Modeling and forecasting trend of COVID-19 epidemic in Iran until May 13, 2020

Ali Ahmadi*,1, Yasin Fadaei2, Majid Shirani3, Fereydoon Rahmani4

Received: 18 Mar 2020 Published: 31 Mar 2020

Abstract
Background: COVID-19 is a new disease and precise data are not available about this illness in Iran and in the world. Thus, this study aimed to determine the epidemic trend and prediction of COVID-19 in Iran.

Methods: This was a secondary data analysis and modeling study. The daily reports of definitive COVID-19 patients released by Iran Ministry of Health and Medical Education were used in this study. Epidemic projection models of Gompertz, von Bertalanffy, and least squared error (LSE) with percentage error were used to predict the number of hospitalization cases from April 3, 2020 until May 13, 2020.

Results: The prediction of the number of patients on April 3, 2020 by von Bertalanffy, Gompertz, and LSE, with 95% confidence interval (CI), were estimated at 44,200 (39,208-53,809), 47,500 (38,907-52,640), and 48,000 (40,000-57,560), respectively. The number of deceased COVID-19 patients was also estimated to be 3100 (2633-3717) individuals by the von Bertalanffy model, 3700 (2900-4310) by Gompertz’s model, and 3850 (3200-4580) by LSE. Making predictions about the flat epidemic curve and number of patients based on Gompertz model, will project 67,000 (61,500-87,000) cases. Based on Gompertz and von models, 7900 (6200-9300) and 4620 (3930-5550) deaths will occur from May 13 to June 1, 2020, respectively, and then the curve will flatten.

Conclusion: In this study, estimations were made based on severely ill patients who were in need of hospitalization. If enforcement and public behavior interventions continue with current trends, the COVID-19 epidemic will be flat from May 13 until July, 2020 in Iran.

Keywords: COVID-19, Coronavirus, Prediction, Modeling, Epidemiology

Introduction
Coronaviruses are a large family of viruses that have been identified since 1965; to date, 7 species of them have been reported to affect humans. These viruses have 3 genotypes of alpha, beta, and gamma. The natural reservoirs of these diseases are mammals and birds, and thus they are considered as zoonotic diseases (1, 2). Severe acute respiratory syndrome (SARS) is caused by a species of coronavirus that infects humans, bats, and certain other mammals, which has led to epidemics in 2002 and 2003 (2-4); Middle East respiratory syndrome (MERS) caused an epidemic in Saudi Arabia in 2012 (5); and more recently their newest variant Coronavirus Disease-2019 (COVID-19) has led to the recent pandemic in China, Iran, Italy, and across the world (6-8). COVID-19 causes a direct and indirect transmission of respiratory diseases with a wide range of symptoms, including cold symptoms, respiratory/fever symptoms, cough, shortness of breath, kidney failure, and even death (9). The complete clinical manifestation of COVID-
19 is not clear yet and its transmission risk is not completely understood (10). However, the virus is believed to be transmitted mostly via contact, droplets, aspirates, feces. Generally, everyone is prone to this viral disease. The mean incubation period of COVID-19 was 5.2 days (4.1–7 days) and the basic reproductive number \( R_0 \) was reported 2.2 (95% CI: 1.4 to 3.9) (11). In another study, the mean incubation period ranged from 0-24 days, with the mean of 6.4 days. The \( R_0 \) of COVID-19 at the early phase, regardless of different prediction models, was higher than SARS and MERS, and the majority of patients (80.9%) were considered asymptomatic or mild pneumonia (12). The case fatality ratio was 2% (12), 2.3% (8), 3.46% (13), and elderly men with underlying diseases were at a higher risk of death (13). As of March 29, 2020, COVID-19 pandemic was declared by the World Health Organization (WHO) in more than 100 countries (most prevalent in the United States, Italy, China, Spain, Germany, Iran, and France) (14). In Iran, the first case of COVID-19 was February 19, 2020 in Qom, and we used the reported data until March 29, 2020 in Iran. Until March 29, 2020 (The date when this manuscript was being conducted), according to the daily reports, 38 309 cases of Covid-19 and 2640 related deaths were reported in Iran (15, 16). As of February 29, 2020, all schools and universities and as of March 7, 2020 almost all public places and shrines have been closed. On March 2, 2020, a team of WHO experts landed in Tehran, Iran, to support the ongoing outbreak in the country (17). Currently, people are referring to health centers and hospitals, and the public is almost alarmed by the epidemic of panic and inaccurate reporting in cyberspace. Important questions in people’s mind are as follow: How many people have COVID-19 in Iran? What is the status of COVID-19 epidemic curve in Iran? When will the epidemic will go and how it ends? We cannot answer these questions with certainty, but they will be investigated in terms of pathogenic agents (coronavirus), host conditions, behavior (human), and environmental factors of coronavirus transmission, daily reports of definitive COVID-19 patients released by Iran Ministry of Health and Medical Education, and the use of modeling given the assumptions and the percentage of error. Although the models are different, multiple and changeable in nature and do not insist on the correctness of the forecasts, the decision-making conditions for health policymakers and authorities are more transparent and helpful (18). This study aimed to model and determine the epidemic trend and predict the number of patients hospitalized due to COVID-19 in Iran using mathematical and statistical modeling.

