http://dx.doi.org/10.5588/ijtld.20.0196

CORRESPONDENCE

Predictive models for COVID-19-related deaths and infections

A.G. Gerli,[1] S. Centanni,[2] M. Miozzo,[3] G. Sotgiu[4]

[1]Management Engineering, Tourbillon Tech srl, Padova, Italy; [2]Respiratory Unit, ASST Santi Paolo e Carlo, San Paolo Hospital, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy; [3]Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy. Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; [4]Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical, Experimental Sciences, University of Sassari, Sassari, Italy.

Correspondence to: Giovanni Sotgiu, Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Via Padre Manzella 4, Sassari, Italy. E-mail address: gsotgiu@uniss.it

Article submitted 30 March 2020. Final version accepted 7April 2020

Dear Editor,

The key role that robust models can play in predicting the incidence of infections and deaths has been highlighted since the beginning of the COVID-19 pandemic. Appropriate and early estimates of the impact of COVID-19 can help each country to implement preventive infection control measures. Furthermore, these models can enable policymakers plan the national and regional healthcare systems required (such as building intensive care units [ICUs]) and design tailored action plans.

An exponential model to predict the burden of infected patients was recently developed for Italy.[1] This computational approach proved reliable for the first 15 days of the epidemic but resulted in a significant overestimation of the level of infection after this period. We found several pre- and post-containment[2] R0 estimates (R0 being the average number of people who will catch the disease from a single infected person) for the first COVID-19 outbreak in the literature. Pedersen and Meneghini[3] predicted pre- and post-containment R0 values of respectively 2.59 and 1.9, at a reduction rate of 27%.

We attempted to fit this model to the Italian data, assuming peak infection day to be 21 March, when the estimated R was 1.27; we then applied a 27% reduction to get the new R value after containment (0.93). Next, we calculated R values using two other models, the first assuming the peak day to be 1 April and another on 15 April. Our values were respectively 1.22 and 0.76, and 1.18 and 0.81, with a reduction rate of 38% and 31%. The exponential model was assessed using the incidence of Italian and Chinese infections (Figure A). Although the first part of the curve in this model fits the trend in confirmed cases in Italy, the model leads to an overestimation of the incidence after 17 days (R^2 = 0.9991, Akaike's Information Criterion [AIC] = 6, before the overestimation). When used for cases in China, the exponential model resulted in a similar problem (i.e., overestimation) after 17 days, although model reliability was acceptable (R^2 = 0.9924, AIC = 6).

We then adopted a more reliable and robust data mining approach: the model was fitted with data retrieved from the Italian and Chinese epidemics, and several curves (e.g., exponential curves, third-degree polynomial curves, 5-parameter logistic [5PL] asymmetrical sigmoidal and Gaussian),[4,5] adapted to the public data sets, were explored.[6] In particular, the Chinese data[7] were divided into two sections—before and after the peak infection day (14 February 2020; 6,464 cases) to better understand the curve trend. Although 15,136 new cases were recorded on 13 February 2020,[3] only 37 had been notified the day before (this was due to the more comprehensive testing approach and delays in recording and reporting).

Data on infection growth up until the peak infection day can be fitted into a third-degree polynomial curve; thereafter, it fits into a 5PL asymmetrical sigmoidal curve following parametric growth (80% of aggregated cases at peak day, assumed to be at 20% of estimated aggregated outbreak duration; 90% of total cases at 24.4% of expected duration; 94% of total cases at 30% of expected duration; 97.5% at 40% of expected duration). The 80/20 per cent rule is applicable to this model, i.e., 80% of aggregated cases will occur in 20% of total outbreak time. Data on Chinese deaths could also be fitted to a mixed third-degree polynomial (based on a starting number of 10 aggregated deaths) up to the peak day, and then to a 5PL asymmetrical sigmoidal curve following parametric growth (45% of aggregated cases at peak day, which is at 20% of estimated aggregated outbreak duration; 50% of total cases at 20,8% of expected duration; 63,3% of total cases at 24% of expected duration; 81,4% at 30% of expected duration; 92,8% at 40% of expected duration; 99,5% at 80% of expected duration). Both curves are shown in Figure B. Using this model, it is possible to predict the trend in infection with accuracy from 17 days after the outbreak has begun. The correlation between expected and occurred deaths is high.

Our case scenario describing the Italian epidemic started on 22 February using third-degree polynomial curve was extrapolated to the first 17 days. Up to 20 March, only a difference of 2.7% between confirmed and expected cases and a difference of –1.3% between confirmed and expected deaths were found (Figure C). We then plotted the two curves of expected cases and deaths according to two different scenarios: peak day on 1 April and on 15 April (Figure D).

We conclude that the proposed predictive model, based on the biological assumption that "herd immunity" can reduce contagiousness in the population after an exponential increase, is valid, and results in improved strategic decision-making by limiting the spread of the SARS-CoV-2 virus, as well as by reducing the mortality rate. Nevertheless, the present health emergency can be adequately resolved only if the healthcare system is ready to adapt to the increased number of cases (mainly patients with severe disease).

