medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Chasing the ghost of infection past: identifying thresholds of change during the COVID-19 infection in Spain

Luis Santamaría[1] & Joaquín Hortal[2]

[1]Estación Biológica de Doñana (EBD-CSIC), C/ Américo Vespucio 26, Isla de la Cartuja, E41092 Sevilla, Spain. email: luis.santamaria@ebd.csic.es

[2]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC). C/José Gutiérrez Abascal 2, 28006 Madrid, Spain. email: jhortal@mcncn.csic.es

Abstract

COVID-19 pandemic has spread worldwide rapidly from its first outbreak in China, with different impacts depending on the age and social structure of the populations, and the measures taken by each government. Within Europe, the first countries to be strongly affected have been Italy and Spain. In Spain, infection has expanded in highly populated areas, resulting in one of the largest nationwide bursts so far by early April. We analyze the evolution of the growth curve of the epidemic in both the whole of Spain and Madrid Autonomous Region (the second largest conurbation in Europe), based on the cumulative numbers of reported cases and deaths. We conducted segmented, linear regressions on log-transformed data to identify changes in the slope of these curves and/or sudden shifts in the number of cases (i.e. changes in the intercept) at fitted breaking points, and compared their results with a timeline including both key events of the epidemic and containment measures taken by the national and regional governments. Results were largely consistent in the four curves analyzed (reported infections and deaths for Spain and Madrid, respectively), showing two major shifts in slopes (growth rates) at 14-15 and 26-29 March that resulted in 37-65% reductions of slope, and originated in infection on 4-5 and 16-18 March (for case detections) and 14-23 February and 5-6 March (for deaths). Small upward shifts in the progress of the disease in Madrid were not associated with significant changes in the intercept of the curve, and seem related with unevenness in case reporting. These results evidence an early deceleration in the spread of COVID-19 coinciding with personal hygiene and social distancing recommendations, as well as the general awareness of the population; and a second, stronger decrease when harder isolation measures were enforced. The combination of both breakpoints seemingly led to the start of the contention of the disease outbreak by early April, the limit of our time series. This highlights the importance of adopting public health strategies that include disseminating basic knowledge on personal hygiene and reduced social contact at the onset of the epidemic, and the importance of early enforcement of hard contention measures for its subsequent contention.

Keywords: Segmented regression, breakpoints, Comunidad de Madrid, COVID-19, infection growth curve, social distancing effectiveness, Spain

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/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

Introduction

COVID-19 infection has rapidly spread worldwide since its first outbreak in Wuhan (China) in mid December 2019. The global number of confirmed cases has gone over one million on 3rd April 2020 (John Hopkins University Coronavirus Resource Center, see Dong et al. 2020), barely 3 months after its first report on 31st December. Individuals infected with COVID-19 remain asymptomatic for 5-6 days, while presenting enough viral load to be infective after 1-2 days of infection (Linton et al. 2020, Lai et al. 2020). Severe cases require hospitalisation 3-15 days after the appearance of the first symptoms, which are similar to other infectious respiratory illnesses. This, together with the initial unawareness of the population, led to a high transmission rate of the infection, which spread rapidly to neighbouring countries, the Middle East and Europe, and then the rest of the world (see https://nextstrain.org/ncov).

An increasing number of countries was progressively affected, and they responded differently depending on the WHO and local expert advice at the moment, the structure and resources of their public health systems, their R+D+i capacity (which determined the number of PCRs available for testing contagions from blood samples, among other things), and their ability to implement social distance measures. The diversity of policy responses, together with the preexisting differences in spatial aggregation, social behaviour and age structure of their populations, provide an unique array of test cases to understand how different levels and combinations of preventive quarantine and social-distancing measures affected the spread of the pandemic.

COVID-19 arrived to mainland Spain in early February (first recorded hospitalisation dates back to 15th February; Table S1). During the first 2-3 weeks of February, COVID-19 infection reached Spain at least three times, via UK and Italy – as evidenced by the presence of three different genetic clusters identified by nextstrain (Hadfield et al. 2018; last accessed 8th April). Different from Italy, where infections were concentrated in the North, the combination of these three introductions with early, unnoticed community transmission resulted in consecutive outbreaks in distant, highly populated areas of the Basque Country and Navarra (North), Madrid (Center), Catalonia (North East), Andalusia (South) and Valencia (East) (see timeline in Figure 5 below, and Table S1). The spatial structure of the Spanish populations has played a role in the particularly rapid spread of the pandemic in some regions the country. Its impact has been harsher in the big conurbations of Madrid (around 6.4M people; second most populated Metropolitan area of the EU, after Paris) and Barcelona (c. 5.4M) - as well as in Álava, Navarra and La Rioja (c. 1M in total), following the early infection of healthcare workers from Txagorritxu Hospital. Balearic and Canary archipelagos also received infections from the early onset of the pandemic, so it is reasonable to assume that by early March COVID-19 infections were widely distributed throughout the whole country.

Several factors make Spanish data one of the most fair accounts of the effects of the pandemic at the country and regional levels, together with Italy and, specially, South Korea. Although the lack of enough tests that has been pervasive for most countries (except South Korea), Spain has achieved one of the highest infection test ratios per capita (Clark et al. 2020), thanks to the early mobilisation of most PCR machines available in universities and research centres for either COVID-19 testing or COVID-19 research. Importantly, only cases testing positive in the PCR make it to the official statistics – and (similar to Italy but different to other European countries) all deaths testing positive are registered as caused by COVID-19 infection, including those associated with previous pathologies or happening outside hospitals (e.g. in private homes and nursing homes). These data provide an underestimation of the total population infected and the number of fatalities– due to the limited number of tests; although for the number of deaths this is partly compensated by the lethality associated to other

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

pathologies that is attributed to COVID-19 when PCRs render positive tests. However, the relatively homogeneous intensity of testing and the stability of criteria for disease attribution throughout the time period of this analysis probably result in unbiased estimators for the spread of the pandemic. It is therefore safe to assume that the number of reported cases of infection and the number of deaths are reasonably good proxies for the advance of the pandemic.

Here we characterize the growth curve of COVID-19 infections in the whole of Spain, from the onset of the pandemic in early February through the establishment of increasingly more restrictive social and governmental restrictions to mobility and personal contact. We also perform he analyses for the Madrid Autonomous Region (Madrid hereafter), a highly populated area with good public transportation and a high daily commuting rate, which represented the country's largest focus of the pandemic – as it represents a prime example of the spread of the virus in a large, mostly panmictic population through time, and the effect of social-distancing measures thereupon.

