Estimating the Incubation Period of COVID-19 (SARS-CoV-2) Virus

Narayana Darapaneni
Director - AIML
Great Learning/Northwestern University
Illinois, USA

Sandeep Nair
Student - AIML
Great Learning
Mumbai, India

Viraj Shah
Student - AIML
Great Learning
Mumbai, India

Sameer Sawant
Student - AIML
Great Learning
Mumbai, India

Vandit Deshmukh
Student - AIML
Great Learning
Mumbai, India

Anwesh Reddy Paduri
Research Assistant - AIML
Great Learning
Mumbai, India

***Abstract* - Our intention is to estimate the incubation period for the novel coronavirus disease (COVID-19 / SARS-CoV-2) that emerged in Wuhan, Hubei province, China, in 2019. We also wish to undertake a comparison of our results with few other published studies. Understanding the incubation period is very important for health authorities as it allows them to introduce more effective quarantine systems for people suspected of carrying the virus, as a way of controlling and hopefully preventing the spread of the virus. As part of this study, 195 cases have been chosen for which data was available from various sources. This includes individuals of 22 countries. Their ages range from 2 to 85 years of age with a median age of 45 years. 57% of the people were male while there were 43% were female. The data has only those cases where information about the time interval of exposure and symptom onset is available. Here we have tried to identify the incubation period using the following three methods: Weibull Distribution, Log-normal distribution, Kaplan Meier Distribution. These results have also been compared with the incubation periods resulting out of few other prominent studies.**

Keywords – Incubation period, Covid-19, Comparative study, Weibull Distribution, Log-normal distribution, Kaplan Meier Distribution.

I. INTRODUCTION

At the turn of 2020, a new form of coronavirus (COVID-19) was found to be the source of infection responsible for an epidemic of viral pneumonia in Wuhan, China, a region in which the first patients began to show symptoms in December 2019. At the time of writing (July 26, 2020) over 1.6 Cr people globally have caught the virus, of whom more than 6.5 Lacs have died. The novel virus, causing severe acute respiratory disease, is thought to be from the same family as Middle East Respiratory Syndrome (MERS) coronavirus and Severe Acute Respiratory Syndrome (SARS) coronavirus, however SARS-CoV-2 has its own unique features which makes it more dangerous. This means that central epidemiological parameters, which includes the incubation period, are being urgently researched in real-time from case reports while the epidemic is continuing. The incubation period of a virus represents the time span from the probable earliest contact with a source of transmission and the earliest recognition of the first symptoms. Accurately estimating the length of incubation period is essential for effective contemporary public health measures to be taken. If health authorities know what the incubation period is then they will know for how long a healthy individual should be monitored and have their movement restricted (quarantine period). Correctly estimating the incubation period will also help us to comprehend how infectious COVID-19 is, make estimations of the size of the pandemic, and decide on the best course of action. With insufficient data available to definitively state what the incubation period for this virus is, the World Health Organisation (WHO) is working with a broad range of 0-14 days, the European Centre for Disease Prevention and Control (CDCDC) is working with a range of 2-14 days, and a number of studies have made the assumption that the incubation period is similar to that of the MERS and SARS coronaviruses [11] [9].

Various infectious/viral diseases have a variety of incubation periods. Determining the incubation period for COVID-19 is no simple matter, as there is no consistency in the quality of the available data. One reason for this is that generally we can only discover the times when the patient was in contact with persons carrying the virus, and then assume that the incubation period runs from the earliest date of exposure to the appearance of clinical symptoms or medical diagnosis. This way of calculating the incubation period may well be responsible for overestimation. This paper represents an effort into estimating the incubation period of SARS-CoV-2 using three different methods as well as comparing our findings with few other studies that has been conducted.

As the COVID-19 outbreak continues to evolve, we are learning more about this new virus every day. Based on various reports the types of infections can be categorized as symptomatic and asymptomatic. [9]

1. Symptomatic: By way of definition, a symptomatic COVID-19 case is a case who has developed signs and symptoms compatible with COVID-19 virus infection.