An inappropriate estimation and prediction based on an inadequate forecasting model can affect the resilience of the national healthcare system and have undesirable consequences. We have established a new model for predicting COVID-19-related deaths and infections that can be easily applied worldwide for better strategic choices in any context. The model can be used to estimate the healthcare burden to implement and scale-up all the most relevant logistic, economic and financial interventions.

References

1 Remuzzi A and Remuzzi G. COVID-19 and Italy: what next? Lancet 2020; S0140-6736(20)30627-9

2 Frieden TR, Lee CT. Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020 Mar 18;26(6).

3 Pedersen, Morten & Meneghini, Matteo. (2020). Quantifying undetected COVID-19 cases and effects of containment measures in Italy. 10.13140/RG.2.2.11753.85600.

4 Ho KM. Effect of non-linearity of a predictor on the shape and magnitude of its receiver-operating-characteristic curve in predicting a binary outcome. Sci Rep 2017; 7(1): 10155.

5 Cerveri P, Forlani C, Borghese NA, Ferrigno G. Distortion correction for x-ray image intensifiers: local unwarping polynomials and RBF neural networks. MedPhys 2002; 29(8): 1759-1771.

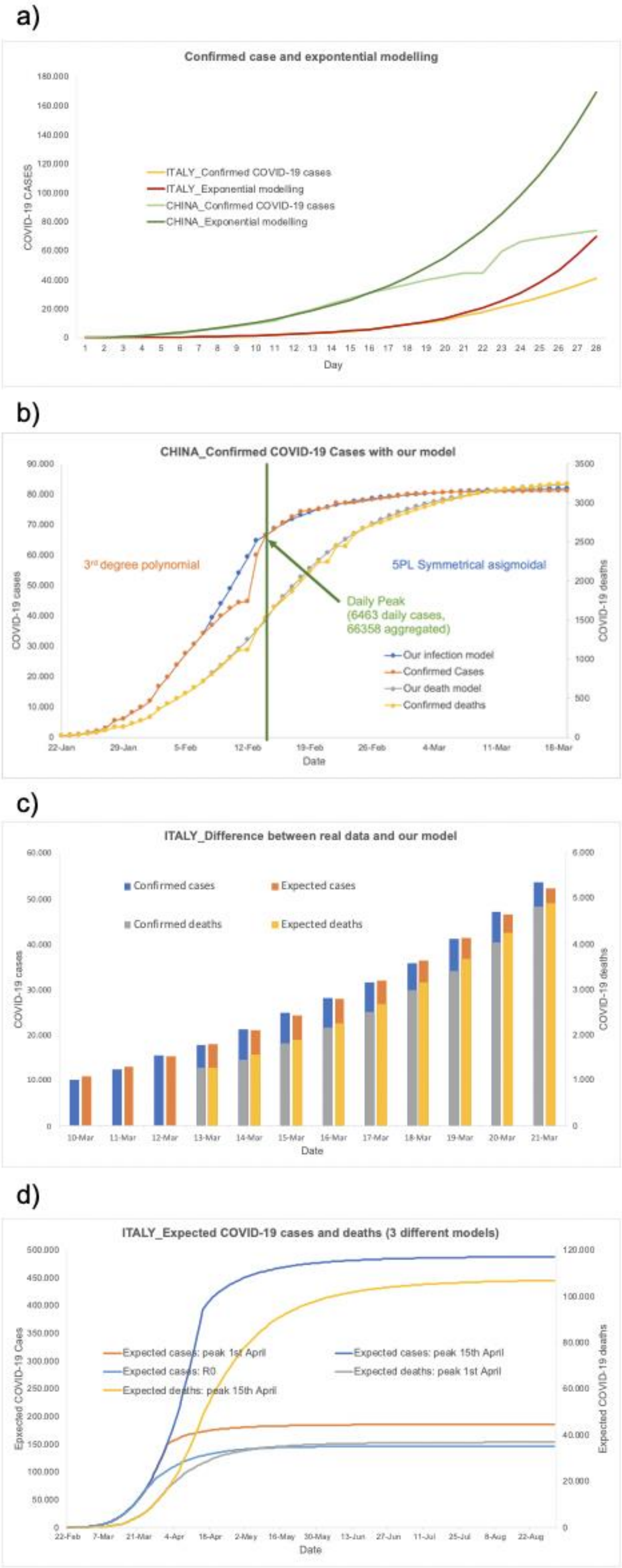
6 Johns Hopkins School of Public Health. Novel coronavirus (COVID-19) cases data. Baltimore, MD, USA: Johns Hopkins School of Public Health, 2020. https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

7 BBC. Coronavirus: Sharp increase in deaths and cases in Hubei, BBC, 13 February 2020. https://www.bbc.com/news/world-asia-china-51482994