The adoption of containment measures by the national and regional governments followed a sustained increment through time, from the recommendation of preventive measures in late February and early March, to increasingly stricter social-distancing measures on 9-10 March, to a nationwide lockdown announced on 13 March and enforced on 15 March, to the the closure of all non-essential economic activities on 31 March (see Figure 5 below, and Table S1). Such sequence of measures was broadly discussed by experts, media and social media, with opinions ranging from qualifying them as exaggerated or unnecessary during the first weeks of the outbreak; to criticizing them as tardy of insufficient in the weeks that followed. Two controversies have been particularly strong: (i) were preventive and soft social-distancing measures useful, or should hard social-distancing measures have been introduced from the early moments (late February to early March)?, and (ii) did the mass events on the weekend of 7-8 March, coinciding with the International Women's Day demonstrations (over 300k attendants in the whole country, 120K in Madrid) and premier football league matches (around 280k spectators in total and 72K in Madrid, respectively) trigger the early spread of the pandemic in Spain's largest cities, specially in Madrid?

Bearing this temporal sequence in mind, we analyze the growth curves of the cumulative numbers of cases detected and the cumulative number of deaths for both the whole of Spain and Madrid, focusing specifically in the changes in the growth rate (i.e. the slope of log-transformed data) of these curves through time. Based on this analysis, we seek to answer two questions: (1) how effective were the different social-distancing measures in reducing infection and mortality rates?; and (2) how significant were the effects of 7-8 March mass gatherings on the expansion of the epidemic, compared with other key events and control measures?

Data and methods

Timeline of events and control measures

Data on the different events that marked the evolution of the pandemic in Spain (e.g. first cases detected, large infection bouts, first deaths) or influenced its perception by the general public, as well as policy measures (e.g. preventive isolation, social-distancing, lockdowns) and putative key events (e.g. large gatherings associated to sport events, political demonstrations and party rallies), were gathered from official sources, national and international media, and scientific publications. Whenever possible, and in all cases for policy measures, we confirmed their date and content from official documents and/or websites from international, national or

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

regional institutions. We include a broad list of events in Table S1 and selected the most relevant ones for the timeline shown, together with the results of the statistical analyses, in Figure 5's graphical summary.

Infection and fatality data

Official data on the (i) cumulative number of cases, and (iii) cumulative number of deaths were obtained from the daily Covid reports of the Spanish Ministry of Health, as compiled by the Worldometer Coronavirus data service (for national data) and the Covid data service of eldiario.es (for regional data). Data were extracted at two levels of aggregation, for Spain as a whole country, and for Madrid Autonomous Region (i.e. *Comunidad Autónoma de Madrid*). For the analyses we included data from the first day in which at least 10 cases or at least 1 death were measured; and extended the analyses to 22-24 days after the onset of social-distancing measures on 13-15/3/20, a period doubling the average infection-to-detection time (10.1 days; see next section), and equaling the average infection-to-death time (21 days; see next section).

Lag time estimates

To estimate the infection date of reported cases, we calculated the infection-to-testing time by combining reported values of incubation time (mean = 5.0 days in Lauer et al. 2020; median = 5.1 days in Linton et al. 2020; mean = 6.4 days in Lai et al. 2020) with time from illness onset to hospital admission for treatment and/or isolation (median = 3.3 days among living cases and 6.5 days among deceased; Linton et al. 2020). Hence, we used an infection-to-testing time of 9 days for living cases and 12 days for dead cases. Based on the proportion of 36% deaths to 64% recoveries reported from 3/3/20 to 6/4/20 (for a total of 57,006 closed cases in Spain), we estimated an average infection-to-testing time of 10.1 days – which, for simplicity, was rounded to 10 days. Similarly, to estimate infection date from day of death, we combined the reported values of incubation time (mean = 5.0 days in Lauer et al. 2020; median = 5.1 days in Linton et al. 2020; mean = 6.4 days in Lai et al. 2020) and time from illness onset to death (mean = 15.0 days in Lauer et al. 2020; mean = 15 days in Linton et al. 2020, as used also by Russell et al. 2020; mean = 17.8 in Verity et al. 2020) - which resulted in a infection-to-death time of 21 days.

Analyses

We fitted a family of segmented (broken-line) regressions with no, one, two and three breaking points (Models 1 to 4, with two, four, six and eight parameters respectively) and compared them using the their adjusted R^2 and goodness of fit. Goodness-of-fit comparisons were based on two criteria: (i) the distribution of the residuals; (ii) the adjusted R^2; and (iii) a F-statistic comparing each model with the next level of restriction – that is, to the model with one breakpoint (hence, two parameters) less (Hank et al. 2020). To ensure residuals' homoscedasticity, we used linear fits on $\log_{10}$-transformed data; similar results were, however, obtained using exponential fits on untransformed data (not shown).

Fitted breaking points provide objective information on the moment at which infection dynamics changed, while slopes provide information of the direction and magnitude of such

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

changes. When analyzing the data from Madrid, we observed discontinuities that suggested that some breakpoints could involve a change in the intercept, rather than in the slope. This would imply a significant shift in values at a given day, followed by a continuous increase at the same growth rate that preceded such day – an scenario consistent, for example, with a sudden increase in infection rate during the mass gatherings of 7-8 March. To test for this possibility, we assessed the fit of an additional model with two breaking points, the first one involving a change in the intercept and the second one involving a change in the slope (Model 5).

Results

Number of cases

For the whole of Spain, the model with two breaking points (Model 3) provided the best fit (Table 1). Fitted breaking points were placed on day 17.9 (14/3/20, estimated infection on 4/3/20) and 30.5 (26/3/20, estimated infection on 16/03/20) (Figure 1). The growth rate of the number of cases decreased by 49% (from 0.15 to 0.08) after the first breakpoint (14/3/20) and decreased again by another 54% (from 0.08 to 0.03) after the second breakpoint (26/3/20).

Table 1. Results of segmented regressions with an increasing number of breaking points, fitted on the total number of cases detected in Spain from 25/2/20 (first day with 10 cases) to 06/04/20.

