# Exponential distribution of large excess death rates in Europe during the covid–19 outbreak in the spring of 2020

José María Martín-Olalla[1, a)]
*Universidad de Sevilla. Facultad de Física. Departamento de Física de la Materia Condensada. ES41012 Sevilla. Spain*

(Dated: 20200920.1931+0200)

Excess death rates $E$ during the spring of 2020 are computed in $N = 340$ level 3 European territorial units for statistics —NUTS3 in Netherlands (44), Belgium (40), France (96), Spain (50) and Italy (110)— from 2020 provisional week deaths, the population numbers for 2020, and observations in previous years (reference or baseline), all of them obtained from Eurostat web page.

The distribution of excess death rates is found to follow an exponential law with empirical complementary cumulative distribution function $Y$ following $\log Y = a + b(E - \langle E \rangle)/s$. In the middle of the distribution $b_1 = -1.356(9); R^2 = 0.998$ are found and for the largest 46 excess death rates $b_1 = -0.645(18); R^2 = 0.992$ is observed. This result suggests that abnormally large excess death rates may develop in statistical regions independently of what happens in the outside.

Distribution of excess death rates are also analyzed on a country basis. A two-sample Kolmogorov-Smirnov test does not reject the null hypothesis "normal values of $E$ for country $i$ and country $j$ come from the same parent distribution" for any pairing. This is an evidence of a common background in the outcomes. In other words statistical differences among countries can be characterized by averages and standard deviations only. Average death rates, sample standard deviations, excess death tolls are: Netherlands $407 \times 10^{-6}, 387 \times 10^{-6}, 8104$; Belgium $667 \times 10^{-6}, 364 \times 10^{-6}, 8436$; France $228 \times 10^{-6}, 377 \times 10^{-6}, 21\,780$ (overseas departments not included); Spain $1059 \times 10^{-6}, 1149 \times 10^{-6}, 48\,931$ (Canary Islands not included); Italy $791 \times 10^{-6}, 1140 \times 10^{-6}, 49\,184$; and $610 \times 10^{-6}, 877 \times 10^{-6}, 136\,435$ for the combined distribution. A Kolmogorov-Smirnov test rejects ($p < 0.001$) the null hypothesis "normal scores of $E$ come from a Gaussian distribution" only for Italian data and the whole set of excess death rates.

Keywords: disease; epidemiology; exponential distribution; power law; Gaussian distribution; Zipf law; rank plot; Kolmogorov-Smirnov

[a)] Electronic mail: olalla@us.es; Twitter: @MartinOlalla_JM;

CC BY NC ND

## 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[1]. The disease impacted largely in several countries of Europe during the spring of the year 2020.[2] Mortality monitors recorded extremely large anomalies in Italy, Spain, France, Belgium and Netherlands among others. Societal responses in the form of nationwide lock-downs were taken by governments at the spring of 2020. Excess death rates relative to a consistent reference value are then a key quantity to understand the impact of the outbreak and the societal responses that tried to prevent the expansion of the disease.[3]

Many empirical quantities shows a normal (Gaussian) distribution where the sample average value and sample standard deviation suffices to characterize the distribution. Many other societal quantities are characterized by the presence of abnormally large events to which average value means close to nothing.[4] A nice example of this is the distribution of population in municipalities from villages to mega cities. In this case the distribution opens wide relative to a Gaussian distribution due to the presence of very large observations.

This work will analyze the distribution of excess crude deaths in European level 3 territorial units for statistics in Italy, Spain, France, Belgium and Netherlands. Helped by population numbers, excess death rates $E$ will be computed for the spring of 2020. The work aims to understand the distribution of events along the 340 territorial units that entered in the analysis with focus in abnormally large excess deaths which characterize outbreaks. Notice that abnormally low (negative) excess deaths are inhibited due to the fact that mortality in modern societies comes mainly from natural causes, however societal response in the form of lock-downs favors negative excess deaths.

## 2. DATA SET

In April 2020 Eurostat set up "an exceptional temporary data collection on total week deaths in order to support the policy and research efforts related to covid–19". The collection disaggregate weekly deaths by age, sex and territorial unit for statistics. Total week deaths in some European countries can then be obtained from Eurostat web page[1]. This work collected data from Italy (ITA), France (FRA), Belgium (BEL) and Netherlands (NED) disaggregated by the level 3 of the Nomenclature of Territorial Units for Statistics (NUTS3), which fits to the concept of province (ITA), arrondissement (BEL), department (FRA) and COROP region (NED). Corresponding values for Spain (ESP) were collected directly from the National Statistics Institute although they are also available at Eurostat web page.

There are a total of 340 level 3 territorial units in the set of which 110 were Italian, 96 were French, 50 were Spanish, 44 were Belgian and 40 were Dutch. European statistical regions are encoded by a five alphanumeric string composed of the ISO-3166 alpha-2 country code and a three character string denoting the level 1, level 2 and level 3 of territorial units. French overseas departments (FRY) and Canary Islands (ES7) did not enter in the collection.

Population numbers for the territorial units were retrieved from Eurostat `demo_r_pjangr3`[2] table which lists values from 2019 to 2014. Population values for 2020 were obtained after rescaling NUTS3 2019 population numbers with nationwide 2020 population obtained from table `tps00001`[3]. For Spain numbers were taken from table `31304` at the National Statistics Institute which includes numbers until 2020 disaggregated by level 3 territorial units. As of 2020 the total population analyzed add up to 201 393 607, 40 % of the European Union population (502 090 235).

Population numbers always refer to January 1 values. Population were obtained for every ISO-8601 week by linear interpolation from January 1, 2020 backwards. Population in 2020 was set constant irrespective of week number.

## 3. METHODS

### 3.1. Accumulated excess death rates

In this work the accumulated death rates from week 10 to week 21 was computed for Netherlands, Belgium, France, and Spain. For Italy the interval run from week 08 to week 20. In 2020 W10 started on March 2 and W21 ended on May 24. Likewise, W08 started on February 17 and W20 ended on May 17. Loosely speaking they are representative of the spring of 2020. These intervals were suggested from visual inspection of the death rate anomaly in 2020 and from the $z-$score anomalies observed in maps produced by EuroMoMo[4].

