medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Excess deaths in Spain during the first year of the COVID–19 pandemic outbreak from age/sex–adjusted death rates

José María Martín-Olalla*

Universidad de Sevilla. Facultad de Física. Departamento de Física de la Materia Condensada. ES41012 Sevilla. Spain

(Dated: 20210627.2316+0200)

The Eurostat records of weekly deaths disaggregated by AGEGRP5 (five–year age groups) and sex group in Spain have been analyzed to build the age/sex–specific weekly death rates and the age/sex–adjusted weekly death rate and to infer the predicted death rate, the excess death rate and the excess deaths during the first year (52–week) of the pandemic. Adjusted rates were computed extending the last available population structure back in time to assess the death rates that would have been observed if the present population had died at the age/sex–specific death rates observed in the past.
Age/sex–adjusted, 52–week death rate had not been as high as the observed rate 10.67×10^{-3} in the past 13 years.
Poisson regression predicted a death rate of 8.81×10^{-3} which makes an excess of 1.86×10^{-3} ($P-\text{score} = 21.2\,\%$ and $z-\text{score} = 11.9$) with an unbiased standard deviation of the residuals equal to 156×10^{-6}. This translates into 88 242 excess deaths (46 695 males and 41 532 females) with an unbiased standard deviation of the residuals equal to 7396 deaths.
COVID–19 deaths (73 520) accounts for 83 % of the total excess. Taking into account the 9772 COVID–19 suspected deaths that occurred in nursing homes and care facilities during the spring of 2020 it is only 4948 (5.6 % of excess deaths) that remains unattributed.
The infection rate during the first year of the pandemic is estimated in 17 % of population after comparing the ENE–COVID seroprevalence, the excess deaths at the end of the spring 2020 and the excess deaths at the end of the first year of the pandemic.

Keywords: mortality excess; z-score; p-score; Eurostat; INE; ONS; Instituto de Salud Carlos III; life expectancy; demography; standard population

CONTENTS

1. Introduction 2

2. Methods 2
   2.1. Data sets 2
   2.2. Statiscial analysis 3

3. Results 4

4. Discussion 7
   4.1. Regression interval 7
   4.2. Infection rate 7
   4.3. Exponential distribution of age–specific excess death rates 9

5. Conclusions 9

6. Conflict of interest statement 10

7. Data availability 10

References 10

Acknowledgments 10

* olalla@us.es; Twitter: @MartinOlalla_JM;

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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1. INTRODUCTION

The illness designated covid–19 caused by the severe acute respiratory syndrome coronavirus 2 (SARS–CoV–2) caught worldwide attention since its identification in late 2019(Zhu *et al.*, 2020). Spain is one of the European countries most impacted by the disease during the spring of 2020(García-Basteiro *et al.*, 2020). Confirmed covid–19 cases climbed up to several hundred thousands —some few thousands per one million population— and confirmed covid–19 deaths to some 70 000 or 1500 deaths per one million population. A myriad of societal measures ranging from a lock–down to a mandatory use of masks were enforced since the outbreak of the pandemic.

Total excess deaths or crude excess deaths or all–cause excess deaths is a key quantity to understand the impact of a pandemic(Aaron *et al.*, 2020). Many efforts and resources are dedicated to provided fast and accurate numbers of this societal quantity (see Islam *et al.* (2021) (OECD countries) and Karlinsky and Kobak (2021) (world wide)). While the observed deaths during a window of time can be accurately known from official records in modern societies the determination of a predicted or reference value from which excesses are computed is a matter of retrospective statistical analysis. Mortality is characteristically different for males and females and for age groups. Therefore the understanding of age–specific, sex–specific death rates is of the outmost importance. Their trends highlight societal issues: either improvement of public health care which leads to smaller specific rates or, events like wars, outbreaks or economic turmoils which lead to greater specific rates. Age/sex–specific mortality rates are the basic quantities to assess statistical measures like the life expectancy which are also object of interest in this context(Trias-Llimos *et al.*, 2020).

In this paper I analyze Spanish official records of weekly mortality and population dated back from the year 2000. The manuscript is devoted the accumuated mortality in the first 52 weeks (364 days or one year) after the pandemic onset. Total excesses will be derived for averaged and adjusted age/sex specific death rates for the last available structure of the Spanish population. Infection rate during the first year of the pandemic will be derived from SARS–CoV–2 seroprevalence.

2. METHODS

2.1. Data sets

The file `demo_r_mweek3.tsv` is released by Eurostat with weekly deaths provided National Statistics Office, the Instituto Nacional de Estadística (INE) for Spain. The file lists weekly deaths in Spain since the year 2000 disaggregated by statistical region (all levels), sex (males, females, both), and age group (total, five–year age groups until the group 90 years old and above). Weeks can be traced by an ordinal in the record or, following the ISO week date, by the calendar week. Notice that the week calendar consists 52–week (364–day) years and 53–week (371–day) years. Calendar weeks are numbered in ordinals with week 01 always containing the first Thursday of the Gregorian year. The file was downloaded on Jun 18th, 2021 and includes the last major revision by INE of the 2020 weekly deaths, released on Jun 17th, 2021. However, it should be noted that 2020 and 2021 weekly deaths are still flagged as "provisional" in the file.

Observed deaths scale to population. The table `31304` released by INE lists resident population values in Spain for January 1st and July 1st since the year 1971 until July 1st, 2020. The data set is disaggregated by statistical region (NUTS3), sex and age until the group 85 years old and above.

Two steps were taken in order to harmonized both sets. First, weekly deaths for the age group 85–89 and the age group 90 or elder were summed up to yield the age group 85 or elder. Second, population numbers for January 1st and July 1st were assigned to their corresponding weeks. Population numbers for every week since 2000 were computed by interpolation. The last available numbers (July 1st, 2020) were extended into 2021 for the lack of a better quantity.

