

### **Summary:**

Excess mortality data avoid miscounting deaths from the under-reporting of Covid-19-related deaths and other health conditions left untreated. Excess mortality is defined as actual deaths from all causes, minus 'normal' deaths. This article assesses the comparability of data on excess mortality between countries and regions, it reviews the available data sources, and compares and contrasts different statistical measures of excess mortality. Now that the first wave of the pandemic is over for most European countries, the time has come for robust European comparisons. Our preferred measure of excess deaths relative to normal deaths, the more transparent and comparable P-score, is calculated for European countries with high rates of excess mortality during the Covid-19 pandemic. For the 'all age' group and especially for the working age group, '15-64', England has had the highest rates of excess mortality. Spain had the highest cumulative P-score over the pandemic weeks for the 'over 85' age group. We address issues of international data and comparability on the extent of deaths among the care home residents. Generally, research is needed into divergent patterns of excess mortality between and within countries. Excess mortality measures can be compared in cross-tabulation and correlations controlling for common features like density by region, to propose policy hypotheses. Modellers should have ready access to transparent, comparable international data to a granular level to test such hypotheses to aid policy making for potential further waves of the pandemic. We suggest how international statistical agencies and national statistical agencies could publish improved measures of excess mortality.

### **Acknowledgements**

We take responsibility for interpretations of data and analysis but are grateful for advice on data and other matters to Jose Manuel Aburto (Dept of Sociology, Oxford University), Ainhoa Alustiza Galarza (HMD), Nick Andrews (Public Health England), Gabriele Ciminelli (Asia school of Business), Adelina Comas-Herrera (Care Policy and Evaluation Centre, Department of Health Policy, London School of Economics and Political Science), Laurie Davies (Mathematics Department, University Duisburg-Essen), Francesca De' Donato (Department of Epidemiology, Lazio Regional Health Service, Rome, Italy), Mark O'Doherty (Public Health Agency, Northern Ireland), Faisal Islam (Economics Editor, BBC), Dmitri Jdanov (HMD), Gareth John (NHS Wales Informatics Service), Ridhi Kashyap (Nuffield College), Amparo Larrauri (Departamento de Enfermedades Transmisibles, Centro Nacional de Epidemiología, CIBER Epidemiología y Salud Pública, Spain), Diogo Marques (Public Health Scotland), Bent Nielsen (Nuffield College), Justine Pooley (ONS), Max Roser (Our World in Data, Oxford University), Charles Tallack (Health Foundation), and Lasse Skafte Vestergaard (Faculty of Health and Medical Sciences and Statens Serum Institut, University of Copenhagen.)

## 1. Why is it important to examine excess mortality data?

Excess mortality is a count of deaths from all causes relative to what would normally have been expected. In a pandemic, deaths rise sharply, but causes are often inaccurately recorded, particularly when reliable tests are not widely available. The death count<sup>1</sup> attributed to Covid-19 may thus be significantly undercounted. Excess mortality data overcome two problems in reporting Covid-19-related deaths. Miscounting from misdiagnosis or under-reporting of Covid-19-related deaths is avoided. Excess mortality data include 'collateral damage' from other health conditions, left untreated if the health system is overwhelmed by Covid-19 cases, or by deliberate actions that prioritise patients with Covid-19 over those with other symptoms. In a pandemic, measures taken by governments and by individuals also influence death rates. For example, deaths from traffic accidents may decline but suicide rates may rise. Excess mortality captures the net outcome of all these factors. Figure 1 illustrates how the degree of Covid-19 recording relative to excess deaths has varied across some European countries. In Belgium, with a broad definition of what constitutes a Covid-19 death, the excess over 100 percent might suggest that most excess deaths are due to Covid-19 and other deaths, such as those due to road accidents, may have declined.

Excess mortality data can be used to draw lessons from cross- and within-country differences and help analyse the social and economic consequences of the pandemic and relaxing lockdown restrictions. For country comparisons (where under-recording may differ), policy-makers should examine *robust measures* expressed relative to the benchmarks of 'normal' deaths. 'Normal' death rates reflect persistent factors such as the age composition of the population, the incidence of smoking and air pollution, the prevalence of obesity, poverty and inequality, and the normal quality of health service delivery. Estimating the virus reproduction rate,  $R$ , is crucial for assessing the rate and nature of relaxation of lockdowns.<sup>2</sup> Excess death figures could help to avoid the measurement biases inherent in other data typically used to estimate  $R$  in epidemiological models.<sup>3</sup>

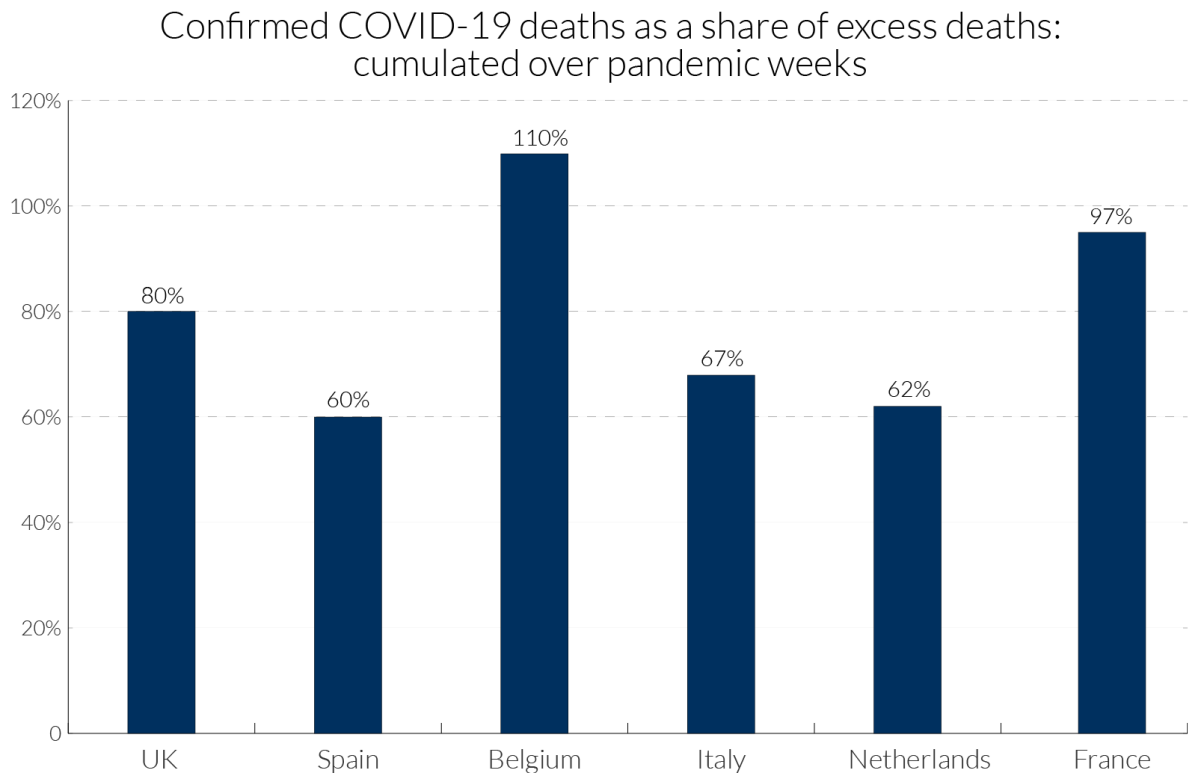
---

<sup>1</sup> See webpage: [COVID-19 Dashboard by the Center for Systems Science and Engineering \(CSSE\)](#), Johns Hopkins University.

<sup>2</sup>  $R$  is the virus reproduction rate, which needs to be kept below 1 to avoid exponential growth of infections.

<sup>3</sup> Details on this can be found in the evidence of Prof. John Edmunds to the UK Science and Technology Parliamentary Select Committee on 7<sup>th</sup> May. He explained that while excess mortality data lag Covid-19 infections, the data are an important check on earlier estimates of the rate of spread of the virus.

**Figure 1: Attributed Covid-19 deaths as a share of excess deaths for poor performers ('all ages'): cumulated over the first wave of pandemic weeks**



Sources: UK - ONS:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales> .

Covid-19 deaths in other European countries - European Centre for Disease Prevention and Control (ECDC):

<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide> .

Notes: 'Lab-confirmed' cases, in all countries except Belgium and France, where 'lab-confirmed plus probable' cases are used. UK taken from Week 13 (week ending 27th March) to Week 23 (week ending 5th June). For non-UK countries, we have assumed a one-week lag in Covid-19 death registrations. This means reported and excess deaths for France, Spain, Italy and the Netherlands are taken from Weeks 10-20; and Covid-19 deaths from ECDC are taken from Weeks 11-21. For Belgium, we have taken Week 11-21 for reported and excess deaths; Covid-19 deaths are taken from Week 12-22.

## 2. How is excess mortality measured and who measures it?

National statistical agencies publish actual weekly deaths and averages of past ‘normal’ deaths. For example, the Office for National Statistics (ONS) reports ‘normal’ deaths for England and for Wales as the average of the previous five years’ deaths. However, there are no published benchmarks for more granular or disaggregated data, such as sub-regions or cities. Using the weekly historical data, researchers could calculate such benchmarks with some effort.

The ratio or percentage of excess deaths relative to ‘normal’ deaths, the P-score, is an easily understood measure of excess mortality, see Box 1. We argue that national statistical offices should publish P-scores for states and sub-regions. In the U.S., the National Centre for Health Statistics publishes data on excess deaths and a variant on P-scores (see Box 1), defining excess deaths as deviations from ‘normal’ deaths plus a margin adjusting for the uncertainty of the data.<sup>4</sup> These data include counties and states, and are disaggregated by gender, age and ethnicity. The NCHS thus sets an international standard for statistical agencies.

However, to obtain cross-European comparisons requires data collation from individual national agencies to construct P-scores or variant P-scores, which are largely comparable, see section 4.1. Another alternative are the Z-scores compiled by EuroMOMO<sup>5</sup> for 24 states, see Box 1. EuroMOMO’s measures of weekly excess mortality in Europe show the mortality patterns between different time-periods, across countries, and by age-groups. The Z-scores standardise data on excess deaths by scaling by the standard deviation of deaths. EuroMOMO are currently not permitted to publish actual excess death figures by country and do not publish the standard deviations used in their calculations. However, they graph the Z-scores and the estimated confidence intervals back to 2015 providing a visual guide to their variability. In contrast to the P-scores, the Z-scores are a measure that is less easily interpretable. Moreover, if the natural variability of the weekly data is lower in one country compared to another, then the Z-scores could lead to exaggeration of excess mortality compared to the P-scores. Strictly, the Z-scores are not comparable across countries, though see the caveats in section 4.1.

At least five separate journalistic endeavours have recently engaged in the time-consuming effort of collating and presenting more transparent excess mortality data, see Table 1. The Financial Times plots numbers of excess deaths, and the P-score or percentage of deaths that are above normal deaths. The Economist shows figures and graphics for excess deaths but not P-scores. However, the published estimates of P-scores in newspapers give only a recent snapshot, missing the context of historical variability provided by EuroMOMO. And we only have P-scores for some countries, regions and cities. A third measure of excess mortality is per capita excess mortality, where excess deaths (actual deaths minus ‘normal’ deaths) are divided by population, see Box 1, is used by the BBC (Table 1).

---

<sup>4</sup> See webpage: “[National Center for Health Statistics](#)”, Centers for Disease Control and Prevention (CDC), US Government.

<sup>5</sup> [EuroMOMO](#) is a European mortality monitoring entity, aiming to detect and measure excess deaths related to seasonal influenza, pandemics and other public health threats. Official national mortality statistics are provided weekly from the 24 European countries and regions in the EuroMOMO collaborative network, supported by the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO).

**Table 1: Sources of comparative excess mortality data for Europe, the UK and the US, and other countries**

Source and metafiles	Measure reported	Period and type of data	Benchmark	Disaggregation?	Locations compared	First publication date
<i>Multi-national institutions</i>						
<a href="#">The Human Mortality Database</a> (HMD) <i>Comprehensive, transparent metafile for data sources and coverage.</i>	Death counts and death rates by country.  <i>[The raw data allow P-scores to be calculated.]</i>	Weekly, 2000-2020 for many. At least from 2015 for all, except Germany (2016).  Occurrence data for the death count, except the UK, which is registration data.	The average benchmarks for earlier years <i>can be calculated</i> from the earlier data e.g. 2015-2019.	By age groups: 0-14, 15-64, 65-74, 75-84, 85+.  By gender (F, M total).	22 countries: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Scotland, Slovakia, Spain, Sweden, UK: England and Wales, UK: Scotland, and the USA.	<i>Regularly updated.</i> Open access on website.
<a href="#">Eurostat</a>  <i>Transparent metafile for data sources and coverage.</i>	Number of weekly deaths.  <i>[The raw data allow P-scores to be calculated.]</i>	Weekly, 2000-2020  Eurostat recommends data of occurrence data for death counts, but accepts date of registration. May vary by country.	Historical average of deaths for that week over 2016-2019.	Three levels of regional breakdowns (NUTS levels): major socio-economic regions (e.g. countries); major sub-national regions; and small subnational regions (e.g. cities).  By age group: 5-year groups, 20 in all.  By gender (F, M total).	22 countries: Austria, Belgium, Bulgaria, Czech Republic, Denmark, England and Wales, Estonia, Finland, France, Germany, Hungary, Iceland, Italy, Luxembourg, Netherlands, Norway, Portugal, Scotland, Slovakia, Spain, Sweden, and the USA.  Sub-national regional data available at both NUTS Level 2 (major regions) and NUTS Level 3 (smaller, higher-resolution regions) for most countries.	<i>Regularly updated.</i> Downloadable on website.
<a href="#">European Mortality Monitoring Project</a> (EuroMOMO)  <i>There is no metafile for data sources and coverage. The underlying data</i>	Z-scores by country for 2015-2020; total ( <i>summing all countries</i> ) weekly and cumulated excess deaths and pooled number of deaths for 2016-2020. Excess deaths are <i>not reported</i> for individual countries. Expected	Weekly data.  Week ends on Sunday.  Occurrence data for the death count, including the UK.	Deviation in mortality from an <i>expected</i> level. See Box 1 for a description of the method and how the expected level is modelled.	All ages and by age groups, recently expanded: 0-14, 15-44, 45-64, 65-74, 75-84, 65+, and 85+	UK and its constituent nations and regions, 24 <i>European countries</i> : Austria, Belgium, Denmark, Estonia, Finland, France, Germany (Berlin), Germany (Hesse), Greece, Hungary, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden and Switzerland.  <i>[Note: the fraction of the population covered by the country level data is not given, e.g. "Italy" in fact only covers 14% of the population, see text.]</i>	Began in 2008. Since 2016 supported by European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO) Regional Office for Europe.

