Retrospective Methodology to Estimate Daily Infections from Deaths (REMEDID) in COVID-19: the Spain case study

David García-García[1], María Isabel Vigo[1], Eva S. Fonfría[2], Zaida Herrador[3], Miriam Navarro[2], Cesar Bordehore [2,4*]

[1] Department of Applied Mathematics, University of Alicante, Alicante, Spain.

[2] Multidisciplinary Institute for Environmental Studies "Ramon Margalef", University of Alicante, Alicante, Spain.

[3] National Centre of Tropical Medicine, Instituto de Salud Carlos III, Madrid, Spain.

[4] Department of Ecology, University of Alicante, Alicante, Spain.

Corresponding author:

Dr. Cesar Bordehore

Department of Ecology, University of Alicante, Campus San Vicente del Raspeig, 03690, Spain

Tel: +34 652995524, Fax: +34 965903625

cesar.bordehore@ua.es

Keywords: COVID-19, MoMo, mortality, deaths, epidemiology, Case Fatality Ratio, first infection, infections inference

Abstract

The number of new daily infections is one of the main parameters to understand the dynamics of an epidemic. During the COVID-19 pandemic in 2020, however, such information has been underestimated. Here, we propose a retrospective methodology to estimate daily infections from daily deaths, because those are usually more accurately documented. The methodology is applied to Spain and its 19 administrative regions. Our results showed that probable infections were between 34 and 42 times more than the official ones on 14 March, when national government decreed the national lockdown. The latter had a strong effect on the growth rate of virus transmission, which began to decrease immediately. Finally, the first infection in Spain may have occurred on 11 January 2020, around 40 days before it was officially reported. In summary, we state that our methodology is adequate to reinterpret official daily infections, being more accurate in magnitude and dates.

1. Introduction

The key parameter to understand and model the evolution of the COVID-19 pandemic is the number of daily infections. Despite its importance, however, it is difficult to estimate reliable data due to the bias of official figures[1]. For symptomatic patients, there are delays of 5.79 days from infection to symptom onset and 5.82 days from symptom onset to diagnosis[2]. In addition, the presence of asymptomatic people and those with mild symptoms hinder their detection. For example, 86% of infections were undetected in Wuhan, China, prior to 23 January 2020, when travel restrictions were imposed[3]. Additionally, not all patients with COVID-19 compatible symptoms are tested, especially at the beginning of the pandemic. So, it is widely assumed that the reported infections are just a fraction of the actual ones. In Spain, a national seroprevalence study based on 38,000 random tests estimated that only 8.1% (CI 95%: 7.4%-8.6%) of the actual infections had been detected in Spain[4].

Another problem is due to the inconsistency of daily infection time series and the variability of the case definitions. For example, on 13 January 2020, China changed the parameters to confirm new cases and ~15,000 new infections were counted in a single day, while daily infections during the previous week were less than 3,500[5]. In Spain, the method of counting official numbers of infections has been modified several times since the pandemic was decreed. On 22 April 2020, a total of 213,024 cases were reported from positive PCR and antibodies tests[6]; however, total cases decreased to 202,990 the next day and thereafter, because only positive PCR were included in the total cases[7]. Accumulated cases on a given day are obtained by adding the new cases that day to the accumulated cases of the previous day. That has not always been the method, however,

because some Spanish regions reviewed and modified the cases in previous days[8,9]. This inconsistency of time series hampers any kind of accurate analysis.

High fidelity time series of each parameter of an epidemic are crucial to run reliable epidemiological models. The number of new infections per unit time is one of the essential parameters. The official data do not reflect the actual day of infection nor the real number of infections. To overcome this limitation, we propose a retrospective methodology to infer daily infections from daily deaths, believing the number of deaths to be a more reliable parameter than the official number of daily infections[1]. These estimated infections are closer to the date of infections, avoiding delays from the incubation period and symptom onset to diagnosis. Then, these time series of daily infections would help to better understand the pandemic dynamics and to quantify the effectiveness of the different containment strategies, avoiding the bias of the official data. In addition, those numbers can be used to feed epidemiological models with more realistic data. Finally, we will apply this methodology in Spain as a whole and each of its 19 administrative regions (17 autonomous communities and 2 autonomous cities).

2. Methods

2.1 REMEDID

We named this methodology REMEDID, **RE**trospective **M**ethodology to **E**stimate **D**aily **I**nfections from **D**eaths and it is described as follows. Date of infection (*DI*) of an

individual may be estimated from the death date (*DD*) by subtracting the incubation period (*IP*) and illness onset to death (*IOD*) periods, hence

$$DI + IP + IOD = DD.$$

Therefore, *DI* can be estimated from *DD* as far as *IP* and *DD* are known; however, neither *IP* nor *IOD* are fixed values. On the contrary, they are random variables that can be approximated by probability distributions. From dozens of cases in Wuhan, Linton et al.[10] approximated *IP* for COVID-19 with a lognormal distribution, X_{IP}, with a mean of 5.6 days and median of 5 days, and *IOD* with a lognormal distribution, X_{IOD}, with a mean of 14.5 days and median of 13.2 days. Therefore, the infection to death period is a random variable, X_{IP+IOD}, that follows the distribution $X_{IP} + X_{IOD}$.

Although the addition of two lognormal distributions does not follow any commonly used probability distribution, its probability density function (PDF) can be estimated convolving the PDF of the two variables[11]. If *g(t)* and *h(t)* are the PDF of *IP* and *IOD*, respectively, then their convolution defines the PDF of X_{IP+IOD},

$$f(t_0) = \int_{-\infty}^{+\infty} g(t_0 - t) \cdot h(t)\, dt,$$

where t_0 is a positive real number representing days from infection. *f(t)* shows a mean of 20.1 days, and a median of 18.8 days. Figure 1 shows *g(t)*, *h(t)*, and *f(t)*, as well as a lognormal PDF with the same mean and median as *f(t)* for comparison. Note that the probability that X_{IP+IOD} is less than or equal to 33 days is 0.95.

[Figure 1]

Given a time series of deaths produced by the illness, *x(t)*, we estimate the infection time series that produced such deaths, *y(t)*. If we assume a case fatality ratio (CFR) of 100%, *y(t)* would represent all daily infections. To calculate *y(t)*, we use the following likelihood-based estimation procedure.

Because the relative likelihood that a given infection will produce a death *t* days later is *f(t)*, the infections at a given time t_0 can be estimated as

$$y(t_0) = \int_{t_0}^{+\infty} x(t) \cdot f(t - t_0)\, dt.$$

In practice, *x(t)* is a discrete time series and could be written as *x(n)*, where *n* is an integer representing entire days. Let *F(n)* be a discrete approximation to *f(t)* as follows:

$$F(n) = \int_{n}^{n+1} f(t)\, dt.$$

Then,

$$y(n_0) = \sum_{n=n_0}^{+\infty} x(n) \cdot F(n - n_0). \qquad (1)$$

A similar analysis has been previously used to estimate a time-variable effective reproduction number for the severe acute respiratory syndrome (SARS) in 2003[12]. For a different CFR, the associated/estimated infections are estimated from *y(n)* in Equation 1 as

$$\text{Inferred infections}(t) = \frac{y(n)}{CFR} \times 100. \qquad (2)$$

Equations 1 and 2 summarize the REMEDID algorithm.

2.2. Data

2.2.a. Official numbers of COVID-19 infections and deaths

COVID-19 daily infections and deaths for Spain are reported by the Centro de Coordinación de Alertas y Emergencias Sanitarias (CCAES), which depends on the Ministry of Health, Social Services and Equality. Time series for the 19 regions of Spain can be downloaded from the national organization that compiles and publishes the data from all regions, the Instituto de Salud Carlos III (ISCIII; https://covid19.isciii.es/). For consistency, time series for Spain are estimated by adding the time series of the 19 regions. We used ISCIII data up to 16 April, and thereafter the number of new infections and deaths appeared in the daily situation reports published by CCAES (https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm). In those reports, daily data were replaced by deaths during

the week after 23 May. From that date, we obtained daily infections by dividing the weekly infections by 7 and used the 7-day running means. Those data will not show frequencies higher than 1 per 7 days, but are still useful to study the pandemic dynamics. Finally, we obtained daily infections from the first documented case on 20 February to 6 June 2020 and daily deaths from the first official death on 4 March 2020 to 6 June.

To determine a more realistic CFR in Spain, a longitudinal seroprevalence study is being conducted by ISCIII in 3 phases. The first phase was completed from 27 April to 11 May and the results published on 13 May[4]. In that study, ~60,000 people were randomly selected and ~38,000 of them agreed to be interviewed and tested. The study found that 5.0% (CI 95%: 4.7%-5.4%) of the participants presented SARS-Cov2 IgG antibodies. Extrapolation of this result to the ~47 million of the population in Spain yields 2.35 million infections on 11 May, which, with 26,920 accumulated COVID-19 deaths on 11May, would give a CFR of 1.14% (CI 95%: 1.06%-1.22%). We assumed this value in our analysis for Spain. The CFR of each region was estimated analogously and summarized in Table 1. Asturias, Extremadura, País Vasco, and La Rioja had CFRs greater than 1.5%. In contrast, Andalucía, Islas Baleares, Canarias, Ceuta, Melilla, and Murcia had CFRs less than 0.8%.