With these, 52–week, all–cause deaths were computed. This quantity accumulates the deaths observed on a given week and on the preceding 51 weeks, for a total of 364 days or one year in the context of this manuscript. The 52–week interval for week n starts in the week $n+1$ of a 52–week preceding year or in the week $n+2$ of a 53–week preceding year, unless $n = 52$ or $n = 53$ in which case the interval starts at week 01 or 02 of the same year.

The 52–week accumulated values were then scaled by the population at the given week to get the 52–week accumulated death rate.

In addition to that confirmed COVID–19 cases and deaths are numbered in Spain by the Centro Nacional de Epidemiología (CNE), equivalent to the CDC after reports from local authorities. The dataset is disaggregated by sex, NUTS3 region and 10–year age groups on a daily basis. These records were harmonized to read weekly counts.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.2. Statiscial analysis

The pre–pandemic period in Spain ended on Sunday March 1st, 2020 at the end of the 1000th week of the 21st century. One year (364 days) later it came Sunday February 28th 2021 at the end of week 08–2021 or the 1052nd week in the 21st century.

To account for the seasonality of human mortality 52–week accumulated death rates at week 08 on every year since 2001 were computed from the data set. A bivariate generalized Poisson regression with a linear rate served to infer the predicted 52–week death rate for the week 08–2021. The predictors of the regression were the ordinal week numbers. The responses were the observed 52–week death rates. The regression was restricted to the eight years prior the pandemic onset: from 2013 to 2020. The unbiased standard deviation of the residuals was recorded.

The bivariate regression was performed for every age group and every sex.

The age–specific, sex–specific analysis was summarized in two complementary ways.

First the weight–averaged sex–specific sum of age–specific values for 2021 were computed as:

$$P_i = \sum_j p_{ij} \frac{n_{ij}}{N_i}. \tag{1}$$

where p_{ij} are age–specific, sex–specific quantities for sex i and age group j; n_{ij} are the last available population number for the age group and sex and N_i is the population number for sex i. The ratio n_{ij}/N_i weights p_{ij} according to the shares of population of the age group within the sex group. In Eq (1) p_{ij} is, as an example, the age–specific, sex–specific death rate predicted value for the year 2021.

Likewise age/sex weight–averaged values can be computed as:

$$P = \sum_{i,j} p_{ij} \frac{n_{ij}}{N}. \tag{2}$$

where N is the last available population number in Spain. The age/sex–specific quantity is weight–averaged according to the shares of a age and sex group in population pyramid.

As an alternative the age–adjusted, sex–specific weekly death rate defined as:

$$D_i = \sum_j d_{ij} \frac{n_{ij}}{N_i}. \tag{3}$$

where the age–specific, sex–specific weekly death rates d_{ij} are weighted accordingly. Likewise the age/sex–adjusted death rate was computed as

$$D = \sum_{ij} d_{ij} \frac{n_{ij}}{N}. \tag{4}$$

Notice that the rate d_{ij} depends on the week ordinal, therefore they refer to different years across the catalogue. However the population numbers n_{ij} and N are those given by the last available records of population. Therefore adjusted weekly death rates propagate the population pyramid back and forth in time so that D or D_i only track the changes in the specific mortality. Compared to that, the unspecific death rates —total deaths scaled by total population— track changes in specific mortality and in population composition.

Usually in age–harmonised analysis the weights are given by a standard, invariant population pyramid, often distant in time. As an example the European Standard Population(Eurostat Task Force, 2013) is widely used to facilitate the comparison between countries, see for instance Islam *et al.* (2021) . However the purpose of this harmoinsation in this manuscript is the comparison of mortality data throughout the 21st century. Therefore I opted to extend the population structure in 2021 back in time to compute the death rate that should have been observed in Spain during 2021 if the age/sex–specific death rates for 2021 had been those observed in the preceding years. In this way excesses deduced from this age/sex–adjusted rates are meaningful in 2021.

Weight–averaged excesses and age/sex–adjusted excesses differ conceptually an order of operations. For the weight–averaged excesses, the predicted values and excesses of the observed age/sex–specific death rates for the year 2021 are first computed. Then, they are are weight–averaged (see Eq. (1) and Eq. (2)) by the 2021 population pyramid. On the contrary, for age–adjusted excesses the order is reversed: first the weight–average of observed death rates (see Eq. (3) and Eq. (4)) are computed for the 2021 population pyramid; then the modelling yields adjusted predictions and excesses for the year 2021.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

