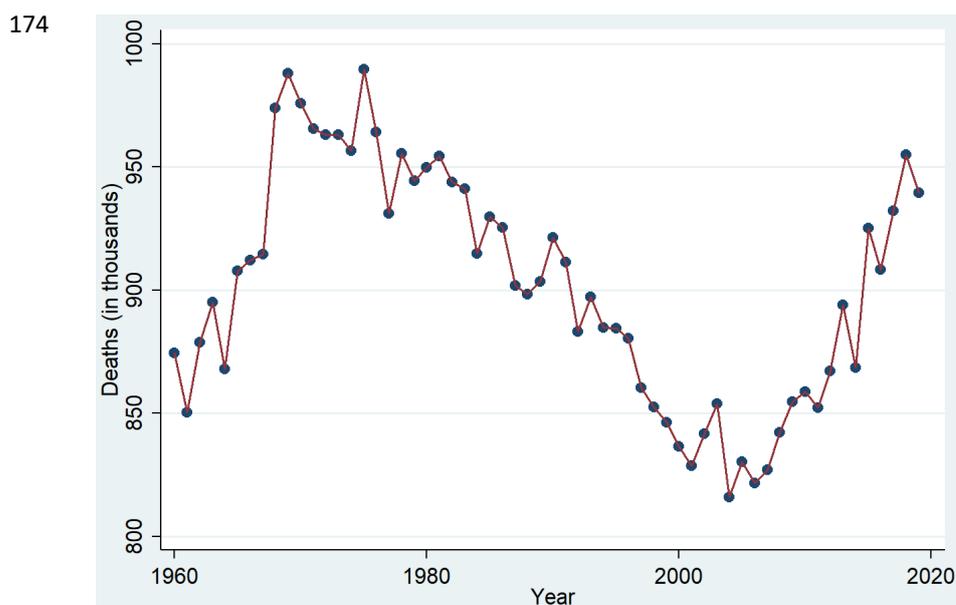
169 annual fluctuations and longer-term upward and downward movements, as already indicated

170 above. Figure 1a shows the situation before the pandemic in the years 1960 to 2019, using

171 Germany as an example. It can be seen that the number of deaths tended to rise until the mid-

172 1970s, then fell until the mid-2000s and has since been rising again.

173 **Figure 1a: Annual number of registered deaths in Germany before Covid-19**



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

177 Expected value in focus year (t) = constant + slope factor * t + residuals.

178 The constant and slope are determined by means of the most common method, Ordinary Least
179 Squares (OLS) regression, which is also implemented in simple spreadsheet programs, i.e. in widely
180 available standard software. How many previous years are included in the estimation is
181 theoretically a question of balance: the more years are included, the better random fluctuations
182 between the years are evened out, but the slower the expected value reacts to changes in the
183 trend. Hence, in the following it is empirically determined which number (n) of previous years
184 provides the most accurate estimate and is therefore recommended.

185 But how is the quality of the expected value to be assessed[5]? First, the **Mean Absolute**
186 **Percentage Error (MAPE)** of the estimation procedure must be determined. To do this, the absolute
187 value of the respective deviations of the observed value from the expected value is added up and
188 divided by the number of corresponding years (1965 to 2019 = 55 years):

190 $MAPE = 1/55 * \text{Sum (1 to 55 previous years) over } | \text{observed value/expected value} - 1 |$

192 This first parameter should be relatively low; but it will be a few percentage points above zero,
193 since the annual fluctuations resulting mainly from the occurrence or absence of more severe
194 influenza waves in winter and heat waves in summer cannot be predicted. Even if an "ideal value"
195 can hardly be obtained here, the relative performance of an estimation method compared to

196 another estimation method can nevertheless be easily identified on the basis of the average level
197 of the estimation error (MAPE): The lower the estimation error, the better the method.

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

204

205
$$\text{Bias} = 1/5 * \text{Sum (1 to 5 previous years) over (observed value/expected value - 1)}$$