Model	Intercepts	Slopes				Breakpoints (#days)			Goodness of fit				
	a1	b1	b2	b3	b4	u1	u2	u3	R2	Adj.R2	F	df1,df2	P
M1	1,6	0,1017							0,9431	0,9402	323,1	2,39	**5,3341E-25**
M2	1,1	0,1509	0,0532			20,4			0,9973	0,9970	372,6	2,37	**3,0549E-25**
M3	1,1	0,1546	0,0788	0,0321		17,9	30,5		0,9989	0,9987	26,1	2,35	**1,1746E-07**
M4	1,1	0,1546	0,0894	0,0707	0,0306	17,4	23,3	31,6	0,9990	0,9987	0,9	2,33	0,42

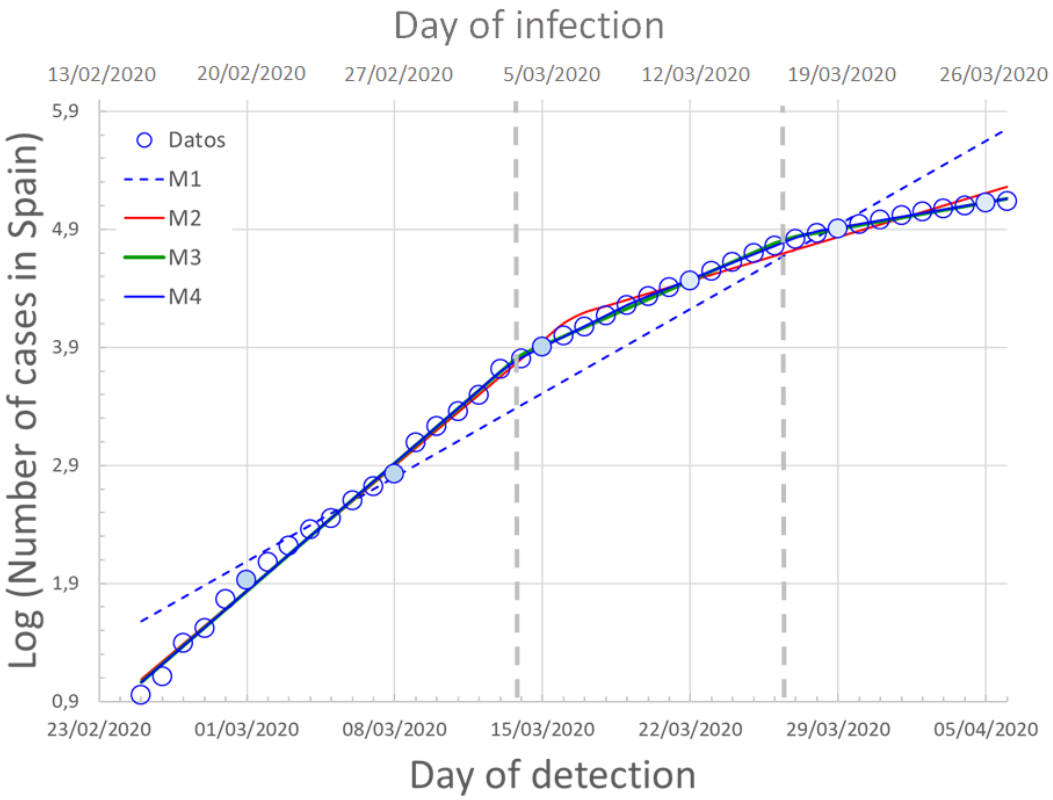

Figure 1. Segmented regressions fitted on the total number of cases detected in Spain. Estimated time

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

of infection is provided in the upper X-axis. Filled points indicate Sundays. Broken vertical lines indicate the breaking points of the best fit (Model 3), which is shown with a thicker line.

The analyses performed on the number of cases from Madrid are consistent with the results for the whole country. The model with two breaking points (Model 3) provided the best fit (Table 2). Fitted breaking points were placed on day 12.7 (14/3/20, estimated infection on 4/3/20) and 26.6 (28-29/3/20, estimated infection on 18-19/3/20) (Figure 2). The growth rate of the number of cases decreased by 65% (from 0.18 to 0.06) after the first breakpoint (14/3/20) and decreased again by 59% (from 0.08 to 0.03) after the second breakpoint (18-19/3/20).

Table 2. Results of segmented regressions with an increasing number of breaking points, fitted on the total number of cases detected in Madrid Region from 25/2/20 (first day with ten cases) to 6/4/20.

Model	Intercepts	Slopes				Breakpoints (#days)			Goodness of fit				
	a1	b1	b2	b3	b4	u1	u2	u3	R2	Adj.R2	F	df1,df2	P
M1	1,9	0,0896							0,9002	0,8943	153,4	2,34	**9,6533E-18**
M2	1,2	0,1761	0,0486			13,5			0,9951	0,9944	307,4	2,32	**1,2891E-21**
M3	1,2	0,1780	0,0611	0,03051		12,7	26,6		0,9965	0,9958	6,1	2,30	**0,006029**
M4	1,2	0,1848	0,1489	0,06005	0,0305	9,0	13,6	26,7	0,9968	0,9958	1,2	2,28	0,33

Model	Intercepts		Slopes		Breakpoints (#days)		Goodness of fit				
	a1	Δa	b1	b2	u1	u2	R2	Adj.R2	F	df1,df2	P
M2	1,2		0,1761	0,0486	13,5		0,9951	0,9944			
M5	1,3	0,15	0,1577	0,0475	7,1	14,1	0,9954	0,9945	1,088	2,30	0,35

An inspection of the values and fits (Figure 2) shows that the apparent jump in the number of cases detected on 9/3/20 (estimated infection on 28/02/20) was caused by the combination of a decrease during the weekend (7-8/3/20) and an increase the following Monday – which kept the point in line with the previous and posterior values. Indeed, the only fitted model that identified a change of slope (Model 4, breaking point at day 9.0, i.e. on 10/3/20, estimated infection 29/2/20) showed a 19% decrease in the growth rate (from 0.18 to 0.15) at such point – although it provided a non-significant improvement in goodness-of-fit relative to a more-parsimonious model without such breaking point (Model 3). Similarly, the model with two breaking points involving a change of intercept and a change of slope (Model 5), which resulted in a 11% increase in the intercept (from 1.3 to 1.45) on day 7.1 (8/3/20, estimated infection 27/2/20), did not result in a significantly better fit compared to Model 2 (Table 2). Here it is worth noting that a similar 'decrease-and-jump' in the number of cases was observed one week before, from Saturday 29/2/20 to Monday 2/3/20, although dates before 1/3/20 were not included in the analysis owing to the low number of registered cases (below the ten-cases threshold).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

