Health Science Monitor
2023; 2(1): 13-20
Published online (http://hsm.umsu.ac.ir)

Original Article

Estimation and prediction of the prevalence rate of COVID-19 disease based on multilayer perceptron artificial neural networks model

Mehdi Kazempour Dizaji[1,2*], Arda Kiani[3], Mohammad Varahram[1], Atefe Abedini[4], Ali Zare[2], Rahim Roozbahani[5], Niloufar Alizedeh Kolahdozi[2], Syeyd Alireza Nadji[6], Mohammad Ali Emamhadi[7], Majid Marjani[5]

[1] *Mycobacteriology Research Center (MRC), National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[2] *Department of Biostatistics, National Research Institute of Tuberculosis and Lung Disease (NRITLD) , Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[3] *Tracheal Diseases Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[4] *Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[5] *Clinical Tuberculosis and Epidemiology Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran.*
[6] *Virology Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[7] *Department Forensic Medicine, School of Medicine Shahid Beheshti University of Medical Sciences, Tehran, Iran*

***Corresponding author:** Mehdi Kazempour Dizaji, **Address**: National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran, **Email**: mkazempour@sbmu.ac.ir, **Tel**: +982127122080

Abstract

***Background & Aims*:** Nowadays, with the coronavirus disease-2019 (COVID-19) pandemic, millions of people have been infected with the coronavirus, and most countries in the world have been unable to treat and control this condition. The aim of this study was to estimate and predict the COVID-19 prevalence rate based on multilayer perceptron artificial neural network (MLP-ANN) model.

***Materials & Methods*:** In this cross-sectional study, based on the information of 4,372 patients with COVID-19 referred to Dr. Masih Daneshvari Hospital in Tehran, the prevalence rate of this disease was estimated. In addition, considering the role of the health measures and social restrictions, the trend of this index based on the MLP-ANN model was predicted.

***Results*:** According to the results of this study, the prevalence of COVID-19 increased by an average of 7.05 per thousand people daily during the 48 days from the onset of the epidemic, and it reached about 341.96 per thousand people. Based on the MLP-ANN model with a lack of attention to the health measures by individuals in the community and failure to reduce social restrictions by the government, the COVID-19 prevalence increased by an average of 1.03 per thousand people per day. While in the case of attention to the health measures by the people and continued social restrictions by the state, the prevalence of this disease decreased by an average of 2.13 per thousand people, daily.

***Conclusion*:** The study on the prevalence of COVID-19 disease and prediction of the trend of this index provides researchers with useful information about the role of the health measures and social restrictions in controlling this disease.

***Keywords*:** COVID-19, Prevalence index, Perceptron artificial neural network, Prediction

Received 15 December 2022; accepted for publication 02 January 2023

Introduction

Coronavirus is a member of a large family of viruses, ranging from the common cold virus to the causative agent of severe acute respiratory syndrome (SARS) (1). Symptoms of the coronavirus disease 2019 (COVID-19) in humans include fever, cough, and shortness of breath (dyspnea), as well as the loss of sense of smell and taste (2, 3). While COVID-19 disease causes mild symptoms in most cases but in some cases, it progresses to pneumonia and multi-organ failure (4). To date, more than five million people worldwide have been diagnosed with COVID-19, and this trend is increasing (5). Globally, the United States has suffered the most in the pandemic, with a quarter of confirmed COVID-19 cases and a third of mortality rates (6). Russia, Brazil, and the United Kingdom are also ranked second to fourth among countries affected by the Coronavirus (5, 7). Spain, Italy, and Germany are also ranked fifth to eighth, respectively (5, 6). Although the virus is less deadly than other emerging viruses of the Corona family, such as SARS and Middle East respiratory syndrome coronavirus (MERS), it has spread rapidly and exhibited special pathogenic behaviors that make it very difficult to control (8-10).

The World Health Organization (WHO) has announced that the new disease is a pandemic and also named it COVID-19 (11). The most common and severe manifestation of this infection is pneumonia. This group of patients usually develops dyspnea on average five days after the onset of their disease and acute respiratory syndrome occurs in 4% of these patients (3, 12). About 85% of confirmed patients with COVID-19 experience mild to moderate disease and approximately 13% develop severe symptoms (12, 13). The disease also becomes critical and dangerous in less than 6% of patients (12).