2.2.b. Excess of expected deaths from MoMo

Expected deaths by any cause, with an error band representing a 99% confidence interval, are modelled by the Mortality Monitoring (MoMo) surveillance system for 22

countries in Europe (www.euromomo.eu). In Spain, MoMo depends on the ISCIII[13] and data are available at https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html#nacional. MoMo also collects daily deaths from any cause from the Spanish Instituto Nacional de Estadística (INE) and the notaries and civil registries. Daily deaths within the error estimates of the expected deaths are considered usual. Any deviation can identify unusual mortality patterns, as is the case of the COVID-19 pandemic. The useful variable here is the excess deaths (*ED*), estimated as the difference between actual and expected deaths. The error band for *ED* can be estimated from the error band of the expected deaths by subtracting the latter from the actual deaths. By cautiously assuming that all significant *ED* is caused by COVID-19, that can be used to estimate non-reported COVID-19 deaths. MoMo usually takes several days to completely update its dataset so, the last week of the time series is not used. As a result, the MoMo time series used spanned from 4 March (first official COVID-19 death) to 1 June 2020 (data downloaded on 8 June).

Because on 11 May MoMo *ED* in Spain was 43,389 deaths, the CFR for the country was estimated similarly as Section 2.1 to be 1.85% (CI 95%: 1.71%-1.96%). The CFR for each region was estimated similarly and provided in Table 1.

3. Results

3.1. COVID-19 vs. MoMo

The COVID-19 official deaths and MoMo *ED* time series overlap for the period from 4 March to 1 June 2020 for Spain and its 19 regions (Figure 2). Table 1 shows the accumulated values before 11 May, which were used to estimate the CFR for Spain and its 19 regions accordingly to the first phase of the National Seroprevalence Study that ended that day[4]. In Spain, MoMo *ED* showed 16,216 accumulated deaths more than the official COVID-19 deaths, which was beyond the error band. That difference comes basically from the four regions with more official COVID-19 deaths (Madrid, Cataluña, Castilla-La Mancha, and Castilla y León), with 5,227, 5,842, 2,427, and 1,630 more accumulated deaths, respectively. This gives a CFR estimated from MoMo *ED* that is double the CFR estimated from COVID-19 deaths for these regions. The estimated values of CFR do not show great changes for the other regions, except for Ceuta and Melilla, which increased their CFR from MoMo *ED*. This result is probably biased due to their low populations and number of deaths.

[Table 1]

[Figure 2]

3.2. Infections estimated from COVID-19 deaths

Infections associated with the COVID-19 deaths were calculated from Equation 1 and shown in Figure 3. By definition of Equation 1, $y(n)$ represents all daily infections assuming CFR= 100%. Note that the obtained time series is 33 days shorter than death time series. We did this to ensure that all infections were estimated with more than 95% of their associated deaths (see Section 2). The estimated infections in Spain reached a maximum on 14 March 2020 (Table 2), the day that the national government decreed the state of emergency and national lockdown. It means that the measures adopted had

an immediate effect, but it was not officially observed until the maximum of recorded infections was reached on 30 March 16 days later (estimated from the 14-day running mean time series). This delay is longer than the expected ~11 days from adding the mean incubation period of 5.78 days[2], and the illness onset to diagnosis mean period of 5.2 days[14, 15]. On the other hand, the maximum number of deaths was reached on 2 April, 19 days after the inferred infection maximum. This delay is similar to the 20 days expected from infection to death (Figures 1 and 3).

[Figure 3]

Figure 4 shows the estimated daily and accumulated infections for Spain and its 19 regions. Note that in Spain, estimated daily infections are an amplified version of *y(n)* in Figure 3 according to Equation 2. In Spain, the first estimated infection was on 11 January (Table 2), 40 days before the first infection was officially reported on 20 February. Spain reached the maximum number of estimated daily infections on 14 March, when the state of emergency and lockdown were enforced. On that date, official daily infections were 1,832; however, estimated daily infections were 62,860 (CI 95%: 59,088-67,888), 34 times more (Table 3). Thus, the documented daily infections were only 2.9% (CI 95%: 2.7%-3.1%) of estimated infections. Officially there were 223,054 accumulated infections before 28 April (last day of estimated daily infections from MoMo *ED*, see next section), but with the assumed CFR there were 2,405,617 total cases (CI 95%: 2,598,063-2,261,274). Note that this value is similar to the 2.35 million infections estimated by the first phase of the National Seroprevalence Study (ISCIII 2020) for the period 27 April – 11 May. This 2.40 million is 11 times larger than the official data, suggesting that only 7.8% (CI 95%: 7.2%-8.2%) of the probable infections

were detected. Although detection improved from 14 March to 28 April, it was still far below our estimation of the probable infectee population.

Almost all regions showed a delay of 1 month or more between the first estimated case and the first documented case. The exceptions were Islas Baleares, and Canarias, with shorter delays of 25 and 28 days, respectively (Table 2). Note that these regions are the most isolated in Spain. Although Canarias and Islas Baleares were the first regions reporting cases, their estimated infections ranked 10 and 14, respectively. In contrast, Madrid, Cataluña, and País Vasco were the 4th, 5th, and 8th regions reporting their first case, but their estimated infections ranked as the 1st, 2nd, and 3rd regions, respectively. Most of the regions reached maximum daily infections around 14 March. Also note that Madrid reached the maximum on 11 March when the educational centres were shut down. The last region reaching the peak was Asturias. All regions showed one maximum, except for Galicia and Navarra with two maxima each. In Galicia, the second maximum was lower than the first and reached on 11 April, 24 days after the first one. In Navarra, the second maximum was larger than the first and reached on 31 March, 16 days after the first one. Islas Baleares and Canarias did not show a second maximum, but had inflection points 9 and 15 days after their maxima, respectively.

On 14 March, Asturias had detected 12.1% (CI 95%: 8.7%-16.7%) of estimated daily infections. The other regions, however, detected only 1.1 to 7.9% of the probable infections (Table 3). The accumulated infections in the regions before 28 April were similar to the entire country, which generally under-detected infections (Table 4). Excluding Melilla (57.0%) and Ceuta (27.0%), all regions detected between 6.3% and

16.8% of probable infections. Aragón, Castilla – La Mancha, Castilla y León, Cataluña, Extremadura, and Madrid had the poorest detection (Tables 3 and 4).

[Table 2]

[Table 3]

[Figure 4]

3.3. Infections from MoMo Excess of Deaths

If we assume that MoMo *ED* accounts for both recorded and non-recorded COVID-19 deaths, negative deaths are meaningless and they were set to zero. Then, the associated daily infections can be estimated, as in Section 3.2, with a CFR of 100% from MoMo *ED* for Spain (Figure 5). Note three main differences between this time series and that estimated from official COVID-19 deaths: (1) time series estimated from MoMo *ED* was 7 days shorter due to the updating delay on MoMo statistics; (2) MoMo data present an error band that was inherited by the estimated infections; (3) MoMo *ED* estimated infections reached a maximum of 1,428 (CI 99%: 1,322-1,529), doubling the 720 inferred daily infections from official COVID-19 deaths in Figure 3. This is because maximum MoMo *ED* was 1,569 (CI 99%: 1,462-1,668%) and maximum COVID-19 official deaths was 929, both estimated from the 14-day running mean time series. The maximum of inferred infections was reached on 13 March, just one day prior to the state of emergency and lockdown. The expected and observed delays with respect to official infections and MoMo *ED* were similar to those observed for estimated

infections from official COVID-19 deaths. Error estimates of the estimated infections were computed in a conservative way as follows: larger (smaller) error band is calculated combining the larger (smaller) error estimate of MoMo *ED* and the smaller (larger) value in the confidence interval of CFR.

[Figure 5]

The estimated daily infections were estimated for Spain and its 19 regions by use of the CFR values of each region reported in Table 1. Figure 6 shows such estimated daily infections, as well as the accumulated infections. In Spain, the first estimated infection from MoMo *ED* happened on 11 January, with an error estimate from 5 to 14 January, 37 to 46 days before the official first infection on 20 February (Table 2). The maximum number of daily infections was 77,838 (conservative CI: 67,754- 90,037) reached on 13 March. In Table 3 we report the daily infection on 14 March for comparison with the infections estimated from official COVID-19 deaths, which showed 14,978 fewer daily infections. This difference is remarkable given that the CFR used for MoMo *ED* is larger (the larger the CFR, the lower the estimated infections). Note that if the true CFR, which is unknown, were used, the infections estimated from MoMo *ED* would double those inferred from official COVID-19 deaths, as happened when a CFR of 100% was used (Figures 3 and 5). When compared to the official daily infections, MoMo estimated infections showed 42 times more cases, only 2.4% (2.0-2.7%) of the probable infections were documented for 14 March.

Table 2 shows the estimated date of first infection for Spain and by region. Note that the first cases estimated from MoMo *ED* in Spain were on 11 January and in Madrid on 10 January, which is possible because significant excess deaths in a region may not become significant for the whole country. In general, the maxima of daily infections were closer to 14 March when they are inferred from MoMo *ED* than from official COVID-19 deaths. All regions showed a unique maximum, except for Asturias, Islas Baleares, and Ceuta, which showed two maxima. Also note that those regions were different from those with two maxima in Figure 4. Daily infections on 14 March were comparable to those estimated from official COVID-19 deaths. The CI of MoMo estimated infections is so conservative, that they contained those from official COVID-19 estimated infections, except for Castilla – La Mancha, Cataluña, and Madrid (Table 3). Such differences were expected because those three regions presented significant differences in accumulated deaths between MoMo *ED* and official COVID-19 deaths before 14 May (Table 1). The accumulated infections before 28 April in all regions showed similar values, within the error estimate, between infections estimated from COVID-19 deaths and MoMo *ED* (Table 5).

