Epidemiological characteristics of novel coronavirus infection: A statistical analysis of publicly available case data

Natalie M. Linton[1,*], Tetsuro Kobayashi[1,*], Yichi Yang[1], Katsuma Hayashi[1], Andrei R. Akhmetzhanov[1], Sung-mok Jung[1], Baoyin Yuan[1], Ryo Kinoshita[1], Hiroshi Nishiura[1,2,†]

[1] *Graduate School of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan*

[2] *CREST, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan*

[*] These authors equally contributed to this work.

[†] Correspondence: nishiurah@med.hokudai.ac.jp; Tel.: +81-11-706-5066

Abstract

The geographic spread of persons infected with the 2019 novel coronavirus (2019-nCoV) provides an opportunity to study the natural history of the newly emerged virus. Migration events put travelers at risk of infection for the duration of their exposure to an area where transmission is known to occur. Using publicly available data of the ongoing epidemic of 2019-nCoV where event dates for cases have been shared, the present study estimated the incubation period and other time intervals that govern interpretation of the epidemiological dynamics of 2019-nCoV infections. Our results show that the incubation periods falls within the range of two to nine days with 95% confidence, and the median incubation period is 4–5 days when approximated using the Weibull distribution, which was the best fit model. The median time from illness onset to hospitalization was estimated at 3 days. Based on the estimate of the 95th percentile estimate of the incubation period, we recommend that the length of isolation and quarantine should be at least nine days. We also note that the median time delay of 13.8 days from illness onset to death should be considered when estimating the case fatality risk of this novel virus.

Keywords: epidemiology; incubation period; virus; distribution; emerging infectious diseases

1 Introduction

As of 24 January 2020, 1287 cases of novel coronavirus (2019-nCoV) infections were reported in mainland China, causing 41 deaths. While infections in the first case cluster were initially thought to be mostly due to zoonotic (animal-to-human) transmission—possibly due to wild animals sold at a local seafood wholesale market [1, 2] – the growth of case incidence in Wuhan after closure of the market and exportation of cases across China and internationally shows compelling evidence of increasing human-to-human secondary transmission, fueled by human migration. Cases have now been detected in many other parts of the world [3], including other Asian countries, the United States, and France. This geographic expansion beyond the initial epicenter of Wuhan provides an opportunity to study the natural history 2019-nCoV infection, as migration events limit the windows of risk to the time interval during which the person traveled to the area where exposure could occur.

The incubation period is defined as the time from infection to illness onset. Knowledge of the incubation period of a directly transmitted infectious disease is critical to determine the time period required for movement restriction of healthy individuals (i.e. quarantine period) [5, 6]. We therefore undertook the incubation period estimation for the 2019-nCoV to assess how long exposed persons

must be monitored. The distribution of the incubation period may also aid in understanding the relative infectiousness of 2019-nCoV over the course of infection.

Another important epidemiologic issue in infectious disease is the inherent time delays governing each event of infection, e.g. hospitalization and death, which inform the temporal dynamics of epidemics. That is, the epidemic curve based on the date of hospitalization for each case is better interpreted and analyzed by understanding the time from symptom onset to hospitalization. A published clinical study has already shown that the average time delay from illness onset to admission is approximately 7 days [7], but variations by patients must be carefully monitored. The time from hospitalization to death is also critical in avoiding the underestimation of case fatality risk [8].

Using publicly available data of the ongoing epidemic of 2019-nCoV with known event dates, the present study aims to estimate the incubation period and other time intervals that govern the interpretation of epidemiological dynamics of 2019-nCoV. We perform the estimation of percentile points using a bootstrapping method.

2 Methods

2.1 Epidemiological data

We retrieved information on cases with confirmed 2019-nCoV infection and diagnosis outside of the epicenter of Hubei Province, China, based on official reports from governmental institutes. 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 becomes publicly available. The arranged data are available as the Online Supplementary Material (Table S1). The latest update to the dataset was on 25 January 2020 for cases reported through 24 January.

