

Epidemiological characteristics and incubation period of SARS-CoV-2 during the 2020–2021 winter pandemic wave in North China: An observational study

Tiantian Liu^{1,2} | Zijian Chen¹ | Jin Xu^{1,3} 

¹School of Statistics, East China Normal University, Shanghai, China

²Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa, Israel

³Key Laboratory of Advanced Theory and Application in Statistics and Data Science – MOE, East China Normal University, Shanghai, China

Correspondence

Jin Xu, School of Statistics, East China Normal University, Shanghai 200062, China.

Email: jxu@stat.ecnu.edu.cn

Funding information

Ministry of Education of the People's Republic of China

Abstract

As the emergence of new variants of SARS-CoV-2 persists across the world, it is of importance to understand the distributional behavior of the incubation period of the variants for both medical research and public health policy-making. We collected the published individual-level data of 941 patients of the 2020–2021 winter pandemic wave in Hebei province, North China. We computed some epidemiological characteristics of the wave and estimated the distribution of the incubation period. We further assessed the covariate effects of sex, age, and living with a case with respect to the incubation period by a model. The infection-fatality rate was only 0.1%. The estimated median incubation period was at least 22 days, significantly extended from the estimates (ranging from 4 to 8.5 days) of the previous wave in mainland China and those ever reported elsewhere around the world. The proportion of asymptomatic patients was 90.6%. No significant covariate effect was found. The distribution of incubation period of the new variants showed a clear extension from their early generations.

KEYWORDS

asymptomatic infection, COVID-19, incubation period, SARS-CoV-2

1 | INTRODUCTION

Since the outbreak of the COVID-19 pandemic in December 2019 from Wuhan China, there have been more than 12 million confirmed cases and two million deaths reported all over the world.¹ The SARS-CoV-2 appeared to evolve in various ways to survive in human as mutations had been found such as the British mutation,² the Brazil variant,³ and the South Africa variant,⁴ among others.

The incubation period (IP), defined as the time between infection and symptom onset, serves as a key epidemiological parameter for medical researchers to understand the viral virulence and for health authorities to make appropriate policies. There have been intensive studies of IP of SARS-CoV-2 for the first pandemic wave in China with reported median of IP ranging from 4 to 8.5 days^{5–19} and possible covariate effects such as age,^{10,19} sex,^{19,20} and travel history.¹⁸

The current general policy for the quarantine time by World Health Organization (WHO) and many governments is 14 days.

By late April 2020, mainland China had almost eradicated the domestic pandemic with about 90 consecutive days of domestic cases under 10 per day.²¹ From June 11, 2020, there were several isolated clusters in small scale (with maximum domestic cases of 257) such as in Beijing, Qingdao (in Shandong province), Dalian (in Liaoning province), and Suifenhe (in Heilongjiang province), until the 2020–2021 winter pandemic wave took place in Hebei province in North China. This pandemic wave caused lockdown of three cities (Shijiazhuang, Xingtai, and Langfang) of population of more than 20 million, quarantine of more than 20 000 people, and 1 death.²² All these cases were considered to be imported cases with variants of SARS-CoV-2 from Europe or Russia after generations from the original type in December 2019.²³

In this article, we conducted an observational study based on the published individual-level data of 941 patients to study the distributional behavior of IP of the new variants and some epidemiological characteristics of interest. Our study aims to (i) shed lights on the evolution of the SARS-CoV-2 from the statistical perspective of the distribution of IP and (ii) provide bases for public health policy-making.

2 | METHODS

2.1 | Study design and data

This study is classified as an observational study or public health surveillance and is exemplified from ethical review or informed consent by the institutional review board.

Since January 3, 2020, when the first case of SARS-CoV-2 was diagnosed in Shijiazhuang of Hebei province, the national health commission of Hebei province started to publish individual patient narratives regarding the transmission timeline and the tracing information along with some important covariates such as sex, age, and the status of living with a case, and so on. We collected such data from 941 patients of Hebei province in North China till the last publication on February 4, 2021. The summary of the data sources is provided in Supporting Information Appendix A.