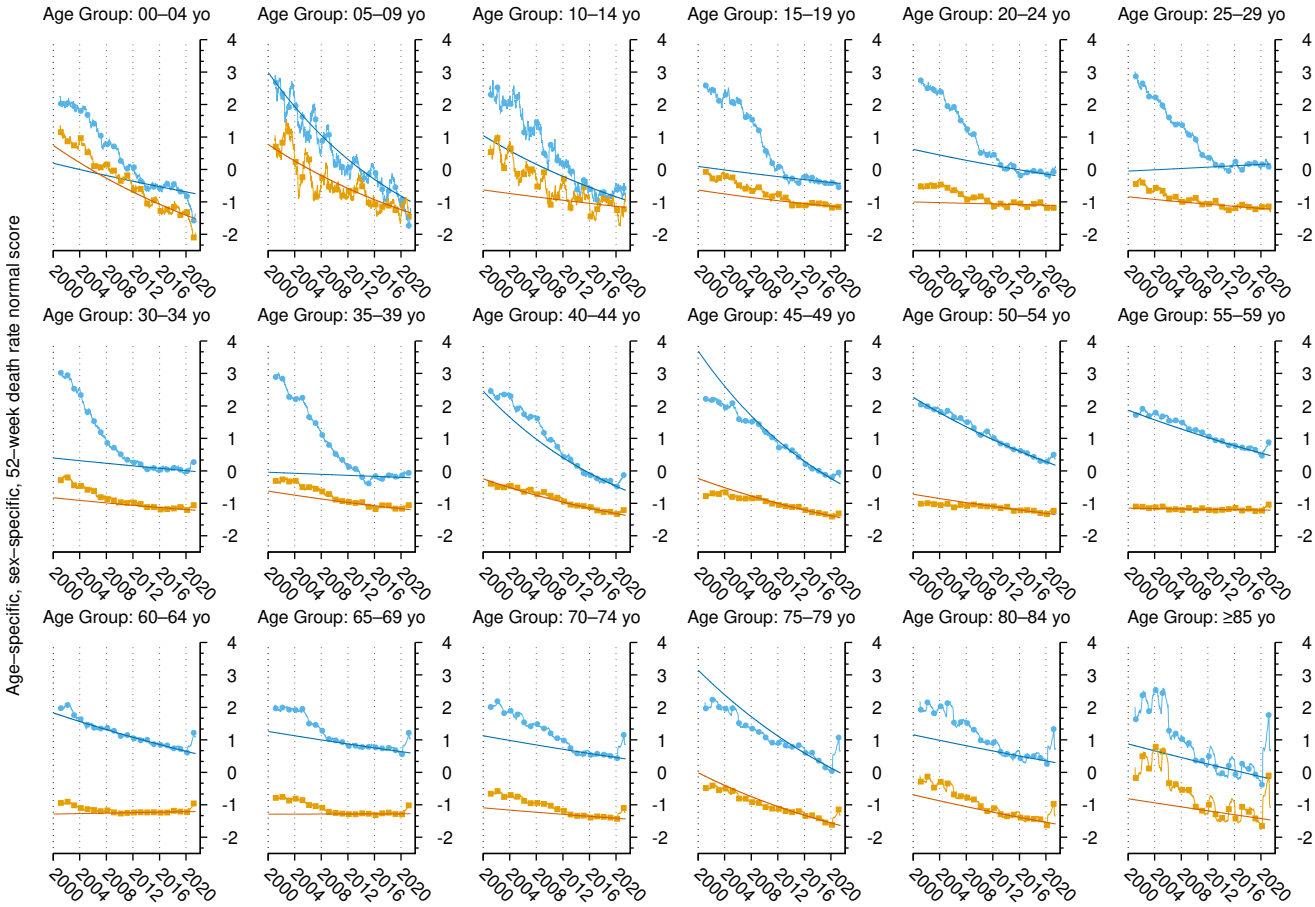

FIG. 1 Age/sex–specific 52–week death rate normal scores. Data points show values at week 08 on every year since 2001 (blue circles, males; orange squares, females). They are joined by weekly values in the remaining weeks of the year (thin light line). Solid dark line displays the results of the Poisson regression for the eight years prior to the pandemic. The normal scores were computed as $(x - E)/s$ where E is the sample average and s the sample standard deviation of the age–specific, sex–unspecific dataset.

3. RESULTS

Figure 1 shows the age–specific, sex–specific, all–cause, 52–week death rate in Spain since the year 2001. Blueish lines and circles refer to the male distribution; orange lines and squares refer to the female distribution. Notice that sex–unspecific sample average and sample standard deviation were removed from the dataset so that the vertical axes span equally through different age groups even though observed death rates are vividly different.

Death rates in Figure 1 show mostly a decreasing trend with increasing time as a result of enhancement in societal and public health issues. For an age group male death rate is larger than female death rate and decreased faster in the preceding years. Some age groups, notably the younger age groups, but also some older groups like the 65–69 age group, show a stalling in the past few years.

Pondering every aspect of this distribution I choose to infer the expected value of death rate in week 08–2021 from a generalized Poisson regression that included the eight years prior to the pandemic. Regression results are shown in Figure 1 by a darker line generally deviating from data points in the first records of the dataset.

Generally speaking the observed death rate in 2021 shows an excursion for male population elder than 30 and female population elder than 40.

Table 1 lists the results of the regression. Each line shows an age–specific, sex–specific analysis which includes the last available population number. First the predicted rate and the excess death rate for the year 2021 are listed.