Symptomatic transmission refers to transmission from a person while they are experiencing symptoms.

2. Asymptomatic: An asymptomatic laboratory-confirmed case is a person infected with COVID-19 who does not develop symptoms. Asymptomatic transmission refers to transmission of the virus from a person, who does not develop symptoms. There are few reports of laboratory-confirmed cases who are truly asymptomatic, and to date, there has been no documented asymptomatic transmission. This does not exclude the possibility that it may occur. Asymptomatic cases have been reported as part of contact tracing efforts in some countries.

As no symptoms are seen in asymptomatic cases, it is not possible to calculate the incubation period in such scenarios. Hence these have been excluded from the study.

II. MATERIALS AND METHODS

The panic engulfing the pandemic has ensured that there is lack availability of quality data. Data available in public domain including news feeds and reports have been used to create this dataset. It contains details of 195 individuals and includes citizens of 22 countries: China (65), Singapore (23), Hong Kong (22), Japan (20), USA (13), Taiwan (10), South Korea (7), Germany (6), Australia (5), Vietnam, France & Malaysia (4 each), Thailand (3), Switzerland, UAE, Spain, Cambodia, Finland, Canada, Sweden, Sri Lanka and Philippines (1 each). 112 of the 195 individuals are male while the remaining 83 are female. The age ranges between 2 and 85 years where as the median age is 45 years. The dataset also contains the date when the individual visited the hospital. Along with the same, the source from where the information has been procured is also updated. The date of exposure to the onset of symptoms ranges between zero and 32 days as per the dataset that we are using.

We have estimated the incubation period using three different methods which are as follows:

A. Weibull Distribution

In probability theory and statistics, the Weibull distribution [16][17] is a continuous probability distribution.

Standard parameterization: The probability density function of a Weibull random variable is:

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

where k > 0 is the shape parameter and λ > 0 is the scale parameter of the distribution. Its complementary cumulative distribution function is a stretched exponential function. 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 = sqrt 2 * sigma).

If the quantity X is a "time-to-failure", the Weibull distribution gives a distribution for which the failure rate is proportional to a power of time. The shape parameter, k, is that power plus one, and so this parameter can be interpreted directly as follows:[3]

A value of k<1 indicates that the failure rate decreases over time (Lindy effect). This happens if there is significant "infant mortality", or defective items failing early and the failure rate decreasing over time as the defective items are weeded out of the population. In the context of the diffusion of innovations, this means negative word of mouth: the hazard function is a monotonically decreasing function of the proportion of adopters;

A value of k=1 indicates that the failure rate is constant over time. This might suggest random external events are causing mortality, or failure. The Weibull distribution reduces to an exponential distribution;

A value of k>1 indicates that the failure rate increases with time. This happens if there is an "aging" process, or parts that are more likely to fail as time goes on. In the context of the diffusion of innovations, this means positive word of mouth: the hazard function is a monotonically increasing function of the proportion of adopters. The function is first convex, then concave with an inflexion point at (e^{1/k}-1)/e^{1/k}, k>1

B. Log-normal distribution

A lognormal distribution [18] is a statistical distribution of logarithmic values from a related normal distribution. A lognormal distribution [8] can be translated to a normal distribution and vice versa using associated logarithmic calculations.