We extracted relevant data from the patient narratives with great care. Specifically, let T_0 denote the date of infection, for example, a date who attended a wedding that turned out to be a transmission occasion. When T_0 could not be precisely recalled, let $[W_0, W_1]$ denote the time window where T_0 belongs, which is either recalled by patient or derived from known information. For example, W_0 is the date when one returns from another city and W_1 is the date when one gets self-isolated or quarantined. Let T_1 denote the date of symptom onset, where symptom includes fever, dry cough, tiredness, and so on, confirmed by physicians. So IP is simply $T_1 - T_0$. At last, let T_2 denote the date of being diagnosed by nucleic acid test. Note T_2 is always observed.

In the later period of this wave, due to mass quarantine, 251 of 941 patients did not have information about the date of infection. That is, neither T_0 nor the window $[W_0, W_1]$ was available. Therefore, only 690 subjects were used in analyzing the distribution of IP. According to the actual observations of T_0 and T_1 , we categorized the data into the following four types. Type I: Both T_0 and T_1 are observed. And T_2 is greater than T_1 . Type II: T_0 is observed. But T_1 is censored by T_2 , that is, one is diagnosed without symptoms. Type III: T_1 is observed. But T_0 is left truncated by the interval $[W_0, W_1]$. Type IV: Neither T_0 and T_1 is observed, that is, T_0 is left truncated by $[W_0, W_1]$ and right censored by T_2 simultaneously. We illustrate the four types of data in Figure 1 and provided examples for each type in Supporting Information Appendix B. At last, we defined the status of living with a case (yes/no, coded as 1/0), where the confirmed case can be the spouse or a relative such as grandma, brother, and so on.

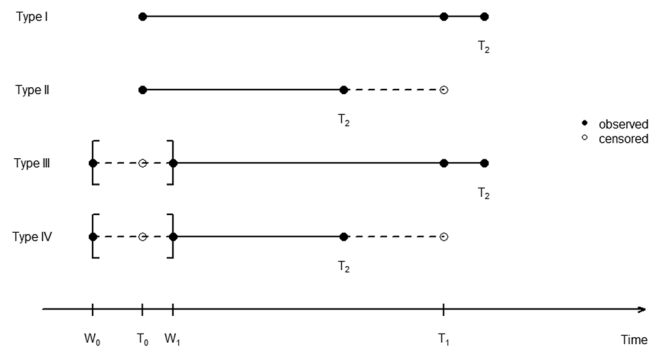


FIGURE 1 Illustration of the four types of observations of T_0 and T_1 with associated $[W_0, W_1]$ (if applicable) and T_2 , where the solid bullet stands for an actual observation and the empty circle stands for censoring

2.2 | Statistical model and analysis

We treated the IP as time-to-event (survival) random variable subject to right censoring. Here event stands for symptom onset and right censoring occurs when a patient is diagnosed without symptoms. The combined Type I and Type II data constituted the main body of such data. For Type III data with left truncation, we treated them as Type II data in a reverse direction, that is, treat IP was censored from left. For Type IV data, we approximated T_0 by W_1 so that to (under)estimate IP in a conservative way as Nie et al.¹⁴ This is in contrast to the liberal estimation by replacing T_0 by W_0 ⁹ and the robust estimation by replacing T_0 by the middle point of W_0 and W_1 .¹⁹

Unlike many statistical analysis methods to estimate the distribution of IP using parametric models such as lognormal, Weibull and exponential models,^{5,12,24} we used the traditional non-parametric Kaplan–Meier method to estimate the probability of being asymptomatic (survival function) and subsequently the median IP.²⁵ We provided a sensitivity analysis using parametric models to examine the impact of modeling.

We used the well-known semiparametric Cox model to assess the significance of covariate effect.²⁵ All tests were two sided with significance level 0.05. All statistical analyses were carried out by statistical software R (version 4.0.4) through the package “Survival” primarily.