Methods

This was a secondary data analysis and mathematical modeling study based on a research proposal approved by Shahrekord University of Medical Sciences (Code of Ethics Committee on Biological Research: IR.SKUMS.REC 1398.254) (19).

For the statistical analysis of definitive COVID-19 patients in Iran, daily reports of the Ministry of Health and Medical Education were used (15). The definitive diagnosis of COVID-19 was made using virus isolates from patients' biological samples in hospitals. When real time polymerase chain reaction (PCR) test turned positive for COVID-19 in patients with respiratory symptoms and confirmed by the reference laboratory in School of Public Health, Tehran University Medical Sciences and Pasteur Institute of Iran (20), they were used for analysis. Patient population growth, epidemic curves, and recovered, and deceased individuals were used to form a conceptual framework of an epidemic and predict the COVID-19 epidemic trend. Quasi classical infectious disease (Susceptible→Exposed→Infectected→Removed: SEIR) model was used (21). Different scenarios were designed and implemented for modeling and forecasting. First, based on a search for reliable sources of disease trends and epidemic curves across the world, the curve of Iran was drawn (10, 18, 22). Focused and scientific group discussion sessions were held with experts on epidemiology, biostatistics, and mathematics, infectious diseases specialists, and health care managers on the topic. Different scenarios were discussed and agreement was reached on the application of final scenarios. To predict the growth of this epidemic, different models were used. In the first scenario, the most optimistic estimation and control of the epidemic was during an incubation period (von model, the most ideal model). In this scenario, traced contacts are isolated immediately on symptom onset (and not before) and isolation prevents disease transmission. In the second scenario, an intermediate and fit-to-data model (Gompertz) was used. In the third scenario, the use of the growth rate is greater than the first and second models, and it is the opposite of the first scenario (LSE). To select the scenarios, fit the data with the models and growth rate of the cases were used. The Gompertz growth, von Bertalanffy growth equation, and curve fitting by LSE method with cubic polynomial for Epidemic forecasts were run in MATLAB software. Models are presented as the following differential equations:

Gompertz Differential Equation:

\[
\frac{dP(t)}{dt} = aP \ln\left(\frac{1}{bP}\right)
\]

Von Bertalanffy’s differential growth equation:

\[
\frac{dP(t)}{dt} = aP \left(\frac{bP}{c} - 1\right)
\]

Cubic polynomial polynomials:

\[
P(t) = ax^3 + bx^2 + cx + e
\]

Where \( p \) represents the number of individuals in each population, \( a, b, c, \) and \( e \) represent unknown parameters and \( t \) time. \( d/dt \) is a derivative of time (23, 24).

The assumptions of the model are as follow:
- Models are based on official reported data and recruitment testing from hospitalized cases.
- Any manipulation and misinformation will affect the model.
- The method for finding patients is fixed.
- The unknown parameters were estimated by running the fminsearch, a MATLAB function, which is a least squares algorithm. The parameters were estimated based on the official reported data of infected, cured, and dead cases. The estimated values of the parameters for different scenarios are reported in Table 1. Also, MATLAB software was used to fit data and solve the equations. MATLAB codes are presented in the supplementary file. Moreover, the percentage
of the root mean square error (RMSE) was used to validate the models and 95% confidence interval (CI) was utilized to calculate the coefficient MATLAB function.