Source and metafiles	Measure reported	Period and type of data	Benchmark	Disaggregation?	Locations compared	First publication date
<i>are not fully transparent.</i>	levels of deaths are not published.					<i>Regularly updated.</i> Data are not downloadable except from charts.
<a href="#">The Health Foundation</a> , UK  <i>Clear description on graphs of data sources and coverage.</i>	Weekly and/or cumulative P-scores; cumulative excess deaths by designated time period for a subset of the RHS locations.	Weekly, 28-Feb-20 to end-May-20.  Occurrence data for the death count except for the UK, which uses registration data.	Baseline differs by country, see their interactive graphs. For the UK, it is the historical average of deaths for that week over 2015-2019. But for Madrid, for example, the average is over 2018-19.	Regional disaggregation in the UK to local authority level, presented graphically.	UK and its constituent nations and regions and local authorities.  <i>European countries:</i> France Italy, Spain and their constituent regions. Sweden, Germany. <i>Cities:</i> London, Madrid, NY City, Paris.	4-Jun-2020.  <i>In two articles. Not updated.</i> Data are not downloadable except from charts.
<i>Journalistic endeavours – some examples</i>						
<a href="#">The Economist</a>  <i>Clear description of data sources and coverage and method on GitHub.</i>	Numbers of deaths, Covid-19-deaths and of excess deaths (actual deaths minus the expected deaths).  <i>[The raw data allow P-scores to be calculated.]</i>	Weekly; approximately monthly in one table.  Occurrence data for most countries. UK based on registration data.	“Expected deaths”, averages ranging from 2 to 5 years, see GitHub.	Some regional disaggregation, see next column.	United Kingdom and its constituent nations and regions and London.  <i>Other countries:</i> Austria, Belgium, Brazil (5 cities: São Paulo, Rio de Janeiro, Fortaleza, Manaus and Recife), Chile (and regions), Denmark, Ecuador, France (and departments), Germany, Indonesia (burials in Jakarta), Italy (and regions), Mexico (Mexico City), Netherlands, Norway, Peru, Portugal, Russia (Moscow), South Africa, Spain (and regions), Sweden, Switzerland, Turkey (burials in Istanbul), United States (and regions).	Started 16 April 2020.  <i>Regularly updated.</i> Open access <a href="#">on GitHub</a> .
<a href="#">The Financial Times</a>  <i>Clear description of data sources and coverage and method on GitHub.</i>	Number of deaths and of excess deaths (actual deaths minus the expected deaths).  <i>[The raw data allow P-scores to be calculated.]</i>	Weekly and cumulative, from beginning of outbreak.  Occurrence data for most countries. UK based on registration data.	Historical average of deaths for that week over 2015-2019.	Regional disaggregation in the UK to its constituent nations and sub-regions in England. Local-level data available for some other countries.	UK and its constituent nations and regions.  <i>European countries:</i> Italy (and regions); Austria; Belgium; Denmark; France (and regions); Germany; Iceland; Netherlands; Norway; Portugal; Russia (cities only); Spain (and regions); Sweden (and Stockholm); Switzerland; Turkey (Istanbul only).  <i>Other countries:</i> Brazil (and regions); Chile (and regions); Ecuador (and Guayas); Indonesia	26 April 2020.  <i>Regularly updated.</i> Open access <a href="#">on GitHub</a> .

Source and metafiles	Measure reported	Period and type of data	Benchmark	Disaggregation?	Locations compared	First publication date
					(Jakarta only); Israel; Peru (and regions); South Africa; USA (and states).	
<p><a href="#">The New York Times</a></p> <p><i>Clear description of data sources and coverage and method on GitHub.</i></p>	<p>Number of deaths and of excess deaths (actual deaths minus the expected deaths).</p> <p><i>[The raw data allow P-scores to be calculated.]</i></p>	<p>Weekly or monthly, differs per country.</p> <p>Occurrence data for most countries. UK based on registration data.</p>	<p>“Expected deaths”, averages ranging from 2 to 5 years, data-dependent, and differing by country, and adjusting reported deaths for trends and seasonal components using a linear model, see GitHub (e.g. 5-years for the U.S. over 2015-2019).</p>	<p>No regional disaggregation. Some cities, see next column.</p>	<p>Austria, Belgium, Brazil (only 6 cities: São Paulo, Rio de Janeiro, Fortaleza, Manaus, Recife and Belem), Denmark, Ecuador (and Guayas), Finland, France (and Paris), Germany, Indonesia (only Jakarta), Israel, Italy (and Bergamo and Milan), Japan (only Tokyo), Netherlands, Norway (and Oslo), Mexico (only Mexico City), Peru (and Lima), Portugal, Russia (only Moscow, St. Petersburg), Spain (and Madrid and Catalonia), South Korea, Sweden (and Stockholm), Switzerland, Thailand, United Kingdom (and London), United States (and 6 cities: Boston, Chicago, Denver, Detroit, Miami, NYC).</p>	<p>30 April 2020.</p> <p><i>Regularly updated.</i> Open access on <a href="#">GitHub</a>.</p>
<p><a href="#">BBC</a></p> <p><i>Citation of data sources. No metafile with links to data or sources.</i></p>	<p>Official COVID-19 deaths.</p> <p>Number of excess deaths (actual deaths minus the expected deaths).</p>	<p>Cumulative over the pandemic; different periods for different countries.</p> <p>Occurrence data for most countries. UK based on registration data.</p>	<p>For most countries, taken as the historical average of deaths for that week over 2015-2019.</p>	<p>Regional disaggregation in the UK to its constituent nations and sub-regions in England.</p> <p>Some cities, see next column.</p>	<p>UK and its constituent nations and regions.</p> <p><i>European countries:</i> Austria, Belgium, Denmark; France; Germany; Italy; Netherlands; Norway; Portugal; Russia (cities only); Serbia; Spain; Sweden; Switzerland; Turkey (Istanbul only)</p> <p><i>Other countries:</i> Brazil (six cities only); Chile; Ecuador; Indonesia (Jakarta only); Iran; Japan; Peru; South Africa; South Korea; Thailand; USA.</p>	<p>18 June 2020.</p> <p><i>Data not downloadable.</i></p>
<p><a href="#">The Guardian</a></p> <p><i>Citation of data sources. No metafile with links to data or sources. Data not downloadable.</i></p>	<p>Number of deaths, official COVID-19 deaths and of excess deaths (actual deaths minus the expected deaths).</p>	<p>Weekly; covering weeks 1-20 of the first wave of the pandemic.</p> <p>Occurrence data for most countries. UK based on registration data.</p>	<p>Historical average of deaths for that week over 2015-2019.</p>	<p>None. No regional data. No age or gender breakdowns.</p>	<p>UK (no breakdown by constituent regions). Spain, Denmark, Sweden, Netherlands,</p> <p><i>European countries:</i> Austria, Germany, Belgium, Italy, France, USA.</p>	<p>29 May 2020.</p> <p><i>Data not downloadable.</i></p>

Sources: The Economist (2020): “[Tracking Covid-19 excess deaths across countries](#)”, 16 April, 2020, and Tozer (2020): “[Measuring the true toll of the pandemic](#)”, 24 April, 2020. For the Economist, Tozer measures excess deaths from the week the first 50 Covid deaths were reported, to around April 12. As of 15 May, The Economist’s J. Tozer and M. González publish the raw country data on [GitHub](#). Also see “[Global coronavirus death toll could be 60% higher than reported](#)”, Financial Times, 26 April, 2020 and Wu et al. (2020): “[46,000 Missing Deaths: Tracking the True Toll of the Coronavirus Outbreak](#)”, The New York Times, 30 April, 2020. See EuroMOMO webpage: “[Methods](#)”.

### Box 1: Measures of excess mortality: P-scores, per capita excess mortality and Z-scores

Denote the number of weekly deaths by  $x$ .

**The P-score** is defined as follows:

$(x$  minus the expected ('normal') value of  $x$  for the population), divided by the expected value of  $x$  for the population.

**A variant P-score** (U.S. National Center of Health Statistics) is defined as follows:

$(x$  minus the upper threshold for the expected value of  $x$  for the population), divided by the upper threshold for the expected value of  $x$  for the population.

- *The upper threshold* is defined as the expected value plus the 2.5% confidence interval for this expected value. This takes into account uncertainty created by the natural variability of  $x$ .

**The per capita excess mortality** is defined as follows:

$(x$  minus the expected value of  $x$  for the population), divided by the population.

**The Z-score** is defined as follows:

$(x$  minus the expected value of  $x$  for the population), divided by the standard deviation for the population of  $x$  around its expected value.

- EuroMOMO *estimate* the expected value of each country's weekly deaths using data for the previous five years, taking seasonal factors and trends into account, and adjust for delays in registration.
- EuroMOMO assume that a Poisson distribution, adjusted for excess dispersion is a good approximation to the underlying probability distribution of weekly deaths.\*
- Graphs published for each country show the weekly Z-scores since 2015 compared to their usual range of -2 to +2, the approximate 95% confidence interval. Around 2.5% of observations would thus usually have a Z-value over 2. The Z-score equals 4 line is also shown, corresponding to a 'substantial increase': under usual conditions, the Z-value would exceed 4 only around 0.003% of the time.
- The graphs show more deviations of Z-scores 'exceeding 2' and 'exceeding 4', than one would expect. The main reason is that to fit the baseline, EuroMOMO chose only the period of the year when additional processes (e.g. Winter influenza and Summer heat waves) leading to excess deaths are not likely to happen. Normal variability is thus measured after excluding these seasons.\*\*

---

\* The Poisson is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate and independently of the time since the last event. The calculation is described in Farrington et al. (1996).



### 3. Key issues for comparing rates of excess mortality across and within nations

There are several reasons for wanting to compare excess mortality between regions or countries. The first is simply to compare the death toll of the first wave of the pandemic. Useful aggregate measures include the count of excess deaths relative to normal deaths, for example, the P-score, and excess deaths relative to population size, see Box 1. The second of these measures has the problem that older populations tend to have higher normal death counts. This measure of excess deaths will overstate the incidence of the pandemic in older compared to younger populations. For the second reason, that of evaluating the effectiveness of policy responses, one needs to dig deeper, and the simple measures above require further interpretation. Countries may differ in the size of the initial source of infection, in their age structure, in the distribution of co-morbidities in the population and the prevalence of dense urban centres, making some countries more vulnerable. Comparing age-standardised mortality can be helpful in controlling for differences in age structures. Finally, the third motivation for comparisons is a purely objective one of improving the scientific understanding of the dynamics of the spread of infections, their incidence and the death rates of those infected. Key to this last endeavour is the production of granular data, i.e. disaggregation of excess deaths data by age, gender, region, and, where possible, socio-economic categories.

A recent controversy in the UK amongst statisticians has served to reinforce the point of our paper, which is that there can be international comparability *now* of excess mortality with aggregate and more granular P-score data. There are already widely available granular data sets on related aspects such as inequality and urban density, which could be combined with such data for illuminating the comparisons across countries and revealing the effectiveness of different types of policy. Ideally there should be transparent definitions of data and comparability of definitions across nations which may involve coordination by existing international bodies for standards of data dissemination. This will evolve over time but does not preclude analysis *now*. Important is the accessibility of data to all, especially modellers in the fields of epidemiology, economics and sociology. Scientific analysis with appropriate data is needed to inform policy *now* because not only may there may be successive waves of the pandemic in each country, but many countries experiencing later pandemic crises have the potential to reignite infections in earlier countries when borders are open, and there may be pandemics in future years.

Turning to the controversy, Spiegelhalter (2020a) in a Guardian article on 30th April 2020 made valid points about data definitional differences and poor collection of data across some countries of Covid-19 infection and mortality rates. We are in agreement on this, but although he discusses the more reliable data on excess mortality, he argues that we will have to wait for months if not years before we can begin making useful comparisons across countries. However, given that the first wave of the pandemic in Europe has neared its end in most countries, *now* is a good time to make international comparisons at least within Europe. Indeed, on 4 May 2020, a letter<sup>6</sup> from three statistics professors, Philip Brown, James Smith and Henry Wynn disputed Spiegelhalter's claims saying: "Yes, there are inconsistencies, underreporting and heterogeneity within countries, but the policies adopted by different countries show very large differences in effects that would seem to dwarf such worries." Their concern was that the article would deflect criticism of the political handling of the crisis (and indeed it had already in their view). They argue that comparisons combined with careful modelling are needed *now* to explain variations in mortality rates and infection rates across locations toward improved policy. They cite for instance a U.S. modelling endeavour, Rubin et al. (2020), the latest of version of which analyses and forecasts US county level data on death rates, taking into account local factors across US counties including population density, incidence of smoking and social distancing as

---

<sup>6</sup> See Letters, The Guardian: [UK behind the curve in curbing Covid-19 deaths](#), 4 May 2020.

measured by cell phone movement data. The statisticians suggest that such modelling tools are appropriate to apply to country comparisons, and critical for modelling testing and tracing to the community level. We emphasise this modelling point more broadly in section 8.

To interpret large differences in excess mortality between nations requires consideration of several factors, and the within-nation deviations in these factors: the average infection rates in preceding weeks, average mortality risk from Covid-19 for those infected (the case fatality rate) and constraints on Covid-19-specific health capacity.

Turning to the first of the factors, consider differences in infection rates. Compare two countries or regions with the same average Covid-19 mortality risk where 1 percent of all adults are infected in A, while 5 percent are infected in B. Then the rate of excess deaths for adults measured by the P-score will be about 5 times as large in B in the weeks following the incidence of the infection. Countries that locked down early and had effective test, trace and isolate procedures kept down the average infection rate and hence the excess death rate.<sup>7</sup> Within countries, infection rates can differ. London's higher excess mortality was influenced by higher initial imports of infections and a higher virus reproduction number given its high density and hard-to-avoid close physical contact on public transport and at work. Thus, countries that have a higher fraction of adults in locations or occupations where the virus can more easily spread will tend to have higher excess death rates.

Mortality risks for infected adults, the second of the factors mentioned above, can differ between and within countries. For example, the percentage increase in mortality risk may be greater for some ethnic groups, or for some co-morbidities such as diabetes or pre-existing lung conditions. Then country differences in the prevalence of obesity and smoking will influence comparative excess mortality.