3 | RESULTS

3.1 | Summary of some epidemiological characteristic

A total of 941 cases from January 2, 2021, to February 3, 2021 (throughout the pandemic wave of 32 days) in Hebei province were collected as described in Section 2. Figure 2 shows the numbers of daily diagnosed cases along with the associated reproduction number R_0 . It is seen that the pandemic wave climbed exponentially in the first 10 days and plateaued at about 90 cases per day for four days

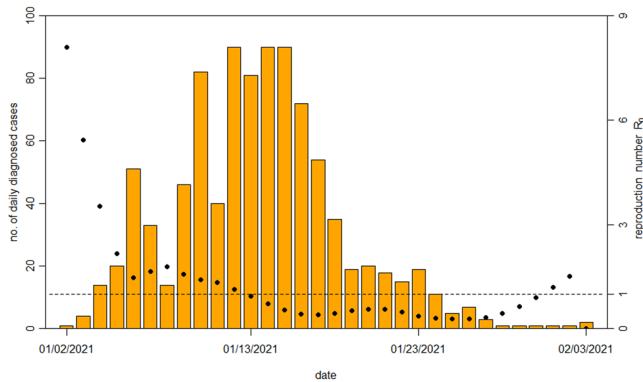


FIGURE 2 Number of daily diagnosed cases (left axis) and reproduction number R_0 (right axis) of the studying pandemic wave

before a rapid drop. The overall R_0 for the wave was around 0.84–1.0, depending on the estimation methods.²⁶ The quick control of the pandemic was largely due to the prompt measures of mass test and lockdown.

Table 1 summarizes some epidemiological characteristics of the data. The findings were as follows. (i) Individuals of age 31–60 were more likely to be infected compared with other age groups. Similar age-based difference was detected by Zhang.²⁷ (ii) More women (58.7%, 552 of 941) got infected than men (41.7%, 389 of 941) in this wave ($p < .001$). (iii) 19.3% (182 of 941) of the patients were living with a case. (iv) The proportion of asymptomatic patients was as high as 90.6% (853 of 941), nearly 10 times of that (9.4%, 88 of 941) of the symptomatic patients. While, this proportion of asymptomatic patients in the previous wave (as of February 11, 2020) was just 1.2%.²⁸ (v) The median time period between symptom onset and being diagnosed was 2 days. Note that results in (iv) and (v) were largely due to the mass test. (vi) About 59% (573 of 941) of the patients had at least one negative outcome of nucleic acid test before being diagnosed lately. About 20.2% (190 of 941) of them even had experienced more than three times of such “false” negative results. This remarkable proportion of “false negative” corroborated with the finding in (iv) and could be attributed to the sensitivity of the test kit and more likely to the prolonged IP when the variants of the virus was not detectable.²⁹

Among all 690 patients with at least one observation of T_0 or T_1 , there were 27, 180, 61, and 422 patients of Types I–IV, respectively. It is noted that in the earlier period of this wave (January 3, 2021 to January 8, 2021) most cases (122 out of 137) were reported in rural areas where the patients (many were local farmers) sought first medical aid in local clinics when the initial symptoms appeared. Some of them could not recall precisely when the infection contact was made and we had to estimate it by a broader time window. These patients became the major sources for Types I and III data. As the pandemic spread into the cities, lockdown was imposed and mass tests were given. We found most asymptomatic cases of Types II and IV data in the later time period.