The basic reproduction number \( R_0 \) was calculated using the following formula (25):

\[
R_0 = 1 + r * T_c
\]

where, \( T_c \) and \( r \) are the mean generation interval of the infected and the growth rate, respectively, \( T_c = 7.5 \), and \( r = 0.1 \). The growth rate of Gompertz model is \( r = 0.1 \); thus, the number of \( R_0 \) is 1.75. All estimations and detections of COVID-19 were made based on the current conditions of laboratory sampling from critically ill and hospitalized patients (tip of the iceberg spread of the disease). In this modeling, asymptomatic patients and those with moderate symptoms from whom no samples were taken were excluded. Also, the data of patients diagnosed based on CT scans were not included in this study. The forecast dates were selected based on the end of New Year (April 3, 2020) holidays in Iran and the onset of epidemic curve flattening (May 13, 2020).

**Results**

Frequency of daily statistics of COVID-19, including definite new cases, number of deaths, and recovered cases in Iran are shown in Table 2. The trend of this epidemic spread in Iran (daily linear and cumulative trend) is illustrated in Figure 1. According to data released on COVID-19 in Iran as of March 29, 2020, the following forecasts for April 3 and May 13, 2020 were reported (Figs. 2, 3, and 4).

According to the Gompertz model, in the most optimistic perspective, the maximum number of infected people until April 3, 2020 is 47,500, with 95% confidence interval (CI: 38,907-52,640) (Fig. 2 A1). The percentage of the root mean square error (RMSE) for Gompertz model is 12%. Based on von Bertalanffy’s growth model (the most ideal model with high isolation of patients and others intervention such...
Forecasting trend of COVID-19 epidemic in Iran

as China experience), the maximum number of COVID-19 patients is 44,200, with 95% confidence interval (CI:
39208-53809), (Fig. 2 A2) and 17% RMSE. According to the method of the least squared error, this value was estimated to be 48 000 (CI: 40000 - 57560), with 19% RMSE (Fig. 4 H).

According to Gompertz and von models on prediction of recovered patients, and the models estimated the number of recovered individuals to be 15 300, with 95% confidence interval (CI: 8600-23000) and 3% RMSE, and 15 950, with 95% confidence interval (CI: 13000-24000), and 3.7% RMSE until April 3, 2020, respectively (Fig. 2 B1, B2).

Moreover, according to Figure 4 G, the maximum population of recovered individuals was estimated to be 15 900, with 95% confidence interval (CI: 13500–19000), according to the method of least squared error.

To predict and estimate the number of deaths due to COVID-19, 2 growth models (Fig. 2 C1, C2) and the least squared error (Fig. 4 I) were used to maximize von’s growth in the most optimistic manner. According to von, Gompertz, and least squared error methods, the number of deaths, with 95% confidence interval, will be 3100, (CI: 2633-3717) with 17% RMSE, 3700 (CI: 2900-4310) with 12% RMSE, and 3850 (CI: 3200-4580) on April 3, 2020, respectively. To make predictions for May 13, 2020, when the epidemic curve will be flat, 3 models of growth with 95% confidence interval were used. Based on Gompertz model, the epidemic curve will be flattened on May 13, 2020 with about 67 000 patients (CI: 61500- 87000), with 10% RMSE (Fig. 3 D). Based on von and Gompertz models, the number of cured cases will be 22 000 (CI: (12500 -32000), with 3.1% MSRE, and 26 000 (CI: 21700-39500), with 3.7% MSRE, respectively (Fig. 3 E2, E1). To predict the number of deaths due to COVID-19 when the epidemic

http://mjiri.iums.ac.ir
Med J Islam Repub Iran. 2020 (31 Mar); 34.27.
curve is flat based on Gompertz and von models, there will be 7900 deaths (CI: 6200-9300), with 10% MSRE, and 4620 deaths (CI: 3930-5550), with 9.1% MSRE, until May 13 and June 1, 2020, respectively. (Fig. 3 F1, F2).

