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)

On April 18, 2020, we launched a webpage (https://ddi.sutd.edu.sg/when-will-covid-19-end/) (screenshot in Figure 1) to share data-driven predictions of next developments and end dates of COVID-19 in different countries and have also been continually updating the predictions daily with latest data. For each country, a simple figure is provided to show the estimated pandemic life cycle together with the actual data or history to date, which in turn reveals the inflection point and ending phase. The predictions were started purely driven by personal curiosity regarding when COVID 19 will end in Singapore where we live and other countries. From the soft launch on April 18 to April 26, the website has received more than 1,000,000 unique visitors from all over the world. The data-driven predictions and visualizations have been circulated by individuals on the internet and used by researchers and government officials. In response to the growing popularity of our daily predictions, this paper explains the motivation, theory, method, and data behind the predictions.

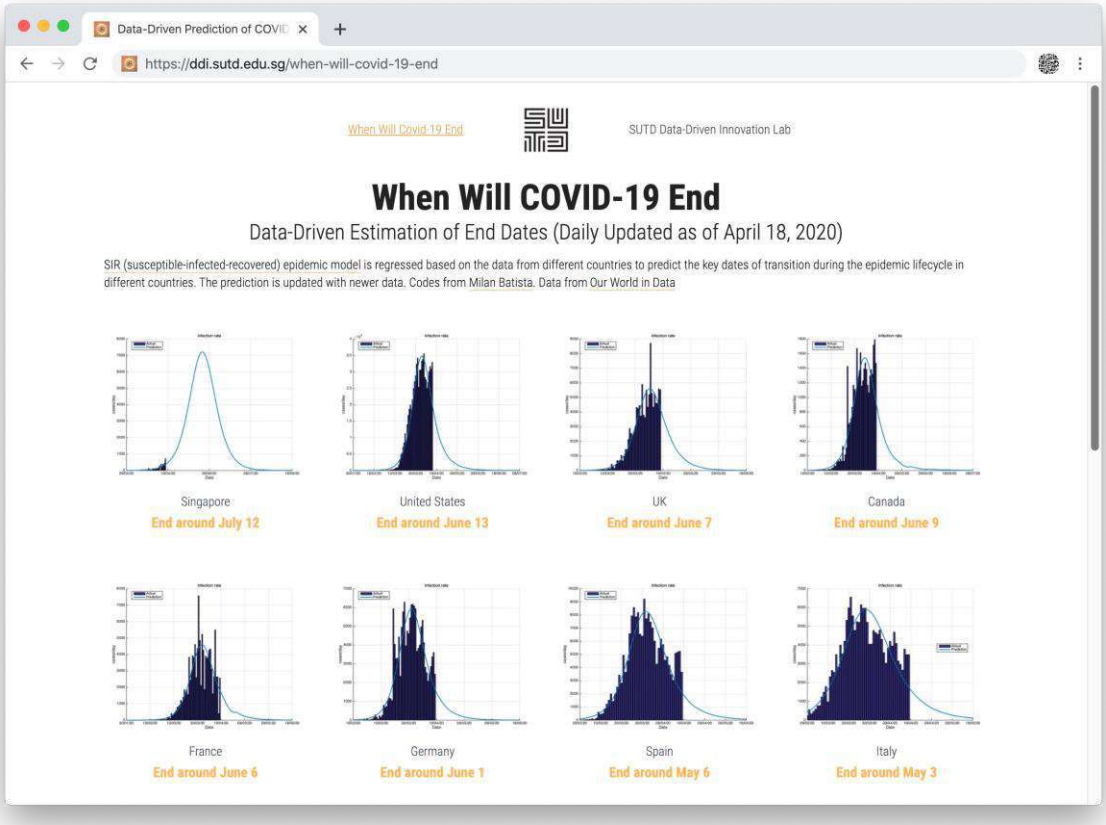


Figure 1. Screenshot of the website and predictions on April 18, 2020

(continuing on next page)

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. 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 predict the end dates of COVID-19 and also continually update the predictions as its evolves and generates more data. Such "predictive monitoring", i.e., the continual monitoring of predicted future events, such as the ending of the ongoing pandemic, using the latest data generated over time, can make the planning, decisions, behaviours and mentality at the present moment more "future-informed" and stimulate precautionary or proactive actions. In contrast, most monitoring practices today focus on reporting actual cases of infection, recovery and death every day, which mainly drives reactive and passive policies and actions, such as locking down a city only when many infections have been reported.