[Figure 6]

4. Discussion

Infection dynamics is key to understand and model the COVID-19 pandemic. Nevertheless, reported infections in Spain (and almost worldwide) are of poor quality qualitatively and quantitatively. Specifically, reported infections are underestimated and dates are about 16 days later than the dates of infections. According to the first phase of the National Seroprevalence Study, only 8.1% of infections were documented in Spain[4].

This misreporting is mainly due to: (1) the lack of testing for asymptomatic and mild symptomatic people[3], which could have been detected with a "test, track and trace" strategy, as done in South Korea, China, and Singapore[16]; (2) deaths outside of the hospitals, with significant cases from nursing homes for the elderly, where about 19,000 deaths with COVID-19 or similar symptoms occurred, but only those diagnosed prior to death were documented as COVID-19 deaths[17]. This poor quality of data hinders our options to study the infection dynamics.

To overcome these difficulties, we propose the REMEDID algorithm, which calculates time series for daily infections from daily death time series. This algorithm needs three inputs to be applied:

(1) PDF of the period from date of infection to date of death. In this study, that was calculated by combining the PDF of two random variables, the Incubation Period and the Illness Onset to Death period, as estimated by Linton et al.[10] for cases in Wuhan, China. The incubation period can be assumed to be similar in China and elsewhere, because it is a characteristic of the virus, but the illness onset to death period could be affected by the type of health care and, thus, by each national health system[2]. There is no such information available for Spain to date. Nonetheless, because it would be desirable to use local distributions of Illness Onset to Death, when this information becomes available for Spain, the calculations could be redone easily by applying the MATLAB code available in the Supplementary Material.

(2) Time series of daily deaths. We use two sources, the official COVID-19 death figures from the CCAES and the MoMo *ED*. We proceed in this way because official

COVID-19 deaths have been underestimated, especially in the regions of Castilla–La Mancha, Castilla y León, Cataluña, and Madrid, because non-hospital deaths and cases during the early pandemic were somewhat undertested (Table 1 and Figure 2). We expect that undocumented COVID-19 deaths appear in the excess of expected deaths from any cause calculated from MoMo[13]. Interpretation of those results must be made carefully, however, due to the following reasons. First, *ED* is a statistical variable calculated from what happened in previous years. Second, some expected deaths, such as those produced by traffic and work accidents, may have been reduced due to the national lockdown that began on 14 March 2020 and gradually has been eased. Nonetheless, the magnitude of that decrease in fatalities is insignificant compared to the daily death toll from COVID-19. Traffic deaths for April 2019 and 2020 were 74 and 30, respectively[18], while work deaths for first trimester of 2019 and 2020 were 109 and 144, respectively[19]. Hence, the reduction in deaths from traffic and work accidents were less than 5 deaths per day, an insignificant number in comparison to the more than 300 COVID-19 deaths per day in average from March to May. On the other hand, some non-COVID-19 deaths may have been influenced by other pandemic-related factors[20], such as breakdown in medical follow-up (especially chronic conditions and among older patients), delays in attending to the healthcare appointments and/or being attended due to the collapse of the Health System, transplant delay and reduction in donors' numbers from 7.2 to 1.2 per day[21], the interventional cardiology activity decreased between 40% and 81%[22]. Thus, MoMo *ED* do not exactly correspond to COVID-19 deaths and must be interpreted carefully. A detailed study of the changes in other sources of mortality is recommended. In any case, it is a valuable source of data that must be explored considering the huge bias in official data of daily infections.

(3) CFR estimate. CFR is defined as the ratio of deaths among total infections. Thus, the lower the documented infections, the larger the CFR will be. For that reason, under detection of infections leads to overestimation of the CFR. For example, in Spain, on 28 April, official data showed 223,054 accumulated infections and 24,275 deaths, a CFR of 10.88%[23]. However, this value was overestimated because infections are underestimated. Preliminary results of the first phase of the National Seroprevalence Study in Spain with random testing showed that 5.0% of the population (~47 million) had already been infected[4]. As CFR also depends on death records, we calculated a CFR of 1.14% (CI 95%: 1.06%-1.22%) from official COVID-19 deaths and a CFR of 1.85% (CI 95%: 1.71%-1.96%) from MoMo *ED*. Similar seroprevalence studies are desirable in other countries, because CFR from official data are so variable. For example, on 28 April CFRs of 0.56%, 7.18%, and 13.53% were reported for Iceland, the world, and Italy, respectively[24]. In addition, we assumed a constant CFR throughout the study, although introduction time dependence on the CFR would enhance the infectee estimates. A constant CFR does not consider who is affected by saturation of the health systems (due to limited human and material resources). Moreover, CFR among severe cases (mostly detected) is much greater than that among asymptomatic and mild cases (mostly undetected). Consequently, the CFR calculation has a high degree of uncertainty. Because information on the CFRs will improve, calculations in this study can be redone easily using the MATLAB code provided as Supplementary Material.

We ran the REMEDID algorithm to provide the estimated time series of infections for Spain and for its 19 regions for the period from 10 January to 28 April 2020. These time series provide valuable information to understand the time evolution of the pandemic. Our main findings for Spain are:

(1) On 14 March, when the national government decreed the state of emergency and national lockdown, estimated infections were between 34 and 42 times larger than officially documented ones.

(2) The National lockdown had a strong effect on the transmission of the virus, as shown by a rapid slope decrease around day 14 March.

(3) The first infection in Spain probably occurred on 11 January, 40 days before than officially reported (20 February).

The REMEDID algorithm has strengths and limitations. First, it uses the number of deaths, which are more accurately recorded than infections. Second, it allows elucidation of the date of the infections estimated during the seroprevalence studies, which only determines the total number of infections. Thus, the REMEDID algorithm complements the seroprevalence studies. Third, the estimated daily infections indicate the probable date of infection better than the official numbers. Note that official infections are delayed due to the inclusion of two periods, i.e., incubation period (from infection to illness onset) and illness onset (symptoms) to diagnosis. Therefore, although the maxima of infections theoretically should have coincided with the state of emergency and the national lockdown on 14 March, official infections in Spain were maximum 16 days later. Infections estimated from death data do not show such delay, however. Determination of the actual day of infection is very important for evaluation of the immediate effect of the implemented countermeasures. It must be said that in Madrid, the estimated infections reached their maxima on 11 March, when regional government warned the population to stay at home, and schools and universities were closed, forcing 1.2 million students to stay at home. Moreover, overall recommendations on disease control and social distancing were given by the Ministry of

Health to the general public since de beginning of March (see https://www.mscbs.gob.es/en/profesionales/saludPublica/ccayes/alertasActual/nCov-China/ciudadania.htm).

Four, the algorithm can be applied to any area, independent of population size, but being cautious with low populations. For example, we applied REMEDID methodology to a city of 84,777 inhabitants such as Ceuta, to a big region such as Madrid with more than 6.6 million inhabitants, and to all Spain with 47 million. This versatility is useful to study different epidemic dynamics, reflected in the variety of shapes and slopes of deaths and infections, showing the heterogeneity of the behaviour of the epidemic at each region, each one with different population size and density, social behaviour and, thus, transmission pattern. For instance, infection evolution in several regions presented two maxima, showing unique dynamics that must be addressed individually. Knowledge of the real epidemic dynamics of different population nodes is key to succeed in modelling attempts, because we could calculate the sole effect of group size[25]. This spatially-explicit information, in combination with population size per node and mobility would allow us to use a metapopulation approach in future models[26-28].

Among the methodology limitations, the most important is that it can only be implemented retrospectively. Thus, these estimates cannot be used to control the pandemic in real time. But considering that different regions or countries are at different stages at the same time, results of this methodology for the first communities could be applied elsewhere. Furthermore, it is useful to enhance models and improve our knowledge of the pandemic dynamics, including the effectiveness of the different measures adopted to flatten the curve and design safe post-lockdown measures. In

addition, more realistic and accurate models could be obtained by use of the more realistic daily number of infections provided by REMEDID, which, in turn, would improve the models outcome and enhance comparisons of different lockdown and post-lockdown measures[29].

We believe that REMEDID methodology applied to daily COVID-19 deaths (if accurately reported) or to MoMo *ED* could be useful to analyse the dynamics of the pandemic retrospectively and more accurately quantify the real daily infections with respect to the official numbers. We only need the CFR and, for greater precision, the PDF of Infections to Death or, alternatively, the PDF of Incubation Period and Illness Onset to Death. This approach could be implemented anywhere, improving our understanding of the dynamics of the pandemic, and the effectiveness of the confinement measures. This is of high importance to prepare strategies to face successfully and reduce the effects of future epidemic episodes.