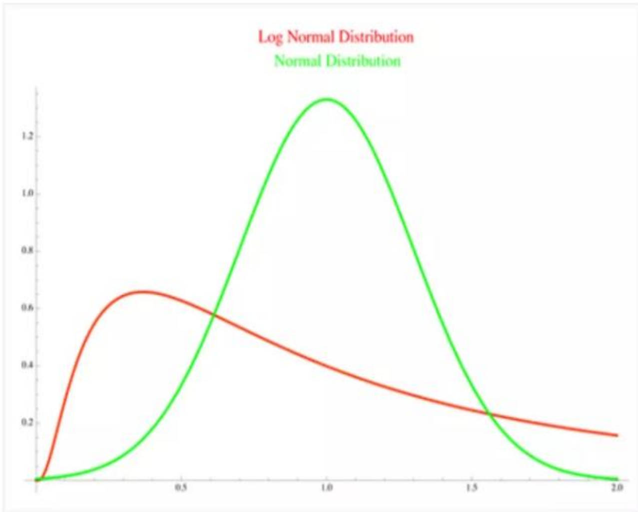

Fig 1: distribution curve of normal and log-normal distribution

In general, most log-normal distributions are the result of taking the natural log where the base is equal to e=2.718. However, the log-normal distribution can be scaled using a different base which affects the shape of the lognormal distribution.

Overall the log-normal distribution plots the log of random variables from a normal distribution curve. In general, the log is known as the exponent to which a base number must be raised in order to produce the random variable (x) that is found along a normally distributed curve.

C. Kaplan-Meier Distribution

The Kaplan–Meier estimator [19], also known as the product limit estimator, is a non-parametric statistic used to estimate the survival function from lifetime data. Survival Analysis is a set of statistical tools, which addresses questions such as ‘how long it would be, before a particular event occurs’; in other words we can also call it as a ‘time to event’ analysis.

In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment. The estimator of the survival function S(t) (the probability that life is longer than t) is given by:

$$\widehat{S}(t) = \prod_{i:\ t_i \le t} \left(1 - \frac{d_i}{n_i}\right), \quad (2)$$

with t_i a time when at least one event happened, d_i the number of events (e.g., deaths) that happened at time t_i, and n_i the individuals known to have survived (have not yet had an event or been censored) up to time t_i

D. Model selection

Akaike information criterion (AIC) (Akaike, 1974) [10] is a fined technique based on in-sample fit to estimate the likelihood of a model to predict/estimate the future values.

A good model is the one that has minimum AIC among all the other models. The AIC can be used to select between the additive and multiplicative Holt-Winters models.

$$AIC = -2 * \ln(L) + 2 * k \quad (3)$$

where L is the value of the likelihood, N is the number of recorded measurements, and k is the number of estimated parameters.

We compared Weibull, lognormal and Kaplan Meier models and found that the lognormal had the best fit by AIC, with Weibull significantly lower and Kaplan Meier strongly dis-favoured.

The estimated lognormal model parameters are μ=1.678(1.453,1.902) and σ=0.505(0.336,0.759)

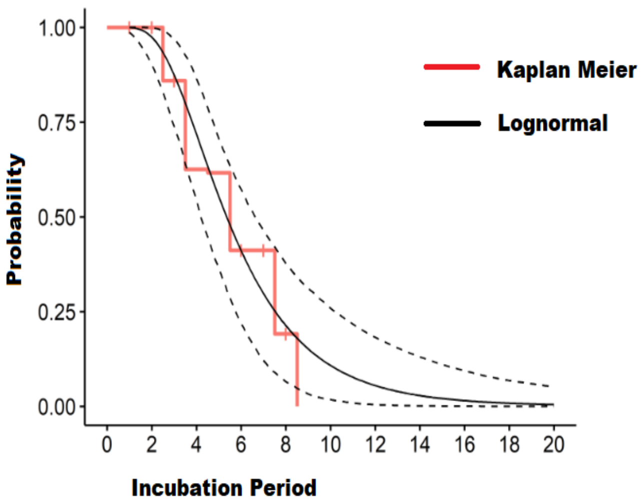

Fig 2: Kaplan Meier vs Lognormal Distribution

A Comparative Study of Our Findings with other Published Studies

As part of this study we have gone ahead and compared our findings with that of few other studies on incubation period of COVID-19.

