ORIGINAL ARTICLE

Forecasting Malaysia COVID-19 Incidence based on Movement Control Order using ARIMA and Expert Modeler

Edre MA, Muhammad Adil ZA, Jamalludin AR
Department of Community Medicine, Kulliyyah of Medicine, International Islamic University Malaysia

ABSTRACT

INTRODUCTION: Coronavirus disease (COVID-19) is a novel pandemic that affects every other country in the world. Various countries have adopted control measures involving restriction of movement. Several studies have used mathematical modelling to predict the dynamic of this pandemic. Forecasting techniques can be used to predict the incidence cases for the short term. The study aims to forecast the COVID-19 incidence using the Auto Regressive Integrated Moving Average (ARIMA) method. **MATERIALS AND METHODS:** Using publicly available data, we performed a forecast of Malaysia COVID-19 new cases using Expert Modeler Method in SPSS and ARIMA model in R to predict COVID-19 cases in Malaysia. We compare 3 different time frames based on different Movement Control Order (MCO) period. We compare the model fit and prediction across models. **RESULTS:** All models show static cases for each MCO 7-day prediction. For prediction until 12 May, the third MCO time frame shows the best model fit for both techniques. Both software shows a stationary trend of cases of below 100. **CONCLUSION:** These MCO models have shown to stabilize the rate of new cases. Further sub analysis and quality of data is needed to improve the accuracy of the model.

KEYWORDS: Malaysia COVID-19, Forecast, ARIMA, Expert Modeler, Movement Control Order

INTRODUCTION

Coronavirus Disease 2019 (COVID-19) is a serious pandemic that threatens to affect every country in the world. As of 5th of June 2020, there are 8266 confirmed cases in Malaysia with 116 mortality.[1] Malaysian Prime Minister announced on 16 March 2020 that Malaysia is adopting a Movement Control Order (MCO) starting from 18 March 2020 to control the disease from spreading.[2] MCO is a non-pharmaceutical intervention (NPI) aimed at flattening the epidemic curve on a large scale, that have shown to reduce early transmission in China.[3] Since 18 March 2020, there are a total of five various phases of MCO in Malaysia.[4] The effect of the MCO has shown a reducing trend in incidence cases.[5] This has prompted the government to loosen the MCO before eventually lifting the order.[4] Many studies have used mathematical modelling to predict the dynamic of this pandemic.[6] This prediction is an important parameter for many decisions making.

Corresponding Author:
Asst. Prof. Dr. Edre Mohammad Aidid,
Department of Community Medicine,
Kulliyyah of Medicine,
International Islamic University Malaysia,
Jalan Sultan Ahmad Shah, Bandar Indera Mahkota,
25200 Kuantan, Pahang, Malaysia
Tel no: +609-5704605
E-mail: edreaidid@iium.edu.my

Public health decision making came from the surveillance. One of the main public health surveillance aim is to make reliable forecast.[7] Forecasting is a mathematical modelling technique to predict future outcome based on the past values and trend. It is widely used not just in economic sectors but also in health.[8] A systematic review on forecasting influenza outbreak have shown some degree of accuracy despite various technique that is being used.[7] These forecasting techniques are useful for prediction of new cases to help planning in prevention and control measure.[7] There are few methods that can be used for forecasting such as compartmental model, deep learning and time series.[8] The most widely popular technique used the basic compartmental model such as SIR model which is based on susceptible, infectious and recovered classes and the functions of each of this compartment change with time. Secondly the deep learning method or neural networks which utilize the emergence of big data. Finally, the time series method that used for short term diseases. Most of current predictions focus on SIR models which is the simplest form of the compartmental models of Susceptible(S), Infected(I) and Recovered(R).[8] However, these models require a lot of assumptions and could be misleading. The more complex models of SIR in fact require more parameters and thus become less reliable in predictions.[9] Thus, the time series forecasting model provides flexible and efficient prediction of COVID-19 cases and its related outcomes to help make better informed decisions.