TABLE 1 Epidemiological characteristics of the study patients

Characteristic	Statistic
N	941
Age	
Median (IQR)—year	46 (30)
Minimum—year	0.5 (6 months)
Maximum—year	91
Distribution— <i>n</i> (%)	
0–10	77 (8.2)
11–20	78 (8.3)
21–30	97 (10.3)
31–40	164 (17.4)
41–50	129 (13.7)
51–60	177 (18.8)
≥60	219 (23.3)
Sex— <i>n</i> (%)	
Male	389 (41.3)
Female	552 (58.7)
Exposure status— <i>n</i> (%)	
Living with a case	182 (19.3)
Others	759 (80.7)
Symptom status before being diagnosed— <i>n</i> (%)	
Symptomatic	88 (9.4)
Asymptomatic	853 (90.6)
Days between symptom onset and being diagnosed	
Minimum	0
Maximum	11
Mean	2.4
Median	2
Number of negative tests before being diagnosed— <i>n</i> (%)	
0	386 (41.0)
1	64 (6.8)
2	146 (15.5)
3	155 (16.5)
4	79 (8.4)
5	48 (5.1)
6	19 (2.0)
7	21 (2.2)
8	16 (1.7)
9–12	7 (0.7)

Abbreviation: IQR, interquartile range.

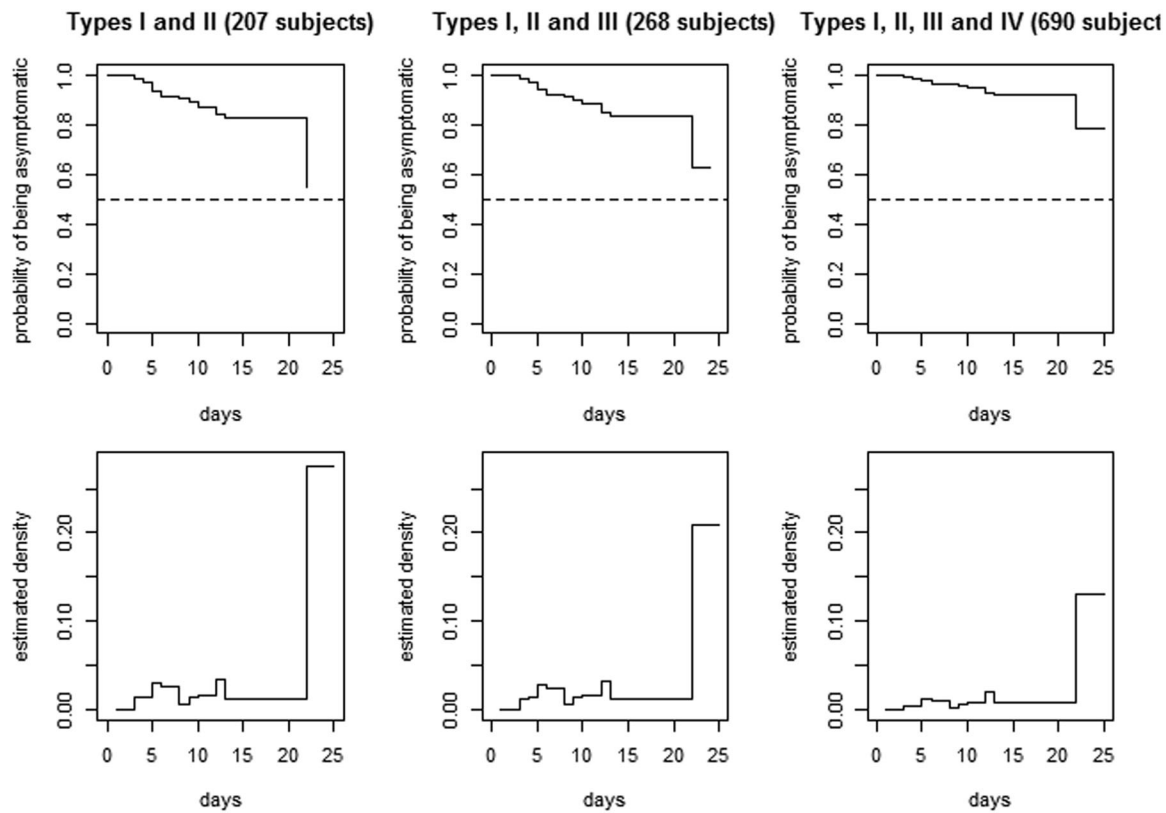


FIGURE 3 First row: Kaplan–Meier estimates of the probability of being asymptomatic based on different aggregated types of data, where the horizontal dotted line at 0.5 does not cross with any estimated survival curves. Second row: the corresponding estimated density of IP. IP, incubation period.