References

1. Holmdahl, S. M. I. & Buckee, D. P. C. Wrong but Useful — What Covid-19 epidemiologic models can and cannot tell us. *N Eng J Med*. DOI: 10.1056/NEJMp2016822 (2020)

2. Fonfría, E.S. et al. Essential epidemiological parameters of COVID-19 for clinical and mathematical modeling purposes: a rapid review and meta-analysis. medRxiv 2020.06.17.20133587; doi: https://doi.org/10.1101/2020.06.17.20133587 (2020)

3. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, Issue 6490, 489-493, DOI: 10.1126/science.abb3221 (2020).

4. Instituto de Salud Carlos III. Estudio ENE-COVID19: Primera ronda estudio nacional de sero-epidemiología de la infección por SARS-COV-2 en España. Informe preliminar, 13 May 2020. Available online at https://www.ciencia.gob.es/stfls/MICINN/Ministerio/FICHEROS/ENECOVID_Informe_preliminar_cierre_primera_ronda_13Mayo2020.pdf (ISCIII, 2020)

5. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 24, published online on February 2020, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200213-sitrep-24-covid-19.pdf?sfvrsn=9a7406a4_4 (WHO, 2020)

6. Centro de Coordinación de Alertas y Emergencias Sanitarias. Actualización nº 84. Enfermedad por el coronavirus (COVID-19), published online on 23 April 2020, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/Actualizacion_84_COVID-19.pdf (CCAES, 2020a).

7. Centro de Coordinación de Alertas y Emergencias Sanitarias. Actualización nº 85. Enfermedad por el coronavirus (COVID-19), published online on 24 April 2020, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/Actualizacion_85_COVID-19.pdf (CCAES, 2020b).

8. Centro de Coordinación de Alertas y Emergencias Sanitarias. Actualización nº 111. Enfermedad por el coronavirus (COVID-19), published online on 20 May 2020, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/Actualizacion_111_COVID-19.pdf (CCAES, 2020c).

9. Centro de Coordinación de Alertas y Emergencias Sanitarias. Actualización nº 112. Enfermedad por el coronavirus (COVID-19), published online on 21 May 2020, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/Actualizacion_112_COVID-19.pdf (CCAES, 2020d).

10. Linton et al. 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** (2), 538, DOI: 10.3390/jcm9020538 (2020).

11. Uspensky, J. V. Introduction to mathematical probability. Ed. McGraw-Hill Book Company, New York and London (1937).

12. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* **160**, 509-516 (2004).

13. León-Gómez, I. et al. Exceso de mortalidad relacionado con la gripe en España en el invierno de 2012, *Gac Sanit* **29**, 258-65 (2015).

14. Tindale, L. et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. medRxiv 2020.03.03.20029983; doi: https://doi.org/10.1101/2020.03.03.20029983 (2020).

15. Ki, M. Task Force for 2019-nCoV. Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. *Epidemiol Health* **42**, e2020007. DOI: https://doi.org/10.4178/epih.e2020007 (2020).

16. Fang Yong, S. E., et al., Connecting clusters of COVID-19: an epidemiological and serological investigation. *The Lancet*, published online April 21, 2020 https://doi.org/10.1016/S1473-3099(20)30273-5 (2020).

17. Radio Televisión Española. Radiografía del coronavirus en residencias de ancianos: más de 19.200 muertos con Covid-19 o síntomas compatibles, accessed online on 1st

June 2020 at https://www.rtve.es/noticias/20200529/radiografia-del-coronavirus-residencias-ancianos-espana/2011609.shtml (RTVE, 2020).

18. Dirección General de Tráfico, Ministry of Interior. Cuadro comparativo de accidentes mortales y fallecidos a 24 h en vías interurbanas, published online on 13 May 2020, http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/accidentes-24-horas/2020/Mayo/CC_-2020_05_13.pdf (DGT, 2020).

19. Ministerio de Trabajo y Economía Social. Estadística de Accidentes de Trabajo, Avance enero-marzo 2020, published online on 18 May 2020, http://www.mitramiss.gob.es/estadisticas/eat/eat20_03/ATR_03_2020_Resumen.pdf (MTES, 2020).

20. Vandoros, S. Excess Mortality during the Covid-19 pandemic: Early evidence from England and Wales. medRxiv 2020.04.14.20065706; doi: https://doi.org/10.1101/2020.04.14.20065706 (2020).

21. Domínguez□Gil, B. et al. COVID-19 in Spain: Transplantation in the Midst of the Pandemic. *Am J Transplant*. Accepted Author Manuscript. doi:10.1111/ajt.15983 (2020).

22. Rodríguez-Leor, O. et al. Impacto de la pandemia de COVID-19 sobre la actividad asistencial en cardiología intervencionista en España. *REC Interv Cardiol* **2** (2), 82-89 (2020).

23. Centro de Coordinación de Alertas y Emergencias Sanitarias. Actualización nº 90. Enfermedad por el coronavirus (COVID-19), published online on 29 April 2020, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/Actualizacion_90_COVID-19.pdf (CCAES, 2020e).

24. Roser, M., Ritchie, H., Ortiz-Ospina, E. & Hasell, J. Mortality Risk of COVID-19. Published online in Our World in Data: https://ourworldindata.org/mortality-risk-covid#the-case-fatality-rate (accessed on 23rd May 2020).

25. Caillaud, D., Craft, M.E. & Meyers, A. Epidemiological effects of group size variation in social species. *J. R. Soc. Interface* 10: 20130206, DOI: 10.1098/rsif.2013.0206 (2013).

26. Levins, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am. Am.* **15** (3), 237–240 (1969).

27. Hanski, I. Metapopulation Ecology. Oxford University Press (1999).

28. Ball, F. et al. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**, 63-67. http://dx.doi.org/10.1016/j.epidem.2014.08.001 (2015)

29. Bordehore, C. et al. Understanding COVID-19 spreading through simulation modeling and scenarios comparison: preliminary results. medRxiv 2020.03.30.20047043, doi: https://doi.org/10.1101/2020.03.30.20047043 (2020)

Table 1. Accumulated COVID-19 deaths and MoMo *ED* for Spain and its 19 regions. Values from 4 March to 11 May. Population data from INE. *: differences beyond the error estimation.

Region	2019 Population	Percentage of infections (ISCIII 2020)	Official data		MoMo data			*ED* minus COVID-19 deaths
		% (CI 95%)	COVID-19 deaths	CFR (CI 95%)	*ED* (CI 95%).	Days with lower limit >0	CFR (CI 95%)	
Spain	47,026,208	5 (4.7; 5.4)	26920	1.14 (1.06; 1.22)	43136 (34560; 50646)	54	1.83 (1.70; 1.95)	16216*
Andalucía	8,414,240	2.7 (2.2; 3.2)	1326	0.58 (0.49; 0.72)	1482 (-990; 3740)	21	0.65 (0.55; 0.80)	156
Aragón	1,319,291	4.9 (3.8; 6.3)	828	1.28 (1.00; 1.65)	893 (-92; 1648)	29	1.38 (1.07; 1.78)	65
Asturias	1,022,800	1.8 (1.3; 2.5)	304	1.65 (1.19; 2.29)	382 (-602; 1157)	10	2.07 (1.49; 2.87)	78
Islas Baleares	1,149,460	2.4 (1.6; 3.5)	210	0.76 (0.52; 1.14)	193 (-651; 803)	6	0.70 (0.48; 1.05)	-17
Canarias	2,153,389	1.8 (1.1; 2.8)	151	0.39 (0.25; 0.64)	149 (-946; 1145)	5	0.38 (0.25; 0.63)	-2
Cantabria	581,078	3.2 (2.1; 5)	203	1.09 (0.70; 1.66)	202 (-400; 688)	9	1.09 (0.70; 1.66)	-1
Castilla - La Mancha	2,032,863	10.8 (9.3; 12.4)	2811	1.28 (1.12; 1.49)	5238 (4088; 6295)	47	2.39 (2.08; 2.77)	2427*
Castilla y León	2,399,548	7.2 (6.3; 8.1)	1914	1.11 (0.98; 1.27)	3544 (2353; 4664)	44	2.05 (1.82; 2.34)	1630*
Cataluña	7,675,217	5.9 (4.9; 6.9)	5621	1.24 (1.06; 1.49)	11463 (8887; 13861)	51	2.53 (2.16; 3.05)	5842*
Ceuta	84,777	1.1 (0.5; 2.3)	4	0.43 (0.21; 0.94)	49 (-177; 118)	3	5.20 (2.49; 11.44)	45
C. Valenciana	5,003,769	2.5 (1.9; 3.2)	1336	1.07 (0.83; 1.41)	1656 (-201; 3350)	28	1.32 (1.03; 1.74)	320
Extremadura	1,067,710	3 (2.2; 4.1)	487	1.52 (1.11; 2.07)	710 (-247; 1504)	23	2.22 (1.62; 3.02)	223

Galicia	2,699,499	2.1 (1.7; 2.6)	595	1.05 (0.85; 1.30)	608 (-926; 2006)	13	1.07 (0.87; 1.32)	13
Madrid	6,663,394	11.3 (9.8; 13)	8720	1.16 (1.01; 1.34)	13947 (12315; 15464)	56	1.85 (1.61; 2.14)	5227*
Melilla	86,487	1.9 (1.2; 2.9)	2	0.12 (0.08; 0.19)	31 (-200; 100)	0	1.86 (1.22; 2.94)	29
Murcia	1,493,898	1.4 (0.8; 2.4)	139	0.66 (0.39; 1.16)	207 (-685; 1028)	8	0.99 (0.58; 1.73)	68
Navarra	654,214	5.8 (4.3; 7.6)	493	1.3 (0.99; 1.75)	657 (-20; 1164)	27	1.73 (1.32; 2.33)	164
País Vasco	2,207,776	4 (3.1; 5.2)	1429	1.62 (1.24; 2.09)	1621 (207; 2663)	30	1.83 (1.41; 2.37)	192
La Rioja	316,798	3.3 (2.4; 4.4)	347	3.32 (2.49; 4.56)	255 (-236; 582)	14	2.43 (1.83; 3.35)	-92

Table 2. Date of first infection and maximum of daily infections for official COVID-19 daily infections (estimated from the 14-days running mean time series), and daily infections estimated from COVID-19 deaths and from MoMo *ED*. Infections are estimated from COVID-19 deaths and MoMo ED with their corresponding CFR and 95% confidence interval detailed in Table 1. *: there were no infection according to the smaller error estimate of infections calculated from MoMo *ED*.

