When Will COVID-19 End? Data-Driven Prediction

Jianxi Luo
Data-Driven Innovation Lab (http://ddi.sutd.edu.sg)
Singapore University of Technology and Design (http://www.sutd.edu.sg)

Updated at 2 AM, April 28, 2020

On April 18, 2020, DDI Lab launched a webpage (https://ddi.sutd.edu.sg/when-will-covid-19-end/) (screenshot in Figure 1) to share data-driven predictions of the next developments and end dates of COVID-19 in different countries and has been continually updating the predictions daily with the latest data. For each country, a simple figure is provided to visualize the estimated pandemic life cycle together with actual data to date, which further reveals the predicted inflection point and ending phase. The prediction was started purely for self-curiosity regarding when COVID 19 might end in Singapore where we live and then has been expanded to cover other countries in response to requests of site visitors. Since April 18, the site has unexpectedly received millions of visitors from all over the world. This paper explains the motivation, theory, method, data and cautions for the predictions.

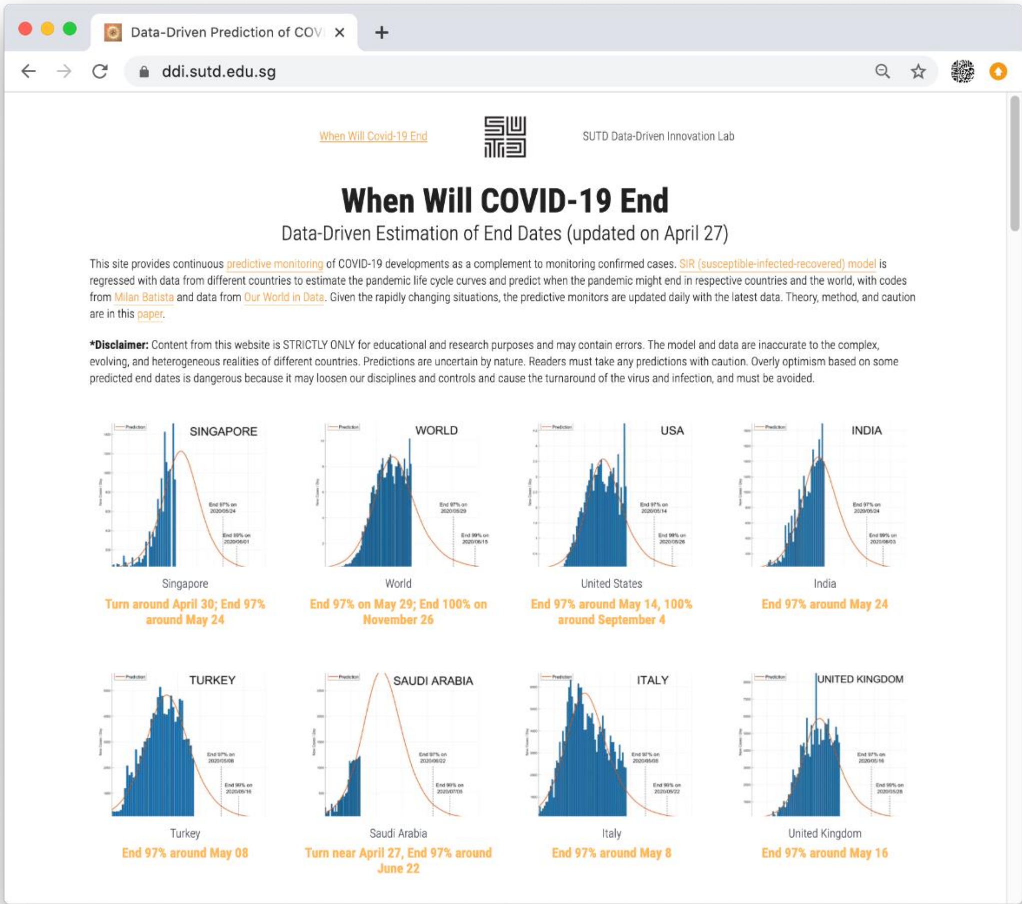

Figure 1. Screenshot of the website and predictions on April 27, 2020

Motivation

All of us around the world today naturally want to know when the COVID-19 pandemic will end. Estimating the end dates has been subconscious for most people as it is mentally needed and essential part of planning during the COVID-19 pandemic, but also naturally difficult to be done well due the uncertainty of future as a result of the complex, dynamic and heterogenous nature of the world we live in and the infectious disease we are facing. Meanwhile, our existing knowledge of historical pandemic process patterns and the continually accumulating data of the current pandemic make it possible to take a model-based and data-driven approach to objective predictions of the end dates of COVID-19 and also continually update the predictions as it evolves and generates more data. Such "predictive monitoring", i.e., the continual monitoring of predicted likely future events, such as the ending of the ongoing pandemic, using the latest data generated over time, might be able to reduce the anxiety from the blindness of future, gauge over-pessimism or over-optimism, stimulate pre-cautionary or pro-active actions, and make the planning, decisions, behaviours and mentality at the present moment more "future-informed". In contrast, most monitoring practices today focus on reporting actual cases of infection, recovery and death every day, which guides reactive and passive policies and actions, such as locking down a city only when many infections have been reported.

