

Daily report

21-03-2020

Analysis and prediction of COVID-19 for different regions and countries

Situation report 5

Contact: clara.prats@upc.edu

With the financial support of



Foreword

The present report aims to provide a comprehensive picture of the **pandemic situation of COVID-19** in the EU countries, and to be able to foresee the situation in the next coming days.

We employ an **empirical model**, verified with the evolution of the number of confirmed cases in previous countries where the epidemic is close to conclude, including all provinces of China. The model does not pretend to interpret the causes of the evolution of the cases but to permit the **evaluation of the quality of control measures made in each state** and a **short-term prediction of tendencies**. Note, however, that the effects of the measures' control that start on a given day are not observed until approximately 5-7 days later.

The model and predictions are based on two parameters that are daily fitted to available data:

- ✓ α : the velocity at which spreading specific rate slows down; the higher the value, the better the control.
- ✓ K : the final number of expected cumulated cases, which cannot be evaluated at the initial stages because growth is still exponential.

Next, we show a report with 8 graphs and a table with the short-term predictions for (1) European Union and its countries, (2) other countries, (3) Spain and its autonomous communities.

We are currently adjusting the model to **countries and regions** with at least 4 days with more than 100 confirmed cases and a current load over 200 cases. The **predicted period** of a country depends on the number of datapoints over this 100 cases threshold:

- ✓ Group A: countries that have reported more than 100 cumulated cases for 10 consecutive days or more → 3 days prediction;
- ✓ Group B: countries that have reported more than 100 cumulated cases for 7 to 9 consecutive days → 2 days prediction;
- ✓ Group C: countries that have reported more than 100 cumulated cases for 4 to 6 days → 1 day prediction.

The whole methodology employed in the inform is explained in the last pages of this document.

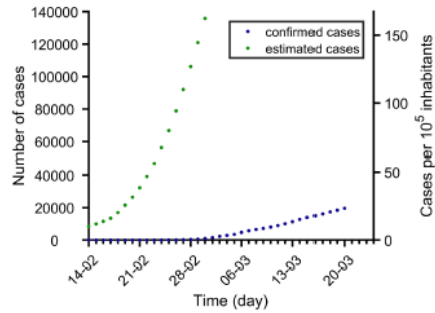
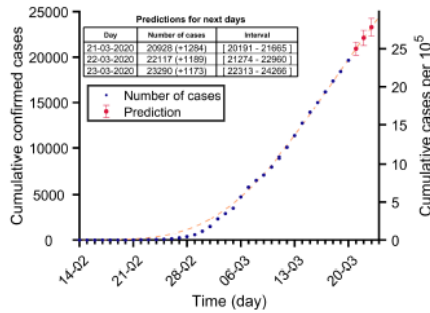
Martí Català, MD
Pere-Joan Cardona, PhD
*Comparative Medicine and Bioimage Centre of
Catalonia; Institute for Health Science Research
Germans Trias i Pujol*

Clara Prats, PhD
Sergio Alonso, PhD
Enric Álvarez, PhD
Daniel López, PhD
*Computational Biology and Complex Systems;
Universitat Politècnica de Catalunya - BarcelonaTech*

Countries' reports details

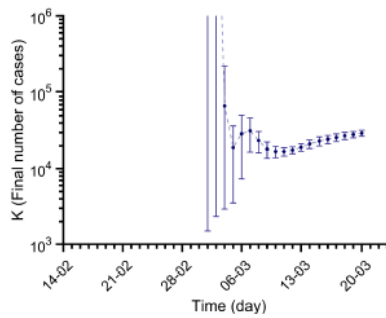
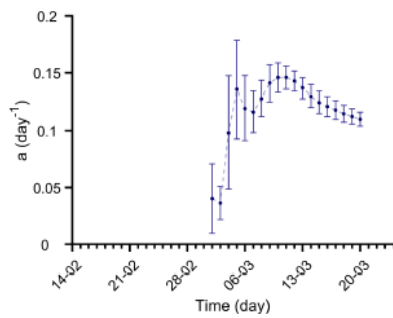
Iran 20-03-2020. Population: 83.7M. Current cumulated incidence: 23/10⁵

Confirmed cases:
data (blue),
model fitted
(dashed line),
predictions (red
points and table)