Study 1: Backer J.et al [1) employed travel histories and symptom onsets for 88 confirmed cases discovered beyond the boundaries of Wuhan, China, in the early stage of the coronavirus outbreak. These authors made an estimation that the incubation period ranged from 2.1 to 11.1 days (2.5th to 97.5th percentile) and that the mean incubation period was 6.4 days (95% CI: 5.6-7.7). This research offers empirical evidence and falls into the previously mentioned incubation periods estimated by both the ECDC and the WHO []. The researchers employed three parametric forms related to the distribution of the incubation period: lognormal distribution, gamma distribution, and the Weibull distribution. They employed uniform prior probability distribution for the exposure interval related to the point of infection for all 88 individuals. The researchers used posterior distribution samples employing the RStan package within R software version 3.6.0 (R Foundation, Vienna, Austria).

Study 2: Natalie ML. Et al [2] examined COVID-19's epidemiological characteristics and incubation period. The researchers harvested information relating to confirmed diagnoses of COVID-19 infection beyond the disease epicentre in Wuhan, China, using official reporting from state institutes and reporting on mortalities both within and outside Wuhan. The data used by the authors was either directly harvested from government sources or from news websites reporting government statements. The data collection process was real-time, so it was added to as further details emerged. The final data selection represented a selection of reported cases up to January 31, 2020. The outcomes of this research concluded that the incubation period falls into a range of 2-14 days (95% CI), with the mean being approximately five days as found by employing best-fit lognormal distribution. Mean time between onset of symptoms and admission to hospital (either for treatment or isolation) was estimated to be between three and four days with no truncation and between five and nine days with right truncation. Based on the 95th percentile estimate for the incubation period, the researchers recommended that exposed individuals should be quarantined for a minimum of 14 days. When making estimates of the risk of fatality in COVID-19 cases, the median time delay between illness onset and death of 13 days (17 days with right truncation) should be taken into consideration.

Study 3: Jiang X et al [12] undertook research making a comparison between incubation periods for MERS, SARS, and SARSCoV-2. The researchers reported that SARS-CoV-2 has an extended incubation period, which has led to modifications in official policy for control and screening. To prevent the virus spreading, any individual who may have been exposed should go into isolation for 14 days, this being the outer limit of predictions for incubation times. Nevertheless, by analysing a large dataset for this research, researchers report that no identifiable difference exists between incubation times for SARS-CoV-2, severe acute respiratory syndrome coronavirus (SARSCoV), and the Middle East respiratory syndrome coronavirus (MERS-CoV), which highlights the requirement for more extensive and better-annotated datasets. This research covered 49 patients with SARS-CoV-2 who had definite dates for first exposure, end of exposure and beginning of symptoms, 153 patients with SARS-CoV, and 70 MERS-CoV patients; this data was amalgamated from seven separate papers. The results indicated that MERS incubates on average 5.8 days (95% CI: 5-6.5), SARS-CoV 4.7 days (95% CI: 4.3-5.1), and SARS-CoV2 4.9 days (95% CI: 4.4-5.5). This demonstrates that the longest incubation period is MERS-CoV, with SARS-CoV2 second longest.

Study 4: Lauer Stephen et al [13] researched the COVID-19 incubation period by looking at diagnosed cases that have been publicly reported. The aim of the study was to ascertain COVID-19's incubation period and to detail its implications for public health. The researchers examined diagnosed cases of COVID-19 occurring between January 4, 2020, and February 24, 2020. The research covered 181 subjects diagnosed with SARS-CoV-2 infection outside Hubei province, China by examining press releases and news reports from 50 different provinces, regions, and nations. The researchers harvested information regarding patient demographics, dates/time of possible exposure, onset of symptoms, onset of fever, and admission to hospital. The researchers estimate that, conservatively, 101/10,000 cases (99th percentile, 482) will experience symptom onset more than 14 days after being quarantined or actively monitored. Nevertheless, the researchers noted that severe cases could be overrepresented in public reporting; it is possible that severe and mild COVID 19 infections have different incubation periods. This research adds to the evidence that COVID-19 is like SARS in having a median incubation period of around five days. This recommendation comes from the research looking at proposed quarantine/active monitoring times for subjects with potential exposure to the virus.