Theory and Methodology

The evolution of COVID-19 is not completely random. Like other pandemics, it follows a life cycle pattern from the outbreak to the acceleration phase, inflection point, deacceleration phase and eventual stop or ending. Such a life cycle is the result of the adaptive and countering behaviours of agents including individuals (avoiding physical contact) and governments (locking down cities) as well as the natural limitations of the ecosystem. However, the pandemic life cycles vary by countries, and different countries might be in different phases of the life cycles at a specific point in time. For instance, on April 21, in Singapore, Prime Minister Hsien-Loong Lee announced the extension of circuit breaker to June 1 in response to the spikes of COVID-19 cases, on the same day when Prime Minister Giuseppe Conte announced Italy's plan to reopen from May 4. Ideally such decisions and planning can be rationalized by well knowing where our own country (together with other countries and the world as a whole) is in its own pandemic life cycle, when the turning point is coming if it has yet come, and most importantly when the pandemic will end. The basis for such actionable estimation is the pandemic's life cycle.

The pandemic life cycle pattern is expected to appear as a S-shape curve when one plots the accumulative count of infection cases over time or equivalently as a "bell-shape" curve of the daily counts over time (see examples in Figure 2). Note that the bell here is not expected to be symmetrical with no expectation of a normal distribution. Such patterns as well as the underlying dynamics have been well studied in various domains including population growth, diffusion of new technologies and infectious diseases, and have theoretically established mathematical models, including the logistic model that describes a general life cycle phenomenon (such as population growth) and the SIR (susceptible-infected-recovered) model that describes the spread of infectious diseases. The context-specific and explainable SIR model is used in our predictions. In this paper we will not repeat to give the details of these models here, because they can be easily found in many mathematics textbooks and public websites. Both models incorporate two main parameters, whose values determine the shape of a specific life cycle curve. The model parameters for a country can be regressed based on actual data from the country. These models are not new and have many open-source Python and Matlab code implementations available online.

