UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Computational Biology and Complex Systems Research Group

CMCiB Comparative Medicine & Bioimage Centre of Catalonia

IGTP Germans Trias i Pujol Research Institute

Analysis and prediction of COVID-19 for EU-EFTA-UK and other countries

Methods

(updated January 2021)

Contact: clara.prats@upc.edu

With the financial support of European Commission *and* Ministerio de Ciencia e Innovación, "la Caixa" Foundation

Foreword

This collection of reports 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 provide some figures and tables with several **indexes and indicators** as well as an **Analysis** section that discusses a specific topic related with the pandemic.

As for the predictions, 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 trends**. Note, however, that the effects of the measures' control that start on a given day are not observed until approximately 7-14 days later.

We show an individual report with 8 graphs and a summary table with the main indicators for different countries and regions. We are 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 whole methodology employed in the informs is explained in these pages.

Martí Català
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
Miquel Marchena, PhD
David Conesa
Daniel López, PhD
Computational Biology and Complex Systems; Universitat Politècnica de Catalunya – BarcelonaTech

With the collaboration of: Daniel Molinuevo, Pablo Palacios, Tomás Urdiales, Aida Perramon, Inmaculada Villanueva

These reports are funded by the European Commission (DG CONNECT, LC-01485746)

PJC and MC received funding from "la Caixa" Foundation (ID 100010434), under agreement LCF/PR/GN17/50300003; CP, DL, SA, MC, received funding from Ministerio de Ciencia, Innovación y Universidades and FEDER, with the project PGC2018-095456-B-I00;

Methods

Methods

(1) Data source

Data are daily obtained from European Centre for Disease Prevention and Control (ECDC)[1] and country official sources (when indicated). 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 individual countries and for the UE+EFTA+UK as a whole:

- ✓ Number of cumulative confirmed cases
- ✓ Number of reported new cases
- ✓ Number of cumulative deaths

Then, two indicators are calculated and plotted, too:

- ✓ Case fatality rate: number of cumulative deaths divided by the number of cumulative 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* after applying a 7-day moving average to the new cases dataset, so that fluctuations (e.g., weekend effect) are smoothed. **Updated methodology to account for weekend effect is discussed and explained in reports #152[2] and #154[3].**

(3) Classification of countries according to their epidemic level: the scale Biocom-Cov

Countries are assigned a degree in the discrete Biocom-Cov scale, which aims to facilitate a simple way of assessing the situation of the country. It is based on the level of daily new cases per 100,000 inhabitants as follows:

Pandemic degree	Daily new incident cases per 10^5 inh.
0	0
1	0-0.1
2	0.1-0.5
3	0.5-1.25
4	1.25-2
5	2-3
6	3-5
7	5-8
8	8-14

[1] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
[2] https://upcommons.upc.edu/handle/2117/331959
[3] https://upcommons.upc.edu/handle/2117/332347

9	>14

(4) Fitting a mathematical model to data

Previous studies have shown that Gompertz model[4] 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 wave that is characterized by an initial exponential growth but a progressive decrease in spreading velocity provided that appropriate control measures are applied. Once in the tail, predictions work but the meaning of parameters is lost.

Gompertz model is described by the equation:

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

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 cumulative cases of the UE and of countries 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*.

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.

The model and its results on predictions in European countries during the first wave has been published in Plos Computational Biology[5].

(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 (3-5 days). 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 bar. For series longer than 9 timepoints, last 3 points are weighted in the fitting so that changes in tendencies are well captured by the model. **Updated methodology to account for weekend effect is explained in report #155[6].**

(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-

[4] Madden LV. Quantification of disease progression. Protection Ecology 1980; **2**: 159-176.
[5] Català et al, 2020, Plos Comput Biol 16(12): e1008431, https://doi.org/10.1371/journal.pcbi.1008431
[6] https://upcommons.upc.edu/handle/2117/332350

survivors[7] 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.

Full methodology and results have been published in Plos One[8].

[7] 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

[8] Català et al, PLoS ONE 16(1): e0243701, 2021. https://doi.org/10.1371/journal.pone.0243701