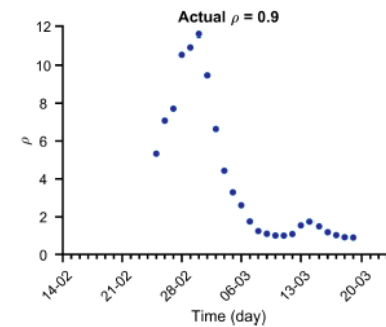
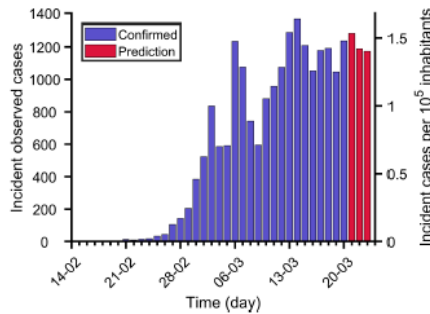
Estimated cases using death rate (see Methods)

Fitted α value using points prior to each date



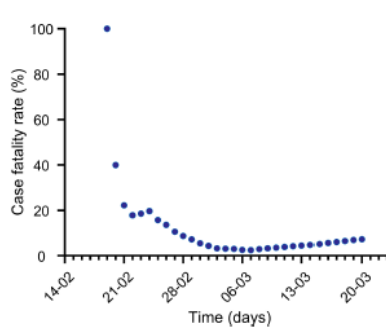
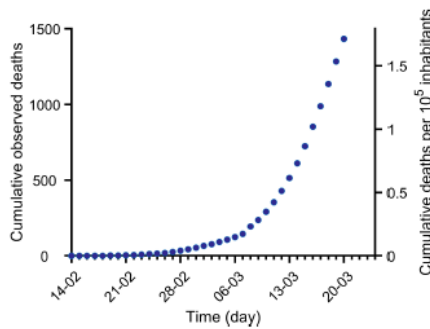
Fitted K value using points prior to each date

Reported and predicted new cases



Evolution of ρ , a parameter related with Reproduction number (see Methods)

Reported deaths



Deaths / cumulated reported cases

(1) Analysis and prediction of COVID-19 for European Union and its countries

HIGHLIGHTS (21-03-2020):

- **Luxembourg shows highest relative cumulative incidence ($84/10^5$ inhabitants), greater than Italy ($79/10^5$ inhabitants), in addition to high spreading rate.**
- **Estonia, Slovenia and Greece show a controlled dynamic, by the moment.**
- **New countries in group B: Ireland, Poland, Romania, Estonia. New country in group C: Luxembourg**

(2) Analysis and prediction of COVID-19 for other countries

HIGHLIGHTS (21-03-2020):

- **United States of America shows a low relative incidence but high spreading rate.**

(3) Analysis and prediction of COVID-19 for Spain and its autonomous communities

HIGHLIGHTS (21-03-2020):

- Highest incidence is found in La Rioja ($178/10^5$ inhabitants), followed by Madrid ($134/10^5$ inhabitants) and Navarra ($101/10^5$ inhabitants)
- Autonomous communities with highest spreading rate are Castilla y León ($\rho=4.84$), Catalunya ($\rho=3.98$) and Euskadi ($\rho=3.53$).

Data obtained from <https://github.com/datadista/datasets/tree/master/COVID%2019> and <https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm>

Methods

Methods

(1) Data source

Data are daily obtained from World Health Organization (WHO) surveillance reports¹, from European Centre for Disease Prevention and Control (ECDC)² and from Ministerio de Sanidad³. These reports are converted into text files that can be processed for subsequent analysis. Daily data comprise, among others: total confirmed cases, total confirmed new cases, total deaths, total new deaths. It must be considered that the report is always providing data from previous day. In the document we use the date at which the datapoint is assumed to belong, i.e., report from 15/03/2020 is giving data from 14/03/2020, the latter being used in the subsequent analysis.

(2) Data processing and plotting

Data are initially processed with Matlab in order to update timeseries, i.e., last datapoints are added to historical sequences. These timeseries are plotted for EU individual countries and for the UE as a whole:

- ✓ Number of cumulated confirmed cases, in blue dots
- ✓ Number of reported new cases
- ✓ Number of cumulated deaths

Then, two indicators are calculated and plotted, too:

- ✓ Number of cumulated deaths divided by the number of cumulated confirmed cases, and reported as a percentage; it is an indirect indicator of the diagnostic level.
- ✓ ρ : this variable is related with the reproduction number, i.e., with the number of new infections caused by a single case. It is evaluated as follows for the day before last report ($t-1$):

$$\rho(t-1) = \frac{N_{new}(t) + N_{new}(t-1) + N_{new}(t-2)}{N_{new}(t-5) + N_{new}(t-6) + N_{new}(t-7)}$$

where $N_{new}(t)$ is the number of new confirmed cases at day t .

(3) Classification of countries according to their status in the epidemic cycle

The evolution of confirmed cases shows a biphasic behaviour:

- (I) an initial period where most of the cases are imported;
- (II) a subsequent period where most of new cases occur because of local transmission.