	First infection			Maximum of daily infections		
	From COVID-19 deaths (CI 95%)	From MoMo *ED* (conservative CI)	Official COVID-19	From COVID-19 deaths	From MoMo *ED*	Official COVID-19
Spain	11 Jan (11-11 Jan)	11 Jan (5 Jan–14 Jan)	20 Feb	14 March	13 March	30 March
Andalucía	19 Jan (18-19 Jan)	12 Jan (5 Jan-21 Jan)	26 Feb	20 March	12 March	30 March
Aragón	21 Jan (20-22 Jan)	21 Jan (9 Jan-30 Jan)	4 March	19 March	16 March	30 March
Asturias	28 Jan (27-30 Jan)	22 Jan (9 Jan-9 Feb)	29 Feb	26 March	13 March, 22 March	26 March
Islas Baleares	26 Jan (25-28 Jan)	12 Jan (6 Jan-3 Feb)	20 Feb	17 March	12 March, 24 March	27 March
Canarias	23 Jan (21-24 Jan)	12 Jan (5 Jan-4 Feb)	20 Feb	13 March	14 March	26 March
Cantabria	29 Jan (27-30 Jan)	19 Jan (8 Jan-7 Feb)	29 Feb	18 March	17 March	26 March
Castilla - La Mancha	19 Jan (19-20 Jan)	14 Jan (9 Jan-19 Jan)	29 Feb	15 March	12 March	3 April
Castilla y León	20 Jan (19-20 Jan)	18 Jan (9 Jan-22 Jan)	28 Feb	15 March	14 March	31 March
Cataluña	18 Jan (18-18 Jan)	17 Jan (6 Jan-20 Jan)	25 Feb	16 March	15 March	30 March
Ceuta	13 Feb (8-17 Feb)	26 Jan (19 Jan-*)	14 March	18 March	22 Feb, 15 March	29 March
C. Valenciana	20 Jan (19-21 Jan)	14 Jan (6 Jan-26 Jan)	25 Feb	16 March	13 March	29 March
Extremadura	25 Jan (24-26 Jan)	17 Jan (9 Jan-31 Jan)	29 Feb	15 March	14 March	29 March
Galicia	25 Jan (24-26 Jan)	12 Jan (5 Jan-31 Jan)	3 March	18 March, 11 April	17 March	29 March
Madrid	11 Jan (11-12 Jan)	10 Jan (6 Jan-12 Jan)	25 Feb	11 March	11 March	30 March
Melilla	5 Feb (2-8 Feb)	24 Jan (18 Jan-*)	12 March	11 March	9 March	31 March

Murcia	29 Jan (27-31 Jan)	17 Jan (7 Jan-6 Feb)	7 March	17 March	12 March	28 March
Navarra	25 Jan (25-26 Jan)	21 Jan (10 Jan-28 Jan)	29 Feb	15 March, 31 March	14 March	30 March
País Vasco	18 Jan (17-19 Jan)	15 Jan (7 Jan-27 Jan)	28 Feb	20 March	16 March	29 March
La Rioja	28 Jan (27-30 Jan)	21 Jan (14 Jan-8 Feb)	1 March	20 March	14 March	31 March

Table 3. Daily infections on 14 March for official COVID-19 daily infections, and daily infections estimated from COVID-19 deaths and from MoMo *ED*. Infections are inferred from COVID-19 deaths and MoMo ED with their corresponding CFR and 95% confidence interval detailed in Table 1. Percentages are official COVID-19 infections as proportions of the estimated infections. *: there were no infections according to the smaller error estimate of infections calculated from MoMo *ED.*

	Daily infections on 14 March			Percentage of detected daily infections (%)	
	From COVID-19 deaths (CI 95%)	From MoMo *ED* (conservative CI)	Official COVID-19	From COVID-19 deaths (CI 95%)	From MoMo *ED* (conservative CI)
Spain	62,860 (59,088; 67,888)	77,455 (67,445; 89,530)	1832	2.9 (2.7; 3.1)	2.4 (2.0; 2.7)
Andalucía	5,877 (4,789; 6,966)	8,460 (3,191; 16,131)	168	2.9 (2.4; 3.5)	2.0 (1.0; 5.3)
Aragón	1,675 (1,299; 2,154)	2,081 (769; 3700)	26	1.6 (1.2; 2.0)	1.2 (0.7; 3.4)
Asturias	373 (269; 517)	520 (61; 1,382)	45	12.1 (8.7; 16.7)	8.7 (3.3; 73.8)
Islas Baleares	700 (466; 1020)	814 (41; 3,042)	14	2.0 (1.4; 3.0)	1.7 (0.5; 34.1)
Canarias	1,168 (714; 1817)	2,216 (141; 9,698)	19	1.6 (1.0; 2.7)	0.9 (0.2; 13.5)
Cantabria	503 (330; 786)	634 (95; 1883)	20	4.0 (2.5; 6.1)	3.2 (1.1; 21.1)
Castilla - La Mancha	5,455 (4,697; 6,263)	7,517 (5,855; 9,330)	112	2.1 (1.8; 2.4)	1.5 (1.2; 1.9)
Castilla y León	5,115 (4,476; 5,755)	5,771 (4,278; 7,341)	54	1.1 (0.9; 1.2)	0.9 (0.7; 1.3)
Cataluña	11,526 (9,573; 13,480)	14,450 (10,873; 18,293)	184	1.6 (1.4; 1.9)	1.3 (1.0; 1.7)
Ceuta	44 (20; 92)	25 (2; 87)	1	2.3 (1.1; 5.0)	4.0 (1.1; 50.0)
C.	3,620 (2,751;	4,546 (1,898; 8,176)	279	7.7 (6.0; 10.1)	6.1 (3.4; 14.7)

Valenciana	4,633)				
Extremadura	976 (716; 1333)	1,080 (392; 2195)	29	3.0 (2.2; 4.1)	2.7 (1.3; 7.4)
Galicia	1,245 (1,008; 1,542)	2,060 (461; 5197)	80	6.4 (5.2; 7.9)	3.9 (1.5; 17.4)
Madrid	21,255 (18,433; 24,452)	24,099 (19,757; 28,885)	604	2.8 (2.5; 3.3)	2.5 (2.1; 3.1)
Melilla	78 (49; 118)	53 (0; 143)	5	6.4 (4.2; 10.2)	9.4 (3.5; *)
Murcia	717 (410; 1229)	753 (83; 3,551)	24	3.3 (2.0; 5.9)	3.2 (0.7; 28.9)
Navarra	908 (673; 1189)	1,444 (709; 2,387)	37	4.1 (3.1; 5.5)	2.6 (1.6; 5.2)
País Vasco	2,100 (1,627; 2,730)	2,831 (1,368; 4,690)	109	5.2 (4.0; 6.7)	3.9 (2.3; 8.0)
La Rioja	280 (203; 373)	327 (35; 685)	22	7.9 (5.9; 10.8)	6.7 (3.2; 62.9)

Table 4. Accumulated infections in Spain and its administrative regions until 28 April for official COVID-19 infections, and infections estimated from COVID-19 deaths and from MoMo *ED*. Infections are estimated from COVID-19 deaths and MoMo *ED* with their corresponding CFR and 95% confidence interval detailed in Table 1. Percentages are official COVID-19 infections as proportions of the estimated infections. *: there were no infections according to the smaller error estimate of infections calculated from MoMo *ED*.