Several studies have used time series method using the Auto Regressive Integrated Moving Average (ARIMA) to predict incidence cases.[10,11] The techniques commonly used are the ARIMA and exponential smoothing (ETS). Various statistical software such as R Studio and IBM SPSS have incorporated automated features to best fit the prediction model. ARIMA has three components which are the autoregressive (p), differencing (d) and moving average (q).[12] These three components provide the framework to model the prediction better. The advantage of using ARIMA is that it can incorporate regressive predictors to improve forecasting accuracy.[13] However, the model is prone to overfit and underfit if careful modelling consideration is not taken such as limited historical data or timepoints.[13]

Auto ARIMA is one of the convenient features in R, giving the ability of the statistical package to choose the best fitting model. On the other hand, Expert Modeler method using ETS has various specific models, broadly classified into seasonal and non-seasonal.[14] The advantage is that the more recent the observation the higher the associated weight.[14] Thus, it is important to compare these forecasting methods as there are paucity of studies comparing the forecasting accuracy and performance between these 2 methods. The aim of this study is to forecast COVID-19 incidence using ARIMA and Expert Modeler method, as well as to compare the result using different time frame dataset and statistical software.

MATERIALS AND METHOD

We used datasets from publicly available domain.[15] The data was captured from the official Malaysian Ministry of Health (MOH) press release daily on COVID -19. The datasets contain numbers of newly reported confirmed cases extracted from the daily Ministry of Health press conference and official press release. We compare the result of ARIMA using two methods namely the Auto ARIMA and Expert Modeler method in RStudio and IBM SPSS, respectively. The analyses automatically select the best fitting model. The Auto ARIMA function in R uses a variation of the Hyndman-Khandakar algorithm. The model is based on MLE, unit root test and minimization of AICc.[16] Whereas, for the Expert Modeler method, the forecasting package in SPSS will automatically choose the best fit model based primarily on ETS in this analysis as there are various smoothing techniques such as Holt, Winter's Additive and so on.[17] In addition, the Expert Modeler also considers seasonality patterns.[14]

We also compare forecast result based on three different time frames. The timeframe used which are from MCO 1 (18 March 2020), MCO 2 (1 April 2020) and MCO 3 (15 April 2020) until the end of phase 4 MCO (12 May 2020). In each time frame, we compare the models using the statistical method mentioned above. The model fit and predicted incidence until 12 May 2020 was compared across all models and techniques. Additionally, we also compared the predicted 7 days of new cases starting from MCO 1, MCO 2 and MCO 3 with the actual number of new cases, taking the x-axis as days since the first COVID-19 case in Malaysia.

RESULTS

Model fit was based on parameters such as R squared (R^2), root mean squared error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE). In general, the smaller the value the better the fit except for R squared. The results are tabulated in Table I below. Both methods show the MCO3 model shows better fit in terms of R-squared, RMSE or MAE parameters.

Expert Modeler

Taking day 1 as the start of MCO 1, the predicted number of cases appears to be stabilized after 2 May 2020 until the end of MCO 4 and below 100. The analysis used expert modeler (simple, non-seasonal). Similarly, the MCO 2 model shows a static number of predicted new cases, below 100 from 3 May until 12 May 2020. The prediction interval becomes wider with time. For the MCO 3 model, the expert modeler selects simple seasonal exponential smoothing techniques to forecast the new cases. Similarly, the number of new cases were also forecasted to be

Table I: Fit parameters and estimates for MCO 1, MCO 2 and MCO 3 models

Parameter	MCO 1		MCO 2		MCO 3	
	AUTO ARIMA	EXPERT MODELER	AUTO ARIMA	EXPERT MODELER	AUTO ARIMA	EXPERT MODELER
R^2	-	0.54	-	0.61	-	0.31
RMSE	35.69	36.08	32.28	33.63	23.45	20.72
MAPE	28.65	29.07	30.81	32.13	35.83	28.06
MAE	27.86	-	26.33	-	20.17	-
Maximum likelihood estimate (MLE)	-239.65	-	-318.08	-	-391.11	-
Alpha (smoothing factor) estimate	-	0.67	-	0.47	-	0.42

below 100 until 12 May 2020 even though there was variability in the prediction interval. Figure 1 shows the three models in comparison.

7-day forecast using Expert Modeler

The 7 days forecasting trend of MCO 3 model has the best fit based on MAPE and R^2. The 7-days predicted number of cases were static for all 3 models, initially increased from MCO 1 to MCO2 and then decreased in the MCO 3 model. The predicted cases did not differ much in MCO 1 and MCO 2 model except for MCO 3 model. Table II summarizes the findings.