Discussion
In this study, the trend of COVID-19 epidemic prediction and estimation of the number of patients, R₀, deaths, and recovered individuals were performed and reported based on mathematical and statistical models. Although this prediction may be associated with random errors, it was made with assumptions about the past trends of the COVID-19 epidemic in Iran as well as the behavior of the people and government interventions (sampling of severe cases and hospitalization). Implementation of government interventions such as social distancing, isolation of patients, and follow-up of those around them based on the epidemic management protocol in Iran is of high importance. Thus, observance of these interventions by the public and the government has an impact on modeling predictions. Moreover, according to a valid scientific report, delay in the onset of symptoms until the isolation of patients plays an important role in controlling the epidemic (18). To control the majority of outbreaks, for R₀ of 2.5, more than 70% of contacts had to be traced, and for R₀ of 3.5, more than 90% of contacts had to be traced. The delay between symptom onset and isolation had the largest role in determining whether an outbreak was controllable when R₀ was 1.5. For R₀ values of 2.5 or 3.5, if there were 40 initial cases, contact tracing and isolation were only potentially feasible when less than 1% of transmission occurred before symptom onset (18). Therefore, efforts should be made to control this epidemic with greater vigor and urgency and to conduct a daily risk assessment. In the current epidemiological situation (26) in the world and Iran, fear control and avoidance of rumors are very important for COVID-19 prevention and control.

There are 3 important and debatable points of view about this epidemic in Iran: First, to avoid tension in the society; second, to properly interpret the COVID-19 case fatality ratio (CFR) in Iran and calculate CFR tactfully; and third, to recommend personal hygiene, including hand washing and avoiding contact with suspected patients, social distancing, discovering unknown cases of infection and early detection, tracing direct contact and isolation of patients, all of which have been emphasized by the health care officials to overcome this disease. Moreover, in interpreting this index, since the denominator of the fraction is only positive cases in hospital beds and the numerator is the number of patients who died of COVID-19. This index should also be calculated until the end of the epidemic period, and if it is until the end of the epidemic and their outcome (death/recovery) is determined, this indicator will approach the real number. The estimated case fatality ratio among medically attended patients was reported to be approximately 2% (12) and the true ratio may not be known for some time (27). The underreporting estimation is very sensitive to the baseline CFR, meaning that small errors lead to large errors in the estimate for underreporting (28).

Underreporting of COVID-19 patients (1.38% / cCFR) and modification of disease mortality had previously been
estimated by the Center for Mathematical Modeling of Infectious Diseases, London School of Health and Tropical (28). If a country has a higher adjusted CFR (eg, 20.02%), it means that only a fraction of cases have been reported (in this case, 1.38 /20.02 = 6.89% cases reported approximately). This formula can accurately estimate the statistics of all patients with COVID-19 (from asymptomatic and mild to severe cases and death).

One article reported that up to 70% of the supply chain could be cut off and the epidemic could be controlled if contact and isolation, quarantine and isolation were appropriately accomplished (18). We think the top priorities in Iran are now circular and comprehensive efforts to conduct epidemiological studies and identify all aspects of the disease (source of disease, reservoir, pathways, infectivity, incubation period, incidence and prevalence, pathogenicity, immunogenicity, herd immunity, causes, epidemic and pandemic pattern, primary and secondary attack rates, response time, time needed for isolation and quarantine, treatment regimens, vaccines and other prevention methods, disease surveillance, and statistical reporting) and evidence-based interventions and epidemic control. The experience of China and South Korea should be used to control this epidemic disease in Southeast Asian countries and in Iran, as South Korea and China were successful in controlling the disease. Cultural conditions are also effective. Given the prediction and modeling of the number of Coronavirus cases in Iran and because the virus is going to circulate in the country for at least a few weeks, we will have an ascending trend in the coming weeks.

We recommend using the WHO guideline to properly manage patients (29). Considering China’s experience and the fact that it took about 70 days for the epidemic curve to flatten in China, and based on a search of scientific texts, this study provides the following recommendations:

1. Up-to-date and accurate data on definitions related to suspected, probable, and definitive people with Coronavirus should be collected at all provincial levels in the health care system.

2. Percentage of completion and accuracy of assessments and data should be monitored precisely and the epidemic curve should be drawn based on district, province, rural, and urban divisions and be provided to provincial and academic headquarters in an updated dashboard format.