	Accumulated infections before 28 April			Percentage of detected infections (%)	
	From COVID-19 deaths (CI 95%)	From MoMo *ED* (conservative CI)	Official COVID-19	From COVID-19 deaths (CI 95%)	From MoMo *ED* (conservative CI)
Spain	2,405,617 (2,261,274; 2,598,063)	2,321,620 (1,888,856; 2,876,594)	223,054	7.8 (7.2; 8.2)	7.4 (6.0; 9.1)
Andalucía	231,741 (188,825; 274,652)	271,546 (77,022; 641,475)	12,435	9.5 (8.7; 10.1)	3.8 (1.6; 13.2)
Aragón	65,539 (50,825; 84,265)	64,758 (20,279; 141,181)	5,420	6.2 (5.8; 6.6)	6.6 (3.0; 20.9)
Asturias	18,956 (13,690; 26,324)	21,343 (1,703; 71,009)	2,497	8.1 (7.5; 8.6)	9.6 (2.9; 120.4)
Islas Baleares	28,516 (19,013; 41,589)	35,230 (1,003; 151,780)	1,905	9.3 (8.6; 9.9)	4.5 (1.0; 156.6)
Canarias	39,248 (23,980; 61,052)	72,612 (2,562; 416,656)	2,183	16.1 (14.9; 17.1)	2.7 (0.5; 76.4)
Cantabria	18895 (12,399; 29,527)	22,891 (2,021; 91,260)	2,314	10.6 (9.7; 11.2)	7.8 (2.0; 88.9)
Castilla - La Mancha	224,600 (193,400; 257,875)	216,943 (158,317; 292,205)	17,988	6.3 (5.8; 6.7)	6.6 (4.9; 9.1)
Castilla y León	175,054 (153,173; 196,943)	171,556 (115,383; 243,356)	18,516	7.9 (7.3; 8.4)	7.7 (5.4; 11.4)
Cataluña	473,710 (393,426; 554,005)	442,121 (312,385; 610,671)	48,779	7.8 (7.2; 8.3)	8.0 (5.8; 11.3)

Ceuta	938 (428; 1,970)	1,131 (32; 4,324)	119	27.0 (24.7; 28.4)	8.5 (2.2; 300.0)
C. Valenciana	127,506 (96,906; 163,213)	129,662 (40,669; 298,873)	11,126	8.2 (7.6; 8.7)	7.1 (3.1; 22.7)
Extremadura	32,715 (23,996; 44,716)	32731 (8,992; 85,082)	3,337	6.6 (6.1; 7.1)	8.2 (3.2; 29.8)
Galicia	57,463 (46,522; 71,149)	64,554 (9,471; 213,653)	9,277	16.8 (15.5; 17.8)	11.8 (3.6; 80.2)
Madrid	763,220 (661,904; 878,031)	743,197 (584,666; 938,270)	61,650	6.5 (6.0; 6.9)	6.5 (5.1; 8.2)
Melilla	1,639 (1,038; 2,506)	2,364 (0; 7,421)	114	57.0 (52.7; 60.1)	4.3 (1.4; *)
Murcia	21,762 (12,436; 37,310)	26,627 (1,842; 156,480)	1,738	12.7 (11.7; 13.4)	5.6 (1.0; 80.7)
Navarra	38,533 (28,573; 50,491)	39,533 (15,937; 82,878)	5,267	10.1 (9.3; 10.7)	10.5 (5.0; 26.0)
País Vasco	90,290 (69,970; 117,380)	88,406 (33,587; 178,011)	13,836	9.5 (8.8; 10.1)	12.7 (6.3; 33.4)
La Rioja	10,596 (7,711; 14,128)	11,395 (1,198; 29,387)	4,553	11.5 (10.6; 12.2)	30.0 (11.6; 285.5)

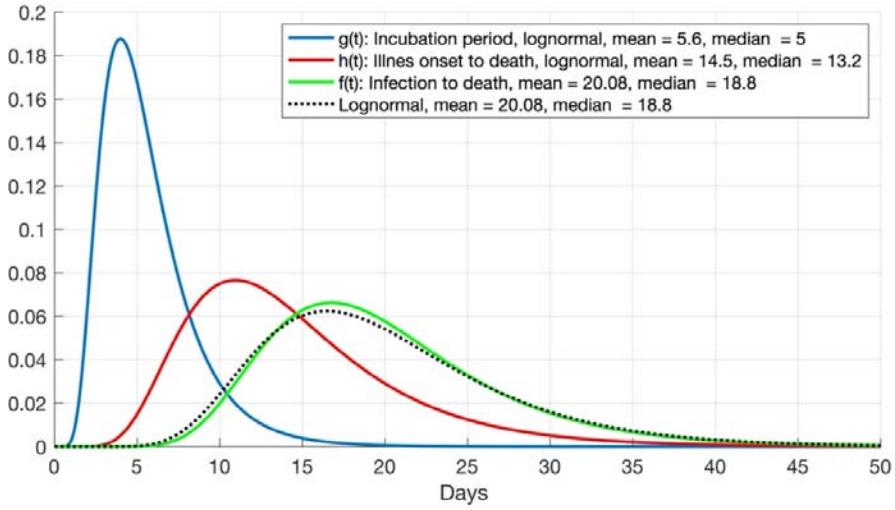
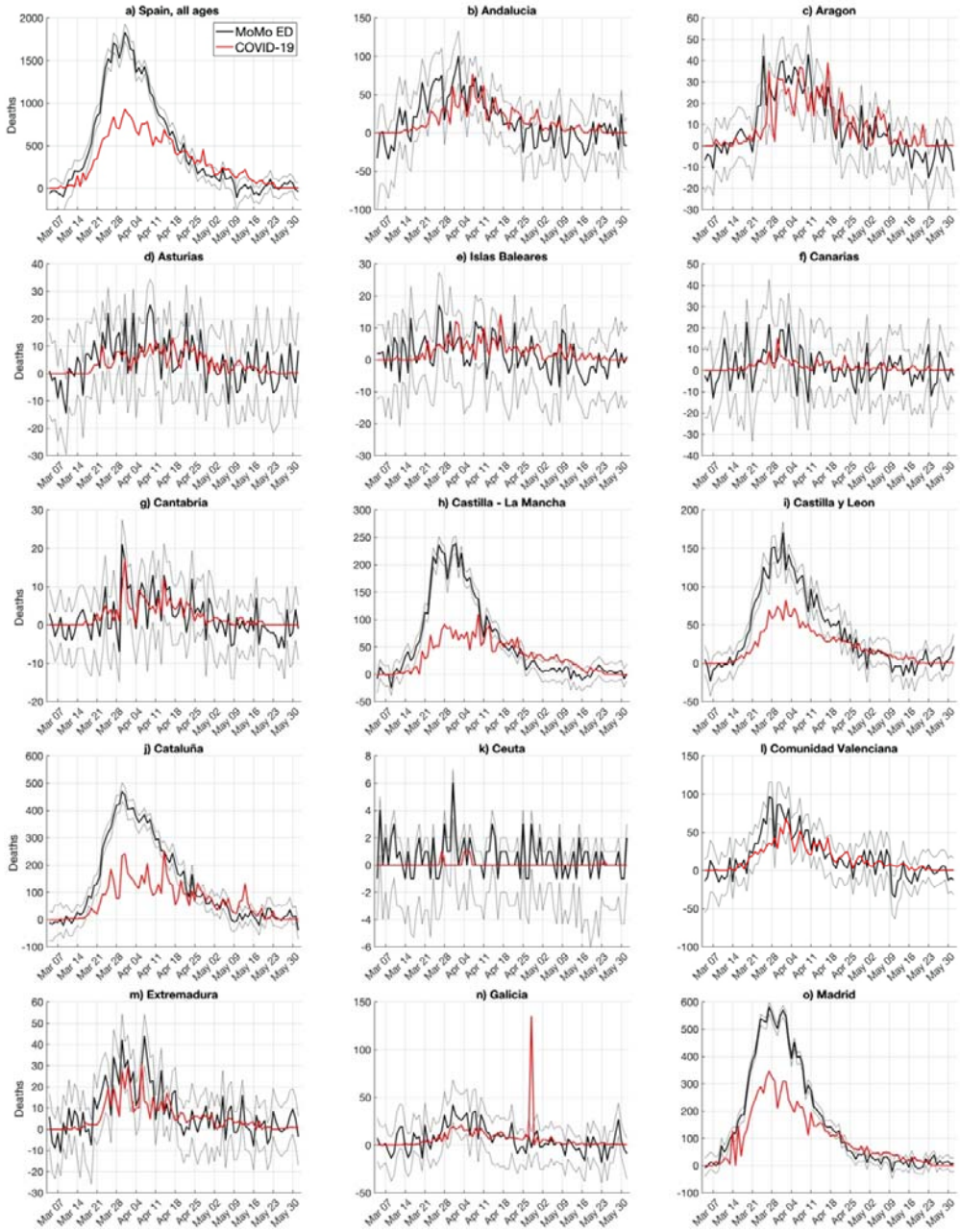


Figure 1. Probability density function (PDF) of incubation period (*IP*, blue line) and illness onset to death period (*IOD*, red line) from Linton et al. (2020). Green line, *f(t)*, is the convolution of both and represents the PDF of infection to death. Black dotted line is a lognormal PDF with the same mean and median than *f(t)*.


a) Spain, all ages
MoMo ED
COVID-19
b) Andalucia
c) Aragon
d) Asturias
e) Islas Baleares
f) Canarias
g) Cantabria
h) Castilla - La Mancha
i) Castilla y Leon
j) Cataluña
k) Ceuta
l) Comunidad Valenciana
m) Extremadura
n) Galicia
o) Madrid
Deaths

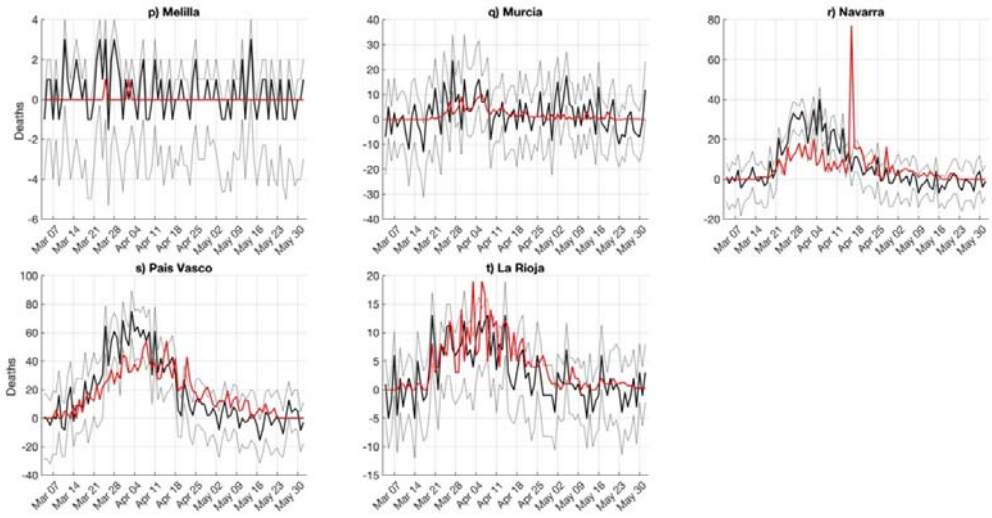


Figure 2. COVID-19 deaths and MoMo Excess of Deaths (*ED*). a) Spain; b-t) Spanish regions. Red lines are official COVID-19 deaths, solid black lines are MoMo deaths, and black dotted lines delimit MoMo death error bands (99% confidence interval).