Notwithstanding multiple proposed medicines and treatment methods for COVID-19, they are still in the experimental phase, and their effectiveness as an effective medication for the treatment of COVID-19 patients is obscure (14-17). There are a lot of unknowns about this virus, and many aspects of this disease are still unidentified, the reasons why many countries have failed to control and treat this condition (10). In response to the current situation of COVID-19 and with consideration of its importance from various aspects of individual and social health, WHO has introduced strategies that include paying attention to the health measures by the people and imposing social restrictions by governments (2, 12). The main goal of these strategies is to reduce the morbidity and mortality rates of COVID-19. Estimation of the disease prevalence index helps to achieve these aims. Such index represents the speed of disease transmission and also indicates the role of health care measures and social restrictions in order to prevent the spread of this disease. The prevalence of the disease is calculated as follows:

$$\frac{\text{The number of definite cases at a given time}}{\text{Average number of diagnostic tests performed during the study}} \times 1000$$

As the definitive cases include the new and old COVID-19 patients, the number of definite cases has been determined by the daily number of new cases, deaths, and number of patients who improved. Predicting the trend of the COVID-19 prevalence will provide the possibility of health planning to reduce the incidence of this disease. Since the initiation of COVID-19 pandemic, many studies have been conducted on health indicators, such as the incidence, prevalence, and mortality rates of COVID-19 (8, 10, 18, 19). However, it is more important to predict their trend than to study health indicators (20).

There are several methods to predict the trend of morbidity and mortality of diseases including COVID-19. One of these methods is the use of statistical models and artificial neural networks (ANNs). With the advances in science and technology, the ANNs models have received great attention to make predictions (21). ANNs have been receiving much more attention from researchers not only in medicine and health studies but also in other sciences and have quickly replaced the classical statistical methods (21, 22). The ANNs with the method that is inspired by the process of learning and information processing in the human brain have the capability of very complex analysis (21).

In the ANNs, there is an input layer that receives information, a number of hidden layers that get information from previous layers, and finally, an output layer that is the result of calculations and problem-solving. The smallest processing unit of ANNs is a neuron; these neurons are present in different layers of the ANNs and are interconnected (21, 22). One of the most popular ANNs widely used is the multilayer perceptron artificial neural network (MLP-ANN) that is based on a computational unit called perceptron (23). The most common type of MLP-ANN consists of a three-layer perceptron, which is shown in Figure 1.

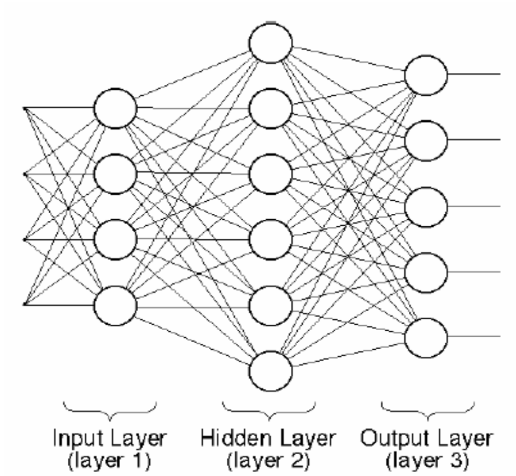

Fig. 1. A typical three-layer perceptron ANN with a hidden layer.

In the ANNs, there is no limit to the type of data studies; thus, these models are capable of reasonable predictions with missing data and data with high levels of censoring, as well as data bias (24-26). ANNs can be one of the most suitable ways to health data analytics without the need for hypotheses, and the structure of data can easily provide an accurate prediction of health indicators (24, 25, 27, 28). Given the increasing application of ANNs for prediction, this study was undertaken to estimate and predict the COVID-19 prevalence rate based on the MLP-ANN model.

Materials & Methods

In this cross-sectional study, in order to estimate and predict the prevalence of COVID-19, the data were collected from 4,372 patients with this disease who referred to Dr. Masih Daneshvari Hospital, a university-affiliated and selected referral center for COVID-19 patients in Tehran, Iran. Data collection was conducted from early March 2020 to late April 2020. Patients' information was collected through their clinical conditions and medical reports, confirmed diagnostic testing for COVID-19, and chest CT findings. Based on this information, the prevalence of COVID-19 was calculated at the above-mentioned time period, and future trends in the prevalence of this disease were predicted.

For the present survey, two different scenarios were considered to predict the prevalence of COVID-19. In the first scenario, it assumed that people in the community pay attention to health measures (e.g. wearing masks and gloves and washing hands regularly) and the government continues to impose social restrictions (e.g. social distancing plans, closure of educational institutions, closing cultural and religious centers, banned of sports gatherings, and reduction of urban traffic and intercity transportation). Afterward, prevalence prediction was made. In the second scenario, it was assumed that the people do not pay attention to health measures, and social restrictions were reduced by the government, then the prevalence of COVID-19 was again predicted.