Auto ARIMA

The Auto Arima function selects the simple exponential smoothing technique for the first dataset from the first MCO. The predicted incidence was static but below 100. For the second MCO dataset, the ARIMA parameter selected was (2,1,0). The predicted incidence trend was within 100 cases per day and shows a stationary trend towards the end. In the third MCO model as in Figure 2, it shows stationary trend as in the first MCO model. The forecast incidence using auto arima all shows below 100 cases.

The forecasting trend of MCO 1 model has the best fit based on RMSE and MAE. The 7-days predicted number of cases were static for all 3 models, initially increased from MCO 1 to MCO2 and then decreased in the MCO 3 model. The predicted cases did not differ much in MCO 1 and MCO 2 model except for MCO 3 model. Table III summarizes the findings.

Table II: Actual versus predicted new cases in all three models using Expert Modeler

Days since MCO	MCO 1[a]		MCO 2[b]		MCO 3[c]	
	Predicted (95% CI)	Actual	Predicted (95% CI)	Actual	Predicted (95% CI)	Actual
1	119 (75, 163)	110	145 (90, 201)	208	126 (66, 186)	110
2	119 (66, 172)	130	145 (85, 206)	217	126 (61, 191)	69
3	119 (59, 180)	153	145 (79, 211)	150	126 (57, 195)	54
4	119 (52, 186)	123	145 (75, 216)	179	126 (52, 200)	84
5	119 (46, 192)	212	145 (70, 221)	131	126 (48, 204)	36
6	119 (40, 198)	106	145 (66, 225)	170	126 (44, 208)	57
7	119 (25, 205)	172	145 (62, 229)	156	126 (41, 211)	50

[a] R^2=0.65, RMSE=21.96, MAPE=66.56
[b] R^2=0.83, RMSE=27.59, MAPE=46.69
[c] R^2=0.84, RMSE=29.99, MAPE=40.20

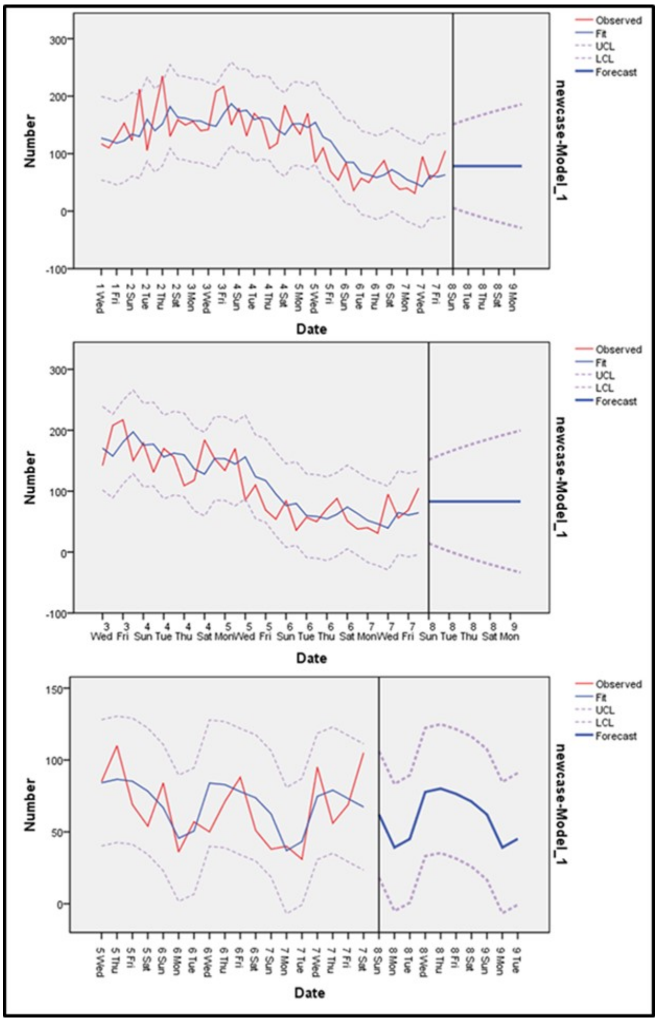

Figure 1: Prediction forecast of new cases until 12 May 2020 based on Expert Modeler

Table IV summarizes the forecasted new cases of COVID-19 in Malaysia based on the three MCO models. All models show predicted cases below 100.

