Statistical distribution of novel coronavirus in Iran

Elham Gholami[1] , Kamyar Mansori[2] and Mojtaba Soltani-Kermanshahi[3]

1. Treatment Deputy, Tehran University of Medical Sciences, Tehran, Iran; 2. Department of Biostatistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran; 3. Social Determinants of Health Research Center, School of Medicine, Semnan University of Medical Sciences, Semnan, Iran.
Corresponding author: Mojtaba Soltani-Kermanshahi, e-mail: msoltani@farabi.tums.ac.ir
Co-authors: EG: manager.elhamm@gmail.com, KM: kamyarmansori@yahoo.com
Received: 11-06-2020, **Accepted:** 07-08-2020, **Published online:** 17-09-2020

Abstract

Background and Aim: The coronavirus disease-2019 (COVID-19) pandemic – novel coronavirus (nCoV) spread worldwide in 2019, and by March 27, 2020, 199 countries, including Iran, were affected. Prevention and control of the infection is the most important public health priority today. The behavior prediction of COVID-19 is a significant problem. Therefore, in the present research, we compared the different distribution of COVID-19 cases based on the daily reported data in Iran.

Materials and Methods: In this research, we compared the different distribution of COVID-19 cases based on the daily reported data in Iran. We focused on 36 initial data on deaths and new cases with confirmed 2019-nCoV infection in Iran based on official reports from governmental institutes. We used the three types of continuous distribution known as Normal, Lognormal, and Weibull.

Results: Our study showed that the Weibull distribution was the best fit to the data. However, the parameters of distribution were different between data on new cases and daily deaths.

Conclusion: According to the mean and median of the best-fitted distribution, we can expect to pass the peak of the disease. In other words, the death rate is decreasing. Similar behaviors of COVID-19 in both Iran and China, in the long run, can be seen.

Keywords: coronavirus disease-2019, distribution, infection, Weibull.

Introduction

Emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Zika virus disease, give their own significant warnings to the public health [1-4]. Despite significant medical researches, however, how, when, and where new diseases will appear is still subject to considerable uncertainty [5]. On December 31, 2019, the Wuhan Municipal Health Commission in Wuhan City, Hubei Province, China, reported a cluster of pneumonia cases with unspecified etiology that had a history of exposure to Wuhan's Huanan Seafood Wholesale Market (a wholesale fish and live animal market selling different animal species). On January 9, 2020, China, CDC reported that a novel coronavirus (2019-nCoV) had been detected as the causative agent and the genome sequence was made publicly available. Sequence analysis displayed that the newly identified virus belonged to the SARS-CoV clade [6]. In an effort to prevent the prevalence, travel limitations were imposed on Wuhan from January 23, and this has since expanded to 12 other cities, and large social gatherings were prohibited [7,8]. The coronavirus

disease-2019 (COVID-19) pandemic –nCoV spread worldwide in 2019, and by March 27, 2020, 199 countries, including Iran, were affected. According to worldwide statistics, the mortality rate is 3.4%. Early symptoms of COVID-19 involve pneumonia, fever, myalgia, and fatigue. To date, no successful vaccine or antiviral agents have been clinically approved for COVID-19. Therefore, prevention and control of the infection is the most important public health priority [9,10].

During the 2019-2020 coronavirus pandemic, Iran reported the first authenticated cases of SARS-CoV-2 infection on February 19, 2020 in Qom [11]. As of March 27, 2020, pursuant to Iranian health authorities, there had been 2378 COVID-19 deaths in Iran with more than 32,000 confirmed infections. Furthermore, on the same date, Iran had the fourth highest number of COVID-19 deaths after Mainland China, Italy, and Spain but ranked first in Western Asia. Accordingly, the mean age and sex ratio (male/female) of patients in Iran was 59 years and 1.4, respectively. Among deaths related to SARS-CoV-2, 59% were male and 41% were female. In early March 2020, non-Iranian government sources stated their assessment of the number of SARS-CoV-2 infections which was much higher than the official numbers [12-16].

There are considerable uncertainties in assessing the risk of this disease, due to a lack of detailed epidemiological analyses. Extensive researches into the 2019-nCoV are needed to fully elucidate its pathway and pathogenic mechanisms and to identify potential

therapeutic targets, which can be effective in developing common preventive and therapeutic measures. The behavior prediction of COVID-19 is a significant problem. Therefore, in the present research, we compared the different distribution of COVID-19 cases based on the daily reported data.