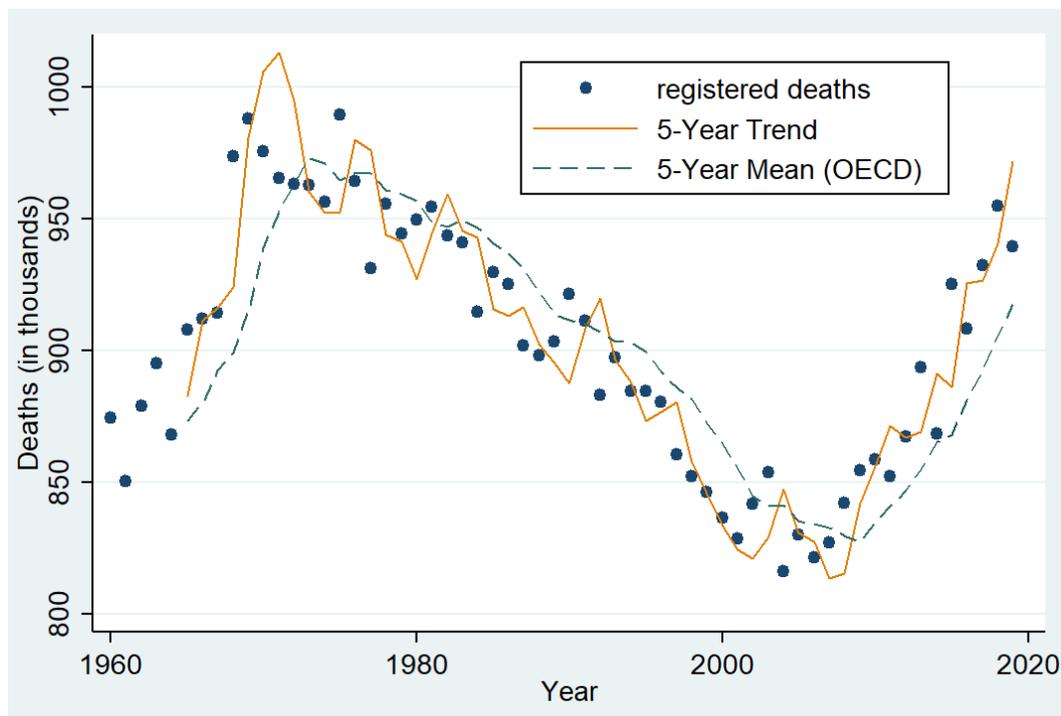
207 Ideally, this second parameter should be as close to zero as possible. A positive bias means that the
208 expected value tends to underestimate the observed value and consequently, on average, shows
209 inflated excess mortality rates. The reverse is true for a negative bias; in this case, excess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

211 Figure 1b compares two methods for forming the expected value based on the death figures in
212 Germany. It can be seen that the five-year mean, as used by the OECD, always lags behind the
213 actual development and estimates the expected values too low when the numbers are rising and,
214 conversely, too high when the mortality is declining. In contrast, the 5-year trend is much closer to
215 the actual development. This is also reflected in the two performance indicators. MAPE (the
216 absolute estimation error) is 1.78 percent on average for the 5-year trend, but 2.29 percent for the
217 5-year mean; the bias is 0.03 percent and 0.27 percent respectively. The differences become even

218 clearer if either ascending or descending phases are considered. For example, taking only the years
219 2010 to 2019, MAPE is 1.96 percent for the trend method and 3.34 percent for the mean method;
220 the bias is -0.03 percent for the former and 3.34 percent for the latter. Thus, in phases with a
221 constantly increasing or decreasing tendency, the bias in the OECD mean method is very
222 pronounced.

223

224 **Figure 1b: Annual number of deaths in Germany from 1960 to 2019 with two different expected**
225 **value forecasts**



227 3) Results

228 First, we tried to find out which trend has the most favourable performance indicators. Table 1
229 shows that, calculated over all countries and years (1960 to 2019), the estimation error (MAPE) is

230 lowest for the 7-year trend. In contrast, the 3-year trend has the lowest bias. In this respect, the 5-
 231 year trend represents a good compromise between the two performance indicators – accordingly,
 232 our further calculations are based on this choice. If we compare the 5-year trend with the standard
 233 method of statistical offices, the 5-year or 4-year mean, we see a particularly marked difference in
 234 terms of bias; this is 9 times higher for the 5-year mean, i.e. the OECD method, and 7 times higher
 235 for the 4-year mean, the Eurostat method, than for our 5-year trend. The difference in the
 236 estimation errors (MAPE) is not as drastic and amounts to 22 and 11 percent.