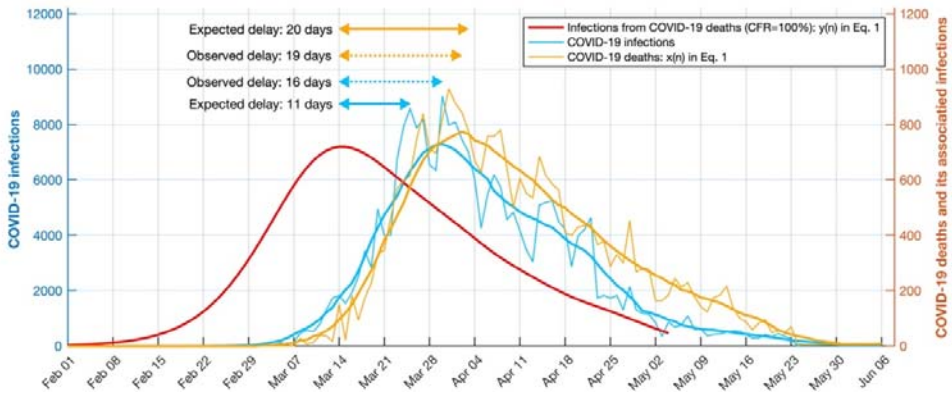
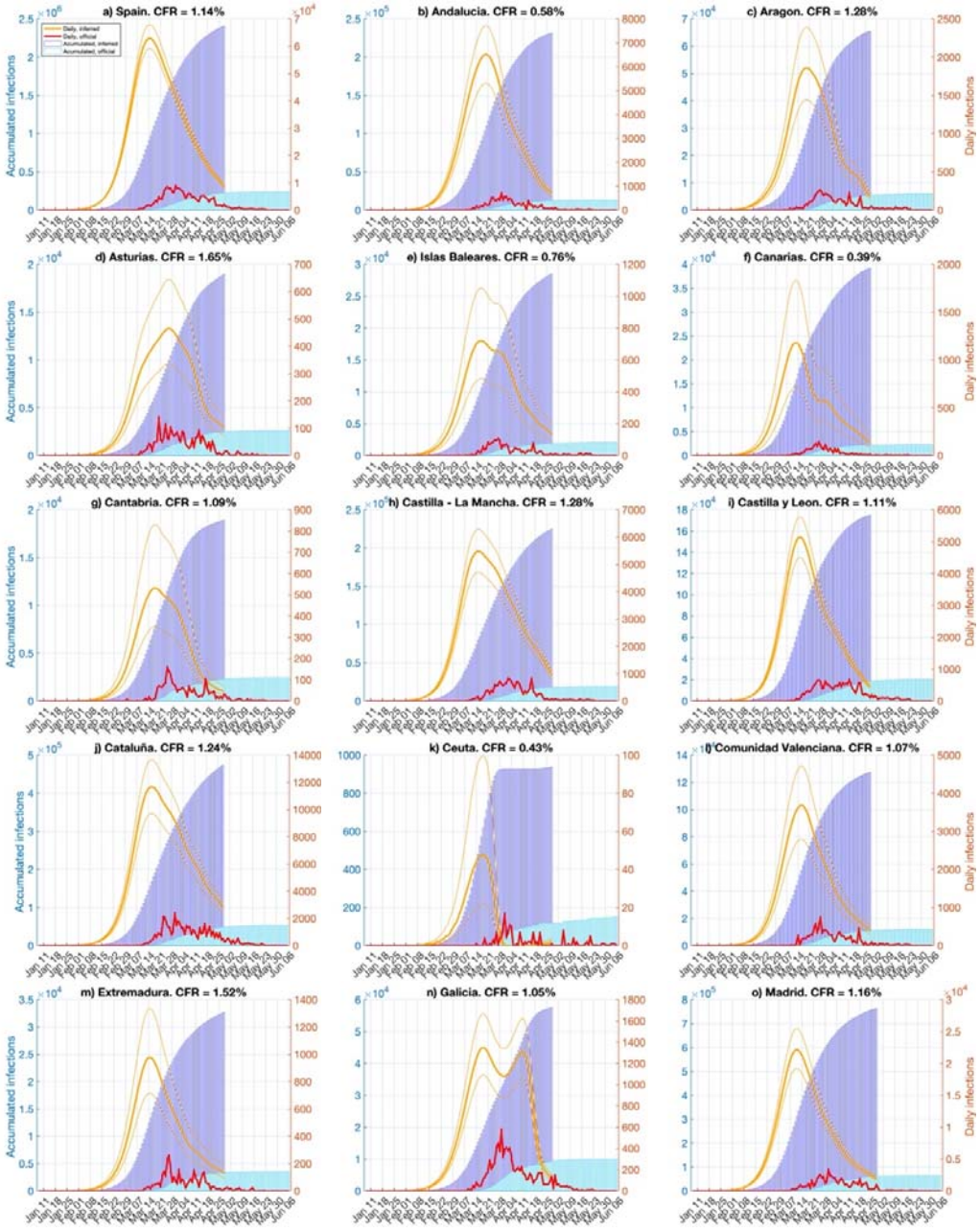


Figure 3. Official COVID-19 infections and deaths, and estimated infections with case fatality ratio (CFR) of 100% in Spain. Left y-axis: COVID-19 daily infections (blue curve). Right y-axis: COVID-19 deaths (orange curve) and its associated/estimated infections *y(t)* from Equation 1 (red curve), i.e., infections with CFR=100%. All curves are for Spain. Blue and orange thin curves are daily data, and thick curves are smoothed by 14-days running mean. Arrows show delay between the maximum of inferred infections and maxima from COVID-19 deaths (orange arrows) and COVID-19 infections (blue arrows). Solid arrows are expected delays, while dotted arrows are observed delays.


a) Spain. CFR = 1.14%
b) Andalucia. CFR = 0.58%
c) Aragon. CFR = 1.28%
d) Asturias. CFR = 1.65%
e) Islas Baleares. CFR = 0.76%
f) Canarias. CFR = 0.39%
g) Cantabria. CFR = 1.09%
h) Castilla - La Mancha. CFR = 1.28%
i) Castilla y Leon. CFR = 1.11%
j) Cataluña. CFR = 1.24%
k) Ceuta. CFR = 0.43%
l) Comunidad Valenciana. CFR = 1.07%
m) Extremadura. CFR = 1.52%
n) Galicia. CFR = 1.05%
o) Madrid. CFR = 1.16%
Accumulated infections
Daily infections

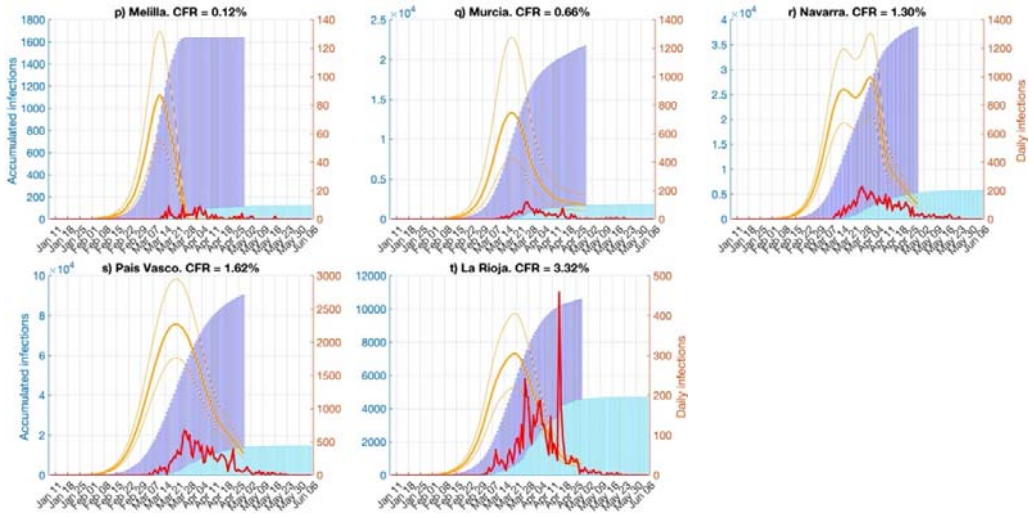


Figure 4. Daily and accumulated infections for official COVID-19 daily infections and daily infections estimated from COVID-19 deaths. Lines are daily infections and refer to the y-axis on the right; bars are accumulated infections and refer to the y-axis on the left. Red lines and cyan bars are official COVID-19 data; orange lines and blue bars are inferred infections with case fatality ratio (CFR) in Table 1. Thin orange lines correspond to the CFR confidence interval.