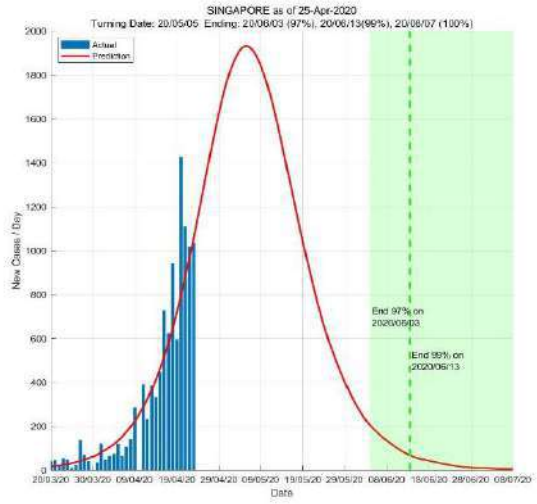
Theory and Methodology

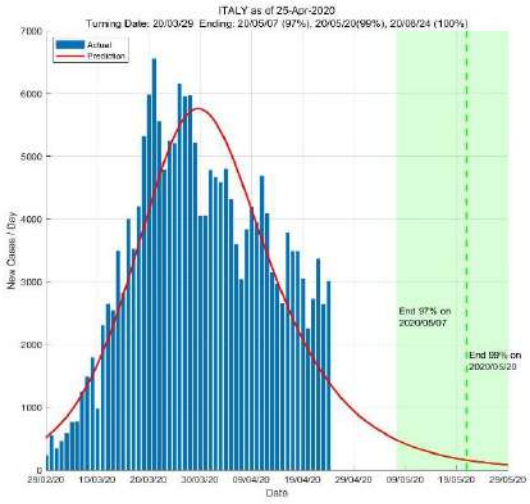
The evolution of COVID-19 is not completely random. Like other pandemics, it follows a life cycle pattern from the outbreak to the acceleration phase, inflection point, deacceleration phase and eventual stop or ending. Such a life cycle is the result of the adaptive and countering behaviours of agents 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 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 case counts over time (see examples in Figure 2). 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 parameters, whose values define 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.

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, one can easily see 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).





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

[2] Milan Batista (2020). fitVirusCOVID19, MIT Central File Exchange. Retrieved April 25, 2020. https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19.