The three scores show the magnitude of the excursion: the $P-$score, the ratio of the excess to the predicted value, given in percent. Then the $z-$score or normal score, the ratio of the excess to the unbiased sample standard deviation of the residuals. Finally the number of years since we observed a death rate larger than the one observed in 2021.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		52–week death rates		52–week scores			52–week counts		
Stratum	Population	2021 all–cause predicted rate	2021 all–cause excess	$P-$score /%	$z-$score	Years back	Excess deaths	COVID–19 undercount	COVID–19 CFR/%
All sexes									
All ages	47 351 567	8.81×10^{-3}	1.86×10^{-3}	21.2	11.9	13	88 242	1.200	2.84
Males									
All ages	23 206 752	9.07×10^{-3}	2.01×10^{-3}	22.2	16.0	12	46 695	1.157	3.17
00–04	1 011 482	632×10^{-6}	-153×10^{-6}	−24.3	−7.0	1	−155		
05–09	1 192 142	66.7×10^{-6}	-20.6×10^{-6}	−30.8	−4.1	1	−25		
10–14	1 309 815	77.4×10^{-6}	11.1×10^{-6}	14.4	1.6	6	15		
15–19	1 250 706	203×10^{-6}	-16.6×10^{-6}	−8.2	−1.6	1	−21		
20–24	1 216 918	316×10^{-6}	19.3×10^{-6}	6.1	1.2	2	24		
25–29	1 286 142	424×10^{-6}	-15.5×10^{-6}	−3.7	−0.8	1	−20	0.077	
30–34	1 380 441	513×10^{-6}	67.3×10^{-6}	13.1	5.7	10	93		
35–39	1 622 728	697×10^{-6}	48.9×10^{-6}	7.0	1.1	8	79	1.525	0.09
40–44	1 971 463	942×10^{-6}	209×10^{-6}	22.2	6.3	6	412		
45–49	1 969 083	1.75×10^{-3}	217×10^{-6}	12.4	6.2	3	428	1.799	0.34
50–54	1 838 221	3.43×10^{-3}	334×10^{-6}	9.7	10.6	4	614		
55–59	1 672 762	5.72×10^{-3}	683×10^{-6}	11.9	9.1	7	1143	1.006	0.79
60–64	1 440 180	8.95×10^{-3}	1.70×10^{-3}	19.0	17.9	12	2444		
65–69	1 173 099	13.9×10^{-3}	2.58×10^{-3}	18.6	13.9	13	3022	1.139	3.59
70–74	1 025 639	21.1×10^{-3}	4.65×10^{-3}	22.1	15.2	11	4768		
75–79	785 232	32.3×10^{-3}	10.2×10^{-3}	32	9.7	11	8018	1.223	12.78
80–84	525 651	63.8×10^{-3}	13.9×10^{-3}	21.7	15.9	13	7287		
≥ 85	535 051	142×10^{-3}	34.4×10^{-3}	24.3	10.6	16	18 404	1.131	29.85
Females									
All ages	24 144 815	8.56×10^{-3}	1.72×10^{-3}	20.1	9.1	14	41 532	1.252	2.54
00–04	955 779	492×10^{-6}	-109×10^{-6}	−22.2	−4.0	1	−104		
05–09	1 120 512	57.1×10^{-6}	-3.54×10^{-6}	−6.2	−0.8	1	−4		
10–14	1 227 471	69.6×10^{-6}	-2.01×10^{-6}	−2.9	−0.2	2	−2		
15–19	1 169 576	104×10^{-6}	-23.4×10^{-9}	0.0	0.0	2	0		
20–24	1 156 904	144×10^{-6}	-13.9×10^{-6}	−9.6	−1.0	2	−16		
25–29	1 252 677	167×10^{-6}	13.9×10^{-6}	8.3	0.9	4	17	0.037	
30–34	1 382 211	228×10^{-6}	36.0×10^{-6}	15.8	3.3	8	50		
35–39	1 638 773	368×10^{-6}	44.4×10^{-6}	12.1	2.2	7	73	1.612	0.05
40–44	1 948 860	577×10^{-6}	69.7×10^{-6}	12.1	5.9	5	136		
45–49	1 936 183	1.02×10^{-3}	81.9×10^{-6}	8.0	4.6	3	159	1.187	0.11
50–54	1 845 222	1.76×10^{-3}	118×10^{-6}	6.7	3.1	3	217		
55–59	1 725 209	2.81×10^{-3}	300×10^{-6}	10.7	4.7	>	518	0.952	0.30
60–64	1 527 042	4.10×10^{-3}	669×10^{-6}	16.3	13.9	19	1021		
65–69	1 287 127	5.97×10^{-3}	1.09×10^{-3}	18.3	10.7	15	1406	1.292	1.61
70–74	1 192 560	9.28×10^{-3}	2.12×10^{-3}	22.8	16.4	11	2523		
75–79	983 312	16.6×10^{-3}	4.62×10^{-3}	28	8.7	9	4543	1.341	6.78
80–84	766 577	38.2×10^{-3}	8.24×10^{-3}	21.5	11.7	12	6314		
≥ 85	1 028 823	119×10^{-3}	23.9×10^{-3}	20.0	7.1	16	24 567	1.241	20.66

TABLE 1 Age/sex–specific, all–cause mortality data in Spain for the first year of the pandemic. First column lists the last available data of Spanish resident population. Then the 52–week death rates: first, the predicted death rate for 2021, then the excess death rate in the year 2021. Thereafter three scores: the $P-$score which is the ratio of the excess to the predicted value, given as a percent; then the $z-$score, which is the ratio of the excess to the unbiased sample standard deviation of the residuals; last, the number of years since the observed death rate was larger than the one observed in the year 2021 with $>$ underscoring a period longer than the record size of 20 years. The last three columns list the 52–week total excess deaths, the ratio of 52–week excess deaths to COVID–19 52–week deaths (the undercount) and the ratio of 52–week excess deaths to 52–week COVID–19and cases, the case fatality risk CFR, expressed as a percentage. All ages refer to the adjusted rates (see Eq. (3) and Eq. (4)).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