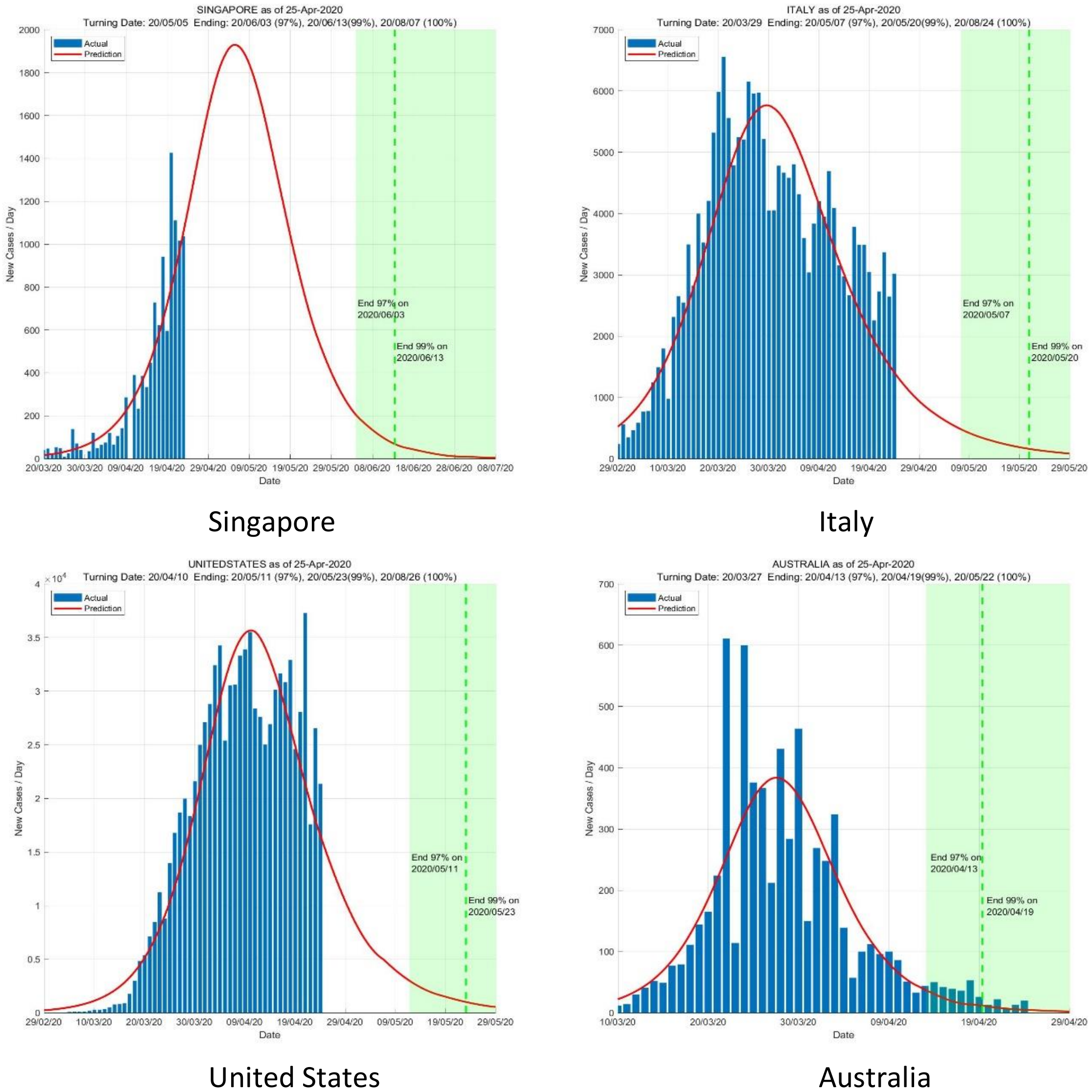

Figure 2. Model-Based Data-Driven Estimation of COVID-19 Life Cycle, Turning and Ending Dates in four countries (Singapore, Italy, United States and Australia) based on data as of 25 April 2020

In this case, to estimate the pandemic life cycle, daily updated COVID-19 data from Our World in Data [1] are used to regress the SIR model of COVID-19 using open-source codes from Milan Batista [2]. Regression is run for individual countries and updated daily with the newest data. The regressed model is used to estimate the full pandemic life cycle and plot the life cycle curve. The initial segment of the curve is fitted with the data to date and the remaining segment of the curve is predicted. With the estimated full life cycle curve (see examples in Figure 2), one can easily observe which phase of the pandemic life cycle a specific country is in (with actual data plotted together), when the inflection point (the peak in the bell-shape curve) is coming (for the interests of the countries still in the accelerating phase), and when the pandemic will end (for the interests of all countries).