FIGURE LEGEND

Figure Models describing COVID-19-related deaths and infection trends in Italy and China. **A)** Confirmed case and exponential modelling: the exponential modelling fits the trend of confirmed Italian (*y* = –320.7698 + 246.9722 EXP(0.2017873 *x)) and Chinese cases (*y* = –5922.295 + 4502.248 * EXP(0.1308463 * x)) only for the first part of the curve (17 days). It overestimates incidence after 17 days. **B)** China: confirmed COVID-19 cases with our model: i) infection growth for confirmed cases up to the peak day can be fitted into a third-degree polynomial curve (*y* = 1504.412 - 803.9791*x + 185.3493*x^2 - 1.338192*x^3), and then into a 5PL asymmetrical sigmoidal curve (*y* = 82729.91 + (20671.76 - 82729.91)/(1 + (x/15.99412)^38.33639)^0.08657033); ii) infection growth for deaths up to the peak day can be fitted into a third-degree polynomial curve (*y* = 28.57353 - 10.4954*x + 3.154154*x^2 - 0.007309942*x^3), and then into a 5PL asymmetrical sigmoidal curve (*y* = 3366.676 + (541.2956 - 3366.676)/(1 + (x/23.86309)^5.758494)^0.6063408). **C)** Italy: difference between real data and our model. Comparison of confirmed infected cases/deaths with the expected data extrapolated from our model (R^2 was respectively 0.9987 and 0.9964; $\chi^2 = 0$). **D)** Italy: expected COVID-19 cases and deaths (three different models): our model applied to Italian cases and deaths, based on peak day on 1 April or on 15 April; number of cases (third-degree polynomial curve): *y* = –275.2647 + 268.0452*x - 35.84856*x^2 + 3.066993*x^3; number of cases (5PL asymmetrical sigmoidal curve) if peak occurs on 1 April: *y* = 186230.2 + (42442.15 - 186230.2)/(1 + (x/26.42601)^40.05356)^0.0828589; number of cases (5PL asymmetrical sigmoidal curve) if peak occurs on 15 April: *y* = 489470.1 + (98216.56 - 489470.1)/(1 + (x/35.30432)^42.19767)^0.07864876; number of deaths (third-degree polynomial curve): *y* = –28.17647 + 32.61064*x - 6.425568*x^2 + 0.4785002*x^3; number of deaths (5PL asymmetrical sigmoidal curve) if peak occurs on 1 April: *y* = 36994.2 + (1764.126 - 36994.2)/(1 + (x/36.61827)^4.696123)^0.7724209; number of deaths (5PL asymmetrical sigmoidal curve) if peak occurs on 15 April: *y* = 107730.4 + (5137.291 - 107730.4)/(1 + (x/50.47383)^4.696123)^0.7724208.

Figure


a)
Confirmed case and expontential modelling
ITALY_Confirmed COVID-19 cases
ITALY_Exponential modelling
CHINA_Confirmed COVID-19 cases
CHINA_Exponential modelling
COVID-19 CASES
0
20.000
40.000
60.000
80.000
100.000
120.000
140.000
160.000
180.000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Day
b)
CHINA_Confirmed COVID-19 Cases with our model
3rd degree polynomial
5PL Symmetrical asigmoidal
Daily Peak
(6463 daily cases,
66358 aggregated)
Our infection model
Confirmed Cases
Our death model
Confirmed deaths
COVID-19 cases
0
10.000
20.000
30.000
40.000
50.000
60.000
70.000
80.000
90.000
COVID-19 deaths
0
500
1000
1500
2000
2500
3000
3500
22-Jan
29-Jan
5-Feb
12-Feb
19-Feb
26-Feb
4-Mar
11-Mar
18-Mar
Date
c)
ITALY_Difference between real data and our model
■ Confirmed cases
■ Expected cases
■ Confirmed deaths
■ Expected deaths
COVID-19 cases
0
10.000
20.000
30.000
40.000
50.000
60.000
COVID-19 deaths
0
1.000
2.000
3.000
4.000
5.000
6.000
10-Mar
11-Mar
12-Mar
13-Mar
14-Mar
15-Mar
16-Mar
17-Mar
18-Mar
19-Mar
20-Mar
21-Mar
Date
d)
ITALY_Expected COVID-19 cases and deaths (3 different models)
Expected cases: peak 1st April
Expected cases: R0
Expected deaths: peak 15th April
Expected cases: peak 15th April
Expected deaths: peak 1st April
Epxected COVID-19 Caes
0
50.000
100.000
150.000
200.000
250.000
300.000
350.000
400.000
450.000
500.000
Expected COVID-19 deaths
0
20.000
40.000
60.000
80.000
100.000
120.000
22-Feb
7-Mar
21-Mar
4-Apr
18-Apr
2-May
16-May
30-May
13-Jun
27-Jun
11-Jul
25-Jul
8-Aug
22-Aug
Date