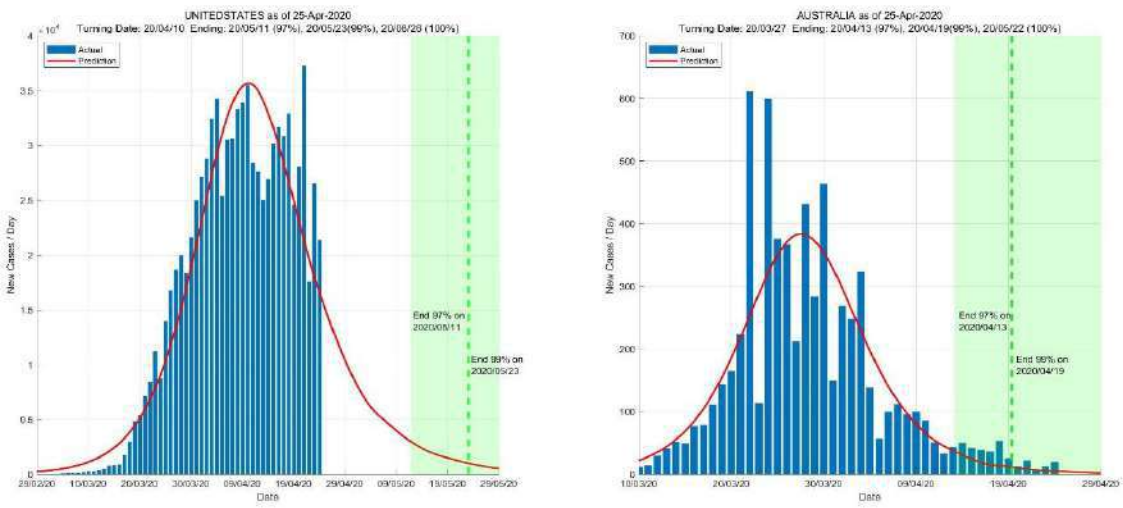


Figure 2. Model-Based Data-Driven Estimation of COVID-19 Life Cycle, Turning and Ending Dates in four countries (Singapore, Italy, United States and Australia) based on data as of 25 April 2020

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 one with the last predicted case in the pandemic life cycle curve. However, practically, estimation of that 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 as the end date from businesses, schools or governments, when most the 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. Our latest predictions provide the following three alternative estimates of end dates in the order of conservativeness.

- The date when the last expected case has identified;
- The date when 99% of the expected total cases have been identified;
- The date when 97% of the expected cases have been identified.

Table 1 (provided in the end of the paper) reports the three alternative estimations of COVID-19 end dates of 131 countries as of 25 April 2020. In any case, prediction and specifying an end date is arbitrary in nature. In fact, one may simply just exploit the estimated life cycle curve, especially its right most tail segment, to sense on which possible dates the pandemic gradually ends to which extents according to his/her own preferences and planning 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 visualize the life cycle curve because it allows easy detection of the inflection point as the peak of the curve to distinguish countries in different phases. For instance, Figure 2 based

on data as of April 25 visually reveal Singapore is still early in the acceleration phase, Italy and United States have already gone over their respective inflection points and approaching their ending phases, and Australia has finished its pandemic more than 99% and is expected to end fully the pandemic around 22 June. Singapore is expected to bend the curve around May 5 and end 97% of COVID 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. Such predictive monitoring should be read together with what are happening in the real world and the governments' policy changes. For instance, Singapore governments' increased restrictions from April may bent it curve earlier than the previous predictions, and the early relaxation of social distance and lockdown in Italy and United States might increase changes of infections and thus delay the ending as predicted now.

Because of the evolving nature of the pandemic, in particular the changing government policies and individual behavers, the curve, inflection and end dates must be continuously re-estimated with newest data from official channels every day for the purpose of predictive monitoring. Especially, for Singapore and other countries that are still early in their own pandemic life cycles, the prediction of the rest of the curve, inflection and ending dates will be more teasing and potentially valuable if done properly, but also expected to be 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, prediction is more accurate because it is based on data covering more 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.

Summary (Temporary)

The model-based and data-driven estimation of pandemic life cycle (visualized) and end date predictions, if properly done, may reduce anxiety 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 website complements with the widely available online dashboards and monitors of daily confirmed, death and recovery cases in different countries. We will continually fine tune the prediction and visualization methodology and update the reports on our research lab website (https://ddi.sutd.edu.sg/when-will-covid-19-end/). When more time is available, I plan to update this paper with more analyses and cases, as well as sharing of learning and reflections from this exercise.

In the meantime, readers must take any predictions, regardless of the model and data used, with caution. Overly optimism based on some predicted end dates might be dangerous

because it may loosen our disciplines and controls and cause the turnaround of the virus and infection rates. Note that the model behind our prediction is only theoretically suitable for one-stage epidemic. The prediction is also conditioned by the quality of the data. The reality is the future is always uncertain. No one predicted the COVID-19 outbreak in October or November 2019, although Bill Gates famously warned about the potential damage of a global infectious disease to the world during a TED 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 to deal with unexpected future events and scenarios.