Lastly, a country's excess mortality is further driven up, and potentially *much* further, by limited Covid-19-specific health capacity. The death rate among infected adults depends on capacity constraints on numbers of hospital beds and staff, numbers of ventilators, PPE, testing and logistical failures in delivery, e.g. to care homes. Given similar initial capacities, a country with a higher average infection rate will be more likely to run into these constraints. By the same logic, given the same high infection rate, a country with lower health capacity would have a higher rate of excess mortality. This is why there is such a focus on 'flattening the pandemic curve'. Different capacity constraints can have different implications for different groups. For example, lack of PPE and testing facilities in care homes will have disproportionately larger effects on mortality for the oldest individuals and this could affect country comparisons.

Covid-19, therefore, interacts with the age distribution, the nature of health service delivery, poverty and inequality, ethnic and occupational structures, air pollution, the relative size of major conurbations and so on. Comparing rates of excess mortality statistics within countries by age groups, by city size and by occupational, social and ethnic groups should generate important insights for future pandemic policy.

Finally, it should be considered whether excess mortality statistics alone are sufficient to measure the impact of a pandemic. The health economics literature has given attention to Quality Adjusted Life Expectancy (QALY) as a criterion for expenditure on health-improving policies. QALYs measure the number of reasonably healthy years a person might expect to live. The number of QALYs lost could supplement the increased death count resulting from the pandemic as a measure of its impact.

---

<sup>7</sup> Transmission and hence rates of infection are also influenced by factors like the nature of social distancing, availability and use of face masks, and cultural differences in the exercise of self-discipline and following of advice.

However, detailed actuarial and medical information is entailed in the complex estimation of the number of QALYs lost. QALYs and the attachment of monetary values to QALYs have long been controversial, see Loomes and Mackenzie (1989), but the concept of a QALY does focus attention on the relative value (by age group) of expected years lost in a pandemic. The excess mortality of working age adults with a normal life expectancy of 30 years might be weighed against the excess mortality of 85-year olds with a life expectancy of 5 years. If the choice is to attach more weight to excess mortality for working age adults this will affect comparisons of countries with different age-specific mortality rates, see section 7.

#### **4. Comparability of statistical measures of excess mortality and other data issues to consider**

We first consider factors influencing whether the statistical measures described in section 2 can be validly compared across countries. We then highlight some important data issues underlying the death count and other statistics that influence comparability.

##### ***4.1 Can we compare the different statistical measures for excess mortality (from all causes) across countries?***

Comparisons between relatively homogeneous countries with moderate population sizes (such as European countries, Japan and Korea) and large countries such as China and the U.S., which span very diverse regions with potentially very different timings and incidence of the pandemic, are necessarily difficult. For the latter, it makes far more sense to compare populous regions or states with nation states of comparable scale.

P-scores, per capita measures of excess deaths and Z-scores use the concept of ‘normal deaths’ in their numerator by comparing raw death figures with what would normally have been expected. Assuming that the data definitions for the death counts, such as the definition of the week, type of death count data collected (registration versus occurrence data, see below) and timeliness of the collection, are identical across countries (which they are not, see the next sub-section), we consider the relative comparability of the statistical measures described in section 2. For any measure, it is clear that cumulating actual deaths and normal deaths over the period of the first wave of a pandemic gives a more robust summary of its impact, as compared to examining only the peak week.

##### *Comparability of P-scores and variant P-scores*

The *P-scores* are robustly comparable across countries, with the caveat that the measure of ‘normal deaths’ is likely to be only approximate (see below). However, the underlying death count data do need to be transparent and fully comparable to make the comparisons valid, see section 4.2.

Normal death rates *already* reflect persistent factors such as the age composition of the population, the incidence of smoking and air pollution, the prevalence of obesity, poverty and inequality, and the normal quality of health service delivery. This makes *P-scores* particularly attractive even if age compositions and other persistent factors differ. Since they measure the *percentage deviation* compared to what is normal, these persistent differences will already be incorporated in the definition of the ‘normal’ death rate.

Variant *P-scores* add an allowance for historic data variability to the normal number of deaths to define an upper threshold (supposedly based on the 95 percent confidence interval around normal deaths). They define excess deaths relative to that threshold and scale by the same threshold to compute a percentage. The variant *P-score* is therefore always a bit below the simple *P-score* but

tracks it closely. Because the variant is more complex, the simple P-score is preferable. It can always be accompanied by an indication of the margin of uncertainty around estimated normal deaths. When cumulated over a number of weeks, that margin of uncertainty falls so that there is then even less difference between the simple and variant measures (see Figure 3).

#### *Comparability of the per capita excess mortality measure*

Scaling excess deaths by population is obviously better than attempting to compare crude excess death counts for countries with vastly different populations. However, countries with older populations will tend to have higher normal death rates. This automatically means that countries like Italy with an older population will have higher measures of per capita excess mortality than countries with younger populations, such as England. Therefore, comparisons of per capita excess mortality need to be made with caution. A possible argument in favour of per capita excess mortality is that total population could be regarded as a rough proxy for the ability of the society to absorb excess deaths. However, on that logic, dividing excess deaths by the working age population would make more sense.

#### *Comparability of Z-Scores*

As explained in Box 1, Z-scores deflate excess deaths by the standard deviation of normal deaths. In principle, given the assumption of the Poisson distribution, see Box 1, Z-scores should not be compared across countries of very different sizes, though they are useful for comparing the profile of weekly excess deaths for an individual country. The reason is that countries with small populations, and therefore more noisy weekly counts of mortality, have higher standard deviations relative to normal deaths than the more populous countries. In practice, due to the inappropriate assumption of the Poisson distribution (see Appendix 1), the excess mortality rankings between countries are more similar to the P-scores than expected.

The Poisson is likely to be poor approximation to the stochastic process for number of deaths, even in what EuroMOMO call *normal* seasons. EuroMOMO exclude Winter and Summer because of systematic shifts in mean deaths due to 'flu, bad weather or heat waves. But it seems extreme to assume there are no systematic shifts in mean deaths throughout Spring and Autumn. If there are excess deaths due to a bad 'flu in Winter, then in Spring below-average excess deaths should result. There are other examples, such as a measles outbreak, or changes in support for the homeless or for care homes (e.g. from fiscal austerity measures), that may affect mortality rates. There could also be time-varying clusters of different influences - such as a varying previous exposure to risks such as smoking - among the most vulnerable age groups. Thus, the constant *mean* assumption is almost certainly wrong. Turning to the weekly *standard deviation* for 'normal' seasons used by EuroMOMO to deflate the Z-score (see Box 1), variations in *systematic factors* such as these which shift the mean will be included in the measure, as well as random noise (see Box 2). Hence, Z-scores include these systematic features in the denominator and numerator. The paradox is that this makes the Z-scores somewhat more comparable for countries of different sizes (see Appendix 1). The Z-scores indicate approximately (given the Poisson assumption) in which weeks excess deaths were statistically significant; hence they can in principle distinguish those countries with few, if any, weeks of excess deaths (e.g. Germany), from countries with many weeks of excess deaths (e.g. Belgium), irrespective of their large population size differences.

Another major defect of Z-scores, compared to P-scores and per capita excess death measures, is that their cumulation over multiple pandemic weeks is problematic. While excess deaths can be cumulated, the standard deviation of normal deaths cannot, and, in any case, EuroMOMO do not report either excess deaths or these standard deviations. This makes it hard to obtain a comprehensive summary of the pandemic's impact from the Z-scores.

## Box 2: Two pieces of evidence against the Poisson assumption used in EuroMOMO Z-scores

We consider two pieces of evidence against the assumption of a Poisson distribution by EuroMOMO. Both show there are common systematic factors driving mortality data.

1. *We examine the correlations of Z-scores within the UK.* If there are systematic sources of variation of death rates, as well as pure noise, these systematic factors for the UK regions are very likely to be *correlated*. On 100 observations, 2015-2019, excluding winter and summer weeks as for EuroMOMO, the correlation matrix is:

	England	Wales	Scotland	N. Ireland
England	1			
Wales	0.345482	1		
Scotland	0.326606	0.205122	1	
N. Ireland	0.298233	0.106424	0.17243138	1

These quite high correlations imply systematic factors common to all regions. Thus, the Poisson distribution cannot be correct as it assumes independence between regions and over time. Moreover, simple regressions between the Z-scores for Wales, Scotland and N. Ireland and that for England, give coefficients, respectively, of 0.32 (0.089), 0.30 (0.088) and 0.29 (0.092), with standard errors in parentheses. In reverse, a multiple regression of the Z-score for England on all the others gives: Wales 0.29 (0.097); Scotland 0.25 (0.099); N. Ireland 0.24 (0.094)

2. *We examine the ratios of Z-scores to P-scores.* If they shared the same concept of normal or expected deaths, the Z/P ratio would equal the ratio of 'normal' deaths to their standard deviation. Under the constant mean Poisson assumption, this ratio would be proportional to the square root of the number of normal deaths. We lack access to EuroMOMO's estimates of the normal number of deaths, but these should be close to the previous 5 years' average. The ranking (high to low) of the estimated Z/P ratios in the peak week of the pandemic for the different countries, should be the same as their ranking by the normal number of deaths. EuroMOMO adjusts the Poisson assumption with a small allowance for extra dispersion but this should not affect the ranking.

The table shows that the expected ranking if the adjusted Poisson assumption were true is far from being confirmed by the evidence. One should expect Belgium to have the lowest Z/P and France the highest, with Italy the second highest, within Europe. Instead, Italy has the lowest, despite its relatively large number of normal deaths. Within the UK, with the exception of Wales, the rankings of ratios of Z/P do follow the rankings by population size and normal death counts. Regions with small populations - hence small numbers of normal deaths - should have *somewhat* noisier death rates since the purely random component of deaths would be larger compared to the systematic component. But only if the systematic component were zero would the ratio of the standard deviation to normal deaths be entirely determined by the normal number of deaths. [Appendix 1](#) spells out the same argument somewhat more formally.

**Table: Peak weeks of excess mortality: country P-scores and Z-scores compared**

Peak weeks	Excess mortality scores		Ratio	'Normal' deaths	Population
	P-score	Z-score	Z/P	number	millions
<b>All age groups, standard P-scores</b>					
England (week 16) (Z: week 15)	116	41.24	0.36	9,787	56.0
Spain (week 14)	154	43.53	0.28	8,118	46.8
Belgium (week 15)	104	30.39	0.29	2,095	11.6
Italy P: (week 13) (Z: week 14)	85	16.94	0.20	11,818	60.5
Netherlands (week 14)	74	23.44	0.32	2,916	17.1
France (week 14)	67	21.72	0.32	11,380	65.3
Rest of UK					
Scotland (week 15)	80	15.8	0.20	1,100	5.4
Wales (P: week 16) (Z: week 15)	77	19.5	0.25	661	3.1
N. Ireland (P: week 17) (Z: week 15)	56	9.38	0.17	301	1.9

## **4.2 Data issues underlying the statistics that influence their comparability**

Even if we deem the P-scores and the population-deflated statistics to be comparable across countries, underlying measurement issues of the death count, strongly affect the comparability across countries. These definitional differences need to be highlighted and made transparent across country data providers and international organisations reporting excess mortality statistics. The transparent reportage of the Human Mortality Database is exemplary in this regard.

### *The accuracy of the basic data collected*

Perhaps the biggest single pitfall for comparability may arise from the accuracy of the raw mortality data. In our VoxEU article (Aron and Muellbauer, 2020a) we highlighted the advantages of excess mortality data over recorded Covid-deaths, see also section 1, assuming that the collection of data on deaths from all causes would be relatively up-to-date and complete.

Yet countries differ in the efficiency of their death registration systems, particularly where those systems are devolved to regional or local administrations. Then, problems in one location can affect or delay the national data, and sometimes the national recording system can be slow to absorb regional information. In a pandemic, it can happen that the capacity of systems is temporarily overwhelmed, most of all in hotspots, often in urban areas. Occasionally the recording methods may be so weak overall, that the observers resort to data on burials.<sup>8</sup>

The most striking recent example of revisions in the raw mortality figures is that for Spain announced on May 27<sup>th</sup>. Raw deaths were suddenly revised up by around 12,000, back to early March. Catalonia, whose capital is Barcelona, accounted for well over half of these increases, followed by the regions of Madrid and Castilla La Mancha. A closer look at the data revisions by age shows that the bulk of the revisions were for those aged 75 or more. This is consistent with news reports of the many deaths in care homes.<sup>9</sup> As we shall see, the upward revision in the Spanish data currently places Spain neck and neck with England as the European country with the highest cumulative P-score for the ‘all ages’ group (Table 2), whereas previous data put England’s all-age P-score well ahead.

### *Lag between occurrences versus registration data on death counts*

Another difference is between the death counts by week of registration of the death and week of actual occurrence of the death. The registration data occur later than the occurrence data. EuroMOMO Z-scores apparently use data by occurrence for all reporting countries, see Table 1.<sup>10</sup> HMD use occurrence data for most countries, with the exception of England and Wales.<sup>11</sup>

The occurrence-data are particularly prone to revision, and with the lags of registration data behind occurrence data often increasing during the height of a pandemic. Comparability in dating the peak week of mortality is sensitive to how the data are recorded. For example, in the UK, the peak week for all underlying regions is week 15 using occurrence data, as for the EuroMOMO Z-scores in Table 2. By contrast, death counts based on registration data for the UK show peak weeks of week 17 for N. Ireland, week 16 for England and Wales and week 15 for Scotland, see Table 2. Figure 2 compares for England the occurrence and registration data in calculated P-scores.

---

<sup>8</sup> For the Indonesian capital Jakarta, burial data are used by the Financial Times, Burn-Murdoch et al. (2020).