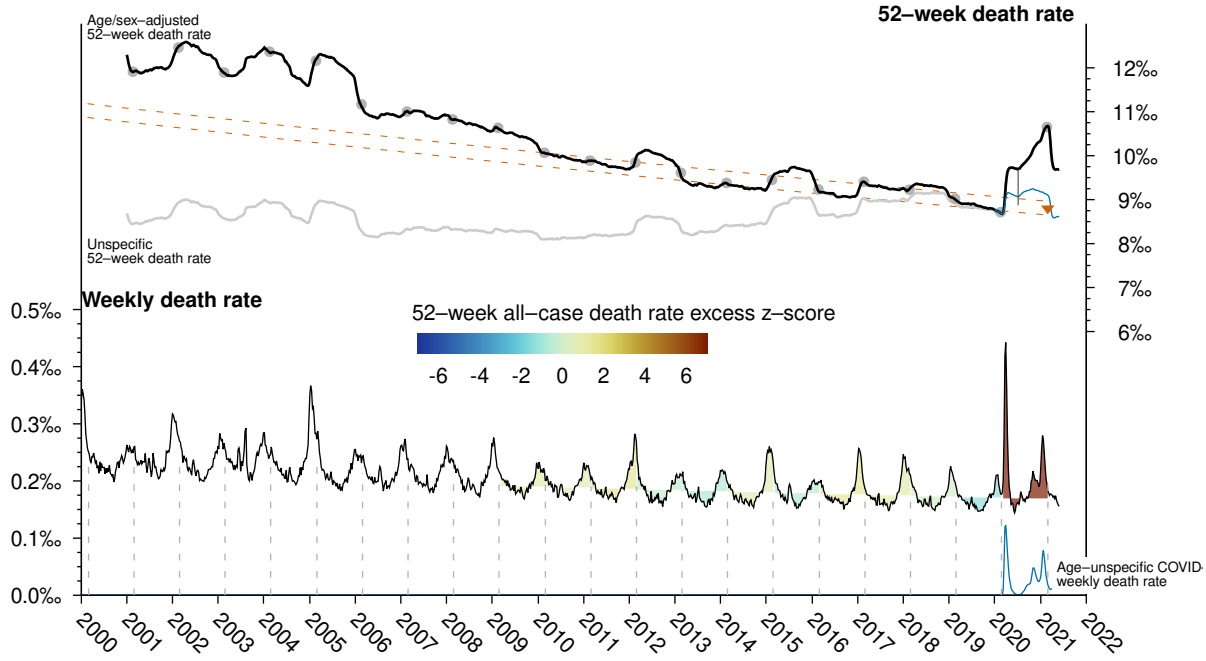

FIG. 2 Bottom: all–cause, age/sex–adjusted weekly per capita death rate and COVID–19 weekly death rate (blue) in Spain since 2000. Top: cumulative 52–week, all–cause, age/sex–adjusted death rate (dark orange) and cumulative 52–week age/sex–adjusted, non COVID–19 death rate (blue). Dots display 52–week values at the week 08 of every year in the 21st century. The down triangle shows the predicted 2021 death rate. The dashed lines around the 52–week death rate are centered at death rate predicted by the Poisson regression and is one unbiased sample standard deviation of the residuals prior to the pandemic onset in width. The vertical thin line notes the excess death rate at the end of week 27–2020. For the sake of comparison the unspecific, 52–week is shown in light gray. The 52–week cumulative death rate predicted by the Poisson model is converted into a weekly average to shade the weekly death rate accordingly. Net shaded area represents the 52–week accumulated death rate. Color depth for shaded areas is given by the $z-$score.

This score is not related to the regression.

The last group of columns shows the excess deaths, the COVID–19 (death) undercount defined as the ratio of excess deaths to COVID–19 deaths and the COVID–19 case fatality risk (CFR) defined as the ratio of excess deaths to COVID–19 cases, expressed as a percentage. Notice that COVID–19 cases were undercount until June 2020 notably in younger age groups, more prone to a mild evolution of the disease. In this ratios the excess death rate is implicitly linked to the disease. While this is questionable from an epidemiological point of view since other cause of deaths may have changed during the pandemic, the point is that excess death rate is a valuable metric to assess the global, societal impact of the pandemic and of the disease to which it is associated.

In addition to age–specific analysis, Table 1 also shows the age/sex–adjusted analysis in the last couple of rows. The averaged and adjusted predicted rates only differed in one count per thousand.

In order to provide a better visualization of the regression method Figure 2 shows the age/sex–adjusted, all–cause weekly death rate (bottom) and the age/sex–adjusted all–cause 52–week cumulative death rate (top, dark orange). For the sake of comparison the unspecific (all–age, all–sex) 52–week death rate is shown in light gray. Notice that the unspecific death rate and the adjusted death rate match to each other at the end of the record as per construction. When they differ the adjusted death rate displays the death rate that should have been observed in the year 2021 had the age–specific, sex–specific death rates been those of the previous weeks/years. The age–unspecific, sex–unspecific death rate displays the actual observed death rate.

Circles in figure 2 display the observed values in week 08 for the years. Dashed lines display the predicted value of the Poisson regression plus and minus one unbiased sample standard deviation of the residuals ($s = 156 \times 10^{-6}$ or 7396 deaths if population numbers are used).

The blue line displays the age/sex–adjusted, non COVID–19 death rate which is obtained after removing the COVID–19 weekly deaths reported by the CNE, shown in the bottom of the figure. The triangle signals the expected, age/sex–adjusted, all–cause death rate for week 08–2021 which is 8.81×10^{-3}, compared to the observed 10.67×10^{-3} rate.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Weekly deaths are shaded by the expected death rate obtained by regression. Therefore the net shaded areas display the net excess death rate. Color depth is given by the normal score of the excess.

4. DISCUSSION

4.1. Regression interval

The goal of computing excess deaths is the goal of determining the predicted or reference value from which excess is computed. One comes with the other. To accomplish that, one requires on the first hand a record of previous observations to infer a prediction; the longer the record, the merrier. When it became evident that mortality record would be under close public scrutiny in 2020 Eurostat requested to national statistical offices to provide, if possible, weekly records of mortality starting in the year 2000. On the other hand causality tends to challenge this idea: the older the observation the less impact may have in the future prediction. As an example Karlinsky and Kobak (2021) considered data from the year 2015 onward in their world mortality dataset while Islam *et al.* (2021) considered data from the year 2016 onward on their analysis.

As observed in Figure 2 the 52–week age/sex unspecific death rate (light gray line) differs the 52–week age/sex adjusted death rate (dark orange line). The latter removes the contributios due to population ageing. Therefore age/sex–adjusted, 52–week is an appropriate quantity for curating the prediction if more than four years are taken into account. The adjusted death rate retains the properties of age–specific analyses in Figure 1: a faster decrease in the earlier years of the record, and a stall in the last ten years or so.

Nonetheless the most radical choice for the prediction value is the observation in the previous year. That suffices for many practical purposes. In Spain the 2020 age/sex–adjusted death rate was 8.70×10^{-3}. In 2021 the observed value was 10.67×10^{-3} which makes an excess of 93 500 deaths. If two years come into play one may consider the averages of the last two observations 8.85×10^{-3}. When three years are available a bivariate Poisson regression can be an option.