Table 1: Various studies with which we compared our findings

Study	**Incubation Period**	**Method for estimating the incubation period.**	**No of cases**	**Location**	**Note**
1	6.4 days (95% CI: 5.6– 7.7), ranging from 2.1 to 11.1 days (2.5th to 97.5th percentile)	Probability density function (PDF). Used a doubly interval-censored likelihood function to estimate the parameter values.	88	Outside Wuhan, China	The study provides empirical evidence and the estimation is within the range for the incubation period of 0 to 14 days assumed by the WHO and of 2 to 12 days assumed by the ECDC [11].
2	5 days (2–14 days with 95% confidence)	Best-fit lognormal distribution	Real-time data	Outside Wuhan, China	Recommend the length of quarantine to be at least 14 days. The median time delay of 13 days from illness onset to death should be considered when estimating the fatality risk.
3	SARS-CoV2 4.9 (95% CI:4.4.-5.5) SARS-CoV and MERS-CoV were also studied	Fitted Weibull, lognormal, and gamma distributions	SARS -CoV2 49	China	long incubation time was reported to be associated with SARS-CoV-2 infection, leading to adjustments in screening and control policies.
4	Median 5.1 days (95% CI, 4.5 to 5.8 days)	Estimated the incubation time using a previously described parametric accelerated failure time model. Doubly interval-censored data. Data reduction technique	181	Regions, and countries outside Wuhan, Hubei province, China	101 out of every 10,000 cases (99th percentile, 482) develop symptoms after 14 days of active monitoring or quarantine. This work provides additional evidence for a median incubation period of approximately 5 days, like SARS.

III. DISCUSSION AND CONCLUSION

Infection with SARS-CoV-2, the virus that causes COVID-19, appears to be highly contagious and is primarily spread by droplets. Containment efforts have emphasized quarantine during the incubation period as the most effective measure to limit spread. Because of the personal and economic toll of this measure and its implication for transmission, we need to maximize our understanding of the incubation period.

Post assessment and analyses we estimate that individual-level incubation periods appear to be log-normally distributed (better than Weibull and Kaplan Meier), with median duration 5.4 (95% CI 4.2, 6.7) days, and that 95% of cases are estimated to have incubation periods between 2.0 (1.3, 2.9) and 14.5 (9.5, 25.2) days.

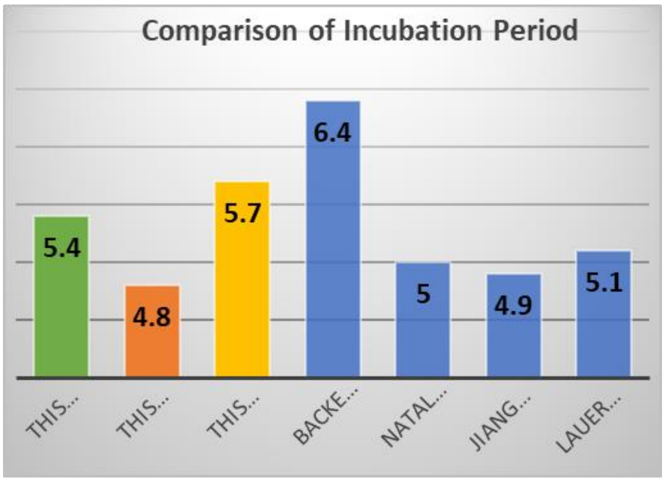

Fig 3: Comparison of incubation periods of our studies with four others

Comparison with the results of few other studies mentioned above confirms that the incubation period of COVID-19 is in line with our assessment. We however found that the incubation period could be as high up-to 35 days in some cases.

Our analysis also suggests that the individual-level variation in incubation times is large. During surveillance, contacts of known cases should be monitored for at least 14 and possibly as many as 35 days before being declared disease-free.

REFERENCES

[1] Backer J. , Klinkenberg D. , and Wallinga J., "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travelers from Wuhan, China," Eurosurveillance, vol. 25, no. 5, 2020.