237 **Table 1: Comparison of estimation errors (MAPE) and bias of different methods for 31 nations.**
 238 **The most favourable values are highlighted in bold. The 5-year trend thus represents the best**
 239 **compromise (marked yellow).**

Method	MAPE (1960 to 2019 mean)	Bias (2015 to 2019 mean)
2Y-Trend	4,54 %	0,23 %
3Y-Trend	3,33 %	0,20 %
4Y-Trend	2,90 %	0,29 %
5Y-Trend	2,77 %	0,32 %
6Y-Trend	2,71 %	0,56 %
7Y-Trend	2,69 %	0,78 %
8Y-Trend	2,70 %	1,01 %
9Y-Trend	2,72 %	1,24 %
4Y-Mean	3,07 %	2,43 %
5Y-Mean	3,38 %	2,86 %

240

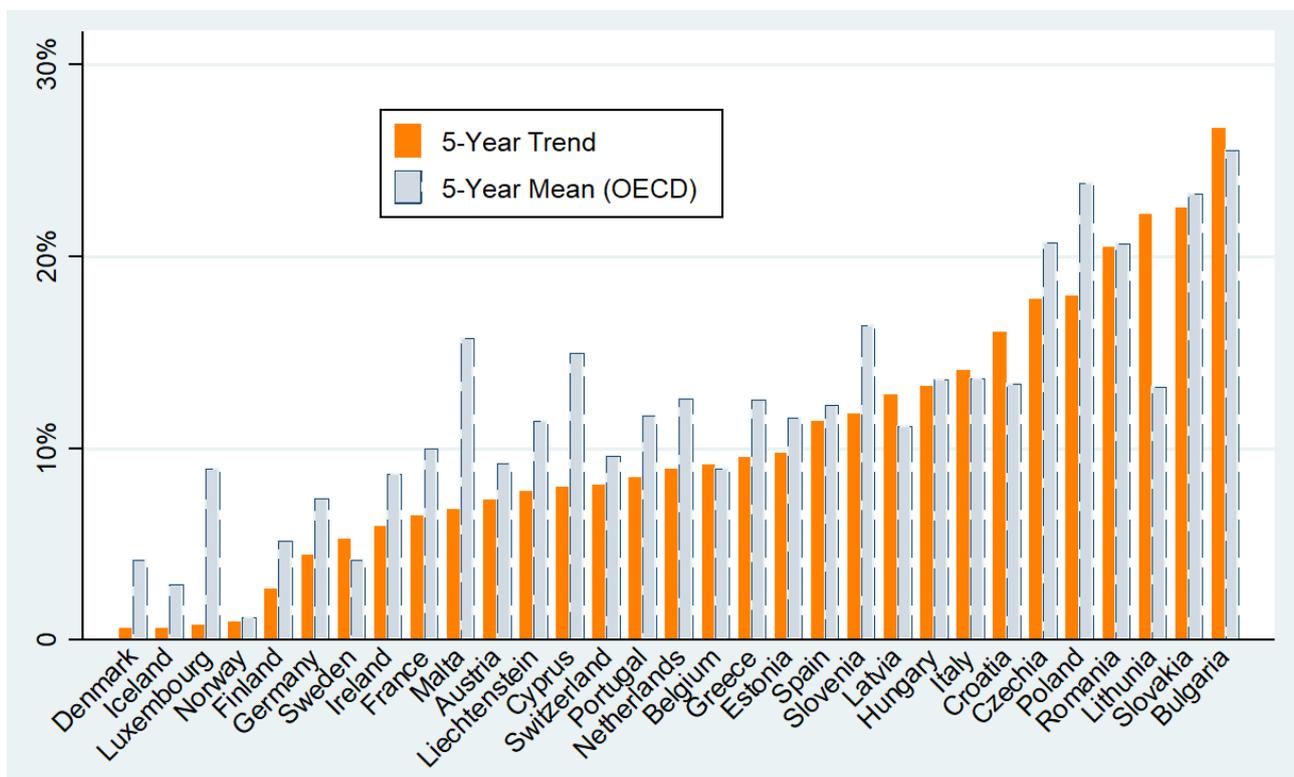
241 In the following, we examine how the use of the 5-year trend, i.e. the method proposed here,
242 affects the estimate of excess mortality in the Covid 19 period, i.e. for the years 2020 and 2021
243 together. For simplicity, only the OECD method (the 5-year mean) is now used for comparison, as
244 the Eurostat method (the 4-year mean) would yield very similar results. For the second pandemic
245 year (2021), we have used the trend or mean from 2015 to 2019 to estimate the expected value so
246 that the calculations are not biased by the increased number of deaths in the first pandemic year
247 (2020). Figure 2 shows the average annual excess mortality over the two-year period 2020 to 2021,
248 ordered from left to right using the excess mortality values by country derived from the 5-year
249 trend. The differences between the two methods are very pronounced for some countries. They
250 result from the fact that the mean method, as already mentioned, does not take into account
251 underlying gradients: If the death figures show an increasing tendency, as is the case in Northern
252 and Western Europe, the mean method overestimates excess mortality; if they show a decreasing
253 tendency, as is partly the case in Eastern Europe, this leads to an underestimation (cf. also the
254 changing gradients in Figure 1 above) [6].