The inflection point is specific as it appears as the peak in the bell-shape curve. However, estimating the “ending date” is not so straight-forward and may be done differently for different considerations. Most theoretically, one can define the end date as the day with the last infection of the pandemic, and thus operationalize the estimation of the end date as the day with the last predicted infection at the right most end of the estimated pandemic life cycle curve. However, practically, estimation of the definitive ending might not be useful to provide guidance for the planning of activities of governments, companies and individuals. One might consider an early date when the predominately most predicted infections (indicated by the regressed pandemic life cycle curve) have been actualized and only a small portion of the total predicted epidemic population is left. The total predicted infection size is the total area under the curve. Our latest predictions provide the following three alternative estimates of end dates in the order of conservativeness.

- The date to reach the last expected case;
- The date to reach 99% of the total expected cases;
- The date to reach 97% of the total expected cases.

Table 1 (provided in the end of the paper) reports the three alternative estimations of COVID-19 end dates of more than 100 countries as of 25 April 2020. The countries with unacceptable model-data fits and lack of statistical significance are excluded. In any case, specifying an end date is arbitrary in nature. With uncertainty and flexibility in mind, one may simply just exploit the estimated life cycle curve, especially its right most tail segment, to sense when the pandemic gradually ends to which extent according to his/her own preferences and needs. Alternatively, estimation as a range of dates might make sense for such uncertain predictions. And the estimated date range is expected to be become narrower as the country continually evolves along the pandemic life cycle curve to its end.

It is noteworthy that the bell-shape curve (rather than the S-shape curve) is chosen to

[1] Our World in Data. https://github.com/owid/covid-19-data/tree/master/public/data

[2] Milan Batista. https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19.

visualize the life cycle because it allows easy detection of the inflection point as the peak of the curve to distinguish countries in acceleration and deacceleration phases. For instance, Figure 2 based on data as of April 25 visually reveals Singapore was early in the acceleration phase, Italy and United States have already gone over their respective inflection points, and Australia has gone through the current pandemic curve more than 99% and is theoretically expected to end fully the pandemic around 22 June. Singapore is expected to bend the curve around May 5, to through 97% of the cycle in the country around June 4, whereas Italy and United States are predicted to end 97% of their pandemics on May 7 and May 11 respectively. The theoretical ends for Singapore, United States and Italy all fall in August.

Caution

Such predictive monitoring for each country should be read together with what are happening in the real world and government policy changes. For instance, Singapore government's strengthened restrictions in April may bend it curve earlier than the previously predicted ones, and the early relaxation of social distancing and lockdown in Italy and United States might increase infection rates and thus delay the ending as predicted now. Also, the predictive monitoring of a country should not be read in isolation, but together with the predictions and real time situations of other countries. No country is in isolation in the world today. The monitoring and control of one country must be coupled with the monitoring and control of other countries.

For example, while the predictive monitoring as of April 26 (refer to Table 1 in the end of the paper) tells the pandemic has "theoretically" ended in China (despite a small number of internal cases reported daily), it also suggests the world will still suffer till the end of 2020 if we remain in our present trajectories of government policies and individual behaviours and without medical cures and vaccines for COVID-19. Given the monitoring and predictions of other countries, the Chinese government is suggested not to open its international ports so soon and not lift the fundament control measures domestically so quickly, until the pandemic nears its end in the world as whole. Although it is the time for all of us to isolate and distance physically from each other, it is also the time that needs more sharing of data, information and knowledge and more close coordination. This is part of the motivation for the DDI COVID-19 prediction site.

Because of the complex, dynamic and heterogenous realities in different countries, the curve, inflection and end dates must be continuously re-estimated with the newest data from official channels every day. That is, the predictions themselves are also needed to be monitored over time, in addition to monitoring the actual cases. Especially, for Singapore and other countries that are still early in their own pandemic life cycles (Figure 2), the prediction of the rest of the curve, inflection point and ending dates will be more teasing and potentially valuable if done properly, but also less relevant to the "real future" to come given that the actual data only cover a smaller and early portion of the total life cycle. By contrast, for Italy and other countries that have passed their inflection points and been approaching ending phases of their present curves, prediction is expected to be more accurate because it is based on data covering more different phases of the life cycle, but also less useful. In such cases, the estimations are more about explaining the history and less about predicting the future. For those countries, a new epidemic wave might come if the governments and individuals lift controls and disciplines too early, especially when the pandemic is still prevalent in other countries.