Specifically, we collected the dates of exposure (entry and/or exit from Wuhan), illness onset, hospitalization, and death. Cases included both residents from other locations who travelled to Wuhan, as well as Wuhan residents who were diagnosed while outside of Wuhan and reported by the governments of the locations where illness was detected. We thus estimated the incubation period by (i) examining visitors to Wuhan and (ii) examining both visitors to and residents from Wuhan who were diagnosed outside of Hubei Province. The former may be more precise in defining the interval of exposure, but the sample size is greater for the latter.

2.2 Statistical model

We used the dates of three critical points of the course of illness (i.e., dates of onset, hospitalization and death) to calculate four time intervals: the time periods (a) from exposure to illness onset (i.e., incubation period), (b) from illness onset to hospitalization, (c) from illness onset to death, and (d) from hospitalization to death. All these intervals were subject to a doubly interval-censored likelihood function to estimate the parameter values (which can be analyzed by using coarseDataTools package of the statistical language R) [9]:

$$L(\theta_g; \mathbf{D}) = \prod_i \int_{E_{L,i}}^{E_{R,i}} \int_{S_{L,i}}^{S_{R,i}} g(e) f(s-e)\, \mathrm{d}s\mathrm{d}e\,.$$

Here, for example in the case of (a), $g(.)$ is the probability density function (p.d.f.) of exposure following a uniform distribution, and $f(.)$ is the p.d.f. of the incubation period independent of $g(.)$. **D** represents a dataset among all observed cases i. Exposure and symptom onset obey the upper and lower bounds, (E_R, E_L) and (S_R, S_L), respectively. For instance, if the date of illness onset is for one day, the respective interval is $(S_R, S_R + 1)$, where S_R is the reported date of illness onset.

We performed a bootstrap method, based on case resampling, to compute the 95% confidence intervals (CI). Likewise, we were able to calculate distributions of (b), (c) and (d). We also assume that the probability density function $f(.)$ follows three different distributions, i.e., lognormal, Weibull and gamma distributions. Akaike Information Criterion (AIC) was used to identify the best fit model for each time interval.

3 Results

Table 1 shows estimated percentiles and AIC values for each combination of time interval and distribution. For the incubation period estimates, the best fit was found with the Weibull distribution for data both excluding and including Wuhan residents. The median incubation period using the Weibull distribution was estimated at 4.6 days (95% CI: 3.3, 5.7) when excluding Wuhan residents ($n = 12$) and 5.0 days (95% CI: 4.1, 5.8) when including Wuhan residents ($n = 31$). Figure 1 shows the cumulative distribution function of the incubation period, and the 5th and 95th percentiles are shown in addition to the median. The 95th percentiles were estimated at 7.3 days (95% CI: 5.6, 8.4) days for non-Wuhan residents and at 7.6 days (95% CI: 6.0, 8.8) when including Wuhan residents.

		Incubation period excluding WR (days)	Incubation period including WR (days)	Onset to hospitalization (days)	Onset to death (days)	Hospitalization to death (days)
Distribution	**Value**	12 cases	31 cases	43 cases	33 cases	34 cases
Lognormal	5%	2.0 (1.0–4.8)	2.3 (1.4–4.5)	0.26 (0.1–0.7)	6.7 (5.4–8.4)	2.2 (1.3–3.6)
	Median	4.2 (2.9–5.7)	4.7 (3.7–5.7)	2.3 (1.4–3.6)	13.8 (11.8–16.0)	7.4 (5.7–9.5)
	95%	8.8 (5.8–10.9)	9.6 (6.4–13.5)	20.6 (13.5–30.3)	28.7 (21.5–36.5)	24.4 (17.6–31.9)
	99%	11.9 (5.9–16.9)	13.0 (7.1–20.7)	50.9 (28.1–93.0)	38.8 (27.1–52.9)	40.1 (26.4–58.2)
	AIC	41.1	53.2	244.1	225.5	224.4
Weibull	5%	2.0 (0.9–4.6)	2.4 (1.3–4.6)	0.1 (0.0–0.5)	4.4 (3.4–6.7)	1.7 (1.0–3.1)
	Median	4.6 (3.3–5.8)	5.0 (4.0–5.8)	2.7 (1.7–4.2)	14.6 (12.5–17.1)	8.3 (6.4–10.5)
	95%	7.3 (5.6–8.4)	7.6 (6.0–8.8)	14.8 (10.1–20.4)	28.9 (21.0–34.9)	20.0 (15.7–23.5)
	99%	8.4 (5.7–10.3)	8.6 (6.3–10.5)	24.3 (15.2–37.1)	35.2 (24.2–43.7)	25.9 (19.8–31.5)
	AIC	37.5	48.7	237.1	244.1	220.1
Gamma	5%	2.0 (0.9–5.3)	2.3 (1.3–4.6)	0.1 (0.0–0.4)	6.1 (4.8–7.9)	1.9 (1.1–3.5)
	Median	4.4 (3.1–5.6)	4.8 (3.8–5.7)	2.7 (1.6–4.1)	14.3 (12.2–16.7)	7.9 (6.2–10.0)
	95%	8.2 (5.6–9.5)	8.7 (6.4–10.8)	14.8 (10.6–19.2)	28.0 (21.7–34.7)	20.9 (16.3–25.4)
	99%	10.2 (5.8–12.8)	10.9 (7.0–14.3)	23.9 (16.4–32.0)	xw	28.8 (21.8–35.7)
	AIC	39.5	51.3	236.4	227.9	220.7