[2] Natalie M.L., Tetsuro K., Yichi Y., Katsuma H., Akhmetzhanov R., Sung-mok J., Baoyin Y., Kinoshita R., Hiroshi N., "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data," J Clin Med., vol. 9, no. 2, p. 538, 2020.

[3] A. Pandey, “Survival Analysis: Intuition & Implementation in Python,” Towards Data Science, 06-Jan-2019. [Online]. Available: https://towardsdatascience.com/survival-analysis-intuition-implementation-in-python-504fde4fcf8e. [Accessed: 08-Sep-2020].

[4] Nishiura H., Jung SM., Linton NM., Kinoshita R., Yang Y., Hayashi K., Kobayashi T., Yuan B., and Akhmetzhanov AR., "The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020," J Clin Med., vol. 9, no. 2, 2020.

[5] Imai N., Dorigatti I., Cori A., Donnelly C., Riley S., and Ferguson NM., "Estimating the potential total number of novel Coronavirus cases in Wuhan City, China.," Imprial College London, London, 2020.

[6] Imai N., Cori A., Dorigatti I., Baguelin M., Donnelly CA., Riley S., et al., "Transmissibility of 2019-nCoV," Imprial College London, London, 2020.

[7] Zhao S., Lin Q., Ran J., Musa SS., Yang G., Wang W., Lou Y., Gao D., Yang L., He D., Wang MH., "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," Int J Infect Dis., vol. 92, pp. 214-217, 2020.

[8] Johnson, N.L., Kotz S., Balakrishnan, N., Lognormal Distributions", Continuous univariate distributions, vol. 1, New York: John Wiley & Sons, 1994.

[9] “Coronavirus Incubation Period (COVID-19) - Worldometer,” Worldometers.info. [Online]. Available: https://www.worldometers.info/coronavirus/coronavirus-incubation-period/. [Accessed: 08-Sep-2020].

[10] “Akaike Information Criterion - an overview | ScienceDirect Topics,” Sciencedirect.com. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/akaike-information-criterion. [Accessed: 08-Sep-2020].

[11] "Q & A on novel coronavirus," European Centre for Disease Prevention and Control (ECDC), 4 Feb 2020. [Online]. Available: https://www.ecdc.europa.eu/en/ novel-coronavirus-china/questions-answers.

[12] Jiang X., Rayner S., Luo MH., "Does SARS-CoV-2 has a longer incubation period than SARS and MERS?," J Med Virol, vol. 92, p. 476–8, 2020.

[13] Lauer Stephen A., Grantz Kyra H., Qifang B., et al.,, "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application," Ann. Intern. Med., pp. M20-0504, 2020.

[14] 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," The New England Journal of Medicine, 2020.

[15] Meili L., Pian C., Qianqian Y., Baojun S., and Junling M., "Transmission characteristics of the COVID-19 outbreak in China: a study driven by data," MedRxiv, 2020.

[16] “Example of a Weibull distribution — astroML 0.4 documentation,” Astroml.org. [Online]. Available: https://www.astroml.org/book_figures/chapter3/fig_weibull_distribution.html. [Accessed: 08-Sep-2020].

[17] “numpy.random.weibull — NumPy v1.10 Manual,” Scipy.org. [Online]. Available:https://docs.scipy.org/doc/numpy-1.10.1/reference/generated/numpy.random.weibull.html. [Accessed: 08-Sep-2020]. [18] https://www.geeksforgeeks.org/python-log-normal-distribution-in-statistics/

[18] kaplanmeier,” Pypi.org. [Online]. Available: https://pypi.org/project/kaplanmeier/. [Accessed: 08-Sep-2020].

[19] E. Chatziapostolou, “Log Rank Test, Kaplan Meier Survival Curve (Python Code),” Analytics Vidhya, 03-Dec-2019. [Online]. Available: https://medium.com/analytics-vidhya/log-rank-test-kaplan-meier-survival-curve-python-code-3fc78da644d5. [Accessed: 08-Sep-2020].