Summary (tentative, to be updated)

The model-based and data-driven estimation of pandemic life cycle, inflection and end dates, if properly done, may reduce anxiety and over-optimism and prepare the mentality of all of us for the next phases of the epidemic evolution, no matter it is going to improve or worsen. Such predictive monitoring will allow the decisions and planning of the governments and companies that must be made now for the future to be more "future-informed". Our site complements with the widely available online dashboards and monitors of daily confirmed, death and recovery cases. We will continually fine tune the prediction and visualization methodology with the latest data and provide daily updates on our research lab website (https://ddi.sutd.edu.sg/when-will-covid-19-end/). This paper will be also updated continually with more learning and reflection down the road.

In the meantime, readers must take any prediction, regardless of the model and data, with caution. Over-optimism based on some predicted end dates is dangerous because it may loosen our disciplines and controls and cause the turnaround of the virus and infection. Although prediction based on science and data is aimed to be objective, it is uncertain by nature. One thing that is certain is that the model, data and prediction are inaccurate and insufficient to fully represent the complex, evolving, and heterogeneous realities of our world. The model is only theoretically suitable for one stage or wave of the epidemic evolution, and relatively more meaningful when applied to data for each single stage if the country has experienced multiple stages (such as Singapore). The prediction is also conditioned by the quality of the data. The data publicly available today is based on tests, which are done differently in different countries and over time periods.

Future is always uncertain. We must keep this in mind when reading any prediction. No one predicted the COVID-19 outbreak in November 2019, although Bill Gates famously warned about the potential damage of a global infectious disease in a talk in 2015. With acknowledging the uncertain nature of the ongoing COVID-19 pandemic and our growing inter-connected and complex world, what are eventually and fundamentally needed are the flexibility, robustness and resilience of people, organizations and governments, as well as sharing and coordination, to deal with unpredictable and unwanted future events.

Table 1. Three alternative estimations of COVID-19 end dates as of 26 April 2020
(sorted by estimated 97% end dates)

Countries	End 97%	End 99%	End 100%
World	30-May-20	16-Jun-20	27-Nov-20
China	27-Feb-20	4-Mar-20	9-Apr-20
South Korea	18-Mar-20	26-Mar-20	27-Apr-20
New Zealand	17-Apr-20	25-Apr-20	11-May-20
Australia	18-Apr-20	27-Apr-20	3-Jun-20
Vietnam	18-Apr-20	25-Apr-20	3-May-20
Iceland	19-Apr-20	25-Apr-20	13-May-20
Austria	20-Apr-20	29-Apr-20	14-Jun-20
Luxembourg	23-Apr-20	3-May-20	6-Jun-20
Niger	24-Apr-20	30-Apr-20	12-May-20
Jordan	25-Apr-20	7-May-20	23-May-20
Djibouti	26-Apr-20	29-Apr-20	8-May-20
Thailand	26-Apr-20	7-May-20	11-Jun-20
Lebanon	26-Apr-20	8-May-20	30-May-20
Switzerland	28-Apr-20	9-May-20	6-Jul-20
Cyprus	29-Apr-20	8-May-20	25-May-20
Uzbekistan	1-May-20	6-May-20	20-May-20
Spain	2-May-20	14-May-20	2-Aug-20
Andorra	2-May-20	13-May-20	4-Jun-20
Germany	4-May-20	16-May-20	5-Aug-20
Croatia	4-May-20	15-May-20	15-Jun-20
Norway	4-May-20	19-May-20	21-Jul-20
Costa Rica	4-May-20	15-May-20	4-Jun-20
Israel	5-May-20	17-May-20	8-Jul-20
Czech Republic	6-May-20	19-May-20	10-Jul-20
Latvia	6-May-20	20-May-20	17-Jun-20
Malaysia	7-May-20	20-May-20	8-Jul-20
France	7-May-20	19-May-20	8-Aug-20
Azerbaijan	7-May-20	17-May-20	15-Jun-20
Turkey	9-May-20	17-May-20	8-Jul-20
Italy	9-May-20	23-May-20	30-Aug-20
Portugal	9-May-20	20-May-20	17-Jul-20
Greece	9-May-20	24-May-20	12-Jul-20
Sudan	11-May-20	16-May-20	23-May-20
Slovenia	11-May-20	28-May-20	13-Jul-20
Philippines	11-May-20	23-May-20	7-Jul-20
Bangladesh	13-May-20	20-May-20	17-Jun-20
Iraq	13-May-20	28-May-20	9-Jul-20
Belgium	14-May-20	26-May-20	1-Aug-20
Macedonia	14-May-20	24-May-20	19-Jun-20
United States	15-May-20	27-May-20	5-Sep-20
United Kingdom	17-May-20	29-May-20	20-Aug-20