Table 1. Bootstrap estimates from 1000 iterations. All cases were diagnosed with laboratory-positive 2019-nCoV outside of Hubei Province. WR: Wuhan residents. AIC $= -2 \cdot L(\theta_*; \mathbf{D}) + 2 \cdot 6$ (all estimates had 6 parameters). Ranges for onset, hospitalization, and death calculated as left = reported date; right = reported date + 1 day. Shaded cells indicate the model with the minimal AIC value.

The median time from illness onset to hospitalization was estimated at 2.7 days (95% CI: 1.7, 4.2) using the gamma distribution, which yielded the lowest AIC value (Table 1). Figure 2A shows the corresponding p.d.f. Time from symptom onset and hospitalization to death were also computed (Table 1 and Figure 2BC). The best-fit models for each interval were the lognormal and Weibull distributions, respectively. The median time from onset to death was 13.8 days (95% CI: 11.8, 16.0) and the median time from hospitalization to death was 8.3 days (95% CI: 6.4, 10.5).

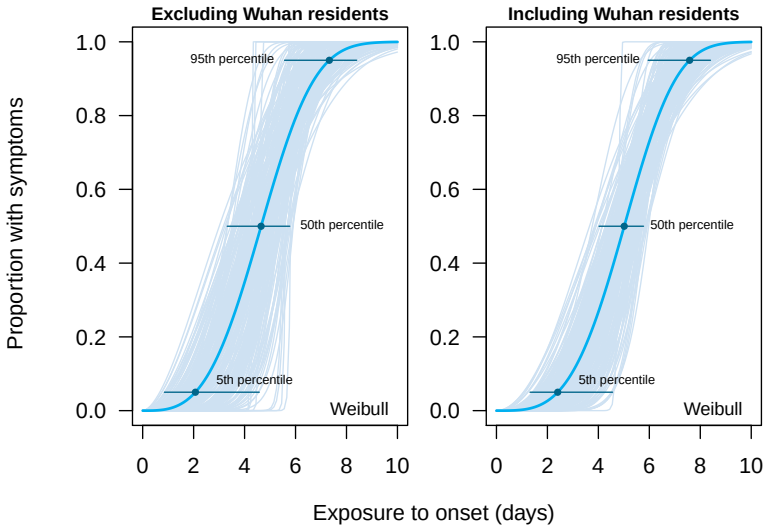

Figure 1. Estimated cumulative distribution for the incubation period of 2019-nCoV from outbreak cases reported in January 2020. Data are from public case reports published by governments outside of Hubei Province, China. Left: excludes Wuhan (Hubei Province) residents from the estimates. Right: includes Wuhan residents in the estimates.