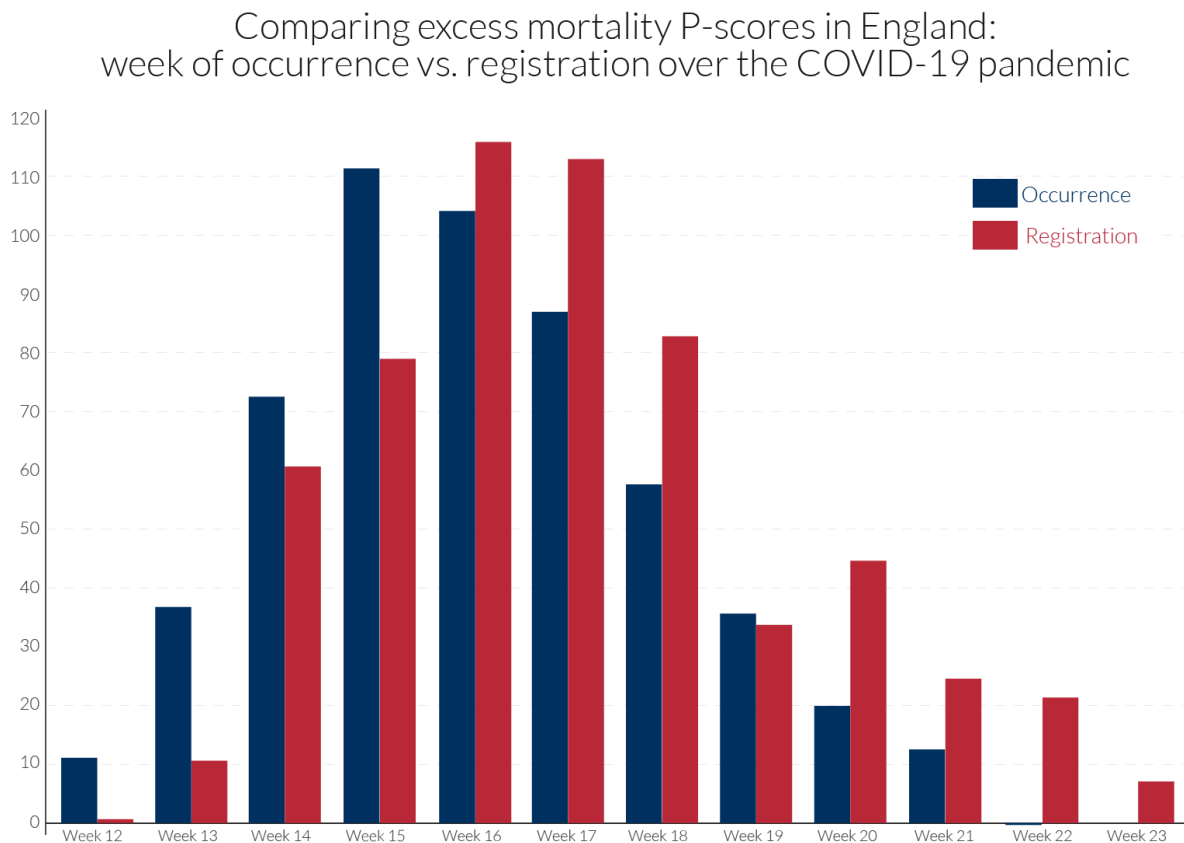
<sup>9</sup> See El Pais: Santaaulalia et al. (2020); Catalan News: ACN (2020a, 2020b).

<sup>10</sup> See their method webpage, Table 1.

<sup>11</sup> See their metafile, Table 1.

It is also important to be cautious when comparing cumulative P-scores across countries if the pandemic has not yet run its full course in some countries.

**Figure 2: Peak of the first wave of the pandemic occurred earlier than when registrations were recorded: contrasting ‘all age’ excess mortality P-scores for England by registration or occurrence data**



Sources: ONS, UK.

### *Measurement of ‘normal deaths’*

The 5-year average could be a crude estimate of normal deaths, e.g. if there are time trends in mortality. If mortality is on an improving trend, normal deaths would be over-estimated by the 5-year average. On the other hand, where populations are increasing or are ageing, the count of normal deaths could also be rising. EuroMOMO use statistical models to adjust for such trends but do not provide their estimates of ‘normal’/expected deaths.

If spring is especially warm as has been the case in Europe in 2020, it is possible that the 5-year average overestimates expected deaths, taking the weather into account. In the latter case, the simple P-score would then underestimate the impact of the pandemic. Also note that not just the effects of the pandemic but of societal reactions, whether driven by government regulation or private behaviour, will be reflected in the death count. Greater social distancing, lower rates of traffic accidents and of deaths due to alcohol abuse as well as ‘collateral damage’ will all affect the death count.

### *Definition of the week*

Countries differ in how they define the week. The mostly widely accepted international definition starts the week on Monday and ends on Sunday. However, of the countries we compare, England, Wales and Northern Ireland start the week on Saturday and ends it on Friday, while all the others, including Scotland follow international practice. This is a relatively minor issue and largely washes out when cumulating excess deaths over multiple weeks, e.g. eleven weeks.

## **5. Why the age distribution matters**

Differences in the age distribution between countries would be irrelevant if mortality risk increased in the same proportion for all. This can never be the case because children have a far lower mortality risk. In countries where children make up a high proportion of the population, the P-scores and excess mortality relative to the total population for the all ages group will be lower.

Looking only at the adult part of the population in a pandemic, there is strong empirical evidence *against* the hypothesis of a proportionate increase in mortality risk at all adult ages. We cannot be sure to what extent this is due to differences in rates of infection or differences in mortality risk once infected.<sup>12</sup> The evidence in section 7 for six countries is for a more than proportionate increase for older adults, i.e. the group of older adults (85+) has a higher P-score than the group of younger adults (15-64). Comparing two countries with the same age-specific P-scores, the country with the higher proportion of older adults would then have a higher all-age adult P-score.

Countries also differ in the age-profile of P-scores. One can see this when comparing the ratio of the P-score for the group of working age adults to that of the group of older adults, e.g. those over 65 or over 85. This ratio is less than 1 everywhere, but some countries have far lower P-scores for working-age adults relative to older adults. To see the implications, take a simple example of two countries with the same age-structure of young and old adults. Suppose the P-score is 1 for the old in both countries, but that country A has a P-score of 0.1 for young adults while that for country B is 0.3. The overall P-score for country B will clearly be higher than for country A. However, if country B also has a higher fraction of young adults, that will attenuate the difference in the overall P-scores between the two countries. Thus, differences in age distributions between countries will affect the measured all-age P-scores and this should be recognised when comparing P-scores.

One could envisage an ‘age-standardised P-score’, adapting the ‘age-standardized mortality rate’, sometimes used to examine the impact of a pandemic. The latter is a weighted average of the age-specific mortality rates per 100 000 persons, where the weights are the proportions of persons in the corresponding age groups of a standard population. The WHO explains the rationale: “Two populations with the same age-specific mortality rates for a particular cause of death will have different overall death rates if the age distributions of their populations are different. Age-standardized mortality rates adjust for differences in the age distribution of the population by applying the observed age-specific mortality rates for each population to a standard population.”<sup>13</sup> A theoretical population, the European Standard Population (ESP), is widely used in Europe to compute age-standardised death rates. This has a particular distribution by age, averaging data from across Europe. The current version from Eurostat was introduced in 2013. The ONS in the UK has also used

---

<sup>12</sup> Spiegelhalter (2020) suggests that for the age group 20-59, the increase in mortality risk compared to normal mortality risk for Covid-19-infected individuals is lower than for those aged 60 and over.

<sup>13</sup> See [Age-standardized mortality rate \(per 100 000 population\)](#), WHO.



age-standardised death rates to compare mortality risk from Covid-19 between the UK regions or between locations with different levels of economic and social deprivation.<sup>14</sup>

However, the 'age-standardized mortality rate' unfortunately conflates variations in normal mortality risk with variations in risk of death during a pandemic. Thus, if the age-standardised mortality rate in 2020 is higher in region A than in region B, this does not necessarily indicate that the Covid-19 mortality risk is higher in A. It may be that normal mortality risk, e.g. based on the average of the previous 5 years, is higher in region A than in B. Age-standardisation removes that part of the difference due to differing *age structures* of the two populations; but it does not remove from normal mortality risk the socio-economic differences, and differences in the incidence of obesity or smoking and in health provision.

An 'age-standardised P-score' would give a better grasp of the increased mortality risk due to Covid-19 than the 'age-standardized mortality rate'. The P-scores for each age group could be computed and the weighted average taken using the age structure of the reference population, rather than of the region or country being considered. It is a better concept because it compares the age-standardised mortality rates during the pandemic period with those normally expected. This type of P-score would provide a provisional answer to the question: 'how different would the overall mortality rate have been with a different age structure of the population?'

There are also potentially *other* ways of standardising aggregate P-scores (or mortality per 100,000 of population) to remove part of the source of between-region or between-country variation. For example, one could standardise by proportions of the population resident in towns and cities classified by common size categories.

The simple aggregate P-score (which weights the age-specific P-scores by the fraction of the population in each age group) and these various standardised aggregate P-scores (which weight the age-specific P-scores by the fraction of the population in each age group in a hypothetical population) have intuitive appeal and be informatively compared across countries. However, one has to be aware of the limitation of any single measure of comparability between countries. Subsumed within the aggregates are implicit value judgements. For example, crucially in the case of a pandemic, there is an implicit assumption that the toll of an older life lost is the same as that of a younger life. However, when a younger life is lost, many more years of life expectancy are lost, and one might want to attach a larger weight to deaths of the young, see section 3.

An important argument of the lockdown sceptics is an extreme version of this last point: "the virus is mainly killing off those that were on their way out anyway", see Kelly (2020). This article quotes a major downward revision of his estimates by British statistician, David Spiegelhalter, who initially suggested that a large number of those dying of Covid-19 would have died in the coming year in any case, but now suggests about 5-15 percent but less than a quarter.<sup>15</sup> On the 11<sup>th</sup> June, cancer specialist Karol Sikora stated for the Telegraph that at least half of those dying of Covid-19 would have died anyway by the end of the Summer of 2020. To try to get a clear position on the issue, Tim Harford (who should be credited for his contribution to the public understanding of data, probability and risk),

---

<sup>14</sup> See the [User guide to mortality statistics](#) of the ONS, UK, and its report on [Deaths involving COVID-19 by local area and socioeconomic deprivation](#).

<sup>15</sup> "I used to think this figure would be quite big but I've reduced my estimate now. I'm not going to put a precise figure on it, but I definitely think the proportion of those who would have died over the next year anyway would be well below a quarter, maybe 5 to 15 percent, rather than 'less than a quarter'." Kelly (2020), Financial Times.

invited actuary Stuart McDonald<sup>16</sup> to comment in the BBC programme “More or Less”.<sup>17</sup> McDonald disagreed with the assertion that a majority would have died in the next 3 months as it was neither supported by the data nor his own research. While it is true that three-quarters of the excess deaths were of people aged 75 and above, and that the majority had one or more pre-existing medical conditions (co-morbidities), in practice, life expectancy is quite high. For example, at the age of 80, life expectancy is 9 years for males and 10 years for females. Co-morbidities add little to this, in his opinion, since four-fifths of this cohort has two or more co-morbidities, and 90 percent have one or more (there is of course variation around the average). He stated that it was hard to find examples of less than two years’ life expectancy. From detailed data in the insurance industry, he suggested that an obese male smoker aged 80, and even with heart or pulmonary disorders, would still have a life expectancy of at least 5 years. This suggests that the pandemic had a huge impact not just on the death count but on life-years lost, properly measured.

Granular data, disaggregating by region, age and gender, as beginning to be provided by Eurostat (see Table 1), allows the observer to apply their own value judgements. These data, combined with medical information at the country level, would be a crucial input in estimates of life-years lost, alongside counts of excess mortality. Granular data are more informative for evaluating the effectiveness of the policy response and for enhancing scientific understanding to inform policies on ending lock-downs and reducing the risk of a second wave of infections.

## **6. What can we learn from a comparison of the P-scores from the ‘all ages’ data**

Cumulation of the P-scores over time is required to get a comprehensive summary measure of the impact of the pandemic. Looking at comparisons over a single week or two, for example, is insufficiently reliable as there is much variation over individual weeks. Different observers choose different periods to define the beginning and end of the pandemic, for instance beginning with the day when the first Covid-19 deaths or first 50 such deaths were registered. In contrast, we frame our comparisons using the same length of period for each country that we are comparing. We use 11 weeks, which is a comprehensive period to measure the extent of the first wave of the pandemic in European countries (not long enough for the US). The actual weeks chosen differ by country: the timing matches the P-scores. Cumulating the P-scores for ‘all ages’ data shows, see Figure 3, that England is slightly ahead of Spain, but that they are ‘neck and neck’. There is also little difference between the two types of P-scores (ordinary and variant) in terms of ranking. Italy, Belgium, the Netherlands and France follow Spain, while within the UK, Scotland, Wales and N. Ireland follow England.

One caveat is that the English data are from registration data and not occurrence data (see section 4.2). Therefore, the timing of the England peak cannot be compared with the timing of the peak for the other European countries which use occurrence data, since registration of death follows after occurrence of death, with a lag.

Examining the detailed P-scores by week for England and the rest of the UK, and the other European countries, it is clear that the peak incidence in Spain is more severe, but more protracted at high levels of deaths in England (Figures 4a and 4b). The same comparison applies to Belgium and Italy, with the

---

<sup>16</sup> Stuart McDonald is an actuary with the Institute and Faculty of Actuaries and a founding member of the group of industry professionals creating a Covid-19 actuaries’ response group, launched in March 2020. For more detailed information on comorbidity, see Edwards and McDonald (2020).