In this study, a set of ANNs with 1 to 20 neurons for the hidden layer was examined in order to determine the appropriate number of neurons suitable for the hidden layer and also to specify the best structure of the MLP-ANN. These models were compared using indices such as the sum of square error and relative error. Among these indices, two three-layer perceptron's ANNs with two neurons for the input layer (one for the study time variable and one for the COVID-19 transmission rate variable over time), 6 and 9 neurons for the hidden layer, and one neuron for the output layer (predicting the prevalence of COVID-19 over time) with the lowest error rate among other networks were selected as the best ANN structure to predict the prevalence of COVID-19. For this study, data analysis was performed using SPSS software version 19 and also STATA software version 14.

Results

In the current study, the COVID-19 prevalence was estimated based on the information of 4,372 patients with COVID-19 referred to Dr. Masih Daneshvari Hospital in Tehran from early March 2020 to late April 2020 (Table 1). According to the study results, although in the early days of the COVID-19 epidemic, the prevalence of the disease was estimated to be less than 100 people per thousand, an upward trend was seen in estimating the disease prevalence. Moreover, the COVID-19 prevalence increased to about 200 per thousand people within 26 days of the onset of the epidemic and about 300 per thousand people within 40 days of the onset of the COVID-19 epidemic. In total, during the 48 days in the initiation of the COVID-19 epidemic, the estimates indicated an upward trend in this index. The daily prevalence of COVID-19 increased an average of about 7.05 per thousand people. During this period, the highest and lowest values of this index were 431.96 and ~3.68 per thousand people, respectively (Table 1).

Table 1. Estimation of the COVID-19 prevalence (per 1000 patients) based on information from patients' clinical records

Time by days from the beginning of the epidemic	Estimation of the COVID-19 prevalence based on information from patients' clinical records	Time by days from the beginning of the epidemic	Estimation of the COVID-19 prevalence based on information from patients' clinical records	Time by days from the beginning of the epidemic	Estimation of the COVID-19 prevalence based on information from patients' clinical records
1	68/3	21	41/164	41	54/302
2	63/5	22	02/173	42	78/307
3	43/7	23	82/184	43	92/315
4	62/9	24	16/191	44	39/321
5	34/11	25	09/196	45	87/326
6	58/16	26	21/203	46	93/330
7	41/20	27	18/211	47	01/335
8	63/28	28	49/220	48	96/341
9	62/32	29	90/230		
10	48/38	30	31/242		
11	45/50	31	56/252		
12	23/60	32	72/257		
13	69/69	33	01/266		
14	18/85	34	82/272		
15	54/102	35	47/279		
16	35/114	36	86/284		
17	12/133	37	04/285		
18	13/138	38	78/287		
19	33/150	39	73/290		
20	03/160	40	54/297		

Estimation of the COVID-19 prevalence based on the MLP-ANNs model during 48 days of the epidemic of this disease showed the high accuracy of MLP-ANN in estimating the COVID-19 prevalence. Comparison of these results in Figure 2 shows the high accuracy of the MLP-ANNs model for the estimation of this index and prediction of its values for the coming days (Figure 2).

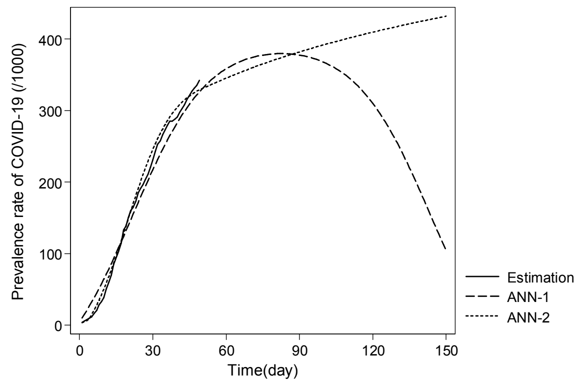

Fig. 2. Estimation of the trend of COVID-19 prevalence rate based on the estimated values of this index from the patients' information and using the MLP-ANNs model, assuming that health measures are taken by people in the community and social constraints continue by government (ANN-1) and also assuming that the community does not pay attention to health measures and the government reduces social restrictions (ANN-2).

As mentioned earlier, in this study, two different scenarios were considered to predict the prevalence of COVID-19 disease. Based on the first scenario, assuming the health measures taken by the community and the continuation of social restrictions by the government, the MLP-ANNs model showed that the COVID-19 prevalence increased until about 80 days of the epidemic initiation. While after that, this index was shown a downward trend, and the prevalence reached 104.25 per thousand people within 150 days from the beginning of the epidemic.