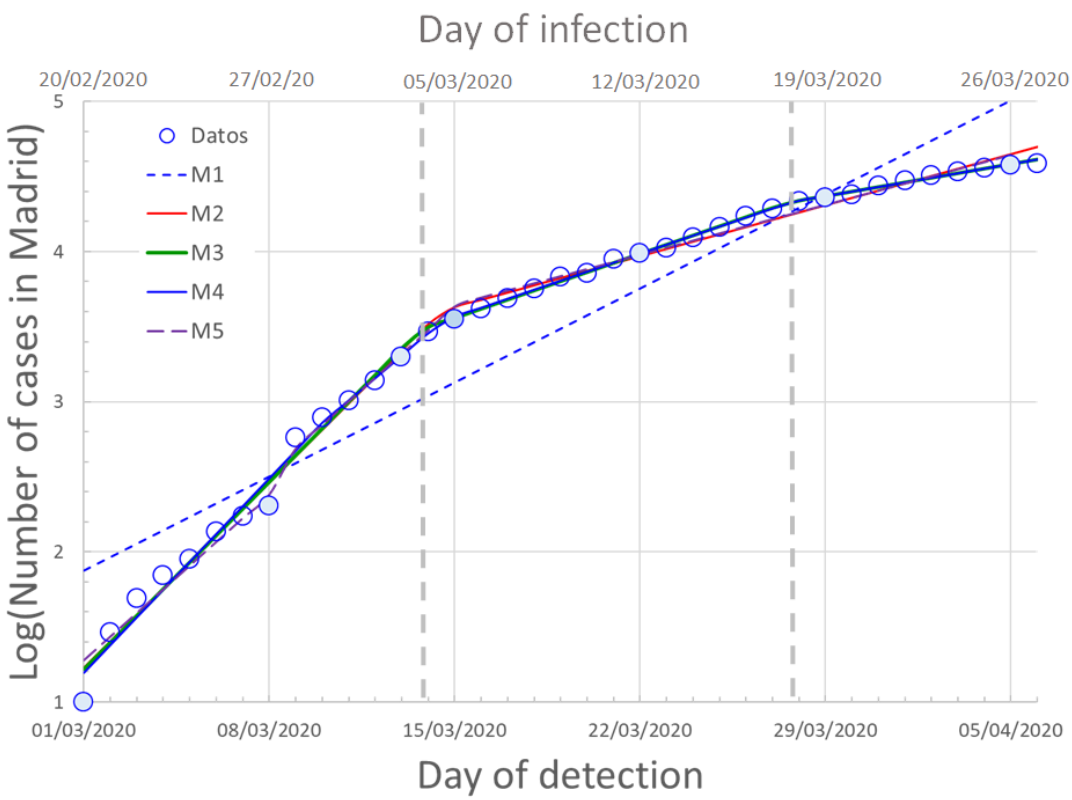

Figure 2. Segmented regressions fitted on the total number of cases detected in Madrid. Estimated time of infection is provided in the upper X-axis. Filled points indicate Sundays. Broken vertical lines indicate the breaking points of the best fit (Model 3), which is shown with a thicker line.

Number of fatalities

The model with three breaking points (Model 4) provided the best fit (Table 3) for the fatalities associated to COVID-19 infection in the whole of Spain. Fitted breaking points were placed on day 3.1 (6/3/20, estimated infection time on 14/2/20), 12.1 (15/3/20, estimated infection on 23/2/20) and 23.6 (26-27/3/20, estimated infection on 5-6/03/20) (Figure 3). The growth rate of the number of fatalities decreased by 37% (from 0.28 to 0.18) after the first breaking point (6/3/20), decreased by another 40% (from 0.18 to 0.11) after the second breaking point (15/3/20), and decreased again by another 62% (from 0.11 to 0.04) after the third breaking point.

Table 3. Results of segmented regressions with an increasing number of breaking points, fitted on the number of fatalities attributed to COVID-19 infections in Spain from 3/3/20 to 6/4/20.

Model	Intercepts	Slopes				Breakpoints (#days)			Goodness of fit				
	a1	b1	b2	b3	b4	u1	u2	u3	R2	Adj.R2	F	df1,df2	P
M1	0,74	0,1166							0,9359	0,9319	234	2,32	**8,1142E-20**
M2	0,19	0,1894	0,0694			14,5			0,9940	0,9932	146	2,30	**3,5854E-16**
M3	0,04	0,2433	0,1387	0,0536		6,3	19,8		0,9977	0,9972	22,6	2,28	**1,4368E-06**
M4	0,000000	0,2832	0,1788	0,1078	0,0411	3,1	12,1	23,6	0,9994	0,9993	39,1	2,26	**1,4512E-08**