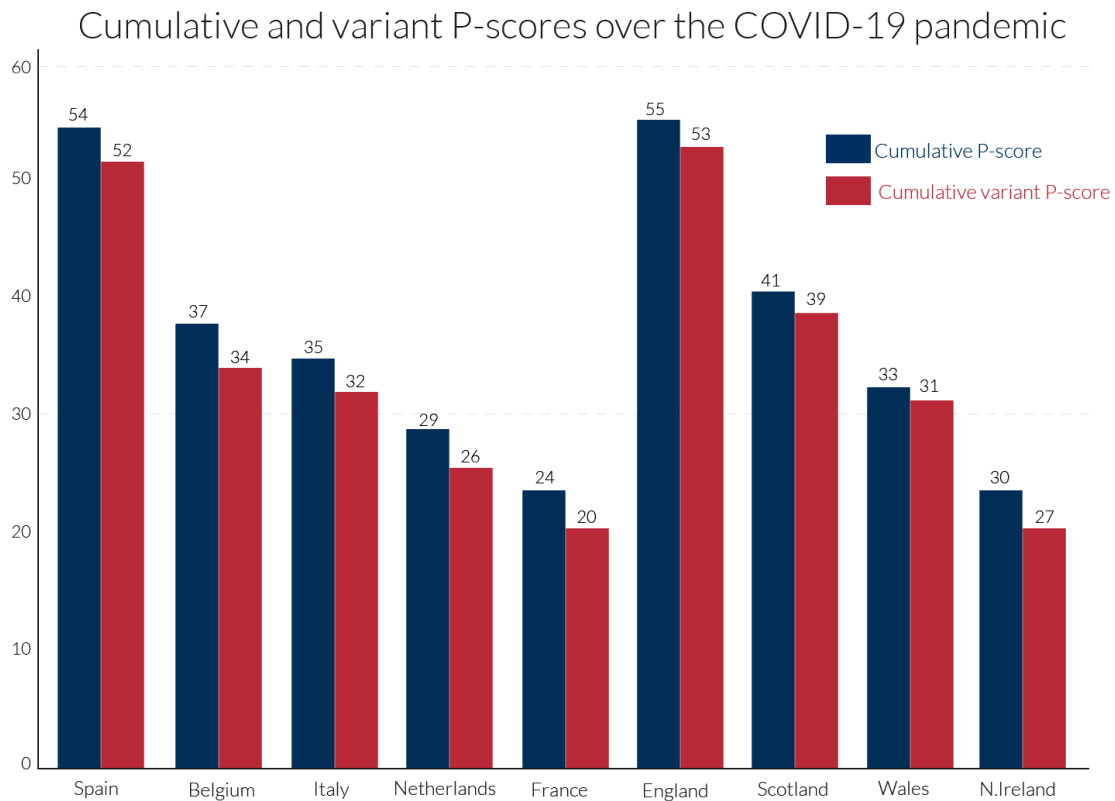
<sup>17</sup> Tim Harford: [Quarantine, Test and Trace and BODMAS: Is it true that Covi-19 mostly ill people who would die soon anyway.](#), More or Less, Radio 4, BBC, 17 June 2020.

latter more protracted. The incidence is quite a bit lower in N. Ireland, which follows Wales and Scotland, behind the England.

The detailed numbers behind the pictures are contained in Table 2. The Z-scores from EuroMOMO are also presented. Since Z-scores are based on occurrence data they provide a more comparable picture for England with the other European countries of the timing of the peak week.

An early proponent of using P-scores of excess mortality - rather than Z-scores - for a meaningful international comparability across countries, and for regions within countries, is the Health Foundation, see Table 1. In two papers, Krelle et al. (2020), published online on the 6<sup>th</sup> May, and Tallack et al. (2020), published on the 4<sup>th</sup> June, present weekly P-scores for the England and Wales, France, Italy, Spain, Sweden and Germany, and for cities, London, Lombardy, Madrid, New York City and Paris. However, of particular interest is their regional disaggregation for some European countries, and sub-regional including to Local Authority areas within the UK. They compare the cumulated P-scores for the UK, France, Italy and Spain from March to early May (March to April for Italy), with the chosen baselines and cumulated weeks visible on the interactive graphs. The comparison of excess deaths as a percentage of normal deaths for regions reveals that the UK showed substantial dispersion between the regions with elevated excess deaths in all regions. The regional dispersion appears to be greater for the other three countries, with some regions especially badly affected (e.g. Lombardy in Italy and the Madrid region in Spain with close to 200% scores) but many little affected (scores below 20 percent). For their dates they report that all regions and countries in the UK had excess death rates close to or exceeding 30 percent, contrasting with only 7 out of 20 regions in Italy, 9 out of 15 regions in Spain and 2 out of 13 regions in France. They also show significant dispersion across UK local authorities (in the top quarter, P-scores exceeded 80 percent, while in the bottom quarter, they averaged around 20 percent), cumulating P-scores from week 12 to 21. They find that almost half of all local authorities had excess deaths exceeding 50 percent of usual deaths.

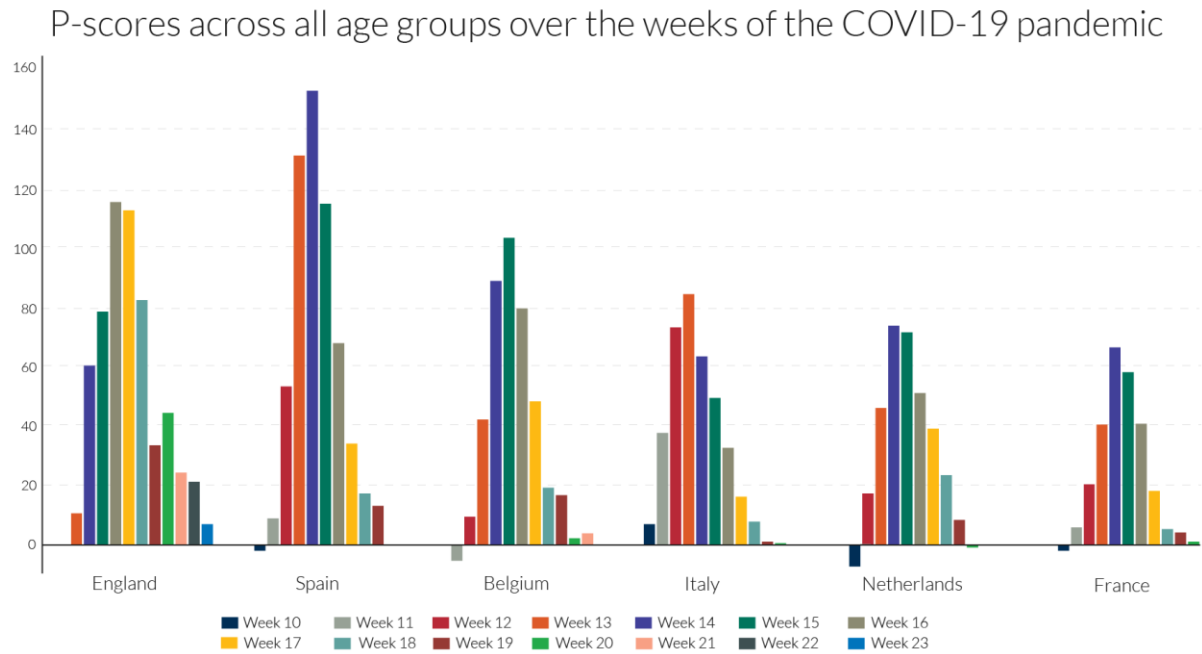
**Figure 3: Cumulative P-scores of excess mortality for poor performers ('all ages') in the first wave of the pandemic**



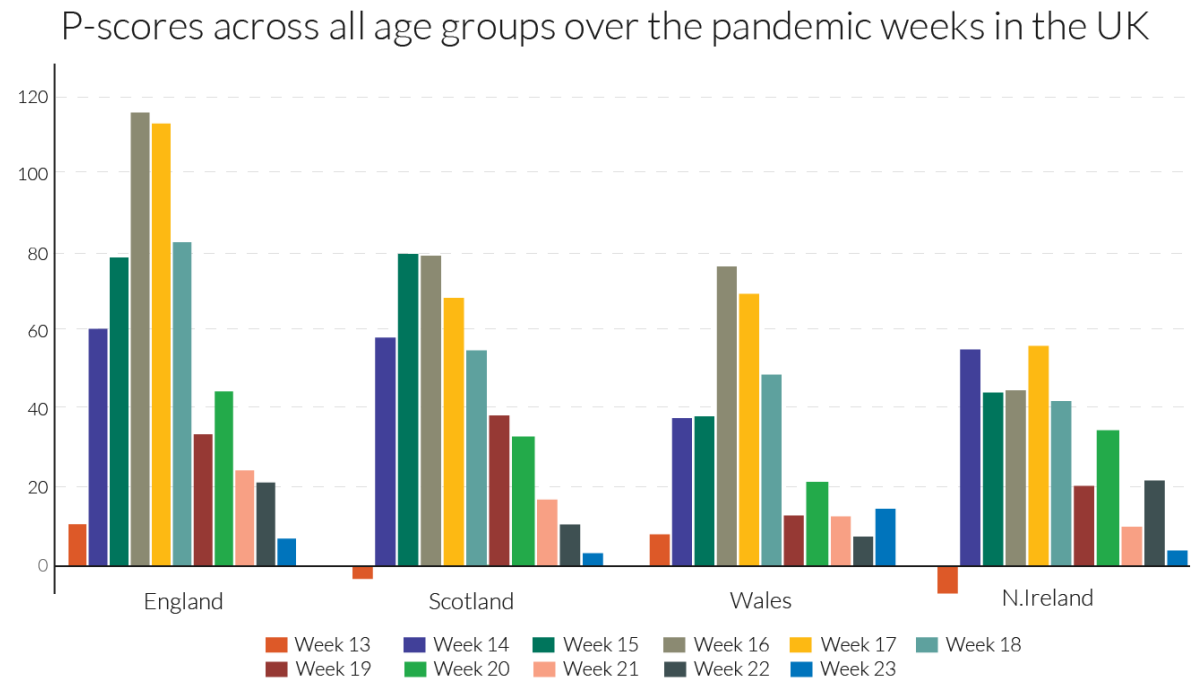
*Sources:* The P-scores and variant P-scores are calculated by the authors using the Human Mortality Database, see meta file: [https://www.mortality.org/Public/STMF\\_DOC/STMFmetadata.pdf](https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf), and the Office for National Statistics for the UK. Cumulative P-scores cover the weeks shown in Table 2.

*Notes:* (i) Incomplete figures for England which is not yet at normal levels. (ii) Deaths by week of registration versus deaths by week of occurrence: HMD use deaths by week of occurrence for all the above countries, except the UK, where deaths by week of registration are used, see section 4.2.

**Figure 4a: Recent weeks of P-scores for poor performers showing peak weeks of excess mortality for ‘all ages’ in the first wave of the pandemic**



**Figure 4b: Recent weeks of P-scores for the UK for ‘all ages’ in the first wave of the pandemic: England, Scotland, Wales and N. Ireland**



Sources: The P-scores are calculated by the authors using the Human Mortality Database, see meta file: [https://www.mortality.org/Public/STMF\\_DOC/STMFmetadata.pdf](https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf), and the Office for National Statistics for the UK.

Notes: (i) Deaths by week of registration versus deaths by week of occurrence: HMD use deaths by week of occurrence for all the above countries, except the UK, where deaths by week of registration are used, see section 4.2. (ii) The country ordering is by cumulative P-scores. The cumulative P-scores are shown in Figure 3 and cover the weeks shown in Table 2.

**Table 2: Our P-scores/variant P-scores and EuroMOMO's Z-scores for poor performers showing peak weeks of excess mortality in the first wave of the pandemic**

<i>All age-groups</i>	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18	Week 19	Week 20	Week 21	Week 22	Week 23	
For week ending: <sup>(iii)</sup>	8-Mar-20	15-Mar-20	22-Mar-20	29-Mar-20	5-Apr-20	12-Apr-20	19-Apr-20	26-Apr-20	3-May-20	10-May-20	17-May-20	24-May-20	31-May-20	7-Jun-20	Cumulative P-Score *
<i>P-Scores [these use data on deaths by week of occurrence– except for the UK which uses data on deaths by week of registration]</i>															
England				11	61	79	<b>116</b>	113	83	34	45	25	21	7	<b>55</b>
Spain	-2	9	54	132	<b>154</b>	116	68	34	17	13	0				<b>54</b>
Belgium		-5	10	43	90	<b>104</b>	80	49	19	17	2	4			<b>37</b>
Italy	7	38	74	<b>85</b>	64	50	33	16	8	1.20	0.48				<b>35</b>
Netherlands	-7	0	17	46	<b>74</b>	72	51	39	24	8	-1				<b>29</b>
France	-2	6	21	41	<b>67</b>	59	41	18	5	4	1				<b>24</b>
<i>Within UK</i>															
Scotland				-4	59	<b>80</b>	80	69	56	39	34	17	11	4	<b>41</b>
Wales				8	38	38	<b>77</b>	70	49	13	22	13	8	15	<b>33</b>
N. Ireland				-7	56	44	45	<b>56</b>	42	21	35	10	22	4	<b>30</b>
<i>Variant P-Scores [these use data on deaths by week of occurrence– except for the UK which uses data on deaths by week of registration]</i>															
England				8	49	60	97	<b>108</b>	70	22	43	22	19	6	<b>53</b>
Spain	-6	4	49	125	<b>148</b>	111	64	32	15	12	0				<b>52</b>
Belgium		-14	2	34	82	<b>94</b>	74	45	15	15	-2	3			<b>34</b>
Italy	3	34	69	<b>81</b>	59	46	28	14	5	-2	0				<b>32</b>
Netherlands	-17	-8	10	41	68	<b>68</b>	46	35	21	6	-5				<b>26</b>
France	-9	0	15	34	<b>61</b>	54	37	16	4	2	-2				<b>20</b>
<i>Within UK</i>															
Scotland				-7	50	<b>68</b>	67	63	49	33	27	14	9	1	<b>39</b>
Wales				4	27	27	<b>66</b>	60	40	4	17	11	1	10	<b>31</b>
N. Ireland				-11	<b>41</b>	33	32	47	26	8	23	3	16	-4	<b>27</b>
<i>Z-scores [these use data on deaths by week of occurrence for all countries]</i>															
For week ending:	8-Mar-20	15-Mar-20	22-Mar-20	29-Mar-20	5-Apr-20	12-Apr-20	19-Apr-20	26-Apr-20	3-May-20	10-May-20	17-May-20	24-May-20	31-May-20		
England	0.57	0.44	5.24	15.15	32.56	<b>41.24</b>	36.08	29.38	20.49	14.42	8.7	6.36	4.36		
Spain	0.73	4.72	17.41	40.47	<b>43.53</b>	32.84	20.06	10.11	4.73	2.69	-1.18	-0.32	-0.06		
Italy	2.62	6.42	11.72	14.73	<b>16.94</b>	13.18	8.89	6.89	4.36	3.32	1.32	1.13	-0.65		
Belgium	0.29	0.85	4.68	11.91	21.01	<b>30.39</b>	20.92	12.03	4.44	4.76	1.69	2.43	2.16		
Netherlands	0.78	2.23	6.58	15.29	<b>21.72</b>	21.23	15.1	11.47	6.14	1.97	-0.02	0.11	-0.08		
France	0.85	1.89	6.33	13.78	<b>23.44</b>	20.06	13.52	4.7	-0.25	-0.79	-1.69	0.12	-3.36		

<i>Within UK</i>														
Scotland	0.47	0.92	1.65	4.82	12.66	<b>15.8</b>	14.68	13.57	10.38	8.25	3.68	2.54	1.28	
Wales	-0.41	0.13	1.59	6.07	11.08	<b>19.5</b>	15.72	12.41	6.96	5.06	2.38	1.25	1.46	
N. Ireland	0.09	1.25	-0.06	2.58	6.11	<b>9.38</b>	6.56	5.52	5.24	3.05	1.97	1.9	0.78	

Source: Z-scores are extracted from the EuroMOMO website, 11-Jun-2020. The P-scores and variant P-scores are calculated by the authors using the Human Mortality Database, see meta file: [https://www.mortality.org/Public/STMF\\_DOC/STMFmetadata.pdf](https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf), and the Office for National Statistics for the UK. Cumulative P-scores cover the weeks shown.