According to the second scenario, assuming that the public did not pay attention to the health measures and the government reduced the social restrictions, the MLP-ANNs model prediction showed the COVID-19 prevalence had a completely increasing trend. Within 90, 120, and 150 days from the onset of the epidemic, the COVID-19 prevalence was predicted to be 381.79, 409.46, and 431.69, respectively.

Based on the MLP-ANNs model, with no attention to the health measures by the community and reduces social restrictions by the government, the COVID-19 prevalence will increase by an average of 1.03 per thousand per day. While this model predicts a reduction in the prevalence of COVID-19 prevalence by an average of 2.13 per thousand people per day, following compliance with health measures by the public and continued government restrictions.

Discussion

With the onset of the COVID-19 epidemic in China and the rapid spread of this disease in other parts of the world, millions of people have been infected with the disease and thousands have died (9, 0, 19, 29). Many countries are trying to control the outbreak of COVID-19 by taking health measures and imposing widespread social restrictions and also to make effective vaccines and medicines for the prevention and treatment of the COVID-19 (3, 12, 14, 30). Since many aspects of COVID-19 are still unknown, and no effective treatment has yet been introduced, the best prevention method for the disease is to pay attention to the health measures and implemented social restrictions (11, 14). Study of health indicators such as prevalence and mortality rates can determine the role of health measures and social restrictions in controlling COVID-19 and also show how much these measures can be effective in reducing the prevalence, morbidity, and mortality of the disease (8, 20, 31). For this reason, since the initiation of the COVID-19 epidemic, many studies have been conducted on health indicators, including the prevalence and mortality rates of COVID-19 (8, 10, 18, 19).

In the present study, the COVID-19 prevalence was estimated based on the information of patients, and also the trend of this index was predicted in two different scenarios based on the MLP-ANNs model. According to the results of this study, the prevalence of COVID-19 showed a completely increasing trend during the 48 days from the onset of the epidemic. In this period, the average prevalence of the disease has increased by 7.05 per thousand, daily. Also, the MLP-ANNs model accurately predicted increasing trend of COVID-19 prevalence. According to this model, the incidence of COVID-19 increased by an average of 6.58 and 6.78 per thousand per day.

In the first scenario, assuming the continuation of social restrictions by the government and attention to the health measures by the community, the MLP-ANNs model predicted that despite increasing COVID-19 prevalence trend, this rate decreased after about 80 days from the initiation of the COVID-19 epidemic. This model in the first scenario predicted that the disease prevalence would decrease by an average of 2.13 per thousand people per day. Hence, for reducing the disease prevalence to less than 5 per thousand people, it would take about 165 days from the onset of the epidemic.

The prediction based on the MLP-ANNs model in the second scenario also showed that if the government does not pay attention to the health measures and reduces social restrictions by the government, the prevalence of COVID-19 will increase. According to the prediction in this scenario, 40 days after the start of the epidemic, although the prevalence of COVID-19 would increase, the acceleration of this process would be somewhat reduced. In fact, increasing or decreasing the prevalence of COVID-19 due to health measures and social restrictions is more because health-care measures and social restrictions affect the reproduction rate of COVID-19 (9, 19). These measures can reduce or increase the disease transmission speed in the community by reducing or increasing the number of cases with this disease, thus affecting the prevalence of COVID-19 (19). Therefore, attention to the health measures (e.g. the use of mask and glove, and regular handwashing) by the people, and also the implementation of extensive social restrictions (e.g. social distancing plan, closure of educational, cultural, and religious centers, and reducing urban and interurban traffic) by the government could cause to reduce the COVID-19 prevalence.

According to the predictions of this study, it is impossible to reduce the prevalence of this disease to zero, at least in a short term, and it takes at least 170 days from the start of the epidemic. Nontheless, if the community does not pay attention to health measures, and the government reduces social restrictions, soon, the prevalence of COVID-19 disease will increase and can prolong the duration of the COVID-19 pandemic.

Conclusion

Taken together, estimating the prevalence of COVID-19 can provide useful information on the reproduction rate of the disease, the role of health measures, and the importance of restrictions in society. However, predicting the trend of this index during the epidemic time can play a key role in health planning to control COVID-19.

Acknowledgments

None declared.

Conflict of interest

The authors have no conflict of interest in this study.

Funding/support

The Shahid Beheshti University of Medical Sciences financially supported this study.

