Daily report 15-03-2020

Analysis and prediction of COVID-19 for the countries of the European Union
Disclaimer: These reports have been written by declared authors, who fully assume their content. They are submitted daily to the European Commission, but this body does not necessarily share their analyses, discussions and conclusions.
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 days later.

The model and predictions are based on two parameters that are daily fitted to available data:
- $a$: 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, for the whole UE and for each country we show a report with 8 graphs and a table with the short-term predictions. Next page explains the information given by each graph.

We are currently adjusting the model to the whole UE and in the EU countries 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 $\rightarrow$ 3 days prediction;
- Group B: countries that have reported more than 100 cumulated cases for 7 to 9 consecutive days $\rightarrow$ 2 days prediction;
- Group C: countries that have reported more than 100 cumulated cases for 4 to 6 days $\rightarrow$ 1 day prediction.

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

Martí Català, MD; Sergio Alonso, PhD; Clara Prats, PhD; Daniel López, PhD
Computational Biology and Complex Systems
Universitat Politècnica de Catalunya - BarcelonaTech

HIGHLIGHTS (15-03-2020):

- Countries in Group A are: Germany, France, Spain and Italy
- Countries in Group B are: Austria, Belgium, Sweden, Netherlands
- Countries in Group C are: Denmark
- High lethality in Italy suggests a likely existence of many cases not diagnosed.
Countries’ reports details

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

Fitted $a$ value using points prior to each date

Estimated cases using death rate (see Methods)

Fitted $K$ value using points prior to each date

Evolution of $\rho$, a parameter related with Reproduction number (see Methods)

Reported new cases

Deaths / cumulated reported cases

Reported deaths

Iran (Islamic Republic of) 14-03-2020 population: 83.7M
EU 14-03-2020 population: 442.4M
Italy  14-03-2020 population: 59.4M
Spain 14-03-2020 population: 46.3M

Predictions for next days

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of cases</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-02-2020</td>
<td>1628</td>
<td>(1661 - 1691)</td>
</tr>
<tr>
<td>30-02-2020</td>
<td>9329</td>
<td>(3000 - 11599)</td>
</tr>
<tr>
<td>07-03-2020</td>
<td>12550</td>
<td>(8656 - 16695)</td>
</tr>
</tbody>
</table>

Number of confirmed cases

Time (day)

Number of cases per 10^5

Number of confirmed cases

Time (day)

Number of cases per 10^5

Number of new cases

Time (day)

Number of new cases per 10^5

K (Final number of cases)

Time (day)

Deaths

Time (day)

Deaths / Confirmed cases (%)

Time (days)
France  14-03-2020 population: 64.7M
Germany 14-03-2020 population: 81.9M

Predictions for next days

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of cases</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-03-2020</td>
<td>5953</td>
<td>(3569 - 5747)</td>
</tr>
<tr>
<td>13-03-2020</td>
<td>5851</td>
<td>(4999 - 6794)</td>
</tr>
<tr>
<td>14-03-2020</td>
<td>7264</td>
<td>(5263 - 9274)</td>
</tr>
</tbody>
</table>

Not enough data

Time (day)

Number of confirmed cases

Number of confirmed cases per 10^5

Number of cases

K (Final number of cases)

Time (day)

a (day⁻¹)

Time (day)

Number of new cases

Number of new cases per 10^5

Time (day)

Deaths

Number of new cases / Confirmed cases (%)

Deaths / Confirmed cases (%)
Sweden  14-03-2020 population: 9.8M
Belgium 14-03-2020 population: 11.4M

Predictions for next days

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of cases</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-03-2020</td>
<td>860</td>
<td>[599 - 1387]</td>
</tr>
<tr>
<td>16-03-2020</td>
<td>1005</td>
<td>[701 - 1420]</td>
</tr>
</tbody>
</table>

Not enough data

Not enough data

Not enough data

Not enough data

Number of new cases

Number of cases per 10^5

K (Final number of cases)

Number of deaths

Number of new cases per 10^5

Deaths / Confirmed cases (%)
Methods

(1) Data source

Data are daily obtained from World Health Organization (WHO) surveillance reports. 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
- Number of 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.
- \( \rho(\text{t} - 1) = \frac{N_{\text{new}}(\text{t}) + N_{\text{new}}(\text{t} - 1) + N_{\text{new}}(\text{t} - 2)}{N_{\text{new}}(\text{t} - 5) + N_{\text{new}}(\text{t} - 6) + N_{\text{new}}(\text{t} - 7)} \)

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

(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.

(4) Fitting a mathematical model to data

Previous studies have shown that Gompertz model 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

---

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

\[ N(t) = K e^{-\ln\left(\frac{K}{N_0}\right) e^{-\alpha(t-t_0)}} \]

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:

- \( \alpha \) 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 (\( \alpha \) 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 \( \alpha \). 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 \( \alpha \) 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 \( \alpha>0.2 \text{ day}^{-1} \), 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.

**(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 \texttt{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.

\textbf{(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-survivors\textsuperscript{2} 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.