Notes: (i) The peak weeks for different countries are in **bold**. The peak weeks for the all age group category is the same for the UK countries. (ii) The country ordering is by cumulative P-scores. (iii) The ONS defines a week as ending on Friday; EuroMOMO define a week as ending on Sunday; for HMD definitions, it is also Sunday for the above countries, except for England and Wales, which is Friday, see the metafile. (iv) Deaths by week of registration versus deaths by week of occurrence: EuroMOMO use deaths by occurrence and HMD use deaths by week of occurrence for the above countries, except in the case of the UK, where deaths by week of registration are used, see section 4.2. (v) Revisions in the raw death count data: there have been recent large revisions in the Spanish death count, see section 4.2. (vi) Which weeks are chosen matter: for example, calculating Italy's cumulative P-score for weeks 9 to 19, instead of weeks 10 to 20, reasonable to do since the pandemic struck first in Italy, gives Italy a P-score of 37 and a variant P-score of 35, putting Italy and Belgium neck and neck.

## 7. Excess mortality for other age groups: 15-64 and 85+

Here, we focus on two age groups, those aged 15-64, containing most of the working age population, and the elderly, those aged 85 or more, many of whom will have been residents in care homes. The evidence here confirms the point made in section 5, that the percentage increase in mortality risk due to the pandemic, measured by the P-score, was higher for older ages. As in section 6, we present the cumulated P-scores over time to get a comprehensive summary measure of the impact of the pandemic for the two age groups. We use the same length of period, 11 weeks, for each country, sufficient to measure the extent of the first wave of the pandemic, though the actual weeks chosen will differ by country as before (see Table 2). What differs from section 6 is that for reasons of data access, 'England and Wales' as an entity are examined here, rather than England alone and other regions of the UK. Cumulating the P-scores for both age groups in Figure 5, shows that in all countries, P-scores are lower for the 15-64 age group than for the 85+ age group. 'England and Wales' lies slightly below Spain for the 85+ age group but is well above it for the working age group of 15-64. In ranking, Belgium, Italy, France and the Netherlands follow Spain and 'England and Wales' for the older age group. But Belgium, France and the Netherlands seem to have sustained far lower deaths than Spain and Italy, and especially 'England and Wales', amongst the working age population group.

It is unclear to what extent these striking differences are due to differences in rates of infection or differences in mortality risk once infected. Over the 11 pandemic weeks, the cumulative P-score for the 15-64 age group in France was negative, though in the middle of the period there were some weeks when it was positive, see Figure 6. This suggests that social distancing and related measures in France may have reduced deaths from other causes for the working age population, which actually saved lives over the first-wave pandemic period. The Netherlands and Belgium also have remarkably low cumulative P-scores for the 15-64 age group and a number of weeks with negative P-scores.

The increase in expected years of life lost, is another measure of the pandemic's impact (section 3). Average life expectancy in the 15-64 age group is obviously substantially higher than the average for the 85+ age group, so many more expected years of life are lost in each excess death among the younger group than among the older. From the higher incidence of deaths among the working age population in England (which dominates the 'England and Wales' figures), it is obvious that England is easily the worst in Europe in terms of expected years of lives lost.

Turning to the timing of the pandemic's incidence, the 'England and Wales' data are from registration data and not occurrence data (see section 4.2). Since registration of death follows after occurrence of death, with a lag, the timing of the England and Wales' peak occurs around one week after its occurrence data, which in turn is later than the peak in most European countries. The timing of the peak week is mostly the same for the two age groups. It is led by Italy in week 13, followed by Spain and France in week 14, the Netherlands in weeks 14-15, Belgium in week 15 and England and Wales in weeks 16-17 (but week 15 according to the occurrence data in section 4.2).

Turning to the detail in Figure 6, the peak incidences for the 85+ age group in Spain and in Belgium are more severe, but for 'England and Wales' the pattern is more protracted at a high level of deaths. The same comparison applies to France and the Netherlands versus Italy, with the last more protracted. Italy initially dominated the headlines for Covid-19-related deaths but ranked fourth for peak excess mortality figures for the over-85s, below Spain, 'England and Wales' and Belgium.

Most disturbing, as noted above, is the comparative story for the 15-64 age-group, where England's relative record in excess mortality in the Covid-19 era is strikingly higher than in the European countries. The 15-64 age-group includes the mass of the working age population. For this age group,



the weekly pattern is rather different than for the over-85s, with 'England and Wales' displaying both a high peak incidence and protracted high level of deaths, followed by Spain and then Italy. Figure 6 shows that not only is England distinctive in the rate of excess mortality in the peak week for the working age group, but the same is true in comparisons of the two weeks before the peak and the subsequent week.

The EuroMOMO graphic visualisations by finer age categories can offer further clues, comparing the 15-44 and 45-64 age groups. Section 3 suggested that comparisons of Z-scores for comparably populous countries yields reasonable approximations in ranking. England and Spain were the only countries with significant excess mortality in the 15-44 age group according to Z-scores, with England far ahead of Spain. Comparisons of Z-scores with less populous states tend to understate excess mortality in the latter, but evidence from the large countries France and Italy suggest that England is a European outlier. While Z-score comparisons with Wales, Scotland and Northern Ireland understate their excess mortality, the differences compared with England are so large that the conclusion that England was exceptional cannot be avoided. For the 45-64 age group, there is evidence of significant levels of excess mortality, at least in the peak weeks of the pandemic, for all the countries in our comparison group of countries with the exception of Northern Ireland. The Z-score evidence is consistent with the P-score patterns in Figure 6 for the 15-64 age group, even if the Z-scores for the smaller countries, Belgium and the Netherlands slightly understate their relative excess mortality. While the Z-scores also understate excess mortality for the 45-64 age group in Scotland, Wales and Northern Ireland, the figures for England are so much higher, that its outlier status is confirmed for this age group as well as the 15-44 age group.

These country differences call for further analysis, especially by age and by regional differences within countries (contrasting, for example, regions with large urban centres and those without). It would be interesting to know to what extent working age excess mortality in London dominated the data for England. It is also possible that cramped housing conditions in London, especially for poorly paid workers, accounts for some of the exceptionalism of the data for England. Regional and country differences by occupational categories should also be illuminating. Aron and Muellbauer (2020b) drew attention to evidence from the ONS for England and Wales of major occupational differences in the incidence of deaths attributed to Covid-19 and in age-standardised death rates. The ONS have matched census data on ethnicity with death records and have found substantially higher Covid-19 death rates for racial minorities, ONS (2020). After adjusting for region, population density, socio-demographic and household characteristics, the raised risk of death involving Covid-19 for people of Black ethnic background of all ages together was 2.0 times greater for males and 1.4 times greater for females, compared with those of White ethnic background. Males of Bangladeshi or of a Pakistani and Indian ethnic background also had a significantly higher risk of death involving Covid-19 than White males (1.5 and 1.6 times, respectively) - once region, population density, socio-demographic and household characteristics were accounted for.

Differences between countries in their institutional populations could be another factor in accounting for differences in excess mortality for those aged 15-64 years. Of the countries in our comparison group, England and Wales (and Scotland) have the highest ratios of prison population to total population, followed by Spain.<sup>18</sup> Further analysis is needed of excess mortality in the prison population

---

<sup>18</sup> See graphics in Our World in Data: [Prison Population Rate](#).

as it is possible that failures to protect inmates from infection in those countries with high infection rates could help explain some of the country differences of excess mortality for those of working age.

### **7.1 Toward comparable international statistics on excess deaths amongst care home residents**

One of the stark differences between countries is how well protected were residents in the care homes. The main elements of what happened in care homes in the UK, France, Italy and in Spain is, by now, well-known. Care home staff had inadequate personal protective equipment (PPE) and inadequate access to Covid-19-tests and residents were not well-shielded from potential infection from visitors and staff. Yet, many elderly patients with the Covid-19 infection were released from hospitals to the care homes to reduce the pressure on hospitals from the volume of new cases, and therefore spread the infection to other residents. It is important to explore comparisons between countries of their excess deaths in care homes, for example at the least, the percentage of cumulative Covid-19 deaths that occurred in care homes. The clues in the rate of excess deaths for the 85+ age group, which show the largest increase in Spain, are consistent with newspaper reports of the disaster that befell many care homes in Spain.<sup>19</sup>

We were not able to find comparable data at this stage for excess deaths of those normally resident in care homes across the European countries. However, considerable strides have been made in improving international comparability through the pioneering work of the International Long-Term Care Policy Network, e.g. Comas-Herrera et al. (2020). For international comparability, counts of deaths of those resident in care homes, plus those *normally* resident in care homes but dying elsewhere (e.g. in hospital), would have to be regularly published. Few if any countries currently do this. To compute the percentage of excess deaths in care homes, or for the comprehensive definition which includes deaths of care home residents outside care homes, requires data for the previous five years to be able to estimate 'normal' deaths.<sup>20</sup> Another issue for international comparability concerns differences in definitions of what constitutes a care home. A focus on those over 65 or 75 years of age to exclude some of the other groups, such as refugees, sometimes included in the care home definition, could help international comparability.

It is interesting that England and Wales have some of the most comprehensive data on mortality in care homes internationally (see Comas-Herrera et al. (2020), and Comas-Herrera and Fernandez (2020) for an earlier analysis of data for England). Comas-Herrera et al. (2020) cite ONS data showing that, from early March to 12 June 2020, excess deaths in care homes in England and Wales numbered 26,745, while the overall excess deaths for England and Wales were 59,138. Thus, about 45 percent of total excess deaths took place *in* care homes. The ONS have not produced data on excess deaths among those normally resident in care homes, however, clearly this would be a higher percentage as some residents may have died elsewhere. It would be useful to know what fraction of excess deaths covered all care home residents (that is, within the home or out of it, say in hospital). The Care Quality Commission (CQC) estimates that 84 percent of total deaths from all causes of care home residents took place *in* care homes in the same period. But this figure includes normal deaths that would have occurred in the absence of the pandemic, as well as the deaths induced by the pandemic (Covid-19 attributed deaths, mis-measured, unattributed Covid-19 deaths and those caused indirectly by Covid-

---

<sup>19</sup> See El Pais: Santaaulalia et al. (2020); Catalan News: ACN (2020a, 2020b).

<sup>20</sup> If there were large changes over these five years in the fraction of the elderly population who were resident in care homes, 'normal' deaths could be adjusted for trends in the care home population.

19, through being untreated, for example). Scaling up the above figure of 45 percent of total excess deaths that took place *in care homes* for England and Wales, by the 84 percent figure, i.e.  $45.2/0.84$ , would give an estimate of 54 percent for the percentage of all excess mortality accounted for by care home residents in England and Wales (whether inside or out of the care home at time of death). This would almost certainly be an underestimate, since the 84 percent is an over-estimate, but the 54 percent estimate gives a lower bound.

To potentially correct the estimate of 84 percent for the normal deaths included in it, and to include deaths of care home residents outside the homes, we consider the specific CQC data on Covid-19-attributed deaths as follows. For the period from early March until the 1 May, the CQC estimate that 72 percent of Covid-19-attributed deaths of care home residents in England and Wales occurred *in care homes*. Their equivalent figure for England alone, for the later period of 2 May to 12 June, is 77 percent. However, if the CQC estimate of 77 percent better represented the fraction of excess deaths of care home residents that took place in care homes than the 84 percent figure used above, then 58 percent (i.e.  $45.2/0.77$ ), would be the estimate of the fraction of all excess deaths accounted for by residents of care homes (whether inside or out of the care home at time of death).

Although Comas-Herrera et al. (2020) examine data sources for 27 countries outside the UK, the only other two countries found with data on excess deaths in care homes are Belgium and France. In Belgium the attribution of deaths to Covid-19 is so widely-defined that the count of Covid-19 attributed deaths actually exceeds the count of excess deaths, see Figure 1 above. For Belgium, Comas-Herrera et al. (2020) report that care home residents accounted for 64 percent of all deaths linked to Covid-19. This suggests that the percentage of excess deaths accounted for by care home residents in Belgium is not far from the 64 percent figure. They report for France that care home residents accounted for 49 percent of Covid-19 deaths. However, since the count of Covid-19 deaths understates excess deaths in France, see Figure 1, it seems likely that a higher percentage of excess deaths occurred among care home residents. For Canada, estimates suggest 81 percent of Covid-19 deaths were among residents in long-term care, but comparable estimates for excess deaths are not available.

We can obtain a little more information for the UK by examining data in Table 4 for the four nations comparing the total excess death count in each with information on the location of Covid-19 attributed deaths. The period covered is weeks 13-23 of the pandemic (for dates, see Table 2). For the UK as a whole, 80 percent of excess deaths have been attributed to Covid-19, though for Wales the percentage was far higher.<sup>21</sup> For the UK, nearly half of excess deaths attributed to Covid-19 occurred in hospital and one quarter in care homes, though many of the hospital deaths were of patients who were resident in care homes. The remaining 20 percent may also be related to Covid-19, as unrecorded or mis-recorded deaths, and those indirectly affected by Covid-19 through other health conditions, such as heart conditions and cancer, being left untreated due to implied capacity constraints in the health service.