255 It can also be seen in Figure 2 that excess mortality during the pandemic in Europe shows a high
256 variance – ranging from 0.8% in Denmark to 27% in Bulgaria. These are the currently much-noticed
257 differences between countries, which are often spontaneously attributed to the measures against
258 the pandemic and their more or less strict compliance. In Germany for example, the rather mild
259 excess mortality is usually assigned to the heavy restrictions during the pandemic while Sweden
260 was criticised for its liberal approach. But in hindsight, we see Germany and Sweden more or less
261 on par, in the 6th and 7th position. In general, it becomes obvious that the lowest excess mortality is
262 predominantly found in Northern Europe, followed by Western Europe, Southern Europe and
263 Eastern Europe, thus clearly following a negative socio-economic gradient. The correlation

264 coefficient between excess mortality and wealth is indeed quite strong (in the negative direction)
265 and highly significant [7].

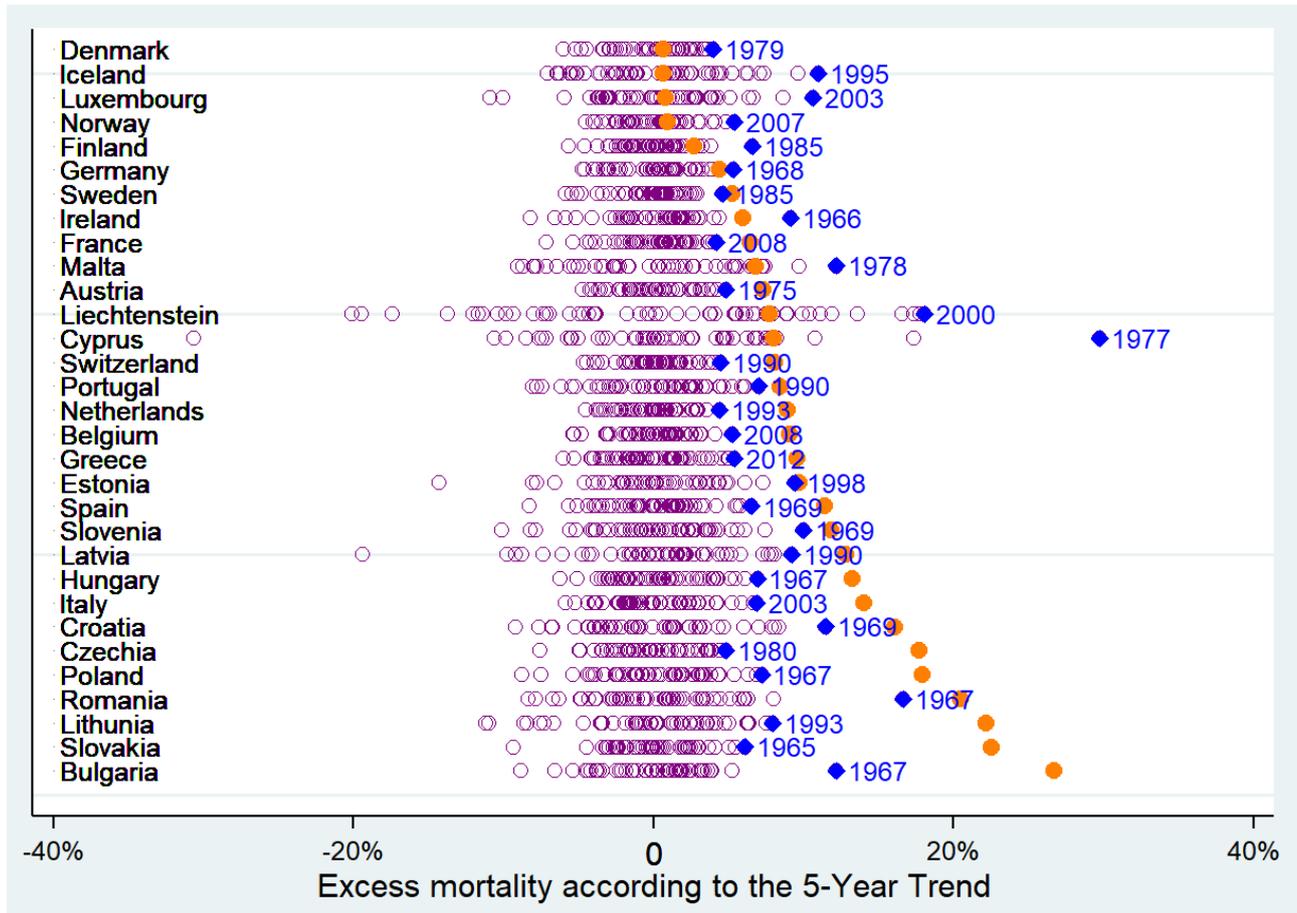
266

267 **Figure 2: Average annual excess mortality in the two-year period 2020 to 2021, i.e. during the**
268 **Covid-19 pandemic. We compare our proposed method (5-year trend) with the**
269 **estimates of the OECD statistical office (5-year mean).**



271 However, we do not know how high the variations in excess mortality were in the past and how
272 much they differed between the countries. In other words: Before we try to interpret death tolls
273 and containment measures during the pandemic, we should analyse the normal mortality
274 fluctuations from pre-pandemic years. Figure 3 attempts to answer this question by comparing the

275 **Figure 3: Average annual excess mortality during the Covid-19 pandemic (orange dots)**
 276 **compared with excess mortality values from 1965 to 2019 (purple rings). Blue diamonds with**
 277 **year labels indicate the respective year with the highest excess mortality in the pre-Covid-19**
 278 **period.**



279 average annual excess mortality during the Covid-19 pandemic with the excess and the lower than
 280 expected mortality from 1965 to 2019. Here it becomes visible that the mortality figures in
 281 countries with smaller populations – due to the law of small numbers, i.e. for statistical reasons –
 282 fluctuate strongly anyways. This applies in particular to Liechtenstein (38,000 inhabitants), Iceland
 283 (335,000), Malta (430,000) and Luxembourg (583,000). However, even in countries with a
 284 comparatively large population, such as Germany, France, Italy and Spain, high fluctuations of
 285 plus/minus five percent are relatively normal. In this respect, it can be stated that in 10 out of 31

286 countries the average annual excess mortalities in the period from 2020 to 2021 do not exceed the
287 range of excess mortalities in the period from 1965 to 2019. Furthermore, it is noteworthy that the