---

[1] https://ec.europa.eu/eurostat/cache/metadata/en/demomwk_esms.htm
[2] See https://ec.europa.eu/eurostat/web/products-datasets/product?code=demo_r_pjangrp3 (accessed September 2020)
[3] See https://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tps00001 (accessed September 2020)
[4] See https://www.euromomo.eu/graphs-and-maps#map-of-z-scores

A baseline or reference value can be obtained analyzing the accumulated death rate in the previous years. For that purpose a Poisson regression —`glm` function with option `family="poisson"` in `R`— with a rate linearly varying with calendar year[5] was performed to get a predicted, baseline, reference death rate for 2020. The accumulated excess death rate of the interval —excess death rate hereafter— was then computed as the difference from the observed value in 2020 to the predicted baseline.

Figure 1 shows as an example the results of the method for the smallest (right) and largest (left) anomalies recorded during the spring of 2020 in level 3 territorial units for every country. Circles display normal scores of the observed accumulated death rates. They are obtained after the observations $O_i$ prior to 2020 are standardized to zero sample average and one sample standard deviation by the normal (linear) transform $(O_i - \langle O \rangle)/s$, $\langle O \rangle$ is the sample average and $s$ is the sample standard deviation. The standard deviation $s$ is listed on every picture. The line is the result of the Poisson regression. It is extended until 2020 to show the predicted baseline obtained from the preceding observed accumulated death rates. The vertical distance from the 2020 observed value to the 2020 predicted baseline is the accumulated excess death rate $E$. Its value is listed on every picture and the excess death toll in 2020 $T$. Every panel also lists the sample standard deviation of the observations until 2019 —which scales the vertical axis— and the sample standard deviation of the residual excesses until 2019: the ratio $E$ to $r$ is the $z-$score of the excess death rate.

Finally, on every picture $T$ displays the accumulated excess death toll in 2020 for the territorial unit.

### 3.2. Statistics

The distribution of accumulated excess death rates $\{E_i\}$ in the spring of 2020 for territorial units will be described by its empirical complementary cumulative distribution function (ECCDF), the probability of finding an excess larger or equal than a given value: $Y = P(X \geqslant x)$.

An estimate for ECCDF can be easily produced from a set of $N$ observations $\{O_i\}$ by ranking them in descending order of $O_i$. The ECCDF of $O \geqslant O_i$ is given by $Y = k/N$, where $k$ is the rank, which is a assigned to $O_k$ the $k$th largest observation. As an example the rank of the largest observation is $k = 1$, then $Y_1 = 1/N$. For the smallest observation the rank is $k = N$ and therefore $Y_N = 1$.

If the ECCDF $Y$ is to be compared to the complementary cumulative distribution function CCDF $F$ of a prescribed continuous distribution then the Kolmogorov-Smirnov (KS) statistic $\sqrt{N} \times |F - Y|$ is the appropriate quantity to analyze. The KS distance $D = \sqrt{N}\max|F - Y|$ assesses whether the null hypothesis "sample distribution $Y$ comes from tested distribution $F$ or $Y - F = 0$" is or is not rejected at a given level of significance. For the standard level $\alpha = 0.05$ the KS distance equals to $D_c = 1.36$, irrespective of $N$.

The impact of $\sqrt{N}$ in $D_n$ implies that larger samples need smaller differences ($\propto 1/\sqrt{N}$) to overshot the critical threshold $D_c$ and reject the null hypothesis. In other words, the smaller the sample the more difficult to reject a null hypothesis. To put a naive example a coin toss experiment sized $N = 3$ can produce uniform empirical scores —three tails or three heads— without rejecting the binomial distribution as a parent distribution. On the contrary uniform empirical scores would be signaling an unfair coin toss if $N = 10$.

Figure 2 shows this procedure for the distribution of population within level 3 of territorial units. From top to bottom and left to right panels show Dutch, Belgian, French, Spanish and Italian distribution of population and the standard Gaussian distribution ($\mu = 0; \sigma = 1$). Notice that the vertical axis of the first five panels displays $Y \times \sqrt{N}$ in logarithmic scale. The KS distance to the Gaussian distribution can be then assessed. On the contrary the last panel displays $Y$ in linear scale. The logarithmic scale enhances the visualization of the largest occurrences while the linear scale allow a better understanding of the absolute differences among empirical distributions and continuous distributions which assess the KS statistic.

Every of the first five panels lists in brackets the population size of the country in 2020, sample size $N$, the sample average $\langle P \rangle$, median population $Q2$, the sample standard deviation, and the maximum KS distance, whose position is noted by a vertical broken line. The critical distance is overshot in France, Italy and Spain, meaning the null hypothesis is rejected, while it still sustains in Netherlands and Belgium. Every of the first five panels also shows a quartile box from which first, second and third quartiles of population can be deduced. Every of the boxes encloses 25 % of the territorial units.

Generally speaking $Y > F$ at the lowest tail of the distribution and $Y < F$ at the highest tail where the distribution opens wide compared to the Gaussian distribution. The first issue suggests the lack of low populated units. This is specifically significant in Belgium, Spain and Italy because the KS distance happens at the lowest end of the distribution (see the vertical broken line). The second issue highlights the existence of abnormally high populated territorial units, which are at plain sight on Figure 2 in the first few ranks of every of the first five panels: NL33C Groot-Rijmond and NL329 Groot-Amsterdam; BE100 Arr de Bruxelles-Capitale/Arr. van Brussel-Hoofdstad and BE211 Arr Antwerpen; FRE11 Nord, FR101 Paris, FRL04 Bouches du Rhône and FRK26 Rhône; ES300 Madrid