Data availability

The raw data supporting the conclusions of this article are available from the authors upon reasonable request.

References

1. Roush S, Fast H, Miner CE, Vins H, Baldy L, McNall R, Kang S, Vundi V. National Center for Immunization and Respiratory Diseases (NCIRD) Support for Modernization of the Nationally Notifiable Diseases Surveillance System (NNDSS) to Strengthen Public Health Surveillance Infrastructure in the US. In2019 CSTE Annual Conference 2019 Jun 3. CSTE.
2. World Health Organization. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 [Internet]. Geneva: World Health Organization; 2020 [cited 2020 Mar 12].
3. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DS, Du B. Clinical characteristics of coronavirus disease 2019 in China. New England journal of medicine. 2020 Apr 30;382(18):1708-20.
4. Hui DS, Azhar EI, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. International journal of infectious diseases. 2020 Feb 1;91:264-6.
5. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)". ArcGIS. Johns Hopkins University. Retrieved 26 May 2020. 26 May 2020.
6. Csse J. Coronavirus COVID-19 global cases by the center for systems science and engineering (CSSE) at Johns Hopkins University (JHU). CSSE, JHU: Coronavirus Resources Center. Diakses dari https://Covid-19virus. jhu. edu/map. html. Diakses pada. 2020 Jul;31.
7. CSSE J. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). 2020.
8. Maitra S, Biswas M, Bhattacharjee S. Case-fatality rate in COVID-19 patients: a meta-analysis of publicly accessible database. medRxiv. 2020 Apr 14:2020-04.
9. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. jama. 2020 Apr 7;323(13):1239-42.
10. Mahase E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. 2020.
11. World Health Organization 2. WHO Director-General’s remarks at the media briefing on 2019-nCoV on 11 February 2020.
12. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England journal of medicine. 2020 Jan 29.
13. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The lancet. 2020 Feb 15;395(10223):514-23.
14. Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nature reviews Drug discovery. 2020 Mar;19(3):149-50.
15. Ahn JY, Sohn Y, Lee SH, Cho Y, Hyun JH, Baek YJ, et al. Use of convalescent plasma therapy in two COVID-19 patients with acute respiratory distress syndrome in Korea. Journal of Korean medical science. 2020 Apr 13;35(14).
16. Lu H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). Bioscience trends. 2020 Feb 29;14(1):69-71.
17. Tanne JH. Covid-19: FDA approves use of convalescent plasma to treat critically ill patients. Bmj. 2020 Mar 26;368(m1256):m1256.
18. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. Jama. 2020 May 12;323(18):1775-6.
19. Team E. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. China CDC weekly. 2020 Feb 2;2(8):113.
20. Aslan IH, Demir M, Wise MM, Lenhart S. Modeling COVID-19: Forecasting and analyzing the dynamics of

the outbreaks in Hubei and Turkey. Mathematical Methods in the Applied Sciences. 2022 Jul 15;45(10):6481-94.

21. Kay JW, Titterington DM, Titterington SD, editors. Statistics and neural networks: advances at the interface. Oxford University Press on Demand; 1999.
22. Warner B, Misra M. Understanding neural networks as statistical tools. The american statistician. 1996 Nov 1;50(4):284-93.
23. Yegnanarayana B. Artificial neural networks. PHI Learning Pvt. Ltd.; 2009 Jan 14.
24. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2001 Apr 15;91(S8):1636-42.
25. Chi CL, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. InAMIA annual symposium proceedings 2007 (Vol. 2007, p. 130). American Medical Informatics Association.
26. Kwon YS, Kim YH, Song JU, Jeon K, Song J, Ryu YJ, et al. Risk factors for death during pulmonary tuberculosis treatment in Korea: a multicenter retrospective cohort study. Journal of Korean medical science. 2014 Sep 1;29(9):1226-31.
27. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. Molecular cancer. 2005 Dec;4(1):1-2.
28. Akl A, Ismail AM, Ghoneim M. Prediction of graft survival of living-donor kidney transplantation: nomograms or artificial neural networks?. Transplantation. 2008 Nov 27;86(10):1401-6.
29. Battegay M, Kuehl R, Tschudin-Sutter S, Hirsch HH, Widmer AF, Neher RA. 2019-novel Coronavirus (2019-nCoV): estimating the case fatality rate-a word of caution. Swiss medical weekly. 2020;150:w20203.
30. Singh R, Adhikari R. Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv preprint arXiv:2003.12055. 2020 Mar 26.
31. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv preprint arXiv:2002.06563. 2020 Feb 16.