Research Article

Characteristics Analysis of Incubation Time of COVID-19

Yanfang Zhang,1 Yibin Zhao,1 Qingqing Ren,1 and Bo Liu2

1Department of Mathematics, Institute of Disaster Prevention, Hebei, Sanhe 065201, China
2Economic College, Hunan Agricultural University, Hunan, Changsha, China

Correspondence should be addressed to Yanfang Zhang; zyf@cidp.edu.cn

Received 1 July 2021; Revised 14 September 2021; Accepted 6 May 2022; Published 21 May 2022

Academic Editor: Leonid Shaikhet

Copyright © 2022 Yanfang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. The statistical characteristics of incubation time of COVID-19 are studied comprehensively in this paper. We studied the mean value and tail characteristic of incubation time. Also, the difference in incubation time is studied in different age groups, different genders, and different contact modes. Methods. We analyzed the data by a lognormal distribution. Tail characteristics of incubation time are studied by the excess function. The difference in incubation time between different age groups, different genders, and different contact modes is studied by one-way ANOVA. Results. The results showed that the mean value of incubation time is about 12.6 days, and the standard deviation is about 6.1 days. The incubation time of 95% of patients was within 24 days. Some patients have a very long incubation time according to the excess function. The incubation time was not related to age and gender but related to the suspected contact modes. Conclusions. The mean value of incubation time is consistent with the actual situation. The above results are a reasonable explanation for the phenomenon of super-long incubation time in some patients and the development trend of epidemic situations.

1. Background

Novel coronavirus pneumonia has been confirmed by medical experts since December 2019 in Wuhan, Hubei. The novel coronavirus 2019 was named by the WHO on January 20, 2020, that is, COVID-19. It has spread rapidly in Hubei Province, and the number of patients is increasing rapidly. The first case in Tianjin came back from a trip to Wuhan. Due to the large population flow on the eve of the Spring Festival, the cases spread rapidly through various channels, and the number of patients is increasing. The first confirmed case of COVID-19 was released on January 21 in Tianjin. The government has actively taken preventive and control measures to control the spread of the virus. Sun et al. [1] systematically explored the effects of lockdown and medical resources on the COVID-19 transmission in Wuhan. In order to understand the development of the epidemic situation and reduce the panic of the masses, scholars researched the epidemic situation. Many medical workers study the symptoms and treatment process of patients [2–7]. By compiling a line-list database of transmission pairs in mainland China. Ali et al. [8] show that the mean serial interval of COVID-19 shortened substantially from 7.8 to 2.6 days within a month (from January 9 to February 13, 2020). Scholars engaged in mathematical works also use a differential equation dynamic system and data analysis methods to analyze and predict the development trend of the epidemic situation [7, 9, 10], Luo et al. [11] devised a pairwise epidemic model with NPI to analyze the COVID-19 outbreak in China. Du and Yuan [12] applied mathematical modeling to investigate the infections of the COVID-19. Muller and Muller [13] present a deterministic transmission dynamic compartmental model for the spread of the novel coronavirus on a college campus for the purpose of analyzing strategies to mitigate an outbreak. In the study of infectious diseases, the analysis of the law of incubation time is of great significance for epidemic prevention, control, and later treatment [14, 15], which is also the main content of this paper.

Based on 136 confirmed cases from January 21 to February 27 in Tianjin, 84 samples were selected as the research data. One-way ANOVA and the lognormal distribution model were used to comprehensively analyze the incubation time characteristics.
2. Analysis of the Epidemic Situation in Tianjin

2.1. Data Sources and Characteristics. The data are from the latest report of the Tianjin government affairs network. There were 136 cases in Tianjin from January 21 to February 27.

In this paper, the data in Tianjin are selected as the research sample because the data sample has the following characteristics.

The information on confirmed cases in Tianjin is relatively complete. The relationship between cases is clear, and the data chain is relatively complete.

Data are clear because the prevention measures are relatively timely and effective and the process of case discovery, isolation, and treatment is scientific and reasonable.

2.2. Data Cleaning. As of March 3, 136 cases of COVID-19 were confirmed in Tianjin. Samples without definite contact time were discarded. The contact date was determined according to the following principles:

(1) If there is a clear contact time, the contact date should be determined by contact time. For example, if the case went to the department store on January 20, then the contact time should be January 20.

(2) If the contact time is a period, the second day shall be the contact time within three days, and the third day should be the contact time beyond three days. For example, if the case traveled to Wuhan from January 21 to January 25, the contact time should be set as January 23.

(3) If the cases live together, the contact time is the same. If the cases are infected by relatives, the contact date shall be set as the date they contact each other.

In addition, for three asymptomatic patients in the sample, the incubation time is defined as the period between exposure to the disease (if transmission occurs) and the onset of clinical symptoms [16]. According to the above principles, 84 case data were selected as study samples.