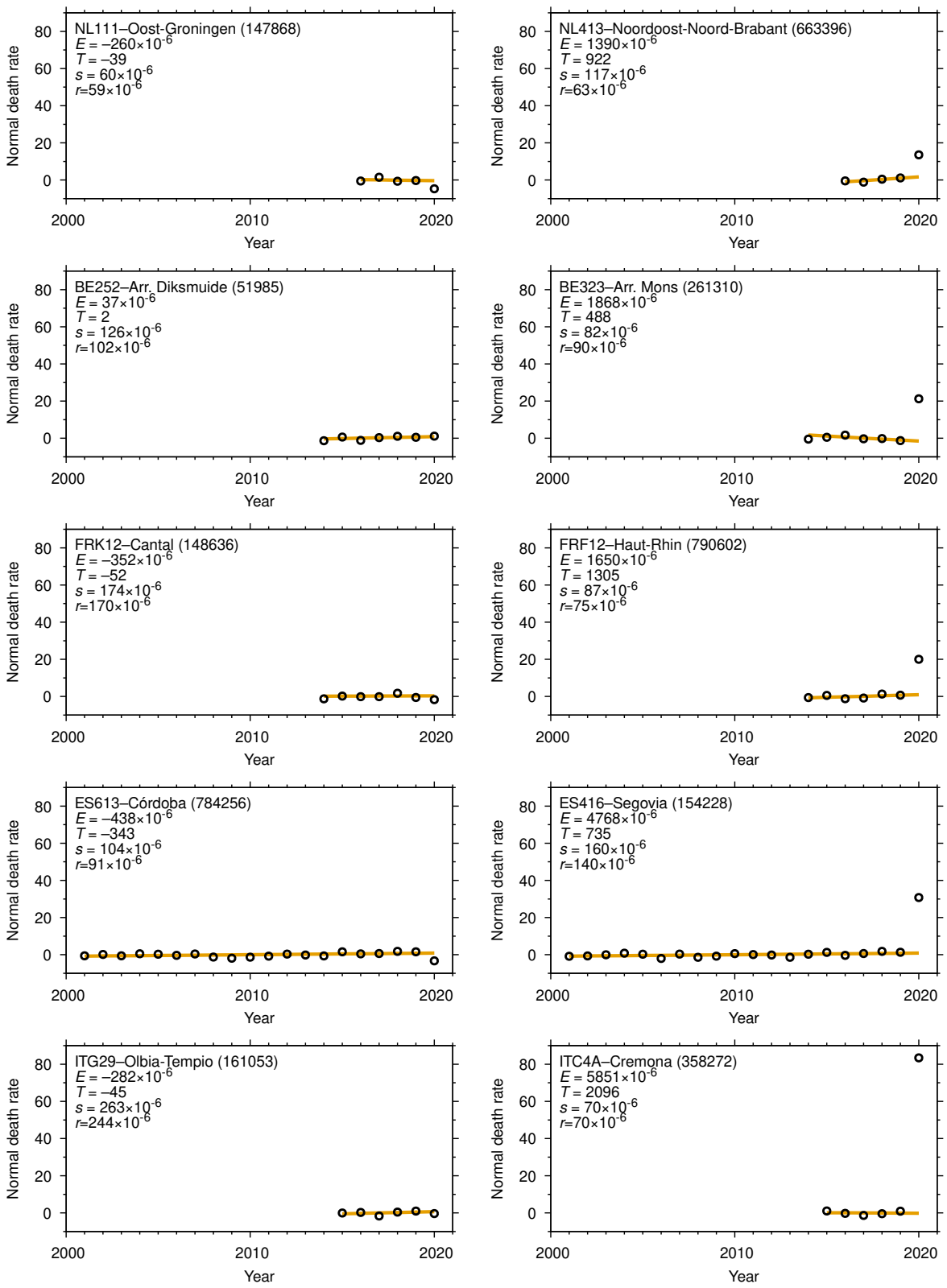


FIG. 1. Smallest (left) and largest (right) accumulated excess death rate recorded in Netherlands, Belgium, France, Spain and Italy (top to bottom) during the spring of 2020. Every picture identifies the territorial unit by its code and name with 2020 population numbers in parenthesis. Circles represent observed values, the line is the predicted values by a Poisson regression with rate linearly varying with calendar year. Vertical axes display the normal scores of death rates: average is removed and the distribution is scaled by the sample standard deviation. The same vertical scale is preserved on every picture. Every panel lists the excess death rate —vertical distance from 2020 observation to reference— the corresponding excess death toll $T$; the sample standard deviation of the observations until 2019 $s$ and the sample standard deviation of the residual excesses until 2019 $r$.