Figure 3 shows the predicted age/sex–adjusted death rate for 2021 (left axis) and the corresponding excess death (right axis) if the Poisson regression started at the year noted in the horizontal axis. The horizontal dashed line signals the result listed in Tab. 1 which included eight years, starting in 2013 and ending in 2020. Vertical error bars signal the unbiased sample standard deviation of the residuals which in the region of interest —when the record stars in the year 2013— is 156×10^{-6} or 7396 deaths if the population is considered. The standard deviation does not speak about the quality of the regression but points out the natural variability of the response over the preceding years. When the last ten/eight years are used in the regression the prediction climbs close to 90 000 excess deaths.

Overextending the range of observations back in time pulls the prediction down and increases the excess deaths. This comes from faster rates of age–specific, sex–specific mortality in the beginning of the 21st century which would predict smaller death rates in 2021. Shrinking the range of observations below five years leads to the same result since the last four year showed a decrease in the age/sex–adjusted death rate.

4.2. Infection rate

A important question to ellucidate is the 52–week infection rate that produced the excess death rate. First column data in Table 2 lists the case rate —the ratio of the accumulated 52–week number of COVID–19 cases to population—. This metric is often challenged by the fact that due to a myriad of societal and health issues a fraction of cases often lies below the radar of epidemiological studies. In this pandemic this was specifically significant during the spring of the year 2020 due to limited testing capabilities.

To overcome this issue Pastor-Barriuso *et al.* (2020) conducted during the spring of the year 2020 a seroepidemiological study (ENE–COVID) within comunity dwelling population which determined the SARS–CoV–2 seroprevalence or infection rate I_s in Spain at the end of the spring of the year 2020. From this quantity it is possible to build an augmented case rate which is composed of the seroprevalence (until week 27–2020) and the accumulated number of case since. This metric is listed in the second data column of Table 2 and mostly doubles the case rate computed from the CNE database.

A third metric can be built up if the infection fatality risk IFR is assumed to be invariant in time. The IFR is the ratio of (excess) deaths to infected population. At the end of week 27–2020 the infected population can be determined by the ENE–COVID seroprevalence(Pastor-Barriuso *et al.*, 2020) and the excess deaths can be determined from the observed 52–week deaths and the deaths predicted by the regression for that week (see vertical thin line

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

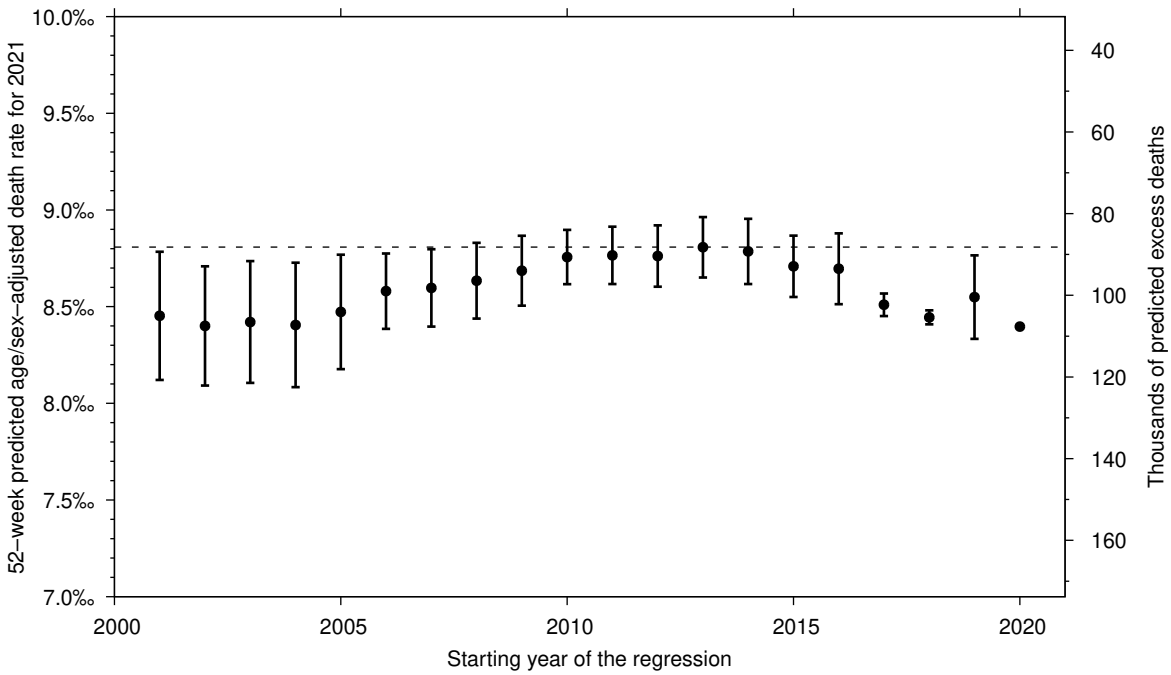

FIG. 3 Predicted age/sex–adjusted death rate in Spain for the year 2021 after the bivariate Poisson regression —predictor is week ordinal, response is age/sex–adjusted 52–week death rate on calendar week 08— if the starting year of the regression is that of the horizontal axis, and the regression extends until the year 2020. For the year 2020 the observed death rate is used; for the year 2019 the average of the last two observed death rates is displayed. Vertical error bars signal the unbiased sample standard deviation of the residuals. The horizontal dashed line displays the value listed in Tab. 1.