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

Countries	Turning Date	End 97%	End 99%	End 100%
World	11-Apr-20	30-May-20	17-Jun-20	9-Dec-20
China	8-Feb-20	27-Feb-20	4-Mar-20	9-Apr-20
South Korea	2-Mar-20	22-Mar-20	31-Mar-20	12-May-20
Cambodia	21-Mar-20	4-Apr-20	12-Apr-20	12-Apr-20
Faeroe Islands	16-Mar-20	10-Apr-20	17-Apr-20	22-Apr-20
Brunei	16-Mar-20	11-Apr-20	21-Apr-20	21-Apr-20
Liechtenstein	20-Mar-20	13-Apr-20	18-Apr-20	18-Apr-20
Australia	27-Mar-20	14-Apr-20	20-Apr-20	23-May-20
New Zealand	29-Mar-20	17-Apr-20	24-Apr-20	10-May-20
Taiwan	24-Mar-20	17-Apr-20	26-Apr-20	7-May-20
Gibraltar	30-Mar-20	18-Apr-20	26-Apr-20	26-Apr-20
Iceland	28-Mar-20	18-Apr-20	25-Apr-20	13-May-20
Guernsey	4-Apr-20	19-Apr-20	25-Apr-20	29-Apr-20
Vietnam	25-Mar-20	19-Apr-20	30-Apr-20	15-May-20
Austria	26-Mar-20	20-Apr-20	29-Apr-20	14-Jun-20
Guam	1-Apr-20	20-Apr-20	28-Apr-20	28-Apr-20
Mauritius	1-Apr-20	20-Apr-20	27-Apr-20	3-May-20
Monaco	30-Mar-20	21-Apr-20	28-Apr-20	28-Apr-20
Luxembourg	27-Mar-20	23-Apr-20	3-May-20	6-Jun-20
Niger	8-Apr-20	24-Apr-20	29-Apr-20	10-May-20
Jordan	26-Mar-20	25-Apr-20	7-May-20	22-May-20
Madagascar	3-Apr-20	25-Apr-20	6-May-20	6-May-20
Montenegro	3-Apr-20	25-Apr-20	6-May-20	19-May-20
Barbados	31-Mar-20	26-Apr-20	30-Apr-20	30-Apr-20
Isle of Man	8-Apr-20	26-Apr-20	2-May-20	8-May-20
Lebanon	25-Mar-20	26-Apr-20	8-May-20	29-May-20
Thailand	28-Mar-20	26-Apr-20	7-May-20	12-Jun-20
Djibouti	16-Apr-20	27-Apr-20	1-May-20	9-May-20
Somalia	20-Apr-20	27-Apr-20	29-Apr-20	1-May-20
Switzerland	29-Mar-20	28-Apr-20	9-May-20	4-Jul-20