288 highest excess mortality during this period does not occur in or around any particular year – i.e. it
289 cannot be explained by other pandemics, such as the Hong Kong flu, which spread worldwide from
290 1968 to 1970 (Viboud et al. 2005).

291

292 **4) Discussion and conclusions**

293 Three main results should be noted and discussed: First, the importance of the term "excess
294 mortality"; second, the need for an evaluation method; third, the accuracy of our calculation
295 method compared to other methods commonly used in the literature.

296 Firstly, it has been shown that relatively high variations in death figures, and thus in excess
297 mortality, already occurred in the past. Our estimation procedure also reveals years with
298 remarkable lower than expected mortality. In the light of these fluctuations, the question arises as
299 to which horizons of meaning should be invoked with the term "excess mortality". Three different
300 notions are conceivable and should not be confused with each other: a) The one established by
301 William Farr (Honigsbaum 2020), used in this text and still most common among statistical offices
302 today, is that of a deviation from a longer-term average or trend, which is understood as
303 "demographic normality". A threshold value is not envisaged here. b) In contrast, the
304 epidemiological and demographic literature often takes the view that all "preventable" deaths
305 should be understood as excess mortality, i.e. also "normal" flu deaths, heat deaths, traffic victims,
306 etc. (e.g. Nielsen et al. 2019; Shkolnikov et al. 2022). As a result, the baseline from which excess
307 mortality is calculated is generally drawn lower and at the same time the normative reference
308 point is shifted. The benchmark is no longer the socially accepted normality before the exceptional

309 event under consideration, but the author's ideas about what he or she considers avoidable and
310 what not. While these ideas may be plausible, they are moving targets in scientific debates and not
311 fixed points of reference across time, space, and culture, that could be used to determine, for
312 example, the magnitude of a pandemic on a historical scale (Gill et al 2022). c) Finally, there is the
313 possible and, in the public mind, perhaps also (mis-)understood meaning of an alert threshold,
314 which is supposed to indicate an extraordinary event. Here, too, the point of reference is
315 demographic normality, except that it is not the zero point of normal fluctuations between excess
316 mortality and lower than expected mortality that would be taken as the baseline, but an upper
317 threshold value – for example, a threshold value of five percent for the annual values of more
318 populous countries [9]. Assuming this threshold, for Covid-19 no alarm would be triggered in 7 out
319 of the 31 countries considered (for excess mortality in the two-year period).