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

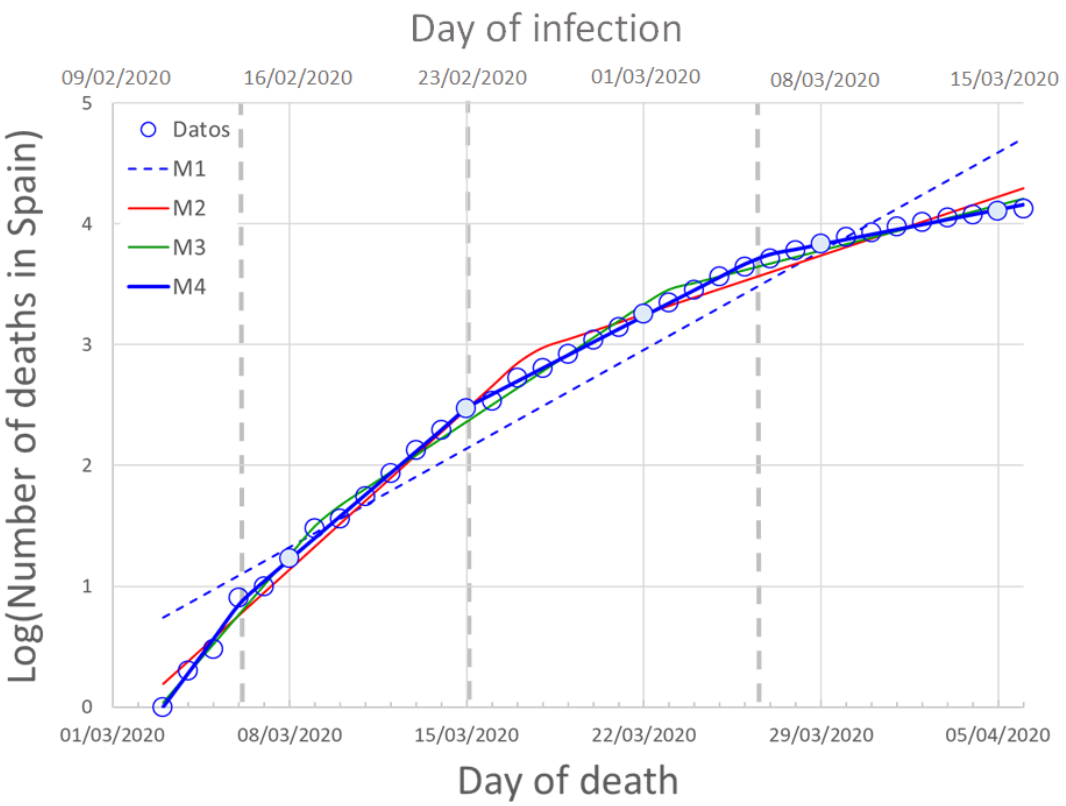

Figure 3. Segmented regressions fitted on the total number of fatalities attributed to COVID-19 infections in Spain. Estimated time of infection is provided in the upper X-axis. Filled points indicate Sundays. Broken vertical lines indicate the breaking points of the best fit (Model 4), which is shown with a thicker line.

The analyses performed on the number of fatalities from Madrid are consistent with the nationwide results. The model with two breaking points (Model 3) provided the best fit (Table 4). Fitted breaking points were placed on day 9.8 (15/3/20, estimated infection on 23/2/20) and 20.7 (26/3/20, estimated infection on 5/3/20) (Figure 4). The growth rate of the number of cases decreased by 56% (from 0.22 to 0.10) after the first breakpoint (15/3/20) and decreased again by 65% (from 0.10 to 0.03) after the second breakpoint (26/3/20). The model with one slope shift (on day 12.4, i.e. 17/3/20, estimated infection on 25/2/20) and one intercept shift (on day 8.2, i.e. 13/3/20, estimated infection on 21/2/20) showed a marginally significant improvement of goodness-of-fit relative to Model 2 (0.10<P<0.05), but a worse fit (larger RSS and adjusted R^2) than the model with two slopes shits (Model 3).

Table 4. Results of segmented regressions with an increasing number of breaking points, fitted on the number of fatalities attributed to COVID-19 infections in Madrid Region from 5/3/20 to 6/4/20.

Model	Intercepts	Slopes				Breakpoints (#days)			Goodness of fit				
	a1	b1	b2	b3	b4	u1	u2	u3	R2	Adj.R2	F	df1,df2	P
M1	0,8	0,8324							0,8992	0,8924	133,7497	2,30	**1,1338E-15**
M2	0,2	0,2081	0,0576			12,3			0,9924	0,9913	170,6796	2,28	**2,0698E-16**
M3	0,1	0,2217	0,0981	0,0344		9,8	20,7		0,9973	0,9966	23,2049	2,26	**1,6492E-06**
M4	0,1	0,2217	0,0998	0,0575	0,0266	9,8	19,2	24,5	0,9974	0,9966	0,7632	2,24	0,48

Model	Intercepts		Slopes		Breakpoints (#days)		Goodness of fit				
	a1	Δa	b1	b2	u1	u2	R2	Adj.R2	F	df1,df2	P
M2	0,2		0,2081	0,0576	12,3		0,8992	0,8924			
M5	0,09	-0,24	0,2322	0,0576	8,2	12,4	0,9939	0,9925	3,22	2,26	***0,056***

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

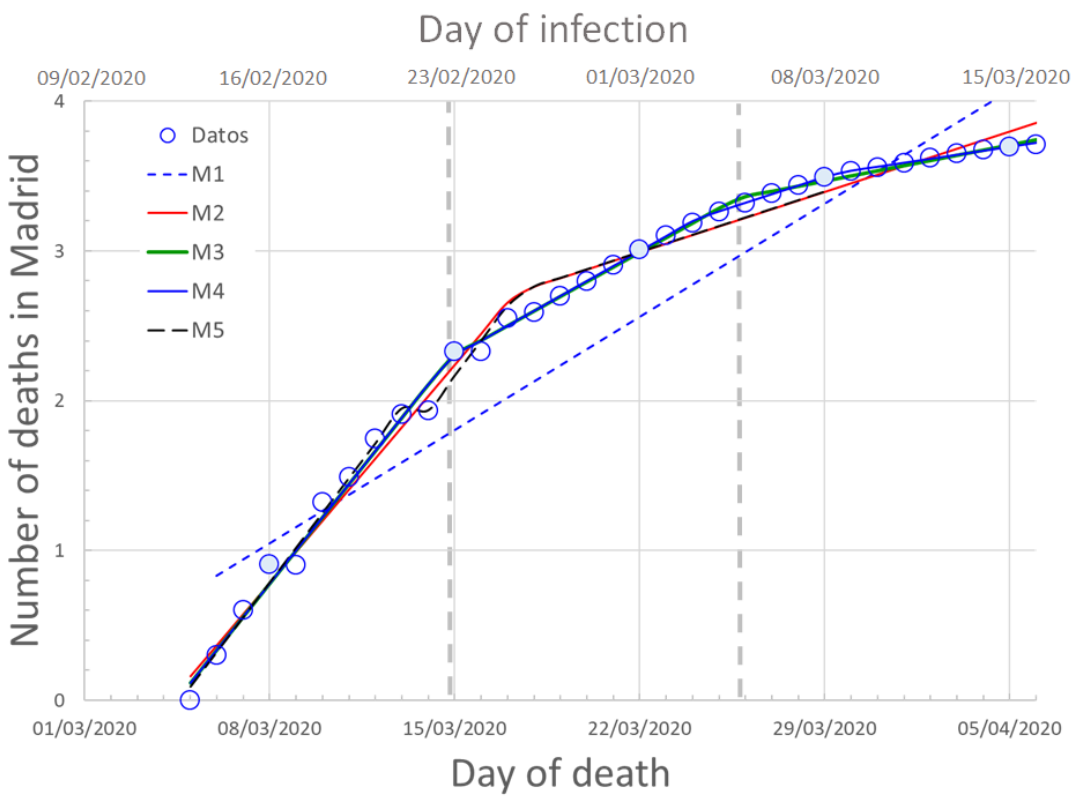

Figure 4. Segmented regressions fitted on the total number of fatalities attributed to COVID-19 in Madrid. Estimated time of infection is provided in the upper X-axis. Filled points indicate Sundays. Broken vertical lines indicate the breaking points of the best fit (Model 3), which is shown with a thicker line.