3. Data should be carefully recorded and analyzed regarding the pathology, time of onset of symptoms, natural course of the disease, and the outcome of the disease to determine effective strategies to prevent and determine the necessity of intervention to control the spread of the disease at different levels. At all levels of the health system (governmental and nongovernmental), medical and diagnostic interventions and their outcomes should be recorded for all patients. These records should be based on the date of onset of symptoms, the date of referral, the date of diagnosis, method of diagnosis, the date of intervention and the outcome (eg, death, recovery, discharge). Such records can widely be used to compare and evaluate the cost-effectiveness of various diagnostic and therapeutic methods. Health system staff should be trained to appropriately and accurately record data, especially by means of web-based networks. This will certainly improve the quality of data recording. All epidemiological indicators that determine the epidemic pattern, including baseline R zero, attack rate, incubation period, index case, primary cases, secondary cases, and GIS mapping should be determined in provinces, cities, and nationwide, and epidemic trends should be monitored.

Access to the results of the analysis and data should be provided for researchers and experts on the basis of specific protocols available for this purpose in the world and Iran, and a thorough critique and creative theories and ideas should be elicited from all university training and research groups. The models used to predict the end of the epidemic and control it should be evaluated as well. The results of our study are inconsistent with Zhuang brief report (30), in which the data were collected from the World Health Organization (WHO). However, this report may not be accurate, as the WHO has not reported it.

Limitations

Given the urgency of the need for valid and transparent models to informed interventions and policies, some further considerations like the no consideration or account of systematic cases, testing coverage and time delay to the test results availability, seasonality, and comorbidities have not been included in this study. However, it may be feasible to consider them in revisions of the models or future studies. Moreover, the progression of disease epidemic across space-time has not been seen in Iran, as we used the parameters of Chinese models in modeling estimates in Iran to calculate $R_0$. In this study, no attempt was made to detect and report cases and deaths in Iran. Although this could have been performed, given that the study has already been designed, conducted, and being reported, and given the exceptional mortality and morbidity situation, it was not feasible to go back and report the cases and deaths due to COVID-19. However, it can be done for the future similar studies. There is no fixed page in Ministry of Health and Medical Education web site for reference of daily reports and they are scattered on various pages of this site. If screening is done in the community along with biological sampling to diagnose COVID-19, the number of cases will certainly be higher and the results of modeling will change.

Conclusion

The actual trend of detecting COVID-19 cases in Iran, which has been based on people’s health behaviors and government interventions, has been increasing. In this study, estimates were based on current trends, social distancing, sampling of severe cases, hospitalization, and tip of iceberg spread disease, and thus asymptomatic, mild, and moderate cases could not be calculated. We used the reports of positive COVID-19 cases in hospitals; thus, the prediction model in this study can be used for patients hospitalized due to COVID-19. Complete reliance on any type of model will lead to systematic and random error, unless modeling provides a prediction with precise and clear assumptions and inputs and outputs. To predict the flattening of the epidemic
curve, 3 models of growth with percentage of the root mean square error (RMSE) were used. Based on RMSE, Gompertz growth model was valid and predicted that the epidemic curve will be around May 13, 2020 with about 67,000 hospitalized patients and 7900 deaths (RMSE= 10%), respectively. This study suggests that government interventions and people’s behaviors determine the persistence of the epidemic, and thus they should be addressed with greater responsibility, accountability, rigor, and quality.

Acknowledgments
We thank the Modeling in Health Research Center, School of Health, Deputy of Research and Technology, Shahrekord University of Medical Sciences. Furthermore, we truly appreciate the Ministry of Health and Medical Education, especially the team of daily information reports of COVID-19 in Iran and reviewers of Medical Journal of the Islamic Republic of Iran.

Also, the authors would like to extend their thanks to Dr. Hadi Raeisi and Dr. Victoria Atabaki Kianpour.

Funding
This study was financially supported by the Deputy of Research and Technology of Shahrekord University of Medical Sciences, Shahrekord, Iran (Grant no. 254).

Conflict of Interests
The authors declare that they have no competing interests.

References
20. the National Committee on COVID-19 in Iran Epidemiology in Ministry of Health and Medical Education. [cited: March 29, 2020]. Available from: http://corona.behdasht.gov.ir