Materials and Methods

Ethical approval

This study used publicly available data and therefore does not require a code of ethics.

Study design

We used the information on deaths and new cases with confirmed 2019-nCoV infection in Iran based on official reports from governmental institutes [10,14]. We collected the data either directly from governmental websites or from news sites that directly quoted governmental statements. The data were collected in real time and thus may be updated as more details on cases become publicly available. The arranged data are available as the Online Supplementary Material (file S1). The latest update to the dataset was on March 27, 2020, for cases reported up to March 26.

We performed a bootstrap method, based on case resampling, to compute the 95% confidence intervals. We used the three types of continuous distribution known as Normal, Lognormal, and Weibull. Akaike information criterion was used to identify the best fit model. We have also presented the median and mean of the best-fitted model to discover the peak.

Normal distribution

A normal (or Gaussian or Gauss or Laplace–Gauss) distribution is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\,e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The parameter μ is the mean or expectation of the distribution (and also its median and mode) and σ is its standard deviation. The variance of the distribution is σ_2. A random variable with a Gaussian distribution is said to be normally distributed and is called a normal deviate [17].

Lognormal distribution

A positive random variable X is lognormally distributed if the logarithm of X is normally distributed [18],

$$Ln(X) \sim N(\mu,\ \sigma^2)$$

Weibull distribution

The probability density function of a Weibull random variable is:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{(-\frac{x}{\lambda})^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Where, k > 0 is the shape parameter and λ > 0 is the scale parameter of the distribution [19].

All statistical analyses were performed using package fitdistrplus in R (version 3.6.3) [20].

Results

We used 36 pieces of daily data from February 21, 2020, in Iran. The smallest number of new cases was 11 (February 22, 2020) and the highest number was 2926 (March 27, 2020). On the other hand, the smallest number of daily deaths was 2 (February 21, 22, and 23, 2020) and the highest number was 157 (March 26, 2020). Figure-1 shows the histogram of daily new cases. The graph also shows three fitted distributions on the data.

Table-1 presents the goodness of fit of three distributions. Accordingly, the best fit belongs to Weibull distribution. With the assumption of Weibull distribution for new daily cases, we can determine the mean, median, and mode of the data as being 897.60, 624.61, and 4.59, respectively.

Figure-2 shows the histogram of daily deaths. The graph also shows three fitted distributions on the data.
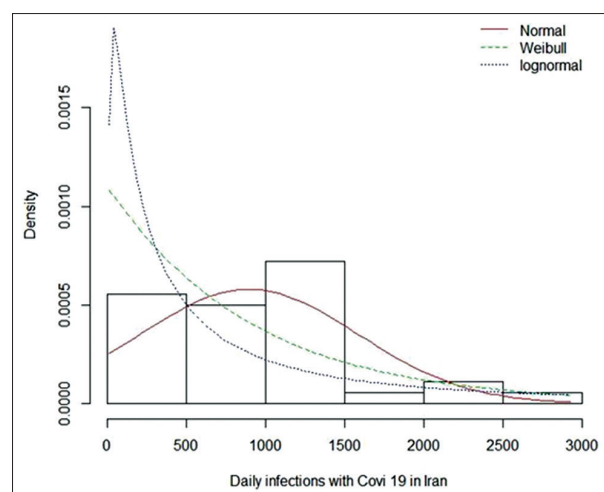

Figure-1: Histogram and three fitted distributions (Normal, Lognormal, and Weibull) on the daily infections with COVID-19 in Iran.
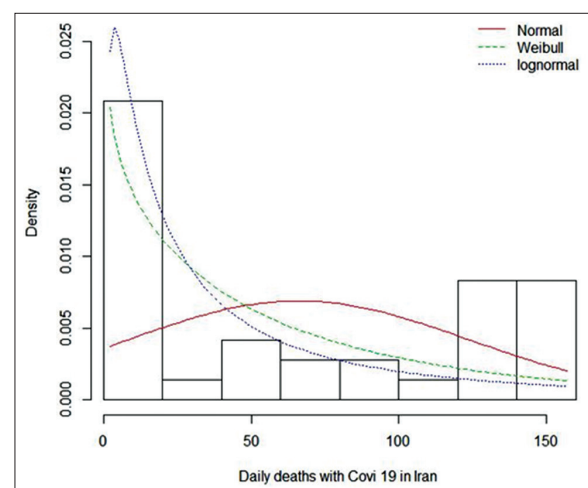

Figure-2: Histogram and three fitted distributions (Normal, Lognormal, and Weibull) on the daily deaths with COVID-19 in Iran.