2.3. Overview of Sample Data Characteristics. According to the simple statistical characteristics of the sample data (Table 1), the skewness and kurtosis of the data are both greater than 1, and the distribution of frequency shows a peak state and tends to the right side. Because the incubation time of some individual cases is too long, the density curve shows right tailing, just as shown in Figure 1.

The numerical characteristics show that the frequency distribution of samples shows a peak state and tends to the right side, which is the characteristic of the lognormal distribution. Therefore, the lognormal distribution will be used to describe the statistical characteristics of incubation time.

3. Lognormal Model of Incubation Time

3.1. Lognormal Distribution

Definition 1 (see [17]). Let $X$ be a continuous random variable with a positive value. If $X$ obeys normal distribution, then $\ln X$ obeys lognormal distribution.

The corresponding probability density function is

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

(1)

The expectation and variance are, respectively,

$$ E(X) = e^{\mu + \frac{\sigma^2}{2}}, $$
$$ D(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}, $$

where $\mu$ and $\sigma$ are the mean and standard deviation of logarithmic variable. It is fit to describe the variables with the heavier upward fluctuation, the slighter downward fluctuation, and asymmetric by the lognormal distribution.

Definition 2 (see [17]). Let $A = \{x: 0 < F(x) < 1\}$, $x_a = \inf A_{x \in A}$, $x^* = \sup A_{x \in A}$. $A$ is called the support of distribution $F$, and $x^*$ and $x_a$ are the upper and lower ends of the distributed support, respectively.

Definition 3 (see [17]). Let $x_1, x_2, \ldots, x_n$ be a sample from population $X$, $F(x)$ is the distribution function of $X$, and $x^*$ is the top end of support. For a fixed large value $u < x^*$, then,

$$ F_u(x) = P(X-u \leq x|X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \ x \geq u $$

(3)

is called the distribution function of the over threshold of the random variable, short for excess distribution. The probability density of excess distribution is

$$ f_u(x) = \frac{f(x+u)}{1 - F(u)}, \ x \geq 0. $$

(4)

Definition 4 (see [17]). $e(u) = E(X-u|X > u)$ is called average excess function of $X$.

The average excess function is used to describe the average value of the random variable whose value is greater than the given threshold. It is an important digital feature used to describe the tail of the random variable, to make the expression of the statistical regularity of the random variable comprehensive and complete.

The average excess function of the lognormal distribution is approximate as follows:

$$ e(u) = \frac{\sigma^2 u}{\ln u - u} \left(1 + o(1)\right) = \frac{\sigma^2 u}{\ln u - u} $$

(5)

Definition 5 (see [17]). If $X \sim F(x)$, $x_p = \inf\{x \in R: F(x) \geq p, 0 < p < 1\}$ is called the $p$ quantile of $F$. If the incubation
time obeys the lognormal distribution, $x_p$ indicates that there are $x_p$ units patient’s latent is within $p \times 100\%$

3.2. Characteristic Description of Incubation Time Based on Lognormal Distribution. According to the analysis of the basic characteristics of sample data in Section 2.3, it can be seen that the sample information has the characteristics of the lognormal distribution, so whether the sample obeys the lognormal distribution is tested by nonparametric test.

Take the quantile of lognormal distribution as $x$ axis and the quantile of sample data as the $y$ axis, we can make the Q-Q chart. The Q-Q chart test shows that the samples obey the lognormal distribution, as shown in Figure 2. The data scatter points in the Q-Q chart are distributed along the diagonal of the first quadrant, indicating that the statistical law of the incubation time is similar to that of the lognormal distribution.

Furthermore, K–S (Kolmogorov–Smirnov) test was used to test that the sample obeys lognormal distribution. The statistic $D = 0.11895$, $p = 0.1712 \gg 0.05$. That is to say there was no significant difference between the incubation time distribution and the lognormal distribution.

Based on the above results, it is reasonable to assume that the incubation time obey the lognormal distribution. The maximum likelihood estimation method is used to estimate the parameters of lognormal distribution. The estimation of the parameters was $\mu = 2.4278$, and $\sigma = 0.458$.

The correlation test between probability density of the lognormal distribution and sample frequency shows $p \ll 0.05$, coefficient of determination $R^2 = 0.70$. It shows $F = 18.0513 \geq 1$, and the lognormal model can explain 70% variation of sample data. The above indexes and the fitting effect are shown in Figure 3.

It shows that the description of the statistical law of incubation time by the lognormal distribution is reasonable and credible. From equations (2) and (3), we can get $E(x) = 12.59$ and $D(x) = 37.04$.

The average incubation time is about 13 days, indicating that it is feasible to observe 14 days of suspected cases in the form of emergency prevention. However, the big variance (standard deviation is about 6.1 days) indicates that the difference between the incubation time of different individuals is very different, in nonemergency situations, it is more reasonable to observe the number of suspected objects for longer periods. The above numerical characteristics are consistent with the actual incubation time of COVID-19.