DISCUSSION AND CONCLUSION

MCO 3 incorporates enhanced MCO (EMCO) and the normal MCO together with social distancing, hand hygiene and wearing of face masks outside.[4] EMCO is when all residents within the area are forbidden

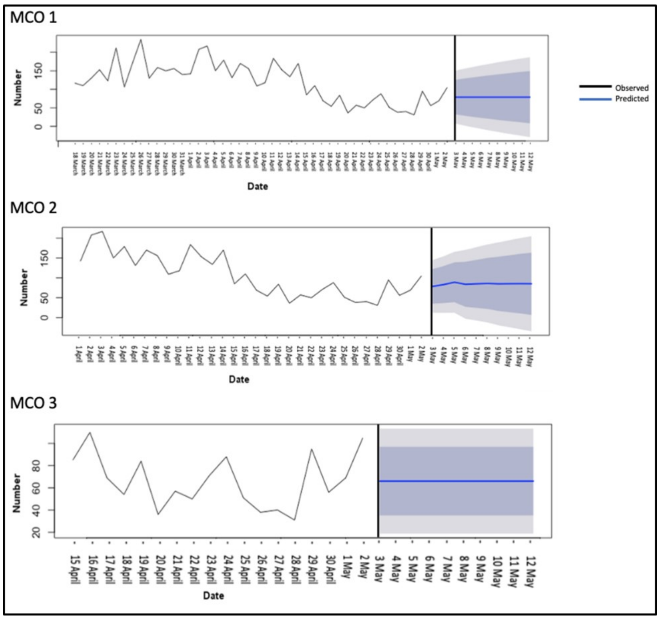

Figure 2: Prediction forecast of new cases using MCO 3 model based on Auto ARIMA

Table III: Actual versus predicted new cases in all three models in Auto Arima

Days since MCO	MCO 1[a]		MCO 2[b]		MCO 3[c]	
	Predicted (95% CI)	Actual	Predicted (95% CI)	Actual	Predicted (95% CI)	Actual
1	117 (73, 161)	110	145 (90, 200)	208	126 (66,185)	110
2	117 (63, 171)	130	145 (84, 206)	217	126 (61,190)	69
3	117 (55, 179)	153	145 (79, 211)	150	126 (56,195)	54
4	117 (48, 186)	123	145 (74, 217)	179	126 (52,200)	84
5	117 (41, 193)	212	145 (70, 221)	131	126 (47, 204)	36
6	117 (35, 199)	106	145 (65, 226)	170	126 (43,208)	57
7	117 (29, 205)	172	145 (61, 230)	156	126 40, 211)	50

[a]RMSE=22.03, MAE=6.62
[b]RMSE=27.62, MAE=11.60
[c]RMSE=29.99, MAE=15.17

Table IV: Forecasted new cases from 3 May 2020 until 12 May 2020

MAY 2020	Number of new cases, n					
	MCO 1		MCO 2		MCO 3	
	AUTO ARIMA	EXPERT MODELER	AUTO ARIMA	EXPERT MODELER	AUTO ARIMA	EXPERT MODELER
3	79	78	78	83	66	62
4	79	78	83	83	66	39
5	79	78	89	83	66	45
6	79	78	84	83	66	78
7	79	78	85	83	66	80
8	79	78	86	83	66	76
9	79	78	85	83	66	71
10	79	78	85	83	66	62
11	79	78	85	83	66	39
12	79	78	85	83	66	45

from getting out their homes during the order, visitors outside the area cannot enter into the area subjected to the order, any business transaction is not allowed in the area and all supplies must be given to the enforcers from outside the barb wires in order to be given to the residents inside the area.[18] The prediction here is before the conditional MCO is enforced, where the conditions of the MCO are relaxed such as allowing opening of economic sectors with strict standard operating procedures.[18]

For auto ARIMA, the rate stabilized after phase 3 MCO, which is similar to the situation in a study in Italy.[19] Interestingly, for the Expert Modeler method it incorporates the element of seasonality as the historical data appears to have a seasonal trend. This is largely due to the appearance of clusters that require EMCO rather than the normal MCO. In addition, the stepping down of the red zone to yellow zone allows the change from EMCO to the normal MCO. These active public health interventions allow some degree of fluctuations in the number of cases but still does not allow exponential increase in the cases. Lipsitch (2020) reported SARS-CoV2 seasonality in terms of environment, human behavior, host's immune system and depletion of susceptible host.[20] The auto arima chose the 0,1,1 model, thus differencing of 1 and moving average of 1. It indicates that the rate stabilizes. This may indicate that after nearly seven weeks of MCO, the number of recoveries slightly exceeds the new cases and this also explains the stabilized graph may indicate effective social distancing measures with tolerable daily cases, an early sign of controlled epidemic. ARIMA analysis comparing multiple countries show that more strict intervention leads to a stationary trend of incidence.[21] Manual ARIMA requires a more selective model and approach to have the best prediction based on the type of outcome selected. It was reported that European countries that use total cumulative cases will have a slightly difference forecast of future cases as compared to new daily cases prediction.[22] Thus, any forecasting is data-driven, and care is needed in interpretation.