320 Secondly, it has been demonstrated that different calculation methods lead to quite different
321 results. It is therefore necessary to calibrate the calculation method used against a past that is
322 considered normal before applying it to an exceptional event. The prevailing neglect of this
323 necessity is all the more astonishing as the methodological reviews cited above show considerable
324 differences in results due to the choice of method. However, considering that excess mortality
325 rates are relevant to public controversies and thus politically sensitive, one should a priori choose a
326 transparent and easily comprehensible method in order to avoid the suspicion of "cherry-picking"
327 – and corresponding alarmist or anti-alarmist motives. Most of the recently published reviews that
328 refer to the problem of method selection remain rather vague in their recommendations for
329 remedial action. Nepomuceno and colleagues (2022), for example, continue to leave the decision
330 on method to the discretion of the researcher: "Moreover, the method used and the reference
331 period (which may or may not include leap week years) should both be chosen carefully." (p. 21)

332 Only Jonas Schöley (2021) uses past data – similar to ours – to assess the accuracy of different
333 methods, but unfortunately in a way that the results are not directly comparable to ours [9].

334 Thirdly, we have proposed and tested a calculation method with comparatively good results. We
335 were able to demonstrate that it has a significantly lower bias and higher accuracy compared to
336 the methods used by Eurostat and the OECD. Importantly, this is achieved without higher demands
337 on data availability and mathematical understanding, unlike other methods described in the
338 literature (e.g. WHO 2022). We therefore believe that it represents a good compromise between
339 ease of verification and statistical accuracy. In particular, the bias is very low and seems to be
340 competitive with computationally much more sophisticated methods (the data in Schöley 2021,
341 Figure 4, seem to indicate this, even if they are not directly comparable).

342 Concerning the deviation of expected from observed values, the accuracy can perhaps still be
343 improved; the estimation error (MAPE) could possibly be diminished: Here it would be particularly
344 interesting to find out whether there is a certain systematics in the fluctuations over the years, as
345 could be revealed, for example, with an ARIMA-procedure commonly used in econometrics for
346 overlapping waves and trends (Shumway/Stoffer 2017). In terms of causality, it would also be
347 worth considering how the remarkably lower than expected mortalities shown in Figures 1 and 3
348 come about [10]. In the epidemiological literature, mortality displacement effects are occasionally
349 reported, i.e. that in the case of heat or flu waves, severely pre-weakened persons die and
350 therefore lower than expected mortalities can occur afterwards (Armstrong et al. 2017). To the
351 best of our knowledge, these effects have never been studied in larger demographic contexts over
352 longer periods of time.

353 However, all methods for assessing excess mortality should be evaluated with regard to their
354 estimation error and bias using standardised procedures, in order to be able to exploit their

355 advantages over other measurement methods (such as direct disease diagnosis). This is
356 particularly important since pandemics – throughout history (Cohn 2012) – have generated
357 considerable controversy and therefore any appearance of non-transparency and political bias on
358 the part of science should be avoided as far as possible. Particularly the methods of the statistical
359 offices – as OECD and Eurostat – are in urgent need of improvement due to their high bias and
360 broad impact. Excess mortality as a summary indicator of a pandemic crisis is too relevant for
361 health policies and public health interventions to risk potentially misleading calculations and
362 interpretations.