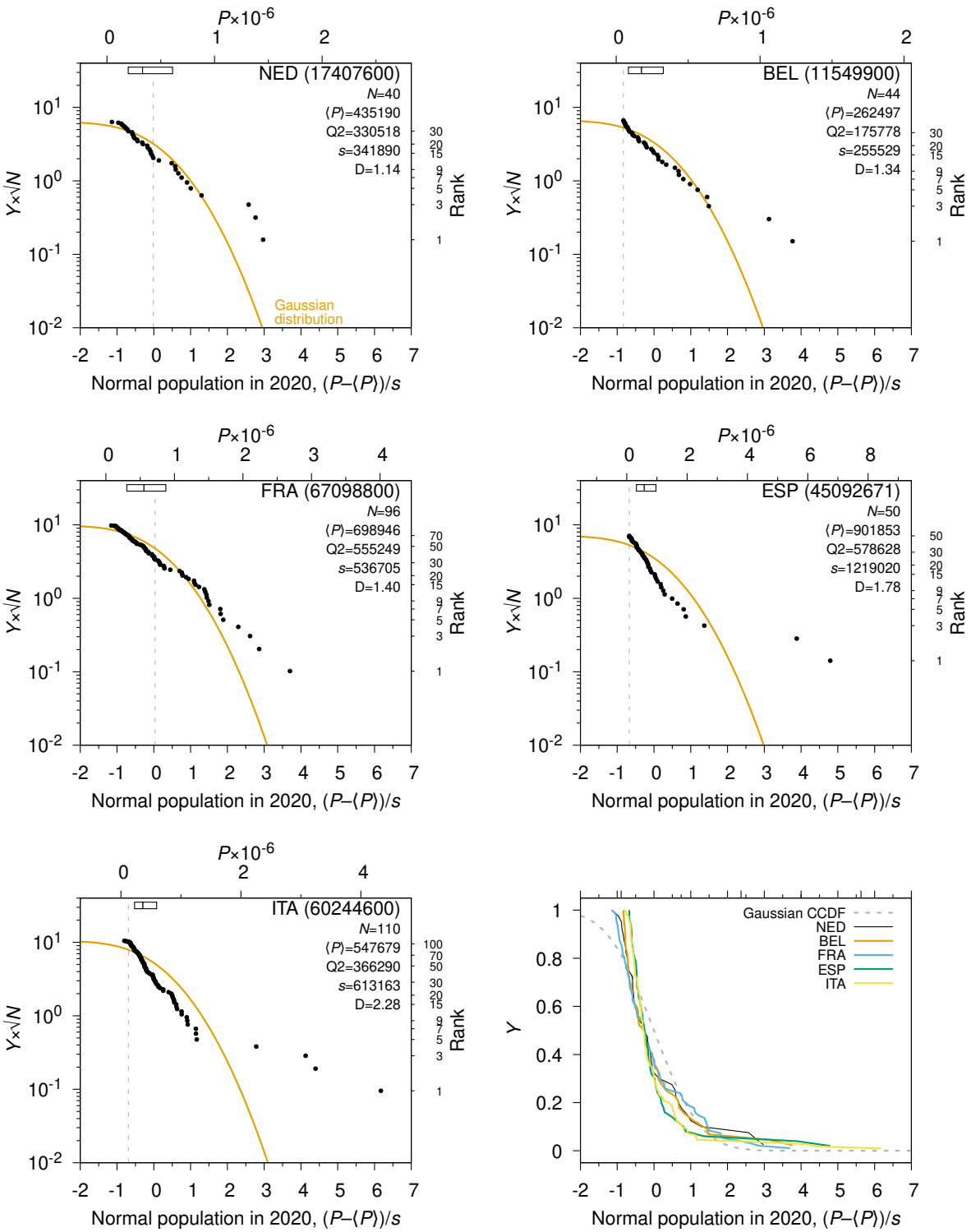


FIG. 2. Distribution of population along level 3 of European territorial units. From top to bottom and left to right Netherlands, Belgium, France, Spain and Italy. Last panel shows them all in a single plot. First five panel shows $\sqrt{N} \times Y$, where $Y$ is the empirical complementary cumulative distribution function $Y_i = k_i/N$, with $k_i$ the rank position in descending order and $N$ is the sample size and the normal values of population in 2020 (bottom axis) and population in 2020 (top axis). They also show the standard Gaussian distribution (orange line). Last panel displays $Y$ and normal values of population in 2020. Every of the first five panels shows ISO-3166-3 codes and population in brackets; number of territorial units; average population of territorial units; median population of territorial units; sample deviation of territorial units and the KS distance to the Gaussian distribution. They also show the quartile box and the position of the KS distance. Canary Islands (ES7) and French overseas departments (FRY) were not included in the analysis.

| | Netherlands | Belgium | France | Spain | Italy | All five |
|---|---|---|---|---|---|---|
| Number of level 3 territorial units | 40<br>11.8 % | 44<br>12.9 % | 96<br>28.2 % | 50<br>14.7 % | 110<br>32.4 % | 340 |
| Population (2020) | 17 407 585<br>8.6 % | 11 549 888<br>5.7 % | 67 098 824<br>33.3 % | 45 092 671<br>22.4 % | 60 244 639<br>29.9 % | 201 393 607 |
| Excess death toll | 8104<br>5.9 % | 8436<br>6.2 % | 21 780<br>16.0 % | 48 931<br>35.9 % | 49 184<br>36.0 % | 136 435 |
| Median excess $Q_2$ ($\times 10^6$) | 353 | 637 | 136 | 632 | 382 | 382 |
| Average excess $\langle E \rangle$ ($\times 10^6$) | 407 | 667 | 228 | 1059 | 791 | 610 |
| Standard deviation $s$ ($\times 10^6$) | 387 | 364 | 377 | 1149 | 1140 | 877 |
| Excess death to population ($\times 10^6$) | 466 | 730 | 325 | 1085 | 816 | 677 |
| KS distance to Gaussian distribution | 0.52 | 0.39 | 1.08 | 1.12 | 2.05 | 3.11 |

TABLE 1. Descriptive statistic values level 3 of territorial units per country and all five combined. First, the number of territorial units and population numbers for 2020; then, excess death toll in the spring of 2020; followed by median excess death rate, averages excess death rate and standard deviation; the ratio excess death toll to population —population weighted average excess death rate— and the KS distance to the Gaussian distribution. Values in blue ink highlight the rejection of the null hypothesis at the standard level of significance ($p < 0.05$). French overseas departments (FRY) and Canary Islands (ES7) did not enter in the analysis.

and ES511 Barcelona; ITI43 Roma, ITC4C Milan, ITF33 Napoli, ITC11 Torino. Save for these few highly populated territorial units the distribution looks like exponential $\log Y = a + bP$ with little curvature in the middle of the population compared to the Gaussian distribution.

## 4. RESULTS

Figure 3 mimics Figure 2 but it now takes the accumulated excess death rate $E$ in the spring of 2020 as the observed quantity. Every of the first five panels displays in a semilogarithmic plot the ECCDF times sample size $Y \times \sqrt{N}$ against the normal score of $E$ (bottom axis) obtained after removing the average excess death rate $\langle E \rangle$ and scaling by the sample standard deviation $s$. The top axis display values of $E$. The sixth panel brings $Y$ against the normal score of $E$ in linear scale.

Descriptive values for the excess death rate statistics are given in Table 1 which lists sample size $N$; the median value of the excess death rate $Q_2$; the sample average value of excess death rate $\langle E \rangle$; the sample standard deviation $s$; the weighted average of excess death rates $E_w$ —the ratio of excess death to population— and the KS distance to a Gaussian distribution. The null hypothesis is rejected for the Italian distribution only. Dutch and Belgian excess death rates find excellent agreement with the Gaussian distribution and extremely low KS distances. Table 1 also lists population numbers and excess death toll.

Figure 4 shows the ECCDF for the full set of $N = 340$ territorial units displayed in Figure 3. Its descriptive values are listed in Table 1 (right most column).

The null hypothesis "$Y$ comes from the Gaussian distribution $F$" is rejected $D_c = 3.11$. It is perceptible in the figure two linear regimes $\log Y = a + bE$ with a crossover around $E = 1200 \times 10^{-6}$ or $k = 46$. The figure displays descriptive values of the bivariate analysis $\log Y = a + b(E - \langle E \rangle)/s$. The lower tail of the distribution shows a rate $b_1 = -1.356(9)$, not far from the empirical slope of the Gaussian distribution. The higher tail finds a rate $b_2 = -0.645(18)$ which is close to one half of $b_1$ and deviates from the Gaussian distribution.

Table 2 lists excess death rates for the top 46 excess death rates and the bottom 50 excess death rates. Descriptive results for the largest and smallest excess death rates are listed in Table 3 and Table 4.