Discussion

The use of segmented regressions provided an objective procedure to identify thresholds of change during the evolution of COVID-19 pandemic in the whole of Spain, and the large conurbation of Madrid. The results of the four variables analyzed, involving two different spatial extents (whole country and regional), and two different lag times (10 days for cases, 21 days for fatalities), showed a consistent temporal pattern (Figure 5) that can be divided into three consecutive phases :

(1) A first phase, in the early moments of the epidemic, characterized by sudden (apparent) jumps in the number of infections, particularly conspicuous in Madrid, which were not accompanied by increases in the infection rate. This phase coincided with the detection of the first cases, imported from abroad; and the two jumps in the number of infections were synchronous with specific events of group infections (at sport events and nursing homes, respectively; red points 4 and 7 in Figure 5). Owing to the lag times mentioned above, the effects of this phase become perceivable between March 1 and March 8. During this phase, however, breaking points signaling one (Madrid) and two (Spain) consecutive decreases in the growth rates of the number of deaths were identified (37-40% and 56% decreases in Spain and Madrid, respectively). These breaking points could not be linked to any specific event or policy measure, and were probably related to the improvement of clinical procedures (detection and hospital treatment) following the detection, hospitalization - and, in some cases, death of the first cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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) A second phase, 1-2 weeks later, showing a highly consistent pattern of decreases in the growth rates of both the number of cases and the number of fatalities (49-65% and 40-56% reductions, respectively) at both geographical extents (Spain and Madrid). This inflection point, detected in the four variables analyzed, corresponds to infections taking place in the first week of March, following the issuing of preventive isolation measures (blue points 3-4 in Figure 5) and the increased awareness of the public opinion (as the perception of the increases in cases and deaths described in phase 1 emerged), but preceding the issuing of most social-distancing measures by the central and regional governments (blue points 5-7 in Figure 5). Owing to the lag times mentioned above, these decreases become perceivable between March 13-14 (for the number of cases) and March 26-27 (for the number of deaths). The perception of slowdown at March 13-14 was reinforced by the impact of the breaking points in the number of deaths detected during phase 1 (for infections taking place after February 23) – underscoring the decoupling of cause and perception (see below).

(3) A third phase, 2 weeks later, showing a consistent pattern of decrease in the growth rates of the number of cases detected at both spatial extents (54 and 59% reduction in Spain and Madrid, respectively), but no change in the growth rate of the number of deaths. This decrease coincides closely with the issuing of strong social-distancing measures (nationwide lockdown and border closure). Owing to the 21-days lag time between infection and death, it is likely that the period of time analyzed (extending to 21 days after the lockdown) was insufficient to detect similar changes in the number of deaths.

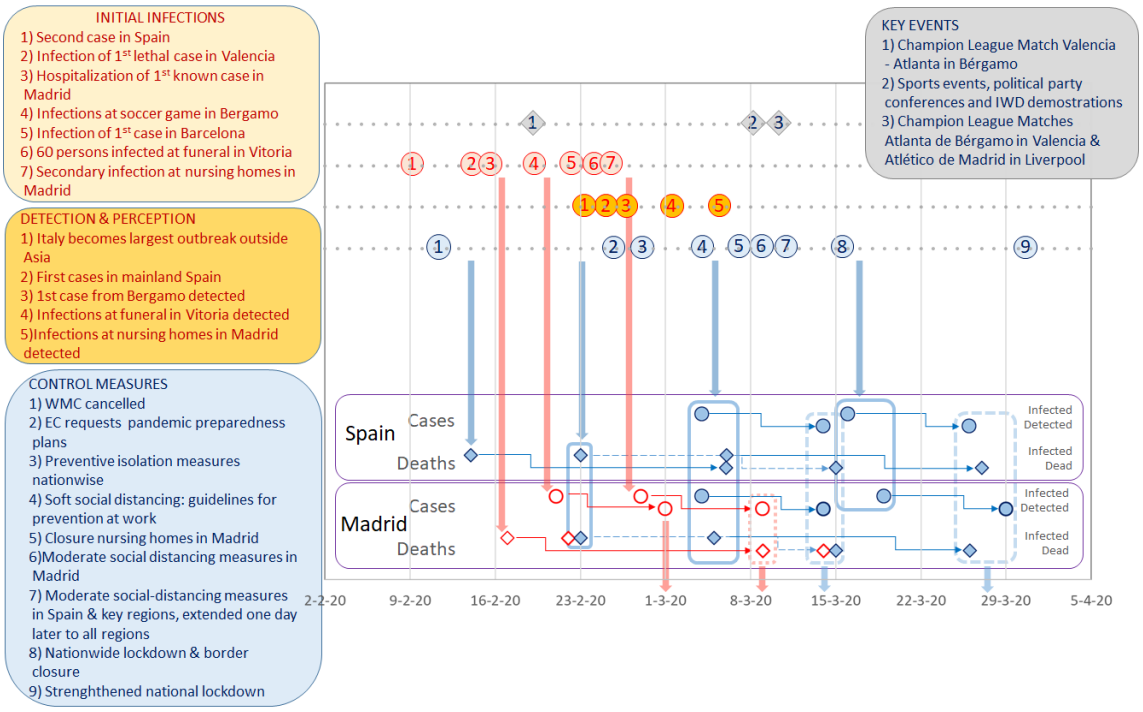

Figure 5. Timeline of the key events for the spread of COVID-19 in Spain, the increased awareness of the Spanish population, and Control Measures taken by the government. WMC stands for World Mobile Congress, and EC for European Commission. Upper vertical arrows indicate proposed causal links between events in the timeline and estimated infection dates of identified breaking points. Horizontal arrows link estimated infection dates and detection dates (for cases or deaths) of identified breaking points. Lower vertical arrows indicate the dates at which changes in infection dynamics (i.e. breaking points in the numbers of cases or deaths) could be perceived for the first time. Colour indicate increases

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

(red) or decreases (blue) in the slope (growth rate) or intercept (sudden increases in values) of the numbers of cases and/or deaths.