Table-1: Goodness of fit of three distributions (Normal, Lognormal, and Weibull) on daily new COVID-19 infections in Iran.

Distribution	Parameters	Estimation	AIC
Normal	Mean	897.97	576.61
	SD	688.10	
Lognormal	Meanlog	6.13	580.88
	Sdlog	1.58	
Weibull	Shape	1.00	565.61
	Scale	899.48	

AIC=Akaike information criterion

Table-2: Goodness of fit of three distributions (Normal, Lognormal, and Weibull) on daily new COVID-19 deaths in Iran.

Distribution	Parameters	Estimation	AIC
Normal	Mean	66.00	398.38
	SD	57.89	
Lognormal	Meanlog	3.44	381.86
	Sdlog	1.48	
Weibull	Shape	0.88	376.76
	Scale	62.35	

AIC=Akaike information criterion

Table-2 presents the goodness of fit of three distributions. Accordingly, the best fit belongs to Weibull distribution. With the assumption of Weibull distribution for daily deaths, we can state that the mean, median, and mode of data are 66.44, 41.11, and 0, respectively.

Discussion

On the March 11, 2020, the director of the WHO said, "In the days and weeks ahead, we expect to see the number of cases, the number of deaths, and the number of affected countries climb even higher. We have therefore made the assessment that COVID-19 can be characterized as a pandemic" [21]. In the present study, we demonstrated the distribution of COVID-19 data in Iran with various types of continuous distributions. One of the distributions we used was the Weibull distribution. The Weibull distribution is related to a number of other probability distributions; in particular, it interpolates between the exponential distribution (k = 1) and the Rayleigh distribution (k = 2 and $\lambda = \sigma\sqrt{2}$). Its complementary cumulative distribution function is a stretched exponential function [22]. We used real daily data (positive infections and deaths) reported by the Ministry of Health [10,14]. Our study showed that the Weibull distribution was the best fit with the data. However, the parameters of distribution were different between new cases and daily deaths data. The shape parameter on the daily deaths data was <1; then, the rate fortunately decreased. On the other hand, the shape parameter on the new cases data was approximately 1. Thus, as mentioned before, the Weibull distribution can change to exponential distribution. We could not find any similar works to compare our results. If we use China COVID-19 data (63 daily data), we can see the same distribution. On the other hand, if we only use 36 initial China COVID-19 data (similar to our data), the Lognormal distribution has the best fit. It is then possible to see the similar behaviors of COVID-19 in both countries in the long run.

According to the mean and median of the best-fitted distribution, it is expected the peak of the disease will pass although it is New Year in Iran, now and the people travel around and visit each other a lot. Contrary to government warnings, many people still set off on journeys. On the other hand, the World Health Organization reported an incubation period for COVID-19 of between 2 and 10 days [23]. Other peaks should be expected if the health instructions are not followed. There are considerable uncertainties in assessing the risk of this event, due to lack of detailed epidemiological analyses. Outbreaks of infectious diseases require investigation by combined teams from departments such as human health, animal health, and wildlife [4].

Conclusion

Our study has shown that the Weibull distribution was the best fit with the data. The death rate is decreasing, and according to the results, we can expect to pass the peak of the disease. Furthermore, similar behaviors with respect to COVID-19 have been observed in both Iran and China in the long run.

Availability of Data and Materials

The data that support the findings of this study are available from: https://www.worldometers.info/coronavirus/.

Authors' Contributions

KM, MS and EG: Study concept and design. EG: Acquisition of data. MS: Analysis and interpretation of data. MS: Drafting of the manuscript. EG, MS and KM: Critical revision of the manuscript for important intellectual content. MS: Statistical analysis and supervision. All authors read and approved the final manuscript.

Acknowledgments

Competing Interests

The authors declare that they have no competing interests.

Publisher's Note

Veterinary World (Publisher of International Journal of One Health) remains neutral with regard to jurisdictional claims in published institutional affiliation.

References

1. Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M.,

Kolesnikova, L., Fouchier, R.A. and Berger A. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.*, 348(20): 1967-1976.