## 5. DISCUSSION

The discussion of the previous results is conditioned by the distribution of level 3 territorial units within countries. While the main goal of territorial units for statistics is to provide homogeneous territorial units the outcome is far from being optimum. For instance in the sixth panel of Figure 2 finds similar empirical distribution of normal scores of NUTS3 population in Spain and Italy —although they vividly differ in sample size— and in France, Belgium and

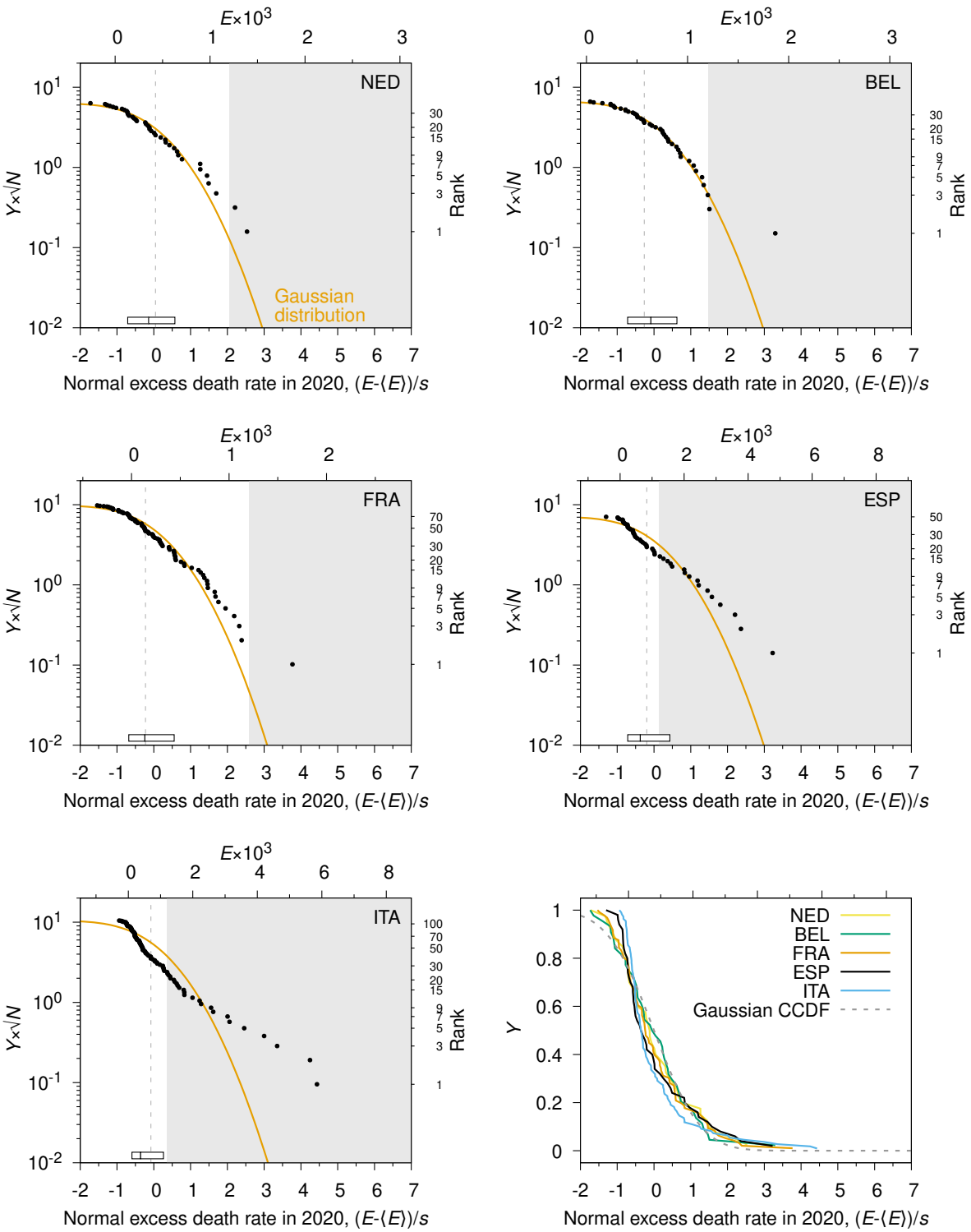


FIG. 3. Distribution of excess death rates $E$ during the spring of 2020 along level 3 of European territorial units. From top to bottom and left to right Netherlands, Belgium, France, Spain and Italy. Last panel shows them all in a single plot. First five panels show $\sqrt{N} \times Y$, where $Y$ is the empirical complementary cumulative distribution function $Y_i = k_i/N$, with $k_i$ the rank position (right axis) from largest to smallest and $N$ is the sample size against the normal values of excess death rate (bottom axis) and excess death rates (top axis). Last panel displays $Y$ and normal values of $E$. Every of the first five panels shows ISO-3166 alpha-3 codes; the quartile boxes; and the position of the KS distance (vertical broken line). Gray background highlights $E \geqslant E_c$ with $E_c = 1203 \times 10^6$ (see Figure 4). Descriptive statistic values are listed in Table 1. Canary Islands (ES7) and French overseas departments (FRY) were not included in the analysis.