Stratum	Case rate/%	Augmented case rate/%	27–2020 excess rate	08–2021 excess rate	Infection rate/%
Overall	4.9	10	825×10^{-6}	1.86×10^{-3}	17
Males					
all ages	6.3	11	828×10^{-6}	2.01×10^{-3}	17
30–39	6.6	11	37.6×10^{-6}	57.4×10^{-6}	7
40–49	6.3	11	99.8×10^{-6}	213×10^{-6}	11
50–59	6.3	11	158×10^{-6}	500×10^{-6}	17
60–69	5.8	10	618×10^{-6}	2.09×10^{-3}	17
70–79	5.5	9	3.21×10^{-3}	7.06×10^{-3}	11
≥80	8.1	11	10.1×10^{-3}	24.2×10^{-3}	27
Females					
all ages	6.8	11	821×10^{-6}	1.72×10^{-3}	18
30–39	7.6	12	16.4×10^{-6}	40.6×10^{-6}	13
40–49	7.1	12	28.0×10^{-6}	75.8×10^{-6}	14
50–59	6.9	11	99.1×10^{-6}	206×10^{-6}	11
60–69	5.4	10	288×10^{-6}	862×10^{-6}	15
70–79	4.8	9	1.37×10^{-3}	3.25×10^{-3}	12
≥80	8.3	11	8.62×10^{-3}	17.2×10^{-3}	22

TABLE 2 COVID–19 case and infection rates in Spain during the first year of the pandemic. The first data column displays the accumulated case rate as obtained from the CNE records. The second data column lists the SARS–CoV–2 seroprevalence(Pastor-Barriuso *et al.*, 2020) until the end of the spring 2020 plus the accumulated number of COVID–cases thereafter. Third and fourth data columns list the excess death rate at week 27–2020 and at week 08–2021 from which the infection rate (fifth data column) is computed after removing 19 681 COVID–19 related deaths that ocurred in institutionalized population.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

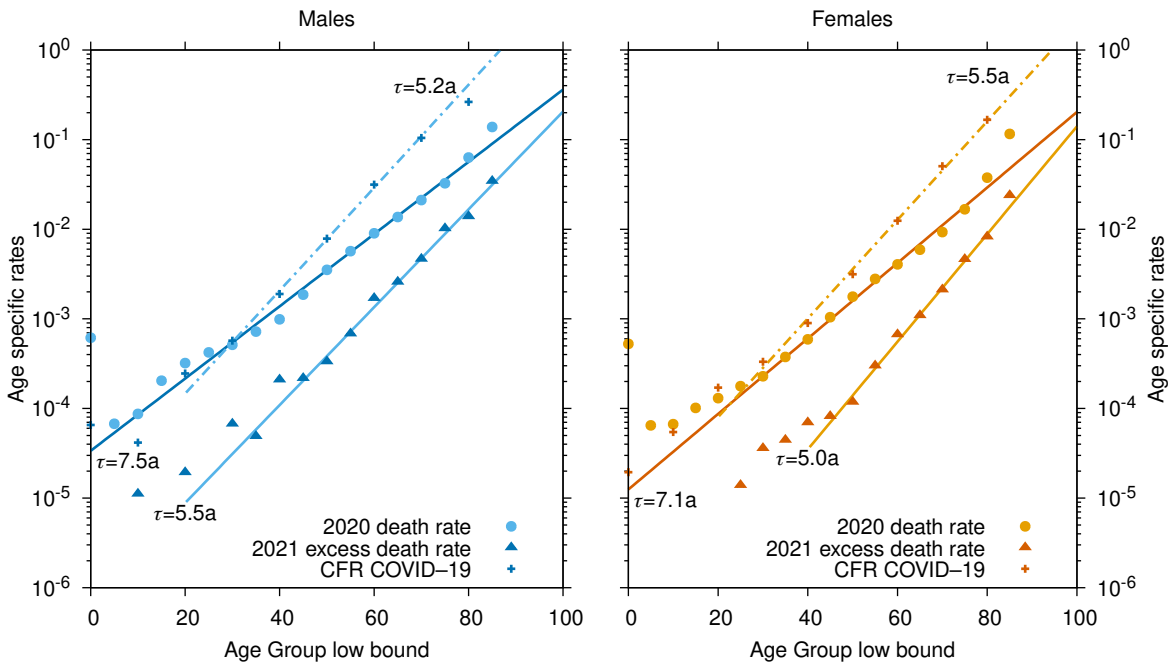

FIG. 4 The Gompertz exponential law for males (left) and females (right) in the 2020 death rates (circles), the 2021 excess death rates (triangles) and COVID–19 case fatality risk (CFR, crosses). Lines show the exponential fitting with numbers showing the increase in age years to double the rate.

in Figure 2). The infection rate on week 08–2021 is then $I_{08} = S_{27} \times E_{08}/E_{27}$. As noted by Pastor-Barriuso *et al.* (2020) (supplementary method) a number of 19 681 COVID–19 related deaths that occurred within institutionalized population must be removed from E_{27} and E_{08} to assess the excess death rates within comunity dwelling population, which the seroprevalence refers to. The infection rate is listed in the sixth data column in Table 2. In many cases this metric climbs to three times the 52–week case rate reported from the CNE.

4.3. Exponential distribution of age–specific excess death rates

Figure 4 shows age/sex–specific death rates versus age (left for males, right for females) in semilogarithmic scales. Metrics included in the figure are: (1) the pre–pandemic 2020 death rate (circles), with a doubling age equal to $\tau = 7.5$ a (males) and $\tau = 7.1$ a (females); (2) the post–pandemic 2021 excess death rate (triangles), with faster doubling ages: $\tau = 5.5$ a (males) and $\tau = 5.0$ a; (3) the case fatality ratio (CFR) for COVID–19 (crosses) defined as the ratio of COVID–19 deaths over COVID–19 cases (see Table 1), its doubling age is $\tau = 5.2$ a (males) and $\tau = 5.5$ a (females).

The fact that doubling ages are similar for the 2021 excess death rate and the COVID–19 CFR is an evidence of the linkage between the disease and the excess death rate, with little variability across sexes.

Sex–specific, age–adjusted regressions yielded excess death rates equal to 1.72×10^{-3} (females) and 2.01×10^{-3} (males) or 41 532 (females) and 46 695 (males). Sex–specific excess death rates differ by a factor 1.17 while the sex–specific predicted values only differ by a factor 1.06. This is an evidence of a larger impact of the pandemic in the male population.