In terms of seasonal changes of the virus, much is not known and need to be explored if it happens that we are facing the same virus as time goes by, or a different strain either new or mutated. But all the models show no huge trends that can suggest virus mutation for the past 7 weeks. Lifting of MCO should consider biological, social, physical and political wellbeing. Since the forecasts are still double figure, complete lift of MCO could result in a surge of cases and clusters especially among the susceptible host. MCO should therefore be relaxed in stages with strong community empowerment. This is why the federal government recently opted for the conditional MCO (CMCO).[18]

The short-term forecasting using 7 days as the forecasting period is useful to make informed public health decisions at the policy level. For instance, the announcement of MCO was done a few days before the end of the 2-week period. First, it signifies half of the incubation period, allowing for asymptomatic cases to be screened before having overt symptoms. Second, literature has suggested that the doubling time of seven days is a signal for public health intervention to be done swiftly to reduce the transmission.[23]

Due to the data-driven nature of this analysis, the forecasting of new cases should be carefully interpreted. This is based on new cases but not based on case date of onset, which can tell much more about the dynamics of the disease. Similarly, prediction using incidence rate may be better in knowing the actual scenario. A geoanalysis according to hotspots or district can be more valuable in determining MCO outcome selectively. The forecast shows cases under 100. However, it is noted that the prediction interval is important because it does cover the possible range of new cases that correspond to the actual number of cases.[12] Here there are some variabilities noted. Future analyses should focus on adding other predictors such as host immunity and contact time.[24,25] The statistical inference here may be best suited if it can be practically applied in the real world. Future studies need to include other outcome parameter that would be helpful as a marker for exit strategy for MCO, such as number of beds occupied in COVID referral hospitals such as in a recent paper in China.[26]

Adding to the limitation of the forecasting lies in the data quality. The reported data of the confirmed cases and the population at risk are not random as the relevant authorities only did targeted sampling.[27] Hence what we observed are the partly probability of certain areas or population to be tested. The cluster groups tested varies in size, from markets to foreign workers within the same vicinity.[27] Hence the distribution of cases reported may not follow any natural distribution and does not have any real predictable trend. Since forecasting rely on the past data, the prediction is as good as the trend.

In conclusion, the MCO models in this analysis have shown to stabilize the rate of new cases. Further sub analysis and quality of data is needed to improve the accuracy of the model. The basis of forecast is not only to give highly accurate predictions, but also to prepare the government to make well-informed decisions to control COVID-19 and ease the way for the new-normal in the community. It is recommended that the quality of data is improved in terms of timeliness and suitable parameters such as patient profile made transparent so that the forecasting becomes more representative in the future.

CONFLICT OF INTEREST

We declare that there is no conflict of interest.

ACKNOWLEDGEMENTS

We would like to acknowledge Mr Erhan Azrai for the COVID-19 dataset that was publicly available through dataworld.

REFERENCES

1. Abdullah NH. Kenyataan Akhbar KPK 5 Jun 2020 – Situasi Semasa Jangkitan Penyakit Coronavirus 2019 (COVID-19) di Malaysia. In: From the Desk of the Director-General of Health Malaysia [Online]. Available at: https://kpkesihatan.com/2020/06/05/kenyataan-akhbar-kpk-5-jun-2020-situasi-semasa-jangkitan-penyakit-coronavirus-2019-covid-19-di-malaysia/. Accessed 5 June, 2020
2. Prime Minister’s Office of Malaysia. The Prime Minister’s Special Message on COVID-19 – 16 March 2020. In: Prime Minister's Office of Malaysia Official Website. [Online]. Available at: https://www.pmo.gov.my/2020/03/perutusan-khas-yab-perdana-menteri-mengenai-covid-19-16-mac-2020/. Accessed 5 June, 2020
3. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MU, Li B, Cai J, Xu B, Yang Q, Wang B. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science. 2020 May 8;368 (6491):638-42.
4. Prime Minister’s Office of Malaysia. Teks Ucapan: Pelan Jana Semula Ekonomi Negara (PENJANA). In: Prime Minister's Office of Malaysia Official Website [Online]. Available at:https://www.pmo.gov.my/category/speeches/. Accessed 5 June, 2020
5. Salim N, Chan WH, Mansor S, Bazin NE, AmaranS, Faudzi AA, Zainal A, Huspi SH, Khoo