3.2 | Estimation of the distribution of IP and the median

We estimated the probability of being asymptomatic based on different aggregated types of data (first row of Figure 3), from which the estimated median IP was found to be greater than 22 days, which was the lower limit of the 95% confidence interval. (The corresponding log-likelihood values are -128.6 , -130.3 , and -152.4 , respectively.) The actual median could be much larger as seen in Figure 3 that the horizontal dotted line at 0.5 does not cross with any estimated survival curves. The minimum observed IP was 3 days. Second, we derived the empirical density of IP from the Kaplan–Meier estimate of the survival curve and plotted the corresponding estimated densities in the second row of Figure 3. It is seen that the densities exhibit a clear trend of the heavy tail.

We compared the estimate of the median IP to 13 reported counterparts over different time periods in Figure 4. The summary of the comparing estimates was provided in Supporting Information Appendix C. First, we noticed that there was a time gap of about 8 months between this wave and the previous wave. Second, the median IP of this wave was significantly prolonged comparing to those in the first wave before May 2020. A meta-

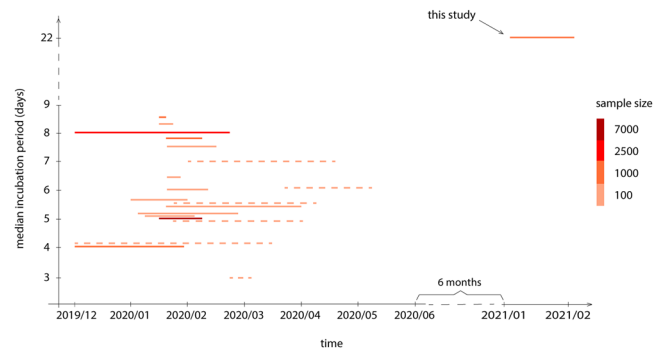


FIGURE 4 Comparison of the estimate of the median IP of the study wave and 13 reported estimates of median IP of the previous pandemic wave in China over different time periods (solid lines). Five estimates of the median IP based on the pandemics in India, Vietnam, South Korea, Singapore, and Argentina were also included (dotted lines). IP, incubation period

analysis based on 7 out of 13 articles with information of confidence intervals showed the median IP of the previous wave was 5.6 days (95% confidence interval [CI], 5.1–8.3). (See Supporting Information Appendix D for the detail.) This indicated possible virological changes of the variants of SARS-CoV-2 that evolved to adapt to the new biological environment.

TABLE 2 Estimates of the median IP and the corresponding 95% CI along with the log-likelihood based on different aggregated types of data under Weibull and lognormal models

Aggregated types	Model	Median	95% CI	Log-likelihood
I + II	Weibull	29.1	20.1, 42.3	-140.4
	Lognormal	37.6	23.0, 61.3	-138.9
I + II + III	Weibull	30.2	20.8, 43.9	-142.5
	Lognormal	39.6	24.1, 65.0	-140.8
I + II + III + IV	Weibull	Infinity	Infinity	NA
	Lognormal	71.5	38.0, 134.6	-162.1

Note: The estimation of the Weibull model under the full aggregated types of data met ill-condition in which the scale parameter is nearly zero due to large proportion of censoring.

Abbreviations: CI, confidence interval; IP, incubation period.

3.3 | Covariate analysis

The covariate analysis by the Cox model found no significant covariate effect (with respect to IP) by sex, age, or living with a case based on different aggregated data. It suggested that the IP of the new variants of SARS-CoV-2 exhibited indiscriminately with respect to the considered covariates. We also split the data into two age groups, namely “less than 60” and “no less than 60.” The log-rank test²⁵ again found no difference in the distribution of IP between the two groups. (See all details in Supporting Information Appendix E.)