Cyprus	5-Apr-20	29-Apr-20	8-May-20	24-May-20
Sudan	22-Apr-20	30-Apr-20	2-May-20	3-May-20
Andorra	29-Mar-20	1-May-20	13-May-20	3-Jun-20
Croatia	2-Apr-20	3-May-20	14-May-20	13-Jun-20
Germany	1-Apr-20	3-May-20	15-May-20	1-Aug-20
Honduras	4-Apr-20	3-May-20	12-May-20	28-May-20
Uzbekistan	13-Apr-20	3-May-20	8-May-20	24-May-20
Costa Rica	30-Mar-20	4-May-20	16-May-20	5-Jun-20
Israel	4-Apr-20	4-May-20	15-May-20	4-Jul-20
Latvia	30-Mar-20	4-May-20	18-May-20	13-Jun-20
Myanmar	16-Apr-20	4-May-20	12-May-20	12-May-20
Norway	27-Mar-20	4-May-20	19-May-20	20-Jul-20
Slovenia	28-Mar-20	4-May-20	16-May-20	15-Jun-20
Spain	2-Apr-20	4-May-20	16-May-20	7-Aug-20
Czech Republic	1-Apr-20	5-May-20	18-May-20	9-Jul-20
Jersey	4-Apr-20	5-May-20	17-May-20	29-May-20
Tunisia	3-Apr-20	5-May-20	17-May-20	10-Jun-20
Azerbaijan	8-Apr-20	6-May-20	17-May-20	14-Jun-20
Ethiopia	12-Apr-20	6-May-20	15-May-20	15-May-20
France	3-Apr-20	6-May-20	18-May-20	5-Aug-20
Cameroon	8-Apr-20	7-May-20	16-May-20	8-Jun-20
Denmark	6-Apr-20	7-May-20	17-May-20	30-Jun-20
Malaysia	31-Mar-20	7-May-20	19-May-20	6-Jul-20
Portugal	6-Apr-20	7-May-20	18-May-20	11-Jul-20
Estonia	1-Apr-20	8-May-20	22-May-20	27-Jun-20
Italy	29-Mar-20	8-May-20	21-May-20	25-Aug-20
Greece	30-Mar-20	9-May-20	25-May-20	13-Jul-20
Jamaica	20-Apr-20	9-May-20	16-May-20	22-May-20
Philippines	7-Apr-20	9-May-20	20-May-20	2-Jul-20
Guyana	10-Apr-20	11-May-20	19-May-20	19-May-20
Malta	4-Apr-20	11-May-20	24-May-20	10-Jun-20
Serbia	15-Apr-20	11-May-20	20-May-20	23-Jun-20
United Arab Emirates	17-Apr-20	11-May-20	19-May-20	22-Jun-20
Burkina Faso	4-Apr-20	12-May-20	27-May-20	23-Jun-20
Iraq	4-Apr-20	12-May-20	1-Jun-20	4-Aug-20
Lithuania	3-Apr-20	12-May-20	26-May-20	2-Jul-20
United States	10-Apr-20	12-May-20	24-May-20	27-Aug-20
Belgium	8-Apr-20	13-May-20	24-May-20	27-Jul-20
Georgia	13-Apr-20	13-May-20	23-May-20	6-Jun-20
Moldova	13-Apr-20	13-May-20	23-May-20	25-Jun-20
Tanzania	23-Apr-20	13-May-20	19-May-20	29-May-20
Algeria	10-Apr-20	14-May-20	26-May-20	2-Jul-20
Ecuador	12-Apr-20	14-May-20	25-May-20	11-Jul-20
Ireland	15-Apr-20	15-May-20	24-May-20	9-Jul-20
Kyrgyzstan	11-Apr-20	15-May-20	27-May-20	18-Jun-20
Macedonia	12-Apr-20	15-May-20	25-May-20	21-Jun-20