Japan	17-May-20	2-Jun-20	18-Aug-20
Moldova	17-May-20	27-May-20	2-Jul-20
Algeria	17-May-20	29-May-20	9-Jul-20
Ukraine	19-May-20	28-May-20	7-Jul-20
Belarus	20-May-20	29-May-20	7-Jul-20
Iran	21-May-20	10-Jun-20	24-Oct-20
United Arab Emirates	22-May-20	2-Jun-20	23-Jul-20
Canada	23-May-20	5-Jun-20	21-Aug-20
Dominican Republic	23-May-20	2-Jun-20	16-Jul-20
Peru	23-May-20	2-Jun-20	28-Jul-20
Cuba	24-May-20	7-Jun-20	11-Jul-20
Ireland	24-May-20	9-Jun-20	31-Aug-20
Denmark	24-May-20	17-Jun-20	19-Oct-20
Russia	25-May-20	5-Jun-20	19-Aug-20
India	25-May-20	4-Jun-20	1-Aug-20
Netherlands	25-May-20	10-Jun-20	9-Sep-20
Singapore	25-May-20	2-Jun-20	16-Jul-20
Oman	26-May-20	6-Jun-20	10-Jul-20
Romania	29-May-20	14-Jun-20	25-Aug-20
Brazil	30-May-20	9-Jun-20	16-Aug-20
Hungary	30-May-20	15-Jun-20	5-Aug-20
Poland	31-May-20	16-Jun-20	30-Aug-20
Finland	1-Jun-20	19-Jun-20	25-Aug-20
Indonesia	5-Jun-20	21-Jun-20	29-Aug-20
Panama	5-Jun-20	23-Jun-20	3-Sep-20
Kazakhstan	5-Jun-20	17-Jun-20	2-Aug-20
Georgia	7-Jun-20	26-Jun-20	29-Jul-20
Guinea	9-Jun-20	20-Jun-20	28-Jul-20
Egypt	9-Jun-20	4-Jul-20	29-Oct-20
Mexico	10-Jun-20	22-Jun-20	7-Sep-20
Chile	11-Jun-20	30-Jun-20	28-Sep-20
Argentina	14-Jun-20	3-Jul-20	15-Sep-20
Pakistan	15-Jun-20	29-Jun-20	12-Sep-20
Sweden	16-Jun-20	6-Jul-20	17-Oct-20
Afghanistan	22-Jun-20	9-Jul-20	4-Sep-20
Saudi Arabia	23-Jun-20	6-Jul-20	28-Sep-20
Kuwait	24-Jun-20	17-Jul-20	20-Oct-20
Bolivia	3-Jul-20	20-Jul-20	15-Sep-20
Colombia	4-Jul-20	26-Jul-20	1-Nov-20
Guatemala	19-Jul-20	18-Aug-20	18-Dec-20
Albania	20-Jul-20	18-Aug-20	29-Oct-20
Bahrain	26-Jul-20	4-Sep-20	6-Apr-21
Qatar	5-Aug-20	26-Aug-20	2-Feb-21