5. CONCLUSIONS

A grand total of 505 490 people died in Spain from March 2nd, 2020 to February 28th, 2021 the first year (52 weeks, 364 days) after the pandemic outbreak. This translates into a death rate equal to 10.67×10^{-3}.

Age/sex–adjusted, 52–week death were built from age–specific (5–year bin) and sex–specific death rates.

Age/sex–adjusted, 52–week death rates had been smaller than 10.67×10^{-3} for the past continued 13 years.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A bivariate generalized Poisson regression for the last eight years of age/sex–adjusted, 52–week death rates predicted 8.81×10^{-3} in the period of time with an unbiased standard deviation of the residuals equal to 156×10^{-6} or 7396 deaths.

This results in an excess death rate equal to 1.86×10^{-3} which translates into 88 242 excess death of which 73 520 deaths (83 %) were previously identified as COVID–19 deaths and the remaining 14 722 (16.7 %) were undercounts. Among them a number of 9772 COVID–19 related deaths (11.1 % of excess deaths) occurred in care facilities and nursing homes during the spring of the year 2020 (see Pastor-Barriuso *et al.* (2020)). Therefore it is only 4948 deaths (5.6 % of excess deaths, a $z-$score of 0.7) that remain unattributed.

It is estimated that 17 % of the Spanish population was infected by the SARS–CoV–2 during the first year of the pandemic.

6. CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

7. DATA AVAILABILITY

Eurostat disseminates the file `demo_r_mweek3.tsv` at the url `https://ec.europa.eu/eurostat/estat-navtree-portlet-BulkDownloadListing?sort=1&file=data%2Fdemo_r_mweek3.tsv.gz` as of June 2021.

INE disseminates the resident population in Spain at the url `https://www.ine.es/jaxiT3/Tabla.htm?t=31304`

CNE disseminates the official records of the COVID–19 disease at `https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos`.

REFERENCES

Aaron, Janine, John Mullbauer, Charlie Giattino, and Hannah Ritchie (2020), "A pandemic primer on excess mortality statistics and their comparability across countries," .

Eurostat Task Force, (2013), *2013 edition Revision of the European Standard Population Report of Eurostat's task force*, Tech. Rep.

García-Basteiro, Alberto, Carlos Alvarez-Dardet, Alex Arenas, Rafael Bengoa, Carme Borrell, Margarita Del Val, Manuel Franco, Montse Gea-Sánchez, Juan Jesús Gestal Otero, Beatriz González López Valcárcel, Ildefonso Hernández, Joan Carles March, José M Martin-Moreno, Clara Menéndez, Sergio Minué, Carles Muntaner, Miquel Porta, Daniel Prieto-Alhambra, Carmen Vives-Cases, and Helena Legido-Quigley (2020), "The need for an independent evaluation of the COVID-19 response in Spain," The Lancet **0** (0), 10.1016/S0140-6736(20)31713-X.

Islam, Nazrul, Vladimir M Shkolnikov, Rolando J Acosta, Ilya Klimkin, Ichiro Kawachi, Rafael A Irizarry, Gianfranco Alicandro, Kamlesh Khunti, Tom Yates, Dmitri A Jdanov, Martin White, Sarah Lewington, and Ben Lacey (2021), "Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries," BJM **373**, n1137.

Karlinsky, Ariel, and Dmitry Kobak (2021), "The World Mortality Dataset: Tracking excess mortality across countries during the COVID-19 pandemic." medRxiv : the preprint server for health sciences , 2021.01.27.21250604.

Pastor-Barriuso, Roberto, Beatriz Pérez-Gómez, Miguel A Hernán, Mayte Pérez-Olmeda, Raquel Yotti, Jesús Oteo-Iglesias, Jose L Sanmartín, Inmaculada León-Gómez, Aurora Fernández-García, Pablo Fernández-Navarro, Israel Cruz, Mariano Martín, Concepción Delgado-Sanz, Nerea Fernández de Larrea, Jose León Paniagua, Juan F Muñoz-Montalvo, Faustino Blanco, Amparo Larrauri, and Marina Pollán (2020), "Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study," BMJ **371**, m4509.

Trias-Llimos, Sergi, Tim Riffe, and Usama Bilal (2020), "Monitoring life expectancy levels during the COVID-19 pandemic: Example of the unequal impact of the first wave on Spanish regions," PLoS ONE **15** (11 November), 10.1371/journal.pone.0241952.

Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, Peihua Niu, Faxian Zhan, Xuejun Ma, Dayan Wang, Wenbo Xu, Guizhen Wu, George F. Gao, and Wenjie Tan (2020), "A novel coronavirus from patients with pneumonia in China, 2019," New England Journal of Medicine **382** (8), 727–733.

ACKNOWLEDGMENTS

In the loving memory of those who died during the first year of the pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20159707; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

JMMO hearthfully thanks Eurostat, the Instituto Nacional de Estadística and the Instituto de Salud Carlos III for making publicly available the datasets. Dr. JM López-Cepero is thanked for pointing Eurostat bulk download url https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&dir=data

This work was performed using free software running on `xubuntu 20.04.1LTS`. Data bases have been imported into `GNU octave-5.2.0` (https://www.gnu.org/software/octave/). Pictures were developed thanks to `gnuplot-5.4.1` (http://www.gnuplot.info/). The manuscript was typeset in `GNU emacs-26.3` (https://www.gnu.org/software/emacs/) assisted by AUCTEX(https://www.gnu.org/software/auctex/). Data in pictures and in tables have been exported directly from `octave`. `R-studio-3.6.3` https://www.rstudio.com/ took care of the generalized Poisson regressions. Bibliography was handled by `Mendeley-Desktop-1.19.8` thanks to Universidad de Sevilla.

This project started on April 18, 2021.