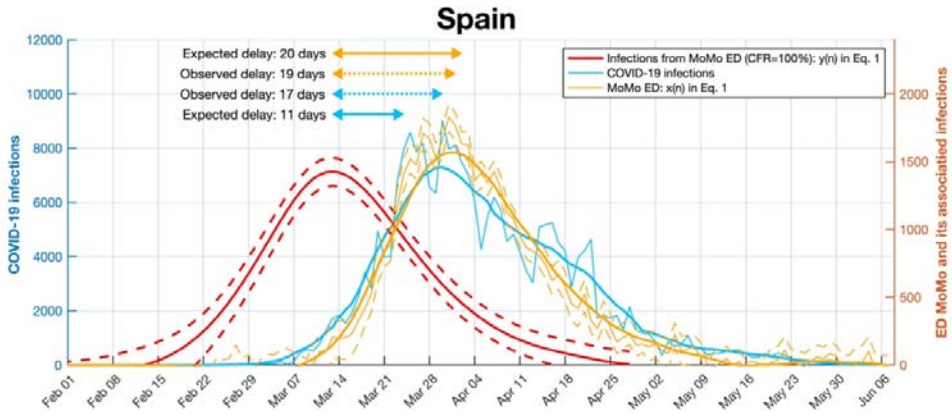
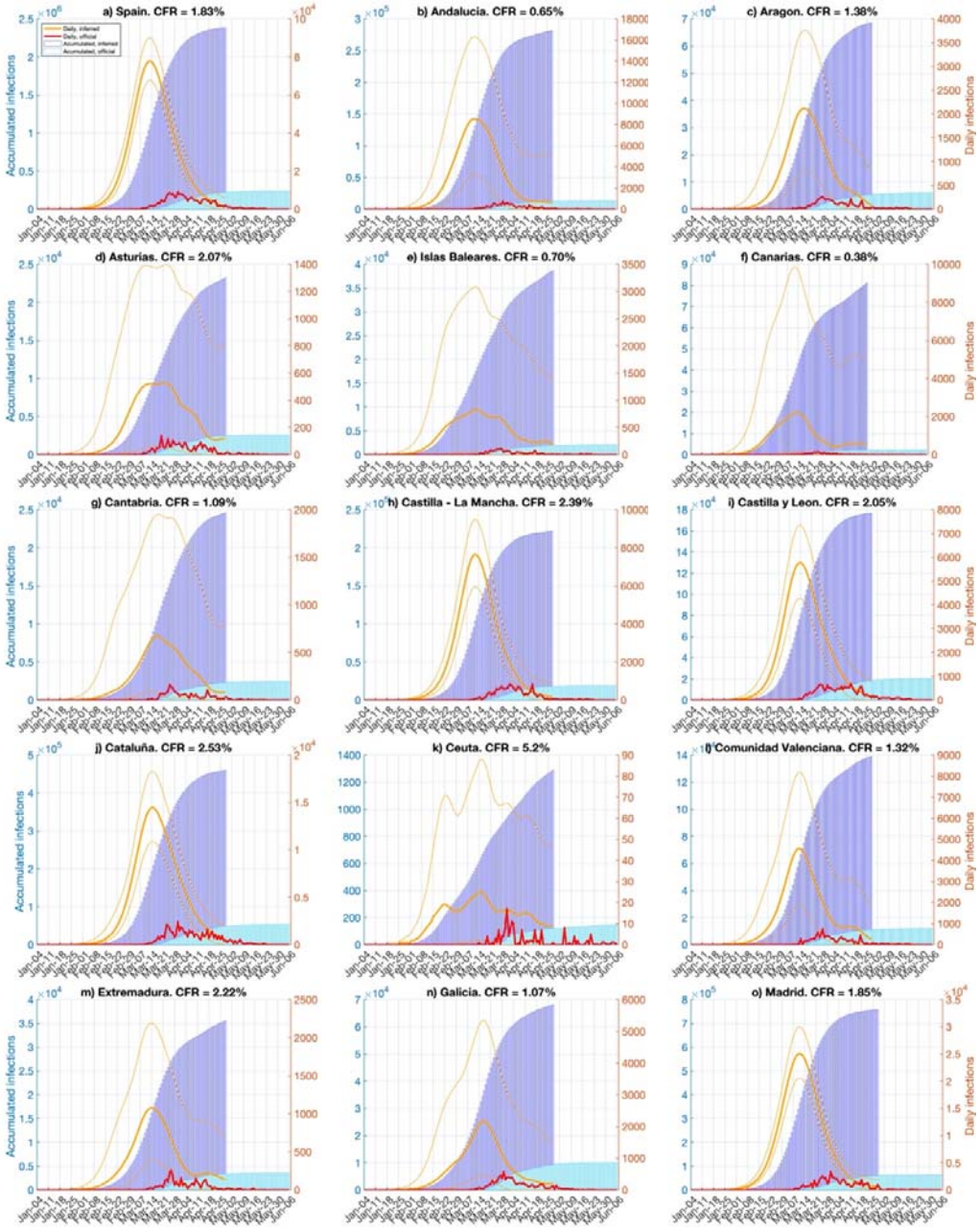


Figure 5. Official COVID-19 infections, MoMo Excess of Deaths (*ED*), and estimated infections with case fatality ratio (CFR) of 100% in Spain. Left y-axis: COVID-19 daily infections (blue curve). Right y-axis: MoMo *ED* (orange curve) and its associated/estimated infections *y(t)* from Equation 1 (red curve), i.e., infections with CFR=100%. All curves are for Spain. Blue and orange thin curves are daily data, and thick curves are smoothed by 14-days running mean. Dashed curves represent the error estimate of MoMo *ED* (orange) and inferred infections (red). Arrows show delays between the maximum of inferred infections and maxima from MoMo *ED* (orange arrows) and COVID-19 infections (blue arrows). Solid arrows are expected delays, while dotted arrows are observed delays.


a) Spain. CFR = 1.83%
b) Andalucia. CFR = 0.65%
c) Aragon. CFR = 1.38%
d) Asturias. CFR = 2.07%
e) Islas Baleares. CFR = 0.70%
f) Canarias. CFR = 0.38%
g) Cantabria. CFR = 1.09%
h) Castilla - La Mancha. CFR = 2.39%
i) Castilla y Leon. CFR = 2.05%
j) Cataluña. CFR = 2.53%
k) Ceuta. CFR = 5.2%
l) Comunidad Valenciana. CFR = 1.32%
m) Extremadura. CFR = 2.22%
n) Galicia. CFR = 1.07%
o) Madrid. CFR = 1.85%
Accumulated infections
Daily infections

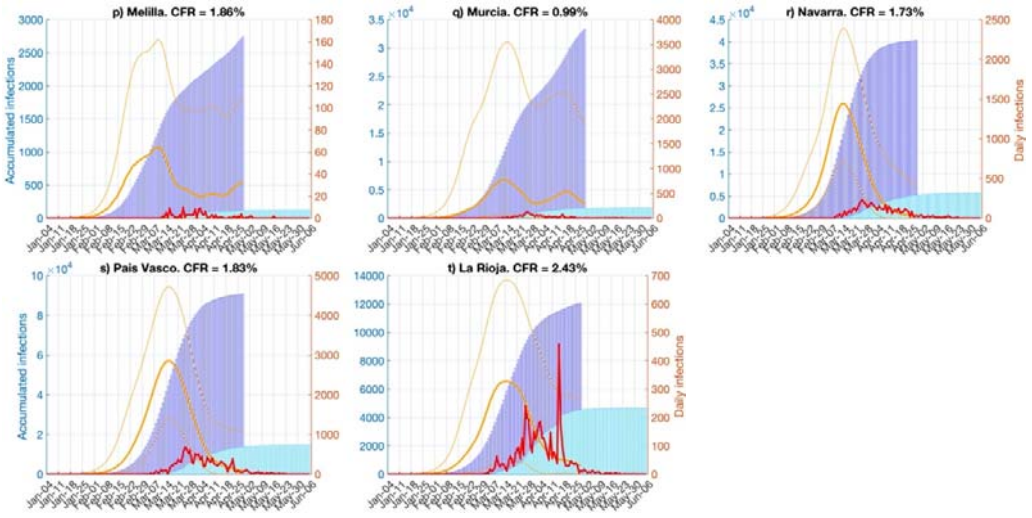


Figure 6. Daily and accumulated infections for official COVID-19 daily infections, and daily infections estimated from MoMo Excess Deaths (*ED*). Lines are daily infections and refer to the y-axis on the right; and bars are accumulated infections and refer to the y-axis on the left. Red lines and cyan bars are for official COVID-19 data; and orange lines and blue bars are for inferred infections with case fatality ratio (CFR) in Table 1. Thin orange lines represent the error estimate of inferred infections.