EJ, Shithil SM. COVID-19 epidemic in Malaysia: Impact of lock-down on infection dynamics. medRxiv. 2020 Jan 1.
6. Arifin WN, Chan WH, Amaran S, Musa KI. A Susceptible-Infected-Removed (SIR) model of COVID-19 epidemic trend in Malaysia under Movement Control Order (MCO) using a data fitting approach. medRxiv. 2020 Jan 1.
7. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza and other respiratory viruses. 2014 May;8(3):309-16.
8. Raissi M, Ramezani N, Seshaiyer P. On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. Letters in Biomathematics. 2019 Oct 16:1-26.
9. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infectious Disease Modelling. 2020 Mar 25.
10. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief. 2020 Feb 26:105340.
11. Perone G. An ARIMA model to forecast the spread of COVID-2019 epidemic in Italy. arXiv preprint arXiv:2004.00382. 2020 Apr 1.
1. Chintalapudi N, Battineni G, Amenta F. COVID-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. Journal of Microbiology, Immunology and Infection. 2020 Apr 13.
13. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018 May 8.
14. IBM. IBM SPSS Forecasting 22. IBM Corporation; 2013.
15. Erhan A. Covid-19 Malaysia - Dataset by Erhanazrai. In: data.world [online]. Available at: https://data.world/erhanazrai/httpsdocsgooglecomspreadsheetsd15a43eb68lt7ggk9vavy. Accessed May 1, 2020.
16. Hyndman RJ, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E. Package ‘forecast’. In: The Comprehensive R Archive Network [Online]. Available at: https://cran. r-project. org/web/packages/forecast/forecast.pdf. Accessed Mar 31, 2020
17. Khaliq A, Batool SA, Chaudhry MN. Seasonality and trend analysis of tuberculosis in Lahore, Pakistan from 2006 to 2013. Journal of epidemiology and global health. 2015 Dec 1;5 (4):397-403.
18. MOH Malaysia. Malaysia: Responding to COVID-19 Pandemic. In: COVID-19 Informations Sessions [online]. Available at: https://apps.who.int/gb/COVID-19/pdf_files/04_06/Malaysia.pdf. Accessed June 5, 2020.
19. Chintalapudi N, Battineni G, Amenta F. COVID-19 disease outbreak forecasting of registered and recovered cases after sixty-day lockdown in Italy: A data driven model approach. Journal of Microbiology, Immunology and Infection. 2020 Apr 13.
20. Lipsitch M. Seasonality of sars-cov-2: Will covid-19 go away on its own in warmer weather. Center for Communicable Disease Dynamics. 2020.
21. Dehesh T, Mardani-Fard HA, Dehesh P. Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models. medRxiv. 2020 Jan 1.
22. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of The Total Environment. 2020 Apr 22:138817.
23. Salim N, Chan WH, Mansor S, Bazin NE, Amaran S, Faudzi AA, Zainal A, Huspi SH, Khoo EJ, Shithil SM. COVID-19 epidemic in Malaysia: Impact of lock-down on infection dynamics. medRxiv. 2020 Jan 1.
24. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G. COVID-19 infection: the perspectives on immune responses.
25. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liu P, Cao X, Gao Z, Mai Z, Liang J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease. 2020 Mar;12(3):165.
26. Liu Z, Huang S, Lu W, Su Z, Yin X, Liang H, Zhang H. Modeling the trend of coronavirus disease 2019 and restoration of operational capability of metropolitan medical service in China: a machine learning and mathematical model-based analysis. Global Health Research and Policy. 2020 Dec;5:1-1.

27. Lim I. Health D-G: Malaysia using targeted Covid-19 testing approach for 'high impact, reasonable cost'. In: Malay Mail [online]. Available at: https://www.malaymail.com/news/malaysia/2020/05/14/health-d-g-malaysia-using-targeted-covid-19-testing-approach-for-high-impac/1866184. Accessed May 16, 2020