Peru	18-Apr-20	15-May-20	25-May-20	13-Jul-20
Bosnia And Herzegovina	7-Apr-20	16-May-20	30-May-20	4-Jul-20
Oman	19-Apr-20	16-May-20	24-May-20	19-Jun-20
United Kingdom	12-Apr-20	16-May-20	27-May-20	14-Aug-20
Mali	21-Apr-20	17-May-20	25-May-20	6-Jun-20
Turkey	14-Apr-20	17-May-20	29-May-20	12-Aug-20
Canada	12-Apr-20	18-May-20	30-May-20	9-Aug-20
Bangladesh	23-Apr-20	19-May-20	30-May-20	15-Jul-20
Japan	14-Apr-20	18-May-20	5-Jun-20	26-Sep-20
Paraguay	10-Apr-20	19-May-20	31-May-20	12-Jun-20
Egypt	18-Apr-20	20-May-20	30-May-20	8-Jul-20
Iran	1-Apr-20	20-May-20	9-Jun-20	22-Oct-20
Russia	24-Apr-20	20-May-20	28-May-20	20-Jul-20
India	20-Apr-20	22-May-20	1-Jun-20	26-Jul-20
Saudi Arabia	27-Apr-20	22-May-20	30-May-20	11-Jul-20
Brazil	21-Apr-20	23-May-20	2-Jun-20	2-Aug-20
Kosovo	18-Apr-20	23-May-20	4-Jun-20	30-Jun-20
Liberia	18-Apr-20	24-May-20	6-Jun-20	14-Jun-20
Netherlands	8-Apr-20	24-May-20	9-Jun-20	8-Sep-20
Cuba	16-Apr-20	25-May-20	8-Jun-20	14-Jul-20
Ukraine	21-Apr-20	25-May-20	6-Jun-20	28-Jul-20
Dominican Republic	18-Apr-20	26-May-20	8-Jun-20	27-Jul-20
Panama	12-Apr-20	27-May-20	12-Jun-20	12-Aug-20
Poland	13-Apr-20	27-May-20	11-Jun-20	19-Aug-20
Romania	13-Apr-20	27-May-20	11-Jun-20	20-Aug-20
Argentina	11-Apr-20	30-May-20	15-Jun-20	12-Aug-20
Hungary	15-Apr-20	30-May-20	15-Jun-20	5-Aug-20
Botswana	10-Apr-20	31-May-20	19-Jun-21	31-May-20
Morocco	24-Apr-20	1-Jun-20	13-Jun-20	30-Jul-20
Finland	11-Apr-20	2-Jun-20	20-Jun-20	26-Aug-20
Gabon	24-Apr-20	2-Jun-20	15-Jun-20	5-Jul-20
Belarus	29-Apr-20	4-Jun-20	16-Jun-20	14-Aug-20
Chile	16-Apr-20	4-Jun-20	21-Jun-20	11-Sep-20
Singapore	5-May-20	4-Jun-20	14-Jun-20	8-Aug-20
Kuwait	22-Apr-20	6-Jun-20	28-Jun-20	10-Nov-20
Indonesia	20-Apr-20	7-Jun-20	24-Jun-20	7-Sep-20
Pakistan	27-Apr-20	9-Jun-20	23-Jun-20	1-Sep-20
Armenia	16-Apr-20	11-Jun-20	29-Jun-20	20-Aug-20
Bolivia	30-Apr-20	11-Jun-20	25-Jun-20	31-Jul-20
Ghana	1-May-20	14-Jun-20	28-Jun-20	16-Aug-20
Mexico	1-May-20	14-Jun-20	29-Jun-20	18-Sep-20
Cote D Ivoire	24-Apr-20	17-Jun-20	6-Jul-20	28-Aug-20
Slovakia	23-Apr-20	17-Jun-20	4-Jul-20	26-Aug-20
Afghanistan	29-Apr-20	18-Jun-20	5-Jul-20	28-Aug-20
El Salvador	23-Apr-20	19-Jun-20	8-Jul-20	9-Aug-20
Sweden	20-Apr-20	19-Jun-20	10-Jul-20	26-Oct-20

Colombia	25-Apr-20	23-Jun-20	13-Jul-20	6-Oct-20
Guinea	17-May-20	24-Jun-20	6-Jul-20	25-Aug-20
Kazakhstan	1-May-20	25-Jun-20	14-Jul-20	27-Sep-20
Guatemala	4-May-20	26-Jun-20	16-Jul-20	8-Sep-20
Kenya	21-Apr-20	2-Jul-20	27-Jul-20	8-Sep-20
Albania	16-Apr-20	6-Jul-20	2-Aug-20	5-Oct-20
South Africa	3-May-20	9-Jul-20	31-Jul-20	29-Oct-20
Qatar	27-May-20	26-Jul-20	20-Aug-20	15-Feb-21
Democratic Republic of Congo	8-May-20	6-Aug-20	5-Sep-20	19-Nov-20
Bahrain	16-May-20	7-Aug-20	9-Sep-20	12-Feb-21