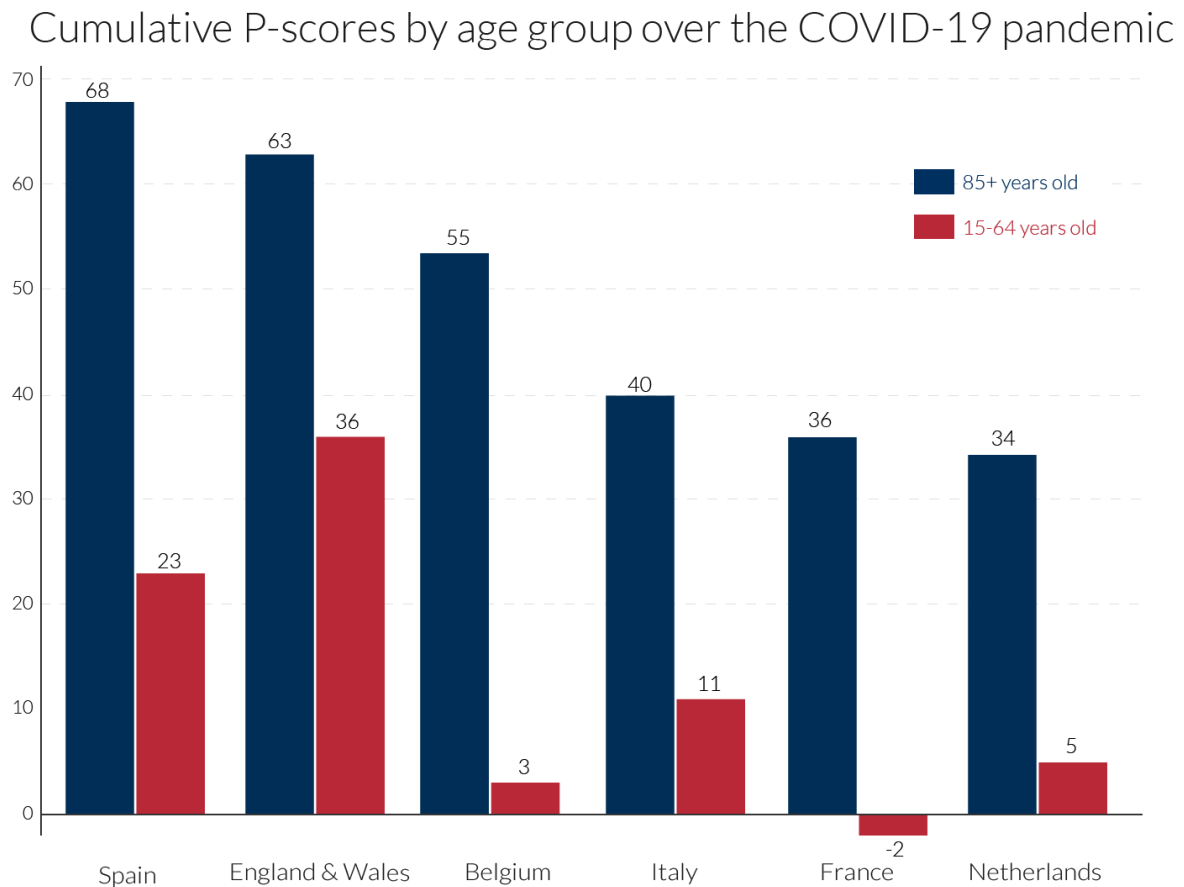
The percentage of excess deaths that took place in care homes from Covid-19 in England, at about a quarter, matches the overall UK figure, but in Scotland and N. Ireland this was sharply higher at 39 and 35 percent, respectively, and in Wales about 30 percent. Concerning the number of Covid-19 deaths, 30 percent of these occurred in care homes in England and in Wales, with 47 percent in Scotland and 43 percent in Northern Ireland. These percentages of Covid-19 deaths are an under-

---

<sup>21</sup> Thus, most excess deaths in Wales in the pandemic period were due to Covid-19, with some mitigating factors from fewer other deaths, e.g. from traffic accidents. This may reflect more transparent recording of Covid-19 deaths. Wales also had a much higher percentage of excess deaths from Covid-19 occurring in the hospitals.

estimate of those normally resident in care homes, because some died in hospital. Hopefully, the compilation of those data will be undertaken by the ONS and the regional health authorities, so that the scale of excess deaths in care homes and its regional variation is properly appreciated.

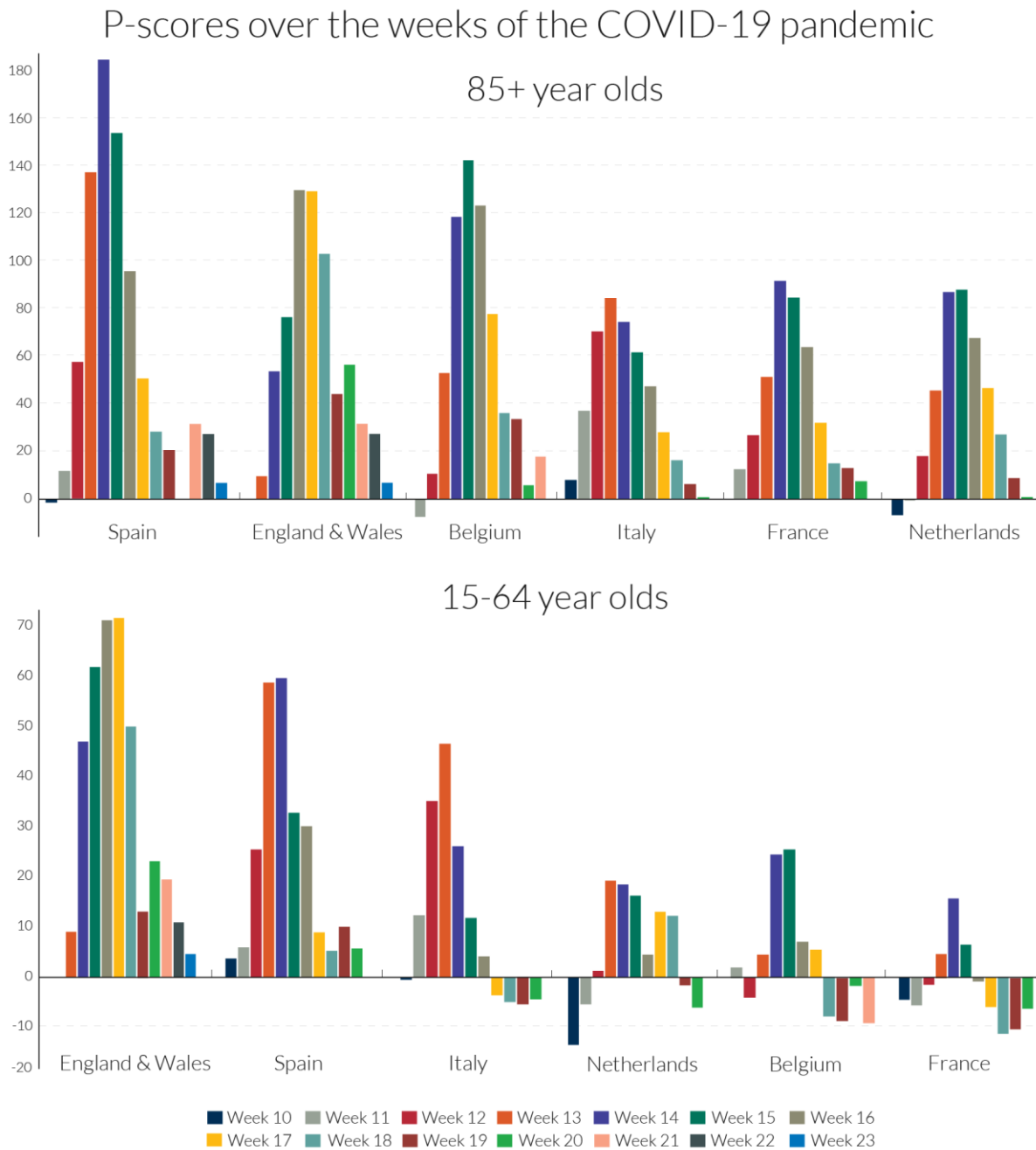
**Figure 5: Cumulative P-scores of excess mortality for poor performers by two age groups in the first wave of the pandemic**



*Sources:* The P-scores are calculated by the authors using the Human Mortality Database, see meta file: [https://www.mortality.org/Public/STMF\\_DOC/STMFmetadata.pdf](https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf), and the Office for National Statistics for the UK. Cumulative P-scores cover the weeks shown in Table 2. The country ordering is by cumulative P-scores for the 85+ age group.

*Notes:* (i) Deaths by week of registration versus deaths by week of occurrence: HMD use deaths by week of occurrence for all the above countries, except the UK, where deaths by week of registration are used, see section 4.2.

**Figure 6: Recent weeks of P-scores for poor performers showing peak weeks of excess mortality by age-group in the first wave of the pandemic**

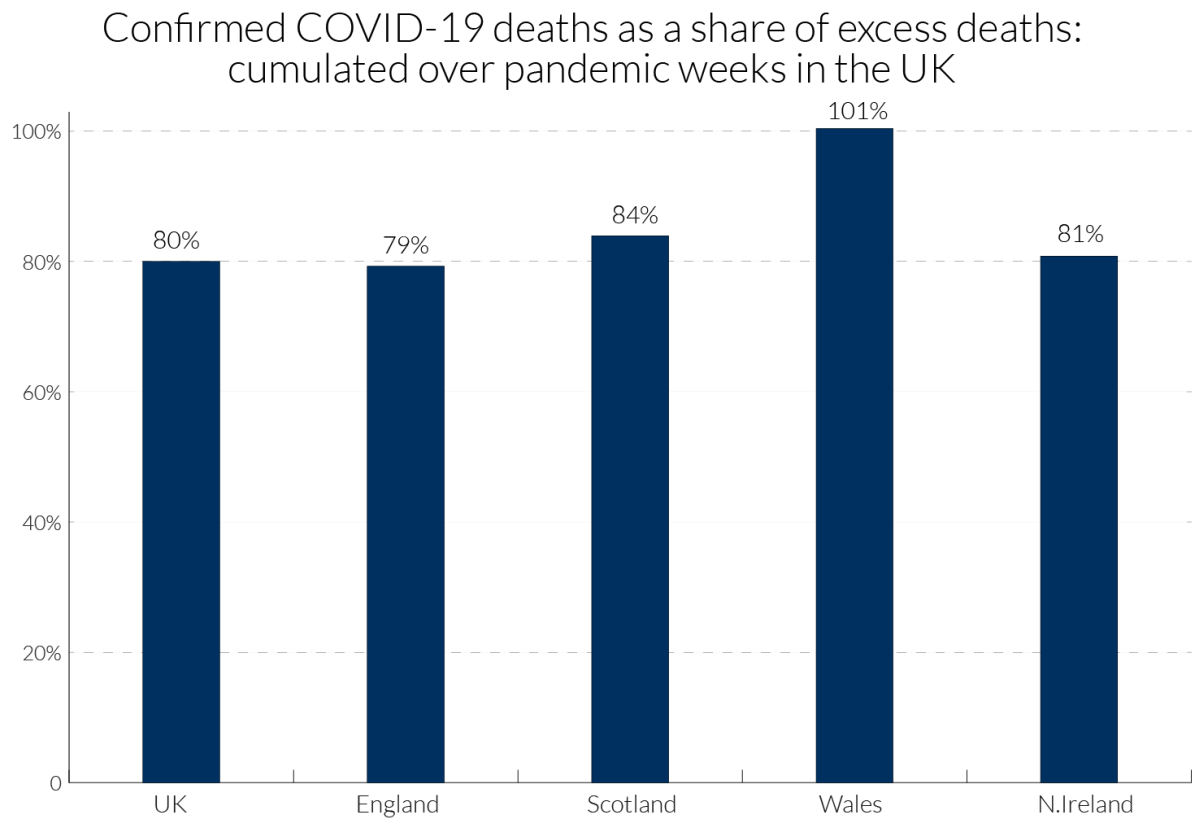


*Sources:* The P-scores are calculated by the authors using the Human Mortality Database, see meta file: [https://www.mortality.org/Public/STMF\\_DOC/STMFmetadata.pdf](https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf), and the Office for National Statistics for the UK.

Cumulative P-scores cover the weeks shown in Table 2. The country ordering is by cumulative P-scores for *each age group*.

*Notes:* (i) Deaths by week of registration versus deaths by week of occurrence: HMD use deaths by week of occurrence for the above countries, except in the case of the UK, where deaths by week of registration are used, see section 4.2. (ii) for France and Italy data for some weeks were absent. (iii) HMD only report for England and Wales. There are no separate data for England and Wales, and no data for the rest of the UK.

**Figure 7: Confirmed COVID-19 deaths as a share of excess deaths for the UK ('all ages'): cumulated over the first wave of pandemic weeks**



**Sources:**

UK - ONS:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales> .

England and Wales - UK ONS:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales> .

Scotland - National Records of Scotland: <https://www.nrscotland.gov.uk/covid19stats> .

N.Ireland - Northern Ireland Statistics and Research Agency (NISRA): <https://www.nisra.gov.uk/publications/weekly-deaths> .

**Notes:** All UK countries taken from Week 13 (week ending 27th March) to Week 23 (week ending 5th June).

**Table 4: Deconstructing the death figures – cumulative deaths over weeks 13 to 23 in the first wave of the pandemic by location of death**

	Reported actual deaths from all causes	'Normal' deaths	Excess deaths	Total COVID-attributed deaths	Total COVID-attributed deaths	Hospital-reported COVID-19 deaths	Hospital-reported COVID-19 deaths	Care home-reported COVID-19 deaths	Care home-reported COVID-19 deaths	Home-reported COVID-19 deaths	Home-reported COVID-19 deaths	Residual COVID-19 deaths	Residual COVID-19 deaths
	<i>number</i>				<i>% of excess</i>	<i>number</i>	<i>% of excess</i>	<i>number</i>	<i>% of excess</i>	<i>number</i>	<i>% of excess</i>	<i>number</i>	<i>% of excess</i>
UK (total)	188,879	124,314	64,565	51,686	80	31,073	48	15,568	24	2,345	4	1,003	2
England	158,958	102,381	56,577	44,631	79	28,310	50	13,334	24	1,999	4	988	2
Scotland	16,535	11,766	4,769	3,990	84	1,846	39	1,861	39	276	6	7	0
Wales	9,232	6,966	2,266	2,292	101	1,485	66	653	29	117	5	37	2
N. Ireland	4,154	3,201	953	773	81	392	41	331	35	39	4	11	1

Sources: UK - ONS: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales> .

England and Wales - UK ONS:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales> .

Scotland - National Records of Scotland: <https://www.nrscotland.gov.uk/covid19stats> .

N.Ireland - Northern Ireland Statistics and Research Agency (NISRA): <https://www.nisra.gov.uk/publications/weekly-deaths> .

Notes: All UK countries taken from Week 13 (week ending 27th March) to Week 23 (week ending 5th June).

## 8. International/national statistical agencies should publish improved measures of excess mortality

Even if we deem the P-scores and the population-deflated statistics to be comparable across countries, underlying measurement issues of the death count strongly affect the comparability across countries. These definitional differences need to be highlighted and made transparent across country data providers and international organisations reporting excess mortality statistics. The transparent reportage of the Human Mortality Database (HMD) is exemplary in this regard.