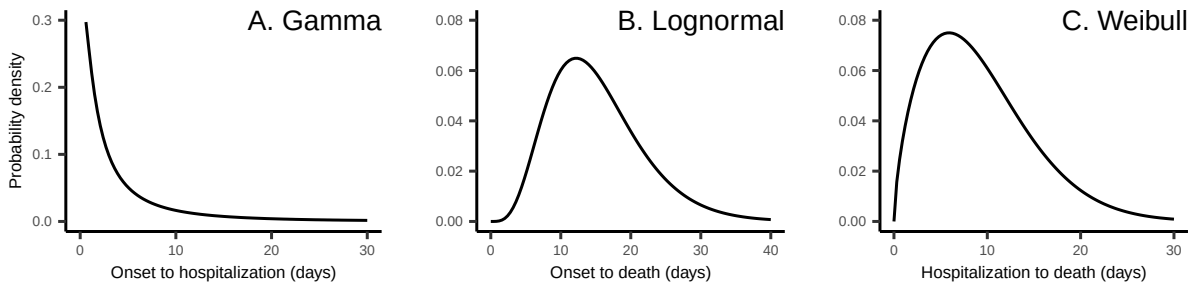

Figure 2. Probability distributions of time from onset or hospitalization to hospitalization or death for 2019-nCoV outbreak cases reported through 24 January 2020. (**A**) Probability density of the time from illness onset to hospitalization in days set to the best-fit gamma distribution. (**B**) Probability density of the time from illness onset to death in days set to the best-fit lognormal distribution. (**C**) Probability density of the time from hospitalization to death in days set to the best-fit Weibull distribution.

4 Discussion

Our results show that 95% of incubation periods fall within the range of 2 to 9 days, and the median incubation period was 4–5 days when the Weibull distribution was used as the best-fit model. The median time from illness onset to hospitalization was approximately 3 days. The median time from illness onset to death was 13.8 days, the delay of which is key to appropriate estimation of the case fatality risk for 2019-nCoV [10].

The present study advances the public discussion on 2019-nCoV infections as both the incubation period and the time from illness onset to death were explicitly estimated using publicly available data. Our estimated median incubation period of 2019-nCoV is comparable to known median values of the incubation period for severe acute respiratory syndrome (SARS)—estimated at 4.0–6.4 days [8, 11, 12]. In addition to empirically showing the comparability to SARS, the present study has also shown that the 95th percentile of the incubation period is around 7–8 days, indicating that a nine-day quarantine period could mostly ensure the absence of disease among exposed healthy individuals.

The time from illness onset to death is also comparable to SARS [8], and the 13.8-day median delay that we calculated indicates that the crude estimation of the ratio of the cumulative number of deaths to that of cases tends to result in underestimation of the case fatality risk, especially during the early stage of the epidemic. During the SARS epidemic in Hong Kong, 2003, the time from illness onset to hospitalization was shown to have shortened as a function of calendar time, reflecting that contact tracing practice had worked out gradually. Moreover, the study on pandemic influenza H1N1-2009 has demonstrated a negative association between the time from illness onset to hospitalization and the basic reproduction number, i.e., the average number of secondary cases generated by a single primary case in a fully susceptible population [13]. While our estimate was approximately 3 days, consistent with high mortality at hospital settings, this may be thus shortened in the future course of the epidemic.

Several limitations of the present study exist. First, the dataset relies on published information, and the defined event date (e.g. the date of illness onset) depends on the decision-making of each governmental authority. Given the novelty of the illness, it is possible that symptom onset and other event data may have been dealt with differently between jurisdictions (e.g., was onset the date of fever or date of dyspnea?). Second, the sample size was limited, and the variance was likely to be biased. Third, we were not able to examine heterogeneity of estimates by different attributes of cases (e.g. age and risk groups).

While several future tasks remain, we believe that the present study has been successful in clarifying the epidemiological characteristics of novel coronavirus infection. The length of quarantine should be at least nine days, and the time delay from illness onset to death of fourteen days must be addressed when estimating the case fatality risk.

Supplementary material: Table S1 Event dates for cases included in the analysis.