The quantile of incubation time is $x_{0.95} \approx 24$, and the results showed that 95% of the patients had an incubation time of less than 24 days. From equation (5), we can get the average excess function of $x_{0.95}$ is $e(x_{0.95}) \approx 6.71$. In other words, for about 5% of patients with incubation time over 24 days, their average incubation time should be about 30 days, which is corresponding to the data of patients with long incubation time.

4. Analysis of the Difference in Incubation Time in Different Age Groups

In order to make a comprehensive understanding of the incubation time of COVID-19, we will analyze the difference in the incubation time by age, gender, and contact modes of the patients.

4.1. Data Processing. According to the statistical characteristics, there are three abnormal points in the data. Before data processing, the following three abnormal points shown in Table 2 are removed. The three abnormal points are all with long incubation time, which is consistent with the current detection of nucleic acid-positive but they are asymptomatic cases.
4.2. Analysis of the Difference in Incubation Time in Different Age Groups. The samples were divided according to the five age groups 0–18 years old, 19–35 years old, 36–50 years old, 51–65 years old, and over 65 years old. Numerical character and the K–S test of normal distribution were carried out for each level. The results are shown in Table 3.

4.3. Statistical Character and Normal Distribution Test of Incubation Time in Different Age Groups. There are only two samples in the 0–18 age group. On the one hand, the results of the two samples are not statistically significant. On the other hand, whether adolescents have strong immunity needs further verification. In view of this, the following analysis will only be carried out for each group aged over 19.

The numerical character of different age groups and the results of the K–S test of the normal distribution are shown in Table 3. It showed that the incubation time in each age group can be considered to obey a normal distribution (P value is greater than 0.05). The homogeneity test of the variance of each age group showed that the P value is greater than 0.05. It can be considered that the variance of each sample is equal. One-way ANOVA showed that the P value is greater than 0.05. Therefore, there is no significant difference between each age group. That is the incubation time has no relationship with age.

4.4. Analysis of the Difference in Incubation Time between Different Genders. The samples were divided into the male group and the female group. Numerical character and K–S test of normal distribution were carried out respectively. The numerical characteristics of males and females are shown in Table 4. The K–S test of normal distribution (Table 4)
showed that the incubation time of different genders obeys a normal distribution (P value is greater than 0.05). The homogeneity test of the incubation time variance of males and females showed that the P value is greater than 0.05. So, the variance of a different gender is equal.

One-way ANOVA was used to analyze the mean incubation time of males and females (P value is greater than 0.05). Therefore, there is no significant difference in the average incubation time between the male and the female. That is, after removing the suspected abnormal points, the incubation time has nothing to do with gender.

4.5. Analysis of the Difference in Incubation Time between Different Contact Modes. According to the conventional pathological analysis, the incubation time of some infectious diseases will be different in different contact modes. For aids and hepatitis B, if the contact modes are different, the virus number is different. Also, the incubation time is different [10].

Transmission routes of novel coronavirus is a problem that is different from that in paper [10]. Therefore, it is worthwhile to study the different ways of infection of COVID-19.

The contact modes will be divided into kinship contact (living with together) and general contact. It showed that in Table 5, there is no significant difference in the incubation time.

The numerical characteristics and the normal distribution K–S test results of two groups with different contact modes (Table 5) show that the incubation time of the two groups of cases can be considered to obey normal distribution. The homogeneity test of the variance of the two groups showed that the P value is greater than 0.05. That is, the variance is equal for the two groups of data.

One-way ANOVA was used to analyze the mean incubation time of the two groups. P value is greater than 0.05, so there was a significant difference in the mean incubation time between the two groups. The incubation time is related to the contact modes. The analysis results of are based on the samples obtained. How the contact modes affect the incubation time is a further problem to be solved.

5. Conclusions

The incubation time characteristic of COVID-19 is comprehensively analyzed by statistical modeling in this paper. The difference between the incubation time of the different groups is analyzed by one-way ANOVA. The conclusion is consistent with the experience. The mean value of super-long incubation time obtained from the over threshold distribution is a supplement to the present experience. Because the data selected in this paper is limited to Tianjin, the analysis results are limited. To analyze whether the incubation time varies in different areas and what characteristics of the diagnosis time will be the focus of the author’s following research.

Data Availability

The excel data used to support the findings of this study are included within the supplementary information file. All the data, models, or codes used during the study are available from the corresponding author by request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (no. ZY20215140), Self financing project of scientific research and development plan of Lang Fang science and Technology Bureau (no.
2020011022), Opening Foundation of Hebei Key Laboratory of Earthquake Dynamics (no. FZ212105), and Earthquake Science and Technology Spark Project of Hebei Earthquake Agency (DZ202000827053).

**Supplementary Materials**

The data used to support the findings of this study are included within the supplementary information file. (Supplementary Materials)

**References**