2. Wolfe, N.D., Dunavan, C.P. and Diamond J. (2007) Origins of major human infectious diseases. *Nature*, 447(7142):279-283.

3. Ventura, C.V., Maia, M., Bravo-Filho, V., Góis, A.L. and Belfort, R. (2016) Zika virus in Brazil and macular atrophy in a child with microcephaly. *Lancet*, 387(10015): 228.

4. Sukumaran, A. and Pradeepkumar, A.S. (2015) One health approach: A platform for intervention in emerging public health challenges of Kerala state. *Int. J. One Health*, 1: 14-25.

5. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. and Yuan, M.L. (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798): 265-269.

6. European Centre for Disease Prevention and Control. (2020) Risk Assessment: Outbreak of Acute Respiratory Syndrome Associated with a Novel Coronavirus, Wuhan, China; First Update. Available from: https://www.ecdc.europa.eu/en/publications-data/risk-assessment-outbreak-acute-respiratory-syndrome-associated-novel-coronavirus. Retrieved on 27-03-2020.

7. New York Times. (2020) Toll from Outbreak Climbs in China as Infections Reach Europe and Australia. New York Times, New York. Available from: https://www.nytimes.com/2020/01/24/world/asia/china-coronavirus.html. Retrieved on 27-03-2020.

8. Read, J.M., Bridgen, J.R., Cummings, D.A., Ho, A. and Jewell, C.P. (2020) Novel Coronavirus 2019-nCoV: Early Estimation of Epidemiological Parameters and Epidemic Predictions. medRxiv.

9. Farnoosh, G., Alishiri, G., Zijoud, S.R., Dorostkar, R. and Farahani, A.J. (2020) Understanding the 2019-novel coronavirus (2019-nCoV) and coronavirus disease (COVID-19) based on available evidence a narrative review. *J. Mil. Med.*, 22(1): 1-1.

10. Worldometer, COVID-19 Coronavirus Pandemic. (2020) Available from: https://www.worldometers.info/coronavirus. Retrieved on 27-03-2020.

11. New York Times. (2020) Iran Reports its First 2 Cases of the New Coronavirus. New York Times, New York.

12. Wikipedia. (2020). Coronavirus Pandemic in Iran. Available from: https://www.en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran. Retrieved on 27-03-2020.

13. Aftab News Agency. (2020) Corona Patients in Iran. Aftab News Agency, New Delhi. Available from: https://www.aftabnews.ir/fa/news/641753. Retrieved on 27-03-2020.

14. Ministry of Health. (2020) Some Characteristics of Covid19 in Iran. Ministry of Health, New Delhi. Available from: http://www.behdasht.gov.ir/uploads/factsheet%204.%20 26.12%20-%20fa.pdf. Retrieved on 27-03-2020.

15. Aftab News Agency. (2020) We have not yet reached the peak of Covid19. Aftab News Agency, New Delhi. Available from: https://www.aftabnews.ir/fa/news/642341. Retrieved on 27-03-2020.

16. Ministry of Health. (2020) Daily Reports of Covid19 in Iran. Ministry of Health, New Delhi. Available from: http://www.behdasht.gov.ir/index.jsp?siteid=1&fkeyid=&siteid=1&pageid=54782&newsview=201158. Retrieved on 27-03-2020.

17. Eric, W.W. (2020) Normal Distribution Function. MathWorld. Available from: https://mathworld.wolfram.com/NormalDistribution.html. Retrieved on 27-03-2020.

18. Norman, L.J., Samuel, K. and Balakrishnan, N. (1994) Lognormal distributions continuous univariate distributions. In: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. 2nd ed., Vol. 1. John Wiley and Sons, New York.

19. Papoulis, P.A. and Unnikrishna, P.S. (2002) Probability, Random Variables, and Stochastic Processes. 4th ed. McGraw-Hill, Boston.

20. Delignette-Muller, M.L. and Dutang, C. (2015) Fitdistrplus: An R package for fitting distributions. *J. Stat. Softw.*, 64(4): 1-34.

21. World Health Organization. (2020) WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 11 March 2020. World Health Organization, Geneva. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Retrieved on 27-03-2020.

22. Rayleigh Distribution MATLAB and Simulink MathWorks Australia. Available from: http://www.mathworks.com.au. Retrieved on 27-03-2020.

23. World Health Organization. (2020) Novel Coronavirus (2019-nCoV) Situation Report-7. World Health Organization, Geneva.

****