363

364 **5) Funding**

365 The study was partly supported by the Volkswagen Foundation, grant Az. 99 026.

366

367 **6) Competing interest**

368 None of the authors is aware of any competing interest.

369

370 **7) Data availability**

371 The spreadsheet file with the respective data and computations is made public on

372 <https://osf.io/5y4sd/files/osfstorage/63a47dc17b5c800255227a92>

373

374 **8) Literature**

- 375 Aburto, J. M., Schöley, J., Kashnitsky, I., Zhang, L., Rahal, C., Missov, T. I., ... & Kashyap, R. (2022). Quantifying
376 impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29
377 countries. *International journal of epidemiology*, 51(1), 63-74.
- 378 Armstrong, B., Bell, M. L., de Sousa Zanotti Stagliorio Coelho, M., Leon Guo, Y. L., Guo, Y., Goodman, P., ... &
379 Gasparrini, A. (2017). Longer-term impact of high and low temperature on mortality: an international
380 study to clarify length of mortality displacement. *Environmental health perspectives*, 125(10), 107009.
- 381 Beaney, T., Clarke, J. M., Jain, V., Golestaneh, A. K., Lyons, G., Salman, D., & Majeed, A. (2020). Excess
382 mortality: the gold standard in measuring the impact of COVID-19 worldwide? *Journal of the Royal*
383 *Society of Medicine*, 113(9), 329-334.
- 384 Canoui-Poitaine, F., Cadot, E., & Spira, A. (2006). Excess deaths during the August 2003 heat wave in Paris,
385 France. *Revue d'épidémiologie et de santé publique*, 54(2), 127-135.
- 386 Cohn, S. K. (2012). Pandemics: waves of disease, waves of hate from the Plague of Athens to AIDS. *Historical*
387 *Research*, 85(230), 535-555.
- 388 De Nicola, G., Kauermann, G., & Höhle, M. (2022a). On assessing excess mortality in Germany during the
389 COVID-19 pandemic. *AStA Wirtschafts-und Sozialstatistisches Archiv*, 16(1), 5-20.
- 390 De Nicola, G., & Kauermann, G. (2022b). An update on excess mortality in the second year of the COVID-19
391 pandemic in Germany. *AStA Wirtschafts-und Sozialstatistisches Archiv*, 16(1), 21-24.
- 392 Ferenci, T. (2022). Comparing methods to predict baseline mortality for excess mortality calculations-
393 unravelling'the German puzzle'and its implications for spline-regression. *medRxiv*.
- 394 Gill, B., Kehler, T., & Schneider, M. (2022). Is Covid-19 a dread risk? The death toll of the pandemic year 2020
395 in long-term and transnational perspective. *Futures*, 142, 103017.

- 396 Honigsbaum, M. (2020). Revisiting the 1957 and 1968 influenza pandemics. *The Lancet*, 395(10240), 1824-
397 1826.
- 398 Karlinsky, A., & Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic
399 with the World Mortality Dataset. *Elife*, 10.
- 400 Kowall, B., Standl, F., Oesterling, F., Brune, B., Brinkmann, M., Dudda, M., ... & Stang, A. (2021). Excess
401 mortality due to Covid-19? A comparison of total mortality in 2020 with total mortality in 2016 to 2019
402 in Germany, Sweden and Spain. *Plos one*, 16(8), e0255540.
- 403 Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development.
404 *Proceedings of the National Academy of Sciences*, 111(51), 18112-18115.
- 405 Levitt, M., Zonta, F., & Ioannidis, J. (2022a). Comparison of pandemic excess mortality in 2020-2021 across
406 different empirical calculations. *medRxiv*.
- 407 Levitt, M., Zonta, F., & Ioannidis, J. (2022b). Excess death estimates from multiverse analysis in 2009-2021.
408 *medRxiv*.
- 409 Mazzuco, S. (2022). Measuring Mortality Crises: A Tool for Studying Global Health. In *Oxford Research*
410 *Encyclopedia of Global Public Health*.
- 411 Nepomuceno, M. R., Klimkin, I., Jdanov, D. A., Alustiza-Galarza, A., & Shkolnikov, V. M. (2022). Sensitivity
412 analysis of excess mortality due to the COVID-19 pandemic. *Population and Development Review*.
- 413 Nielsen, J., Vestergaard, L. S., Richter, L., Schmid, D., Bustos, N., Asikainen, T., ... & Krause, T. G. (2019).
414 European all-cause excess and influenza-attributable mortality in the 2017/18 season: should the burden
415 of influenza B be reconsidered?. *Clinical microbiology and infection*, 25(10), 1266-1276.
- 416 Nunes, B., Viboud, C., Machado, A., Ringholz, C., Rebelo-de-Andrade, H., Nogueira, P., & Miller, M. (2011).
417 Excess mortality associated with influenza epidemics in Portugal, 1980 to 2004. *Plos one*, 6(6), e20661.