| **Top death rate excesses** | | | | **Bottom death rate excesses** | | | |
|---|---|---|---|---|---|---|---|
| Rank | Level 3 territorial unit | $E$ | $z-$score | Rank | Level 3 territorial unit | $E$ | $z-$score |
| 1 | ITC4A–Cremona | 5851 | 83.8 | 340 | ES613–Córdoba | −438 | −4.8 |
| 2 | ITC46–Bergamo | 5634 | 84.0 | 339 | FRK12–Cantal | −352 | −2.1 |
| 3 | ES416–Segovia | 4768 | 34.2 | 338 | FRI11–Dordogne | −326 | −5.9 |
| 4 | ITH51–Piacenza | 4616 | 27.2 | 337 | FRJ27–Tarn | −285 | −2.5 |
| 5 | ITC49–Lodi | 4213 | 26.4 | 336 | ITG29–Olbia-Tempio | −282 | −1.2 |
| 6 | ES417–Soria | 3777 | 14.5 | 335 | NL111–Oost-Groningen | −260 | −4.4 |
| 7 | ITC47–Brescia | 3595 | 73.3 | 334 | FRI23–Haute-Vienne | −252 | −2.5 |
| 8 | ES422–Ciudad Real | 3591 | 36.0 | 333 | FRI31–Charente | −239 | −2.8 |
| 9 | ITH52–Parma | 3139 | 22.9 | 332 | FRL06–Vaucluse | −228 | −2.8 |
| 10 | ES415–Salamanca | 3131 | 29.2 | 331 | ITG14–Agrigento | −226 | −2.8 |
| 11 | ITC48–Pavia | 3077 | 39.2 | 330 | FRJ28–Tarn-et-Garonne | −217 | −2.5 |
| 12 | ES421–Albacete | 2870 | 28.8 | 329 | FRK13–Haute-Loire | −206 | −1.4 |
| 13 | ES424–Guadalajara | 2734 | 18.0 | 328 | ITG2A–Ogliastra | −201 | −0.7 |
| 14 | ITC43–Lecco | 2633 | 20.7 | 327 | FRI21–Corrèze | −193 | −1.2 |
| 15 | ITC18–Alessandria | 2567 | 30.3 | 326 | FRI13–Landes | −192 | −1.6 |
| 16 | ES423–Cuenca | 2456 | 15.5 | 325 | FRL02–Hautes-Alpes | −191 | −2.1 |
| 17 | ES300–Madrid | 2420 | 43.1 | 324 | FRH01–Côtes-d'Armor | −187 | −1.1 |
| 18 | ITC13–Biella | 2258 | 13.0 | 323 | ITI41–Viterbo | −157 | −1.1 |
| 19 | ITC4B–Mantova | 2211 | 16.1 | 322 | ITF11–L'Aquila | −155 | −2.3 |
| 20 | ES411–Ávila | 2162 | 13.1 | 321 | FRI15–Pyrénées-Atlantiques | −135 | −1.2 |
| 21 | ES511–Barcelona | 2024 | 30.8 | 320 | FRI12–Gironde | −134 | −3.1 |
| 22 | ES425–Toledo | 2003 | 15.0 | 319 | FRD11–Calvados | −132 | −1.1 |
| 23 | ITI31–Pesaro e Urbino | 1990 | 13.1 | 318 | FRJ21–Ariège | −131 | −1.6 |
| 24 | BE323–Arr. Mons | 1868 | 20.8 | 317 | FRI34–Vienne | −109 | −0.9 |
| 25 | ITC33–Genova | 1737 | 17.5 | 316 | NL124–Noord-Friesland | −105 | −0.8 |
| 26 | ITC12–Vercelli | 1735 | 15.3 | 315 | FRK11–Allier | −97 | −0.6 |
| 27 | ITC4C–Milano | 1718 | 31.4 | 314 | FRC12–Nièvre | −88 | −0.5 |
| 28 | FRF12–Haut-Rhin | 1650 | 21.9 | 313 | ITG15–Caltanissetta | −85 | −0.8 |
| 29 | ES413–León | 1627 | 12.0 | 312 | NL125–Zuidwest-Friesland | −85 | −0.3 |
| 30 | ITC44–Sondrio | 1579 | 16.8 | 311 | FRK14–Puy-de-Dôme | −83 | −1.6 |
| 31 | ES432–Cáceres | 1577 | 11.5 | 310 | ITF32–Benevento | −78 | −0.7 |
| 32 | ITC20–Valle d'Aosta | 1529 | 10.2 | 309 | ITI15–Prato | −78 | −0.9 |
| 33 | ITC4D–Monza e della Brianza | 1498 | 14.4 | 308 | ES640–Melilla | −78 | −0.5 |
| 34 | ES418–Valladolid | 1486 | 22.7 | 307 | ITG12–Palermo | −65 | −0.8 |
| 35 | ITC17–Asti | 1439 | 29.3 | 306 | ES114–Pontevedra | −64 | −0.9 |
| 36 | ITC15–Novara | 1412 | 15.8 | 305 | NL113–Overig Groningen | −55 | −9.1 |
| 37 | NL413–Noordoost-Noord-Brabant | 1390 | 22.0 | 304 | ITF35–Salerno | −53 | −0.5 |
| 38 | ES412–Burgos | 1355 | 10.9 | 303 | ITI44–Latina | −51 | −0.5 |
| 39 | ITH20–Trento | 1307 | 16.0 | 302 | ITG11–Trapani | −47 | −0.3 |
| 40 | ITH53–Reggio nell'Emilia | 1295 | 14.1 | 301 | FRJ23–Haute-Garonne | −46 | −0.8 |
| 41 | NL423–Zuid-Limburg | 1262 | 14.1 | 300 | ITG2C–Carbonia-Iglesias | −45 | −0.6 |
| 42 | ITC34–La Spezia | 1248 | 6.4 | 299 | FRD13–Orne | −40 | −0.3 |
| 43 | ES230–La Rioja | 1242 | 8.8 | 298 | ITG18–Ragusa | −39 | −0.2 |
| 44 | BE221–Arr. Hasselt | 1216 | 13.1 | 297 | FRG05–Vendée | −35 | −0.5 |
| 45 | ITC31–Imperia | 1214 | 5.0 | 296 | ITF64–Vibo Valentia | −30 | −0.3 |
| 46 | ITC42–Como | 1203 | 15.5 | 295 | FRH04–Morbihan | −29 | −0.3 |

TABLE 2. The 50 largest excess death rates and the 50 smallest excess death rates recorded in the 340 level 3 territorial units. The table lists the rank position, NUTS3 code and name, excess death rate and the $z-$score: ratio of the excess death rate to the standard deviation of the residuals of the Poisson regression (see Figure 1