The impact of the pandemic on deaths has been very strongly related to age and co-morbidity. The proportions of people with one, two or more co-morbidities is highly related to age. The discussion in the previous section highlighted striking differences between countries in age-related P-scores. Publication of P-scores for different age groups in a standard format should therefore be a high priority for international comparability, and HMD is a good source for such data. The evidence is that Covid-19 death rates are substantially higher for men than for women, and how this gender issue varies across countries and over time remains to be explored.

The international NUTS classification of regions provides another comparable frame for international comparisons. As regions differ in their urban/rural structure, comparing regional data can give important insights into risk factors for death rates. Moreover, as the incidence of the pandemic differs in timing and intensity, regional comparisons can throw light on the dynamics of the spread of infections. Eurostat has embarked on a major expansion of regional mortality data according to the NUTS classification, which should greatly aid research.

Another important source of variation across countries has been in the incidence of Covid-19 deaths in care homes. Countries undoubtedly differ in the proportion of older citizens resident in care homes. It would be highly desirable to develop an international standard frame to define what constitutes a care home, perhaps by the size-distribution of the number of residents. Then, comparisons of excess mortality in care homes would be possible. At present, there are limited internationally comparable data on deaths attributed to Covid-19 that occurred in care homes, see Table 4 for a UK comparison, but almost none on excess deaths of those in care homes or normally resident there.

Within countries such as the UK, there have now been several studies comparing the incidence of deaths attributed to Covid-19 by local measures of economic deprivation, occupation and ethnicity. It would highly desirable for parallel studies of excess deaths to be carried out. International comparability is harder in these dimensions given difficulties in standardising categories in measures of deprivation, occupational classification (sometimes not recorded on death certificates, but recoverable from census records) and missing data for some countries on the sensitive issue of ethnicity.

Considerable benefits can be reaped from tabulation, cross-tabulation and correlations, trying to control for common features like density by region, in proposing hypotheses. It is important to allow modellers ready access to transparent, comparable international data to a granular level to be combined with other granular data already available (e.g. on inequality) to test such hypotheses in models. Forecasting P-scores from epidemiological and other models for different scenarios on ending lockdown measures should be an important aid to formulating policy.<sup>22</sup> Granular data by location within and between countries must be produced and made accessible for research and forecasting. An example using granular Italian death registry data is Ciminelli and Garcia-Mandicó (2020).<sup>23</sup> Belloc

---

<sup>22</sup> A study which forecasts the one-year ahead mortality is Denaxas et al. (2020).

<sup>23</sup> They analyse daily death registry data for over 1000 Italian municipalities, which suggest that deaths registered as Covid capture only about half of excess deaths. They find strong evidence that locations where mass testing, contact tracing, and at-home care provision was introduced experienced lower numbers of excess deaths.



et al. (2020) caution against drawing simplistic conclusions from cross-country correlations; they too stress the need for granular, comparable data.

National statistical offices should publish weekly P-scores of excess mortalities for the constituent countries, regions and broad social groupings such as care home residents, to help understand the pandemic and inform policy.<sup>24</sup> We also argue that EuroMOMO should be mandated to produce P-scores as well as Z-scores to aid comparability across countries and be far more transparent on sources and methods. EuroMOMO's five-year graphs of Z-scores visualise the natural weekly variability, helping to interpret the confidence intervals. Similar practice should be followed for published P-scores, including at national statistical agencies.

To end on a cautionary note, excess mortality should also be examined in a longer-term perspective. Spiegelhalter (2020) argues the main impact of Covid-19 may be to shift forward the date of death by a few months for those close to death because of underlying poor health. However, as discussed in section 6, expert actuaries strongly dispute his claim. Moreover, total years of life lost, see section 3, is an alternative indicator of the pandemic's social toll. Even in the extreme and improbable case envisaged by Spiegelhalter, total years of life lost could still show a large upturn. As we saw in section 6, record excess mortality of those of working age in England, making this a particularly telling issue in comparing with other European countries.

If national statistical agencies regularly published monthly, 3-month, 6-month and 12-month moving averages, and weekly P-scores, this would greatly assist our ability to interpret the pandemic data.<sup>25</sup> Provision of timely, regularly updated and comparable granular data on excess mortality by national and international statistical agencies should be high on the agenda. It is not enough to leave this to hard-working journalists.

## References

ACN. 2020a. "[Coronavirus crisis shines spotlight on elderly care homes.](#)" Catalan News, Barcelona, 1 April 2020.

ACN. 2020b. "[Prosecutor investigating handling of Covid-19 in seven Catalan care homes.](#)" Catalan News, Barcelona, 19 April 2020.

Aron, J. and J. Muellbauer. 2020a. "[Measuring excess mortality: England is the European outlier in the Covid-19 pandemic.](#)" VOXEU, Centre for Economic Policy Research, London, 18 May, 2020.

Aron, J. and J. Muellbauer. 2020b. "[Measuring excess mortality: the case of England during the Covid-19 Pandemic.](#)" INET Oxford COVID-19 Research, Economics Department, Oxford University.

Belloc, M., P. Buonanno, F. Drago, R. Galbiati and P. Pinotti. 2020. "Cross-country correlation analysis for research on Covid-19." VOXEU, Centre for Economic Policy Research, London, 28 March 2020.

Burn-Murdoch, J., V. Romei and C. Giles. 2020. "[Global coronavirus death toll could be 60% higher than reported.](#)" Financial Times, 26 April 2020.

---

<sup>24</sup> At more granular levels, the weekly data can become noisy. Averages over longer periods are more informative.

<sup>25</sup> In the case of the ONS, the latest update of their research plans, dated 21 April, 2020, suggests only a limited agenda for investigating excess deaths: see the webpage, [Statement of upcoming analysis on deaths and coronavirus \(COVID-19\)](#).

Ciminelli, G. and S. Garcia-Mandicó. 2020. "[COVID-19 in Italy: An analysis of death registry data.](#)" VOXEU, Centre for Economic Policy Research, London, 22 April 2020.

Comas-Herrera, A. and J-L. Fernandez. 2020. "[England: Estimates of mortality of care home residents linked to the COVID-19 pandemic.](#)" Report available at LTCcovid.org, International Long-Term Care Policy Network, CPEC-LSE, 17 May 2020.

Comas-Herrera A., J. Zalakaín, C. Litwin, A. T. Hsu, E. Lemmon, D. Henderson and J-L Fernández. 2020. "[Mortality associated with COVID-19 outbreaks in care homes: early international evidence.](#)" Report available at LTCcovid.org, International Long-Term Care Policy Network, CPEC-LSE, 26 June 2020.

Denaxas, S., H. Hemingway, L. Shallcross, M. Noursadeghi, B. Williams, D. Pillay, L. Pasea, A. González-Izquierdo, C. Pagel, S. Harris, A. Torralbo, C. Langenberg, W. Wong, and A. Banerjee. 2020. "[Estimating excess 1- year mortality from COVID-19 according to underlying conditions and age in England: a rapid analysis using NHS health records in 3.8 million adults.](#)" 10.13140/RG.2.2.36151.27047.

The Economist. 2020. "[Tracking Covid-19 excess deaths across countries,](#)" The Economist, 16 April, 2020,

Edwards, M. and S. McDonald. 2020. "[The co-morbidity question.](#)" The Actuary, Institute and Faculty of Actuaries, 7th May 2020.

Farrington, C.P., N.J Andrews, A.D. Beale and M.A. Catchpole. 1996. "A statistical algorithm for the early detection of outbreaks of infectious disease." Journal of the Royal Statistical Society A 159: 547-563.

Kelly, J. 2020. "[Spiegelhalter says majority of Covid deaths would not have occurred in coming year.](#)" Financial Times, 22 May 2020.

Krelle, H., C. Barclay and C. Tallack. 2020. "[Understanding excess mortality. What is the fairest way to compare COVID-19 deaths internationally?](#)" The Health Foundation, 6 May 2020.

Loomes, G. and L. McKenzie. 1989. "The use of QALYs in health care decision making." Social Science and Medicine, Elsevier 28(4): 299-308, January.

ONS. 2020. "[Coronavirus \(COVID-19\) related deaths by ethnic group, England and Wales: 2 March 2020 to 15 May 2020.](#)" Office for National Statistics, 19 June 2020.

Rubin, D., G. Tasian and J. Huang. 2020. "[COVID-19 Outlook: Ringing the Alarm Bell for Epicenters, Waving the Caution Flag for Hotspots.](#)" Article, Children's Hospital Philadelphia Policy Lab, 24 June 2020.

Santaeulalia, I, F. Peinado, E. Sevillano and J. Mateo. 2020. "[Scandal over Covid-19 deaths at Madrid nursing homes sparks fierce political row.](#)" El País, Madrid, 10 Jun 2020.

Spiegelhalter, D. 2020b. "[Coronavirus deaths: how does Britain compare with other countries?](#)" The Guardian, 20 April 2020

Spiegelhalter, D. 2020b. "[How much 'normal' risk does Covid represent?](#)", Winton Centre for Risk and Evidence Communication, Cambridge, 21 March, 2020.

Tallack, C. D. Finch, N. Mihaylova, C. Barclay and T. Watt. 2020. "[Understanding excess deaths: variation in the impact of COVID-19 between countries, regions and localities.](#)" The Health Foundation, 4 June 2020.

Tozer, J. 2020. "[Measuring the true toll of the pandemic](#)", The Economist, 24 April, 2020.

Wu, J., A. McCann, J. Katz and E. Peltier. 2020. "[46,000 Missing Deaths: Tracking the True Toll of the Coronavirus Outbreak](#)", The New York Times, 30 April, 2020.

## Appendix 1:

Let  $x(it)$  be the weekly death count in country  $i$  in week  $t$ . It appears that EuroMOMO define<sup>26</sup> the excess death measure  $Z(it)$  as:

$$Z(it) = (x(it) - \mu(it)) / \sigma(it)$$

where  $\mu(it)$  is the predicted value from a model based on historical data up to 5 years ago for seasons of the year less affected by flu and heat waves, and incorporates some trends and seasonals, and where  $\sigma$  reflects the standard deviation of residuals, but is actually computed from a Poisson process modified for longer tails. Each country in the network estimates its own model within a broad methodology and supplies the hub with its weekly estimates.

We think a more transparent and non-parametric measure is the P-score:

$$P(it) = (x(it) - \bar{x}(it)) / \bar{x}(it)$$

where  $\bar{x}(it)$  is the average weekly death count over the previous 5 years.

There is also a parametric variant  $P^{EM}(it)$  which could be defined on EuroMOMO's data using their predicted values for 'normal' deaths as:

$$P^{EM}(it) = (x(it) - \mu(it)) / \mu(it)$$

The Poisson assumption, even modified for longer tails, is nowhere near correct for describing the stochastic process generating  $x(it)$ . The constant mean and independence over time assumptions must be wrong, as explained in Box 2 of the paper, which shows that it is implausible to assume that there are zero systematic mean shifts at all times in the Spring and Autumn. When EuroMOMO measure the standard deviation for 'normal' seasons, variation in these systematic factors as well as random noise will be present.

This suggests a better model of the death count is:

$$x(it) = \beta \times W(it) + \varepsilon(it)$$

where  $W(it)$  is a set of variables which reflect the systematic component of variations in deaths and  $\varepsilon(it)$  is white noise whose distribution can be approximated perhaps by a Poisson or binomial or normal distribution, assuming a constant variance  $\sigma(i)^2$ . Then it is clear that EuroMOMO's estimated  $\sigma$  is an amalgam of the standard deviation,  $\sigma(i)$ , and of the variation of  $W(it)$  around an average value.

Our simple measure of the excess death rate, a P-score, is then:

$$P(it) = [(\beta \times W(it) - \bar{W}(it)) + (\varepsilon(it) - \bar{\varepsilon}(it))] / (\bar{W}(it) - \bar{\varepsilon}(it))$$

using 5-year moving averages for  $\bar{W}(it)$  and  $\bar{\varepsilon}(it)$ .

When a pandemic arrives,  $W(it)$  jumps far from its historical average.  $P(it)$  does a good job in indicating the jump in  $W$ . It is easily understood by non-specialists. The empirical properties of  $P(it)$  can be investigated. One would neither claim that it is serially independent, nor that it has constant variance as that depends on the properties of  $W(it)$ . Econometricians could try to estimate  $W(it)$  with a mix of deterministic variables and state-space terms, to try better to understand the stochastic process driving the death count.

---

<sup>26</sup> See EuroMOMO webpage: "[Methods](#)".

Turning to comparisons between regions within a country, it is obvious that the smaller the population of a region, and in particular the smaller the number of normal deaths, the noisier will be the weekly death count relative to the normal expected value. In other words, the ratio:  $Z(it)/P^{EM}(it) = \sigma(it)/\mu(it)$ , will be lower in smaller regions. One can extend the argument to populous countries compared to those with smaller populations, if overall normal mortality rates are similar. In practice, movements in  $P^{EM}(it)$  will be very similar to movements in  $P(it)$ , especially in pandemics, when the jump in  $W(it)$  dominates the variation in both. As a result of averaging data over sub-populations,  $\sigma(i)/\mu(i)$  at the country vs region level could be argued to vary approximately inversely with the square root of normal deaths for the country and region. This is a result which should not depend on the precise distribution of the white noise, constant variance process for  $\varepsilon(it)$ , i.e. it should not depend on the assumption of Poisson. However, the EuroMOMO estimate of the standard deviation  $\sigma(it)$  is a composite, as noted above, of  $\sigma(i)$  and the variation in  $W(it)$  about its mean. Thus, it will vary far less with the level of normal deaths or population size than would be the case for  $\sigma(i)$  alone. This is because, on a per capita basis, the systematic factors driving  $W(it)$  under normal conditions are likely to be quite similar for different regions of a country. In a pandemic, however, the factors driving  $W(it)$  can diverge more because, for example, infections spread from different starting points and at different rates.

As our paper points out, the rankings of rates of excess deaths in the peak week for the most affected European countries according to Z are quite similar to those from P, even for countries such as Belgium and the Netherlands, which have smaller populations and hence smaller counts of normal deaths than the others. For nations or regions with much smaller counts of normal deaths, the rankings are different as the relative noisiness of weekly death counts compared to normal levels is higher. There is no simple adjustment to convert published Z-scores to P-scores without access to data on normal and actual deaths. In particular, it would be quite wrong to adjust the published Z-scores by the square root of population size of each country to make them more comparable. Comparability is best achieved using the P-scores.