- 418 Pifarré i Arolas, H., Acosta, E., López-Casasnovas, G., Lo, A., Nicodemo, C., Riffe, T., & Myrskylä, M. (2021).
419 Years of life lost to COVID-19 in 81 countries. *Scientific reports*, *11*(1), 1-6.
- 420 Sanmarchi, F., Golinelli, D., Lenzi, J., Esposito, F., Capodici, A., Reno, C., & Gibertoni, D. (2021). Exploring the
421 gap between excess mortality and COVID-19 deaths in 67 countries. *JAMA Network Open*, *4*(7),
422 e2117359-e2117359.
- 423 Schmidt, M., Glasson, J., Emmelin, L., & Helbron, H. (Eds.). (2008). *Standards and thresholds for impact*
424 *assessment*. Springer Publ.
- 425 Schöley, J. (2021). Robustness and bias of European excess death estimates in 2020 under varying model
426 specifications. *MedRxiv*.
- 427 Shkolnikov, V. M., Klimkin, I., McKee, M., Jdanov, D. A., Alustiza-Galarza, A., Németh, L., ... & Leon, D. A.
428 (2022). What should be the baseline when calculating excess mortality? New approaches suggest that
429 we have underestimated the impact of the COVID-19 pandemic and previous winter peaks. *SSM-*
430 *population health*, *18*, 101118.
- 431 Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. In *Time series analysis and its applications* (pp. 75-
432 163). Springer, Cham.
- 433 Viboud, C., Grais, R. F., Lafont, B. A., Miller, M. A., & Simonsen, L. (2005). Multinational impact of the 1968
434 Hong Kong influenza pandemic: evidence for a smoldering pandemic. *The Journal of infectious diseases*,
435 *192*(2), 233-248.
- 436 Vu Manh, C. (2022). Weeks of life lost to COVID-19, the case of the United States. *Immunity, inflammation*
437 *and disease*, *10*(7), e661.
- 438 Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., ... & Murray, C. J. (2022). Estimating
439 excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality,
440 2020–21. *The Lancet*, *399*(10334), 1513-1536.

441 WHO 2022: World Health Organization. Methods for Estimating The Excess Mortality Associated with the
442 COVID-19 Pandemic. [https://www.who.int/publications/i/item/estimating-global-and-country-specific-](https://www.who.int/publications/i/item/estimating-global-and-country-specific-excess-mortality-during-the-covid-19-pandemic)
443 [excess-mortality-during-the-covid-19-pandemic](https://www.who.int/publications/i/item/estimating-global-and-country-specific-excess-mortality-during-the-covid-19-pandemic) [retrieved at 17th of Nov 2022]

444

445 9) References

- (1) Sometimes the number of excess deaths is related to 100,000 inhabitants (e.g. Wang et al. 2022). For comparisons between countries, it should then be noted that there are large differences in normal mortality; i.e. an excess mortality of 100 per 100,000 in a country with a low normal mortality of e.g. 1,000 per 100,000 would account for an excess mortality of 10 percent, while in a country with a higher normal mortality of 2,000 per 100,000 it would only account for an excess mortality of 5 percent.
- (2) In our sample, annual mortality figures across all 31 countries increase from 3.67 million in 1960 to 4.78 million in 2019, or by about 30 percent.
- (3) The lowest prominently published estimate for Germany is 6,300 excess deaths for 2020 and 23,400 for 2021 (de Nicola et al. 2022a and 2022b).
- (4) 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.
- (5) Here we roughly follow the concept of Schöley 2021. Yet, in our case the expected value is in the denominator and not in the numerator, because in this way excess mortality and lower than expected mortality are displayed symmetrically to the expected value. Cf. reference [9] for more detail.
- (6) Therefore, only results from the 5-year trend method will be reported in the following.

- (7) Wealth is measured as GDP per capita in 2020 in logarithmic form. The correlation coefficient is -0.82, with $p < 0.001$.
- (8) The problem of the dependence of the range of variation on population numbers can be adjusted by studentisation; however, the resulting z-scores are not as easily comprehensible in public communication as percentages.
- (9) He 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. Example calculation: Expected value = 100; observed value(1) = 40; observed value(2) = 160 (i.e. minus versus plus 60 deaths). $Schöley(1) = (40-100)/40 = -1.5 = -150\%$. $Schöley(2) = (160-100)/160 = 0,375 = +37\%$. Whereas $Gill\ et\ al.(1) = 40/100 - 1 = -0.6 = -60\%$; $Gill\ et\ al.(2) = 160/100 - 1 = 0.6 = +60\%$. However, the distortion of this order of magnitude only occur for very high deviations, which are relatively rare.
- (10) In Figure 3, it was observed that the amplitude in the lower than expected mortality range is about the same as in the excess mortality range. The amplitude of excess mortalities is generally explained by flu waves in winter and heat waves in summer. But how to explain the approximately equal amplitude in the lower than expected mortality range? Of course, in an estimation method with a bias near zero, excess mortalities and lower than expected mortalities must balance each other out, but one would expect a smaller amplitude in the lower than expected mortality range – a year with high excess mortality would have to be balanced by several years with slightly lower than expected mortality (cf. Vu Manh 2022, figure 1).