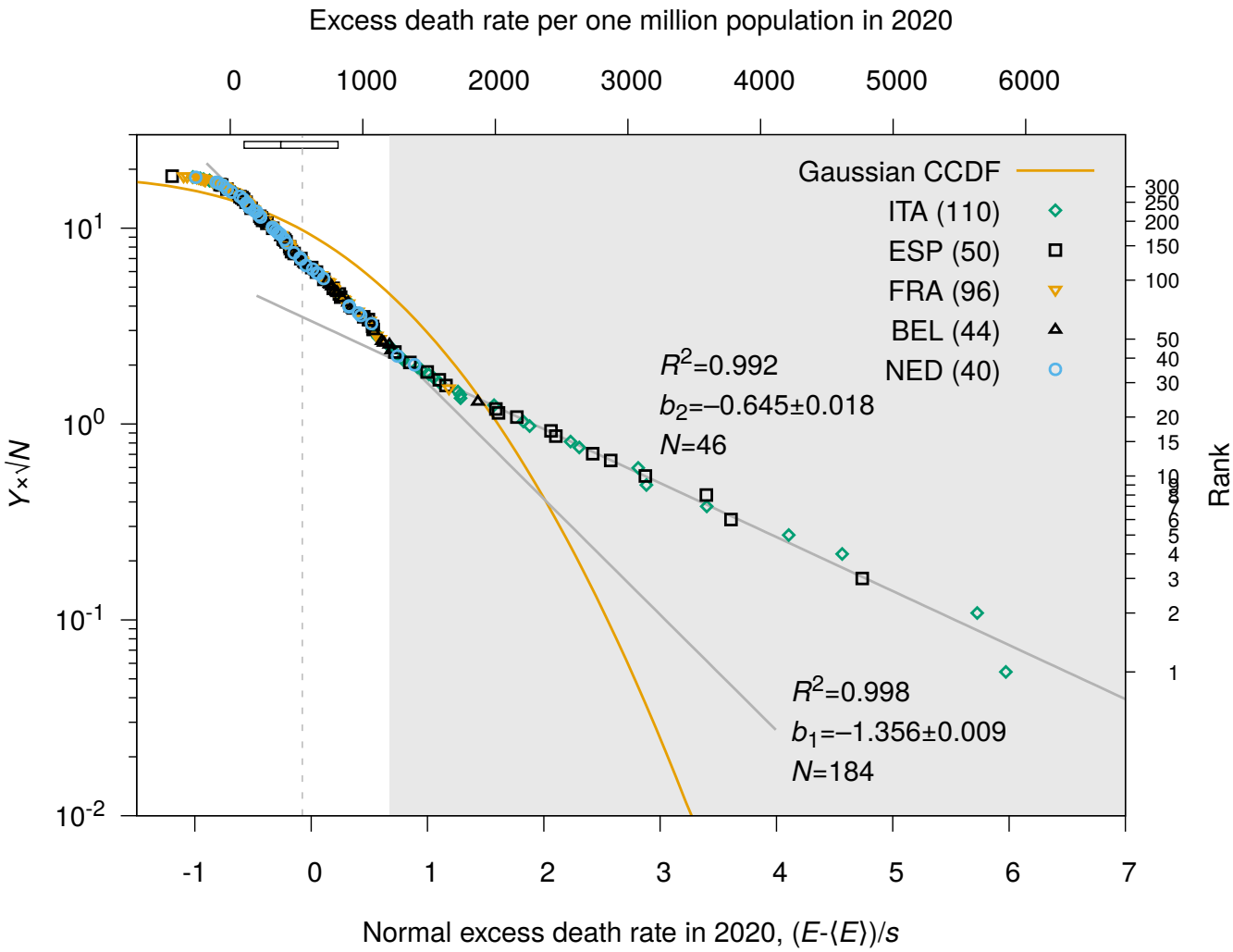


FIG. 4. The empirical distribution of excess death rates in the $N = 340$ level 3 territorial units of Netherlands, Belgium, France, Spain and Italy against the normal of the excess death rate score $(E - \langle E\rangle)/s$. Different color and symbol according to country are used. The Gaussian CCDF is noted by an orange line. Gray background starts at $E_c = 1203 \times 10^{-6}$ where the change in the slope of $Y$ is noted. Gray straight lines show bivariate analysis $\log Y = a + b(E - \langle E\rangle)/s$ for $E \geqslant E_c$ and $E < E_c$. The vertical broken line shows the position of the KS distance to the Gaussian distribution $D = 3.11$ which is computed as the vertical largest difference from the data point to the Gaussian line.

Netherlands. Indeed a two-sample Kolmogorov-Smirnov test finds significant KS distances[5] (with $p \in [0.0007, 0.03]$) if the Spanish or the Italian normal scores of NUTS3 population are tested against the French or Belgian normal scores. This rejects the null hypothesis that every of these possible four pairings comes from the same theoretical distribution.

On the contrary, the sixth panel of Figure 3 finds similar empirical distributions of the normal scores of excess death rates for every of the countries here analyzed. In this case a two-sample Kolmogorov-Smirnov test finds non-significant KS distances (with $p \in [0.101, 0.998]$) for any of the 10 possible pairings. Therefore the null hypothesis "the distribution of normal scores of excess death rates in country $A$ and the distribution of normal scores of excess death rates in country $B$ come from the same distribution" is not rejected for Netherlands, Belgium, France, Spain and Italy.

Notwithstanding this the first five panels of Figure 3 still show differences with Spanish and Italian distributions having almost no curvature in the semilogarithmic plot, therefore showing a marked exponential distribution. This

[5] The two-sample KS distance can be assessed from the sixth panel in Figure 2 for any pair $ij$ by taking the maximum of the vertical distance $|Y_i - Y_j|$ and scaling it with the square root of the harmonic mean of the samples sizes $N_i$ and $N_j$.

| | Netherlands | Belgium | France | Spain | Italy | All five |
|---|---|---|---|---|---|---|
| NUTS3 occurrences | 2 | 2 | 1 | 16 | 25 | 46 |
| | 5.0 % | 4.5 % | 1.0 % | 32.0 % | 22.7 % | 13.5 % |
| | 4.3 % | 4.3 % | 2.2 % | 34.8 % | 54.3 % | |
| Population (2020) | 1 265 099 | 694 554 | 790 602 | 17 204 516 | 14 065 315 | 34 020 085 |
| | 7.3 % | 6.0 % | 1.2 % | 38.2 % | 23.3 % | 16.9 % |
| | 3.7 % | 2.0 % | 2.3 % | 50.6 % | 41.3 % | |
| Excess death toll | 1682 | 1015 | 1305 | 38 701 | 34 858 | 77 560 |
| | 20.8 % | 12.0 % | 6.0 % | 79.1 % | 70.9 % | 56.8 % |
| | 2.2 % | 1.3 % | 1.7 % | 49.9 % | 44.9 % | |
| Excess death rate ($\times 10^6$) | 1329 | 1461 | 1650 | 2249 | 2478 | 2280 |