3.4 | Sensitivity analysis

In addition to the nonparametric Kaplan–Meier method, we used two common parametric models, namely Weibull and lognormal models, to estimate the distribution of IP and the median. It was found in the last column of Table 2 that both parametric models yield smaller log-likelihood values than the Kaplan–Meier method (as reported in Section 3.2), indicating inferior fits to the data. The median IP was at least 29 days (95% CI, 20.1–42.3) under Weibull model and at least 37.6 days (95% CI, 23.0–61.3) under lognormal model, with the lower limits of the CI consistent to the finding of 22 days by the Kaplan–Meier method.

4 | DISCUSSION

Understanding the statistical distribution of IP is critical for studying the pathogenic process, dynamics of transmission, case fatality, and epidemic size. The derived parameters such as median IP and 95% percentile are key for public health policy decision such as isolating infected cases and quarantine period.

In this study, we presented a rather comprehensive analysis of the distributional characteristics of IP based on published data of the

2020–2021 winter pandemic wave in North China, where the variants were thought imported from Europe or Russia.

We found that the median IP was greater than 22 days. It was significantly extended from 5.6 days of the previous wave from December 2019 to April 2020. This large leap was reflected by the recent finding of He et al.³⁰ that 6.9% of population in Wuhan China was asymptomatic throughout the whole period (from infection to self-recovery). It is worth noting that the infection-fatality rate of this wave was only 0.1% (1/941). It might be attributed to the changed virulence of the variants and the timely treatment in the early stage. This finding is corroborated by recent reports of imported cases, many of whom were diagnosed positive after 2–3 weeks of quarantine without any symptom. Our results showed no significant covariate effect of sex, age, and living with a case. In particular, we found no significant difference in the distribution of IP between groups of patients older than 60 and younger than 60.

The strengths of this study are twofold. First, our data set covered a complete pandemic wave of 941 cases and contained accurate patient-level information. Second, by taking advantage of the large sample size, we adopted the nonparametric method to estimate the distribution of IP with high precision. In contrast to the parametric models, our approach is model free and therefore robust to the presence of extreme observations.

Although we made a great effort in ascertaining the dates of infection and symptom onset, there inevitably existed biases due to recall and deduction from tracing information. Our approach here for estimation was conservative such as to underestimate IP by using W_1 to approximate T_0 when analyzing the Type IV data.

In conclusion, the distribution of IP of the new variants of SARS-CoV-2 tends to have a heavy tail toward right with a prolonged median of at least 22 days. An extended period of 3 weeks by a combination of quarantine and self-isolation is recommended for asymptomatic individual from infected regions. Long-term monitoring system is warranted to study the evolution of COVID-19.

ACKNOWLEDGMENTS

This study was supported by Key Laboratory of Advanced Theory and Application in Statistics and Data Science – Ministry of Education, China.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Jin Xu conceived and designed the study, oversaw the collection of the data, took responsibility for the accuracy of the analysis. All authors had full access to all of the data in the study. Tiantian Liu and Zijian Chen collected and analyzed the data. Tiantian Liu, Zijian Chen, and Jin Xu wrote the first draft of the manuscript. All authors critically reviewed and revised the manuscript.

DATA AVAILABILITY STATEMENT

The source of data is provided in the Supporting Information Appendix. The data collected for the study will be made available up on request.