This basic analysis of the growth rates of cases and fatalities evidences the effectiveness of the contention measures taken by Spanish national and regional governments. The confinement of all population that could telework, the suspension of in-class teaching in schools and universities, and the closure of non-essential shops, bars and restaurants resulted in a marked downwards inflection in the curves of both infections and deaths. However, it is also remarkable the existence of an earlier breakpoint, corresponding to a decrease in the rate of infections in the first week of March. This breakpoint was the most consistent (present in the four variables) and the most supported (present in all but one models with one or more breakpoints), suggesting that this was the most important change of dynamic during the period of time analyzed. The dropdown in the rate of infections coincides with an increase of the awareness of the Spanish population (due to the reporting of a rapidly increasing number of cases and deaths, i.e. the delayed perception of the events of phase 1) and the issuing of official recommendations for the prevention and treatment of COVID infections, but precedes the legal enforcement of most social-distancing measures by the regional and central governments (see Figure 5, Table S1). Indeed, it takes place the week before the marked decrease in community mobility registered during the second week of March (from March 8 to March 15) in both Madrid and the whole of Spain (Google 2020). This suggests that a large proportion of the Spanish population changed their behaviour due to the combination of increasing awareness and the dissemination of preventive hygienic and distancing measures by the government and the media – with recourse to the examples of China, South Korea and Italy. It seems therefore fair to argue that, during this week, the population responded swiftly to the voluntary prevention and distancing measures suggested by the authorities and the media, and these were effective to slow down the early spread of the pandemic.

The results also indicate, however, that these measures were not enough to contain the spread of the virus – and the issuing of the first soft social-distancing measures by the regional governments had little impact upon it. The next breaking point, resulting in the strongest reduction if the growth rate of the number of cases, took place closely after the onset of the stronger measures of national lockdown and border closure. As in the previous phase, the population's response was swift and started already with the partial confinement measures, one week before (e.g. Google 2020) - surpassing broadly the expectations of 69 experts who predicted a collapse of the health system on March 25 (Catanzaro 2020; Arenas et al. 2020; Mitjà et al. 2020). These stronger isolation policies were probably instrumental in flattening the curve of infections, though not sufficient to date to bring it to a complete halt. Unfortunately, the data available to date does not allow assessing the effect of the further strengthening of the lockdown (i.e. closure of non-essential economic activities) enforced on 30-31 March, as compared with the previous measures.

What our results fail to show is the expected effects of certain key events pinpointed by the media on the spread of COVID-19 infections. The demonstrations and football matches of 7-8 March did not result in significant increases in neither infection rates nor in the number of infections (as reflected in the numbers of cases and fatalities). An inspection of the stalling-and-jump of the numbers of cases and deaths in Madrid, which triggered such public perception, indicates that they are inconsistent with it – since, given the lag times involved, their effects should have shown up 10 to 21 days later. The stalling-and-jump pattern, which

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

can also be observed in the previous and and following weekends (1st and 15th March, for the number of cases and the number of deaths, respectively), is more likely related to the phase 1 infection bouts shown in Figure 5. Alternatively, they could be caused by droppings in patient attendance to hospitals or case/fatality reporting during the weekend, with the subsequent increase in the following Monday.

With hindsight, it is clear that issuing strong social-distancing measures earlier would have increased their effectiveness, thereby saving more lives and reducing the collapse of the Spanish health system (see e.g. the SIR-based simulations of Casares & Kahn 2020). Our analyses evidence, however, that the responses of the population, media and authorities was slowed down by the perceptual trap created by the prolonged infection-to-detection and infection-to-death lags. It is somehow ironic that the rapid increase in the infections during the last weeks of february (described in phase 1) was only perceivable two weeks later - and, while the increments in number of cases and deaths were attributed to concurrent events that were most likely unrelated, they triggered a swift response precisely at a moment when early containment measures were already starting to work. Similarly, the results of early containment measures were perceived two weeks later, and attributed to a direct consequence of the national lockdown. Fortunately, expert advice to the government was aware of this perceptual trap and insisted on the necessity of stronger social-distancing measures –which have been narrowly sufficient, to date, to reach the objective of flattening the curves.

The spread and growth intensity of the COVID-19 pandemic is being driven by international connectivity (Coelho et al. 2020), which resulted in repeated exposition of the highly interconnected European population. In the case of Spain, this has been particularly important at international hubs such as Madrid or Barcelona, although contagions due to tourism and sport events also initiated the spread at least in Málaga and Valencia (southern and eastern coasts, respectively), spreading the disease and facilitating community transmission throughout the whole Spanish territory (see Table S1). This implies that the risk of new introductions from abroad will regain importance, should the objective of flattening the curve and taming the outbreak succeed in Spain. Our results, however, underscore the importance of the prolonged delays between infection and detection for the early (re)containment of the infection. Though population awareness and widespread adoption of preventive measures will likely slow down the advance of the disease in the future, this does not diminish the importance of implementing a much more active and complete system of early detection. This represents a formidable challenge for Spain, whose public health and R&D systems (the latter being instrumental for supporting testing efforts) were severely affected by the post-2008 austerity measures.

The analyses described above are highly consistent, and they were not strongly affected by changes in the extension of the data series (as more data became available) or in the fitting procedures (untransformed vs log-transformed data). However, their validity is potentially constrained by a number of caveats. We used data on the number of reported cases – i.e. severe cases subject to testing. In general, tests were restricted to those requiring hospitalization or belonging to risk groups, thus underestimating total infection numbers. Similarly, while at the beginning the number of deaths recorded in Spain included all deaths of patients who tested positive or showed a compatible symptomatology, during the peak of the pandemic several regional governments failed to count all deaths outside hospitals if they were not tested either pre- or post-mortem. Given the significant numbers of deaths

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

happening in nurse homes and particular residences, this unevenness in the measurement of COVID-related fatalities could have biased results, eventually flattening the curve. If such effect was significant, however, it should have resulted in changes in the intercept of the number of deaths' curve. Given the lack of evidence for such changes, we believe that our results will stand out if new data coming from autopsies or re-evaluations is available in the future. At any rate, these limitations emphasize that, at present, we can only use the available data as proxies for the actual rates of infection spread and lethality. Therefore, our results must be taken with due caution until more detailed studies surveying seroprevalence in the general populations (including a broad study launched by the Spanish government as we write; Mitchell 2020) are available.

It is also important to note that the relationship of the growth rates and breakpoints in the analysed data with the timeline provided in Figure 5 may be affected by both the variability in the time lags between the time of infection and the times of both diagnosis and death (see introduction), and the irregular reporting of both variables during weekends and among Spanish regions –particularly at the onset of the outbreak. However, the use of cumulative curves smoothes out considerably these effects, and provides in robust estimates of both breaking points and growth rates. Finally, the use of a family of models reaching a maximum of three breaking points, necessary to maintain model parsimony, may have also resulted in the lack of detection of additional breaking points. However, all models showed very high R^2 and adjusted R^2 values; and in three out of four models (all but the number of deaths in Spain), fitting an additional breaking point only resulted in slight, non-significant improvements in model fit. We might have also missed breaking points in the early stages of the infection, when values are lower and data more noisy. Indeed, heteroscedasticity in the original, untransformed data resulted in such a risk of underdetection; but using transformed data achieved homoscedasticity, and more breaking points were actually detected on earlier than on later dates.