TABLE 3. Descriptive values for the 46 largest excess death rates recorded in the 340 level 3 territorial units: the number of occurrences; population; excess death toll; and excess death rate. Percentages are taken first relative to country and second relative to the set of 46 largest excess death rates: 25 Italian territorial units make the 23.3 % of the 110 Italian level 3 territorial units and the 51 % of the 41.3 largest observations.

| | Netherlands | Belgium | France | Spain | Italy | All five |
|---|---|---|---|---|---|---|
| NUTS3 occurrences | 4 | 0 | 24 | 3 | 15 | 46 |
| | 10.0 % | 0.0 % | 25.0 % | 6.0 % | 13.6 % | 13.5 % |
| | 8.7 % | 0.0 % | 52.2 % | 6.5 % | 32.6 % | |
| Population (2020) | 1 005 601 | 0 | 12 400 843 | 1 812 561 | 6 016 422 | 21 235 427 |
| | 5.8 % | 0.0 % | 18.5 % | 4.0 % | 10.0 % | 10.5 % |
| | 4.7 % | 0.0 % | 58.4 % | 8.5 % | 28.3 % | |
| Excess death toll | −106 | 0 | −1721 | −410 | −525 | −2762 |
| | −1.3 % | 0.0 % | −7.9 % | −0.8 % | −1.1 % | −2.0 % |
| | 3.8 % | 0.0 % | 62.3 % | 14.8 % | 19.0 % | |
| Excess death rate ($\times 10^6$) | −106 | | −139 | −226 | −87 | −130 |

TABLE 4. Descriptive values for the 46 smallest excess death rates recorded in the 340 level 3 territorial units: the number of occurrences; population; excess death toll; and excess death rate. Percentages are taken first relative to country and second relative to the set of 46 smallest excess death rates: 24 French territorial units make the 25.0 % of the 96 French level 3 territorial units and the 52.2 % of the 46 smallest observations.

behaviour is less apparent in the French distribution and it is absent in the Belgian and Dutch distributions. In summary distributions differ in sizeable quantities like average, standard deviation, sample size or curvature; other than that they might have been originated from a common, parent distribution.

These results support the analysis of the distribution of $N = 340$ excess death rates reported in the level 3 territorial units of these countries, see Figure 4. A two-sample Kolmogorov-Smirnov test for the full distribution and any of the country distributions does not reject the null hypothesis of them coming from a common distribution ($p \in [0.10, 0.39]$). The full distribution reject the null hypothesis that it comes from a Gaussian distribution $D = 3.11$ (see Table 1).

It is visually perceived in Figure 4 a marked change in the slope of $\log Y$ at $E_c = 1200 \times 10^{-6}$. Observations $E_i > E_c$ (size $m = 46$) are noted in gray background, which is also shown in Figure 3. They are abnormally large events in the following sense: a Kolmogorov-Smirnov test of normality for the $N - m = 294$ observations $E_i \leqslant E_c$ does not reject the null hypothesis ($p = 0.053$) whereas including observations $E_i > E_c$ in the distribution make the test reject the null.

Table 3 shows descriptive values for the distribution of the top 46 excess death rates with Spain displaying the highest percentage of population —38.2 % of Spanish population live in the 16 level 3 territorial units in this set— and the highest percentage of excess death rate —79.1 % of the excess deaths come from the hardest hit territorial units— and Italy displaying the largest excess death rate $2478 \times 10^{-6}$. For the combined analysis, 16.9 % of the population lives in top 46 excess death rate, level 3 territorial units and they scored a 56.8 % of the overall excess death toll with an average death rate equal to $2280 \times 10^{-6}$. Table 4 does the same for the bottom 46 excess death rates.

The hypothesis of an exponential distribution of excess death rates is robust as far as it concerns to the largest

events. With the sole exception of some territorial units in United Kingdom no other country in Europe may have produced events $E > E_c$. Therefore adding more countries to the analysis means adding territorial units with $E < E_c$. The distribution of the events noted in the gray background would not be altered because the rank $k$ is preserved and $Y$ is only shifted by a constant factor from $k/N$ to $k/(N+n)$ where $n$ is the number of new (smaller) events added. In a logarithmic scale this would be noted as a vertical displacement. Albeit in reverse way, this idea is observed in Figure 3: the distribution of the largest events in Spain and Italy still preserve the exponential behaviour.

The exponential distribution appears in the distribution of interevents time of statistically independent phenomena, including mortality. Excess death rates is nothing but a cumulative sum of deaths during a prescribed time interval. It is common-knowledge that the spring outbreak of covid–19 was extremely heterogeneous in Europe as seen in Table 2, Table 3 and Table 4. The observation in Figure 4 suggests that despite this well known clustering, excess death rate could have been developed in isolated conditions: once the disease sets in a territorial unit, once local transmission begins and, perhaps, once lock-downs are enforced, the fate of the outbreak depends on a myriad of societal issues but, to some extend and empirically, it could be independent from what happens in other territorial units.

## REFERENCES


[1]N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, and W. Tan, "A novel coronavirus from patients with pneumonia in China, 2019," New England Journal of Medicine **382**, 727–733 (2020).

[2]A. Campbell and E. Morgan, "Comparisons of all-cause mortality between European countries and regions : January to June 2020," Office for National Statistics , 1–27 (2020).

[3]J. Aaron, J. Mullbauer, C. Giattino, and H. Ritchie, "A pandemic primer on excess mortality statistics and their comparability across countries," (2020).

[4]A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," SIAM Review **51**, 661–703 (2009).

[5]J. M. Martin-Olalla, "Age and sex disaggregation of crude excess deaths during the 2020 spring COVID-19 outbreak in Spain," medRxiv , 2020.08.06.20169326 (2020).


## ACKNOWLEDGMENTS


The author wishes to thank Eurostat and the National Statistics Institutes in Europe for releasing week crude deaths.

This work benefited from `GNU octave-4.2.2` (https://www.gnu.org/software/octave/); `R-3.4.4` (https://www.r-project.org/); `gnuplot-5.2.2` (http://www.gnuplot.info/); `GNU emacs-25.2.2` (https://www.gnu.org/software/emacs/) and AUCTEX(`https://www.gnu.org/software/auctex/`).

This project started on September 12, 2020.