Author Contributions: N.M.L., T.K., A.R.A., and H.N. conceived the study and participated in the study design. All authors assisted in collecting the data. N.M.L., T.K. and H.N. analyzed the data and T.K., H.N., N.M.L. and Y.Y. drafted the manuscript. All authors edited the manuscript and approved the final version.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Peng, W., Xinxin H., Eric, H.Y., Jessica, W.Y., Kathy, L., Joseph, W.T., Benjamin C., Gabriel, L. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Euro Surveillance 2020, 25(3), pii=2000044 (doi:10.2807/1560-7917.ES.2020.25.3.2000044).

2. Center for Disease Control and Prevention. 2019 Novel Coronavirus, Wuhan, China. Available online: (reference link) (accessed on 24 January 2020).

3. European Centre for Disease Prevention and Control data. Geographical distribution of 2019-nCov cases. Available online: (reference link) (accessed on 24 January 2020).

4. Nishiura, H., Lee, H.W., Cho, S.H., Lee, W.G., In, T.S., Moon, S.U., Chung, G.T., Kim, T.S. Estimates of short- and long-term incubation periods of Plasmodium vivax malaria in the Republic of Korea. Trans R Soc Trop Med Hyg. 2007 Apr;101(4):338-43 (doi:10.1016/j.trstmh.2006.11.002).

5. Lessler, J., Reich, N.G., Cummings, D.A. New York City Department of Health and Mental Hygiene Swine Influenza Investigation Team, Nair HP, Jordan HT, Thompson N. Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. N Engl J Med. 2009, 361(27), 2628-36. (doi:10.1056/NEJMoa0906089).

6. Nishiura, H. Determination of the appropriate quarantine period following smallpox exposure: an objective approach using the incubation period distribution. Int J Hyg Environ Health. 2009, 212(1), 97-104. (doi:10.1016/j.ijheh.2007.10.003).

7. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X; Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie. J; Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B. Lancet e. in press, (doi:10.1016/S0140-6736(20)30183-5).

8. Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.M., Thach, T.Q., Chau, P., Chan, K.P., Lam, T.H., Tse, L.Y., Tsang, T., Liu, S.H., Kong, J.H., Lau, E.M., Ferguson, N.M., Anderson, R.M. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet. 2003 May 24;361(9371):1761-6 (doi:10.1016/S0140-6736(03)13410-1).

9. Reich, N.G., Lessler, J., Cummings, D.A.: Brookmeyer, R. Estimating incubation period distributions with coarse data. Statistics in Medicine 2009, 28(22), 2769-84. (doi:10.1002/sim.3659).

10. Ghani AC, Donnelly CA, Cox DR, Griffin JT, Fraser C, Lam TH, Ho LM, Chan WS, Anderson RM, Hedley AJ, Leung GM. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. Am J Epidemiol. 2005,162(5), 479-86 (doi:10.1093/aje/kwi230).

11. Cowling, B.J., Park. M., Fang. V.J., Wu, P., Leung, G.M., Wu, J.T. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. Euro Surveillance 2015, 20(25), pii: 21175 (doi:10.2807/1560-7917.es2015.20.25.21163).

12. Lessler, J., Reich, N.G., Brookmeyer, R., Perl, T.M., Nelson, K.E., Cummings, D.A. Incubation periods of acute respiratory viral infections: a systematic review. Lancet Infect Dis. 2009, 9(5), 291-300 (doi:10.1016/S1473-3099(09)70069-6).

13. Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., Van Kerkhove, M.D., Hollingsworth, T.D., Griffin, J., Baggaley, R.F., Jenkins, H.E., Lyons, E.J., Jombart, T., Hinsley, W.R., Grassly, N.C., Balloux, F., Ghani, A.C., Ferguson, N.M., Rambaut, A., Pybus, O.G., Lopez-Gatell, H., Alpuche-Aranda, C.M., Chapela, I.B., Zavala, E.P., Guevara, D.M., Checchi, F., Garcia, E., Hugonnet, S., Roth, C. WHO Rapid Pandemic Assessment Collaboration. Pandemic potential of a strain of influenza A (H1N1): early findings. Science. e Jun 19, 324(5934), 1557-61 (doi:10.1126/science.1176062)