To summarize, our analyses detected three inflection points in the dynamics of the COVID-19 pandemic in Spain: a first phase of rapid infection, accompanied by a decrease in the mortality rate most likely related to early improvements in the detection and treatment of this novel disease; a second phase of early reduction in the infection rate, chronologically related to public response to voluntary prevention and social-distancing requests by the authorities; and a third phase of major reduction in the infection rate, following the onset of hard social-distancing measures (nationwide lockdown and land border closure). It is also apparent that prolonged infection-to-detection and infection-to-death lags may have caused misperceptions of the impact of certain events (most notably, the 8th March demonstrations and sport events) and the effectiveness of certain measures, while being probably instrumental in triggering the widespread social response during phase 2 of the contagion. Therefore, in addition to the control measures undertaken, a more effective communication strategy that bridges the perceptual gap described could be instrumental in engaging the population into the harsh actions still necessary to fight the pandemic more efficiently.

Funding

The authors and their institutions did not receive at any time payment or services from a third party (government, commercial, private foundation, etc.) for any aspect of the submitted work.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

Disclosure of potential conflicts of interest

The authors have nothing to disclose.

References

Arenas, A. et al. (2020) Situación del CoVID-19 en España. Comentario Científico. https://964f74da-ef80-4d30-819c-e4bc9748a9a1.filesusr.com/ugd/5eba6f_b459def837ba4a38a4495aac6ca278e3.pdf

Casares, M. & Kahn, H. (2020) A dynamic model of COVID-19: Contagion and implications of isolation enforcement. Working paper D.T.2001, *Universidad Pública de Navarra*. Available at http://www2.unavarra.es/gesadj/depEconomia/repec/DocumentosTrab/DT2001.PDF

Catanzaro, M. (2020) El colapso que no fue. elPeriodico 9/4/2020. https://www.elperiodico.com/es/ciencia/20200409/colapso-ucis-no-fue-oriol-mitja-articulo-lancet-7922514

Clark P., Cookson C. & Hughes L. (2020) How the UK got coronavirus testing wrong. *Finantial Times* 27/3/2020. Available at https://www.ft.com/content/fa747fbd-c19e-4bac-9c37-d46afc9393fb

Coelho, M.T.P., Rodrigues, J.F.M., Medina, A.M., Scalco, P., Terribile, L.C., Vilela, B., Diniz-Filho, J.A.F. & Dobrovolski, R. (2020) Exponential phase of covid19 expansion is not driven by climate at global scale. *medRxiv*, 2020.04.02.20050773. Available at https://www.medrxiv.org/content/10.1101/2020.04.02.20050773v1

Dong, E., Du, H. & Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(20)30120-1 Map available at https://coronavirus.jhu.edu/map.html

Fernández, A. (2020a) Coronavirus en España (III): 6 medidas ante el abismo. Available at https://medium.com/@albertofffernandez/coronavirus-en-espa%C3%B1a-iii-6-medidas-ante-el-abismo-315bda1e1b5c

Fernández, A. (2020b) Coronavirus en España (I): Razones para Actuar con Urgencia y Contundencia. Available at https://medium.com/@albertofffernandez/coronavirus-en-espa%C3%B1a-i-razones-para-actuar-con-urgencia-y-contundencia-bba8eafa12f

Hanck, C., Arnold, M., Gerber, A. & Schmelzer, M. (2020) Introduction to Econometrics with R. 7.3 Joint Hypothesis Testing Using the F-Statistic. Available at https://www.econometrics-with-r.org/7-3-joint-hypothesis-testing-using-the-f-statistic.html

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T. & Neher, R.A. (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34, 4121-4123. https://academic.oup.com/bioinformatics/article/34/23/4121/5001388. Nextstrain accessible at https://nextstrain.org/ncov. Repeated accesses.

Google (2020) COVID-19 Community Mobility Report, 29/3/20 version. Available at https://www.gstatic.com/covid19/mobility/2020-03-29_ES_Mobility_Report_en.pdf

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20059345; this version posted April 14, 2020. 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 .

John Hopkins University Coronavirus Resource Center (2020). Available at https://coronavirus.jhu.edu/map.html. Repeated accesses.

Linton, N.M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A.R., Jung, S.M., ... & Nishiura, H. (2020) Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9, 538. https://www.mdpi.com/2077-0383/9/2/538

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *International journal of Antimicrobial Agents*, 105924. https://www.sciencedirect.com/science/article/pii/S0924857920300674

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, doi:10.7326/M20-0504. https://annals.org/AIM/FULLARTICLE/2762808/INCUBATION-PERIOD-CORONAVIRUS-DISEASE-2019-COVID-19-FROM-PUBLICLY-REPORTED

Mitchell, D. (2020) Spanish Government plans to test 30,000 families to discover the spread of the Global Coronavirus pandemic and if containment can be lifted. *EuroWeekly* 07/04/2020. https://www.euroweeklynews.com/2020/04/07/spanish-government-plans-to-test-30000-families-to-discover-the-spread-of-the-global-coronavirus-pandemic-and-if-containment-can-be-lifted/

Mitjà, O., Arenas, À., Rodó, X., Tobias, A., Brew, J., & Benlloch, J. M. et al. (2020) Experts' request to the Spanish Government: move Spain towards complete lockdown. The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30753-4/fulltext#%20

Ravelo, J.L. & Jerving, S. (2020) COVID-19 — a timeline of the coronavirus outbreak. https://www.devex.com/news/covid-19-a-timeline-of-the-coronavirus-outbreak-96396

Russell, T.W., Hellewell, J., Abbott,S., Jarvis C.I., van Zandvoort, K., CMMID nCov working group, Flasche, S., Eggo, R., Edmunds, W.J., Kucharski, A.J. (2020) Using a delay-adjusted case fatality ratio to estimate under-reporting. *Centre for Mathematical Modelling of Infectious Diseases Repository*. https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., …, Ferguson, N. (2020) Estimates of the severity of COVID-19 disease. *medRxiv* 2020.03.09.20033357; doi:10.1101/2020.03.09.20033357. https://www.medrxiv.org/content/10.1101/2020.03.09.20033357v1

@CoronaSurveys Team (2020) Measuring Icebergs: Using Different Methods to Estimate the Number of COVID-19 Cases in Portugal and Spain. https://github.com/GCGImdea/coronasurveys/blob/master/reports/2020-03-29-CaseEstimation.pdf