Once in the stage II, mathematical models can be used to track evolutions and predict tendencies. Focusing on countries that are on stage II, we classify them in three groups:

- Group A: countries that have reported more than 100 cumulated cases for 10 consecutive days or more;
- Group B: countries that have reported more than 100 cumulated cases for 7 to 9 consecutive days;
- Group C: countries that have reported more than 100 cumulated cases for 4 to 6 days.

¹ <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

² <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>

³ <https://www.msbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm>
<https://github.com/datadista/datasets/tree/master/COVID%2019>

(4) Fitting a mathematical model to data

Previous studies have shown that Gompertz model⁴ correctly describes the Covid-19 epidemic in all analysed countries. It is an empirical model that starts with an exponential growth but that gradually decreases its specific growth rate. Therefore, it is adequate for describing an epidemic that is characterized by an initial exponential growth but a progressive decrease in spreading velocity provided that appropriate control measures are applied.

Gompertz model is described by the equation:

$$N(t) = K e^{-\ln\left(\frac{K}{N_0}\right) \cdot e^{-a \cdot (t-t_0)}}$$

where $N(t)$ is the cumulated number of confirmed cases at t (in days), and N_0 is the number of cumulated cases the day at day t_0 . The model has two parameters:

- ✓ a is the velocity at which specific spreading rate is slowing down;
- ✓ K is the expected final number of cumulated cases at the end of the epidemic.

This model is fitted to reported cumulated cases of the UE and of countries in stage II that accomplish two criteria: 4 or more consecutive days with more than 100 cumulated cases, and at least one datapoint over 200 cases. Day t_0 is chosen as that one at which $N(t)$ overpasses 100 cases. If more than 15 datapoints that accomplish the stated criteria are available, only the last 15 points are used. The fitting is done using Matlab's Curve Fitting package with Nonlinear Least Squares method, which also provides confidence intervals of fitted parameters (a and K) and the R^2 of the fitting. At the initial stages the dynamics is exponential and K cannot be correctly evaluated. In fact, at this stage the most relevant parameter is a . Fitted curves are incorporated to plots of cumulative reported cases with a dashed line. Once a new fitting is done, two plots are added to the country report:

- ✓ Evolution of fitted a with its error bars, i.e., values obtained on the fitting each day that the analysis has been carried out;
- ✓ Evolution of fitted K with its error bars, i.e., values obtained on the fitting each day that the analysis has been carried out; if lower error bar indicates a value that is lower than current number of cases, the error bar is truncated.

These plots illustrate the increase in fittings' confidence, as fitted values progressively stabilize around a certain value and error bars get smaller when the number of datapoints increases. In fact, in the case of countries, they are discarded and set as "Not enough data" if $a > 0.2 \text{ day}^{-1}$, if $K > 10^6$ or if the error in K overpasses 10^6 .

It is worth to mention that the simplicity of this model and the lack of previous assumptions about the Covid-19 behaviour make it appropriate for universal use, i.e., it can be fitted to any country independently of its socioeconomic context and control strategy. Then, the model is capable of quantifying the observed dynamics in an objective and standard manner and predicting short-term tendencies.

⁴ Madden LV. Quantification of disease progression. *Protection Ecology* 1980; **2**: 159-176.

(5) Using the model for predicting short-term tendencies

The model is finally used for a short-term prediction of the evolution of the cumulated number of cases. The predictions increase their reliability with the number of datapoints used in the fitting. Therefore, we consider three levels of prediction, depending on the country:

- Group A: prediction of expected cumulated cases for the following 3 days;
- Group B: prediction of expected cumulated cases for the following 2 days;
- Group C: prediction of expected cumulated cases for the following day.

The confidence interval of predictions is assessed with the Matlab function `predint`, with a 99% confidence level. These predictions are shown in the plots as red dots with corresponding error bars, and also gathered in the attached table.

(6) Estimating non-diagnosed cases

Lethality of Covid-19 has been estimated at around 1 % for Republic of Korea and the Diamond Princess cruise. Besides, median duration of viral shedding after Covid-19 onset has been estimated at 18.5 days for non-survivors⁵ in a retrospective study in Wuhan. These data allow for an estimation of total number of cases, considering that the number of deaths at certain moment should be about 1 % of total cases 18.5 days before. This is valid for estimating cases of countries at stage II, since in stage I the deaths would be mostly due to the incidence at the country from which they were imported. We establish a threshold of 50 reported cases before starting this estimation.

Reported deaths are passed through a moving average filter of 5 points in order to smooth tendencies. Then, the corresponding number of cases is found assuming the 1 % lethality. Finally, these cases are distributed between 18 and 19 days before each one.

⁵ Zhou et al., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet; March 9, doi: 10.1016/S0140-6736(20)30566-3