ORCID

Jin Xu  <http://orcid.org/0000-0003-2434-7183>

REFERENCES

- World Health Organization. WHO Coronavirus (COVID-19) Dashboard. Accessed March 19, 2021. <https://covid19.who.int>
- Tang JW, Tambyah PA, Hui DS. Emergence of a new SARS-CoV-2 variant in the UK. *J Infect*. 2020;S0163-4453. <https://doi.org/10.1016/j.jinf.2020.12.024>
- Faria NR, Claro IM, Candido D, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological*. 2021. Accessed March 19, 2021. <https://virological.org/t/genomic-characterisation-of-anemergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>
- Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv*. Published online December 22, 2020. <https://doi.org/10.1101/2020.12.21.20248640>
- Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill*. 2020;25.
- Cheng HY, Jian SW, Liu DP, et al. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern Med*. 2020;180:1156-1163.
- Du ZC, Gu J, Li JH, et al. Estimating the distribution of COVID-19 incubation period by interval-censored data estimation method. *China J Epidemiol*. 2020;41:1000-1003.
- Deng Y, You C, Liu Y, Qin J, Zhou XH. Estimation of incubation period and generation time based on observed length-biased epidemic cohort with censoring for COVID-19 outbreak in China. *Biometrics*. 2020;1-13.
- Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708-1720.
- Kong TK. Longer incubation period of coronavirus disease 2019 (COVID-19) in older adults. *Aging Med*. 2020;3:102-109.
- Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med*. 2020;172:577-582.
- Linton NM, Kobayashi T, Yang Y, et al. 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*. 2020;9:538.
- McAloon C, Collins Á, Hunt K, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*. 2020;10:e039652.
- Nie X, Fan L, Mu G, et al. Epidemiological characteristics and incubation period of 7015 confirmed cases with coronavirus disease 2019 outside Hubei province in China. *J Infect Dis*. 2020;222:26-33.
- Qian GQ, Yang NB, Ding F, et al. Epidemiologic and clinical characteristics of 91 hospitalized patients with COVID-19 in Zhejiang, China: a retrospective, multi-centre case series. *QJM*. 2020;113:474-481.
- Qin J, You C, Lin Q, Hu T, Yu S, Zhou XH. Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Sci Adv*. 2020;6:eabc1202.
- Sun L, Shen L, Fan J, et al. Clinical features of patients with coronavirus disease 2019 from a designated hospital in Beijing, China. *J Med Virol*. 2020;92:2055-2066.
- Xiao Z, Xie X, Guo W, et al. Examining the incubation period distributions of COVID-19 on Chinese patients with different travel histories. *J Infect Dev Ctries*. 2020;14:323-327.
- Yang L, Dai J, Zhao J, Wang YF, Deng PJ, Wang J. Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China. *Epidemiol Infect*. 2020;148:e117.
- Lee H, Kim K, Choi K, Hong S, Son H, Ryu S. Incubation period of the coronavirus disease 2019 (COVID-19) in Busan, South Korea. *J Infect Chemother*. 2020;26:1011-1013.
- The State Council Information Office of the People's Republic of China. Fighting Covid-19 China in action. Accessed March 22, 2021. http://www.gov.cn/zhengce/2020-06/07/content_5517737.htm
- Health Commission of Hebei Province. Status update of SARS-CoV-2 pandemic (20210113). Accessed March 22, 2021. <http://www.hebwsj.gov.cn/html/yqtb/20210114/375502.html>
- Qi S, Zhao X, Hao P, et al. Two reemergent cases of COVID-19 - Hebei Province, China, January 2, 2021. *China CDC Weekly*. 2021;3:25-27.
- Jiang X, Rayner S, Luo MH. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? *J Med Virol*. 2020;92:476-478.
- Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. John Wiley & Sons; 2011.
- Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B*. 2007;274:599-604.
- Zhang J, Litvinova M, Liang Y, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*. 2020;368:1481-1486.
- Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2020;41:145-151.
- Guo YY, Wang K, Zhang Y, Zhang WJ, Wang LY, Pu L. Comparison and analysis of the detection performance of six new coronavirus nucleic acid detection agents. *Chongqing Med*. 2020;49:2435-2439.
- He Z, Ren L, Yang J, et al. Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. *Lancet*. 2021;397:1075-1084.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Liu T, Chen Z, Xu J. Epidemiological characteristics and incubation period of SARS-CoV-2 during the 2020–2021 winter pandemic wave in North China: an observational study. *J Med Virol*. 2021;93:6628-6633. <https://doi.org/10.1002/jmv.27226>